Characterization of the DNA Binding and Structural Properties of the BRCT Region of Human Replication Factor C p140 Subunit*

Masakazu Kobayashi, Francis Figaroa, Nico Meeuwenoord, Lars E. T. Jansen[1], and Gregg Siegal[2]

From the Leiden Institute of Chemistry, Leiden University, Postbus 9502, 2300 RA Leiden, The Netherlands

BRCT domains, present in a large number of proteins that are involved in cell cycle regulation and/or DNA replication or repair, are primarily thought to be involved in protein-protein interactions. The large (p140) subunit of replication factor C contains a sequence of ~100 amino acids in the N-terminal region that binds DNA and is distantly related to known BRCT domains. Here we show that residues 375–480, which include 28 amino acids N-terminal to the BRCT domain, are required for 5′-phosphorylated double-stranded DNA binding. NMR chemical shift analysis indicated that the N-terminal extension includes an α-helix and confirmed the presence of a conserved BRCT domain. Sequence alignment of the BRCT region in the p140 subunit of replication factor C from various eukaryotes has identified very few absolutely conserved amino acid residues within the core BRCT domain, whereas none were found in sequences immediately N-terminal to the BRCT domain. However, mapping of the limited number of conserved, surface-exposed residues that were found onto a homology model of the BRCT domain, revealed a clustering on one side of the molecular surface. The cluster, as well as a number of amino acids in the N-terminal α-helix, were mutagenized to determine the importance for DNA binding. To ensure minimal structural changes because of the introduced mutations, proteins were checked using one-dimensional ^1H NMR and CD spectroscopy. Mutation of weakly conserved residues on one face of the N-terminal α-helix and of residues within the cluster disrupted DNA binding, suggesting a likely binding interface on the protein.

The BRCA1 C-terminal homology (BRCT) domain is an abundant structural unit (1, 2) found in more than 900 proteins from all biological kingdoms as listed in the Uniprot data base.[3] Nearly all of these proteins are involved in the cell cycle checkpoint response to damaged DNA and/or more directly in DNA replication or repair. The BRCT superfamily has been further classified into three subsets. The first consists of a core of highly conserved domains found in proteins such as BRCA1 itself, the *Saccharomyces cerevisiae* Rad9 protein and the p53-binding protein 53BP1. A second distinct and more distantly related set can be found in DNA binding enzymes such as the bacterial NAD-dependent ligases and poly(ADP-ribose) polymerase. Finally, the retinoblastoma tumor suppressor and related proteins may contain a very distant member of the BRCT family (1).

BRCT domains are responsible for a number of important homo- and heterotypic protein-protein interactions. For instance, XRCC1, a protein involved in repair of single-stranded DNA breaks, binds DNA ligase III via its C-terminal BRCT domain (3), whereas its N-terminal BRCT domain specifically binds poly(ADP-ribose) polymerase (4). Other BRCT domains, such as those in the *S. cerevisiae* protein Rad9, function in homotypic (self) protein binding (5). Recently, it has been shown that many BRCT domains posses a phosphoserine-specific, protein binding function (6–8). Finally, a growing number of BRCT domains appear to bind DNA. The best characterized is the BRCT from the large subunit of replication factor C (RFC)[4] (9). The DNA binding characteristics of other BRCT domains, however, remain less well characterized (10, 11).

The BRCT domain consists of roughly 90 amino acids (1, 2). Structural information is available for an increasing number of BRCT domains, including those from XRCC1 (12), BRCA1 (13), 53BP1 (14, 15), DNA ligase III (16), and the bacterial NAD$^+$-dependent DNA ligase (17). Despite the sequence divergence, all of these structures display a conserved fold consisting of a four-stranded, parallel β-sheet surrounded by three α-helices. However, the way in which these domains interact with known ligands varies widely. For instance, the interaction between the C-terminal BRCT domain of XRCC1 and the BRCT domain of DNA ligase III occurs through residues in the α1 helix of each domain (18). BRCA1 also contains two BRCT domains; however, they form an obligate paired structural unit that specifically binds to a phosphoserine containing sequence in the protein BACH1 (19–21). Hence, despite conservation of the three-dimensional structure of each domain, the mechanism by which BRCT domains execute their function differs significantly within the BRCT superfamily.

RFC is a five-protein complex involved in both the replication and repair of chromosomal DNA (22–24). The primary function of RFC appears to be to open the "sliding clamp" protein PCNA and "load" it onto DNA where it serves as a binding platform for a multitude of enzymes and regulatory proteins involved in the replication and repair of DNA. RFC consists of four subunits of between 35 and 40 kDa that share homology with a central region of the fifth subunit, which has a molecular mass of 140 kDa in mammals (referred to as p140). The N-terminal half of the p140 subunit contains sequences unique to RFC, including a region with DNA binding activity (9, 25–27) that is not required for clamp loading (28, 29). This region contains amino acid sequences, which form part of the second, distinct class of BRCT domains (1, 2). Here we demonstrate that this region of p140 does indeed contain a BRCT domain but that additional sequences outside the domain are required for 5′-phosphate-specific, double-stranded DNA binding. Using mutagenesis we define a model of the protein-DNA complex.

MATERIALS AND METHODS

Cloning—The plasmid containing a cDNA clone coding for residues 369–480 of human RFC p140 was a kind gift of Prof. Ullrich Hübscher (Universität Zürich, Switzerland). The three different constructs were generated using standard PCR-based methods and were cloned into pET20b (Novagen) with a C-terminal His$_6$ tag.

Site-directed Mutagenesis—The procedure for generating point mutations was adopted from the QuikChange method (Stratagene) with some modifications. A DpnI-treated, nicked plasmid was transformed into highly competent *Escherichia coli* DH5α cells (10^8 colonies/μg). Plasmid carrying the desired mutation was identified by the unique silent restriction site introduced via the mutagenesis primers and was subsequently verified by DNA sequencing. The (R388A) mutation, was generated using splice overlap extension (30).

Protein Expression and Purification—Proteins were expressed for 3 h at 37 °C in BL21(DE3)/pLysS cells (Novagen). Lysed cells were centrifuged at $20,000 \times g$ for 30 min at 4 °C and the supernatant applied to a 6-ml metal chelate column (Novagen) charged with Ni^{2+}. The column was developed per the manufacturer's recommendations. Fractions containing the His-tagged protein, as judged by SDS-PAGE, were pooled, and EDTA was added to the protein solution to a final concentration of 5 mM to chelate Ni^{2+} leached from the column. The protein solution was concentrated through an Amicon ultrafiltration device (YM10 membrane) to 5-ml volume. The protein was further purified to apparent homogeneity using a 140-ml (1.6×75 cm) column of Superose 12 (Amersham Biosciences) equilibrated with 20 mM Tris, pH7.5, 50 mM NaCl and 1 mM dithiothreitol. All proteins were stably stored at 4 °C.

Preparation of Oligonucleotides—The sequences of all oligonucleotides used in this study are presented in Table 1. The oligonucleotides were synthesized using standard solid state methods including the 5′-PO$_4$ (with the exception noted in Table 1). When labeling was required, 2 pmol of 5′-phosphorylated oligonucleotide were used in an exchange reaction employing phage T4 polynucleotide kinase and 20 pmol of [γ-^{32}P]ATP as substrate. To ensure a very high percentage of double-stranded DNA, the hairpin oligonucleotides were subsequently denatured at 100 °C for 10 min and then slowly cooled to room temperature.

Circular Dichroism (CD) Spectroscopy—RFC p140 proteins were exchanged into a buffer consisting of 10 mM NaCl, 1 mM KH$_2$PO$_4$, 10 mM Na$_2$PO$_4$, pH 7.5, and 2.7 mM KCl. Each protein was diluted to 10 μM in the same buffer. CD spectra were recorded on a Jobin Yvon CD6 instrument at 1-nm intervals over 190–240 nm wavelength at 21 °C in a 0.5-mm path length cuvette. Each of the purified mutants was exchanged into the same K$^+$/Na$^+$ phosphate buffer using PD10 columns (Amersham Biosciences). CD spectra of the mutants (K397E, T415A, R423A, K458E, and wild type) were measured at 21 °C with scanning at 1-nm increments between 195–260 nm in a 0.1-mm path length cuvette. For each protein, five spectra were recorded to obtain an average spectrum from which the background spectrum of the buffer was subtracted.

Stoichiometric Titration—10 μM oligo 1 (in 25 mM Tris-HCl, pH 7.5, 50 mM NaCl and 1 mM dithiothreitol) was mixed with the indicated amount of RFC p140-(375–480) (in the identical buffer) in a final volume of 250 μl. The DNA-protein complex was separated from the free DNA on a Superdex 10/300 GL column (Tricorn) at a flow rate of 1.5 ml/min. The absorbance peaks at 260 nm, corresponding to protein-DNA complex and free DNA, were fitted with a Lorenzian curve and subsequently the area underneath the curves was integrated with a trapezoidal function using SIGMAPLOT 7.0. The fraction of protein-bound DNA was then calculated by dividing the area of the peak representing the complex by the sum of the area of the complex and the free DNA peaks.

Detection of Protein-DNA Complexes—DNA binding was detected using a gel retardation assay. The indicated amount of the various RFC p140 proteins were diluted in a buffer of 10 mM HEPES, pH 7.8, 2 mM MgCl$_2$, 0.1 mM EDTA, 100 μg/ml bovine serum albumin, 15% glycerol, 0.8 μg/ml poly(dI-dC) and 2 mM dithiothreitol. The indicated amount and appropriate volume of 5′-^{32}P-labeled oligonucleotide was added to the mixture to a final volume of 12 μl, which was subsequently incubated on ice for 15 min and then applied to a non-denaturing 8% Tris-glycine acrylamide gel and electrophoresed at 150 V for 30 min at 4 °C in 25 mM Tris-HCl pH 8.5, 200 mM glycine, 1 mM EDTA (TGE) buffer. Radioactivity was detected photographically (X-OMAT, Kodak) or using a PhosphorImager (Bio-Rad) after the gel was dried. The fraction of DNA in the DNA-protein complex was calculated as the percentage of total shift (%) = [complex − background]/[total activity per lane − background] × 100, where total activity = complex + free DNA. Gel retardation was also used to determine the ligand (DNA) requirements for binding but in the form of a competition assay. In these experiments both labeled and unlabeled DNA were premixed and then added to the protein solution. The amount of the competing oligonucleotide used in excess over the labeled oligonucleotide in each assay is indicated in the figure legend. To determine the K_D of the protein-DNA complex, a constant amount of radiolabeled oligo 1 was titrated with an increasing amount of protein. The fraction of protein bound-DNA at each titration point was determined as described above. K_D was determined by iterative fitting of Equation 1 (31),

$$\nu = \frac{(D_{tot} + P_{tot} + K_D - \sqrt{(D_{tot} + P_{tot} + K_D)^2 - 4 \times D_{tot} \times P_{tot})}}{2 \times D_{tot}}$$

(Eq. 1)

using SIGMAPLOT 7.0 assuming a 1:1 protein-DNA complex (see results), where ν is the fraction of the bound DNA, D_{tot} is the total DNA concentration, and P_{tot} is the total protein concentration.

NMR Measurements—Protein samples were prepared for NMR spectroscopy and analyzed as reported earlier (32). The consensus chemical shift index (CSI) score was generated with the program CSI version 1.0 (33) from ^{13}Cα, ^{13}Cβ, and ^{1}Hα chemical shifts values as inputs. Protein concentration varied from 0.05 to 0.1 mM as determined by UV absorbance at 277 nm. One-dimensional ^{1}H NMR spectra of each protein were recorded on a 500-MHz Bruker AVANCE instrument at 298 K (25 °C) and processed with TOPSPIN 1.3 (Bruker).

RESULTS

Identification of Protein Domains Required for DNA Binding Activity—The DNA binding domain of human RFC p140 was initially defined as consisting of residues 369–480 (25). Because this region includes sequences outside the conserved BRCT domain we were interested to find out whether just the BRCT domain-(403–480) itself was sufficient for DNA binding. Based on the published sequence alignments (1, 2) we made three N-terminal deletion constructs whose C termini all coincide with the expected C terminus of the BRCT domain. All three-deletion constructs were C-terminally tagged with six histidines, expressed in *E. coli* and subsequently purified to apparent homogeneity using metal chelate and gel-filtration chromatography. The three proteins behaved somewhat differently on the gel-filtration column (Fig. 1A). Although p140-(375–480) eluted with a retention time expected of a 12-kDa monomer, p140-(403–
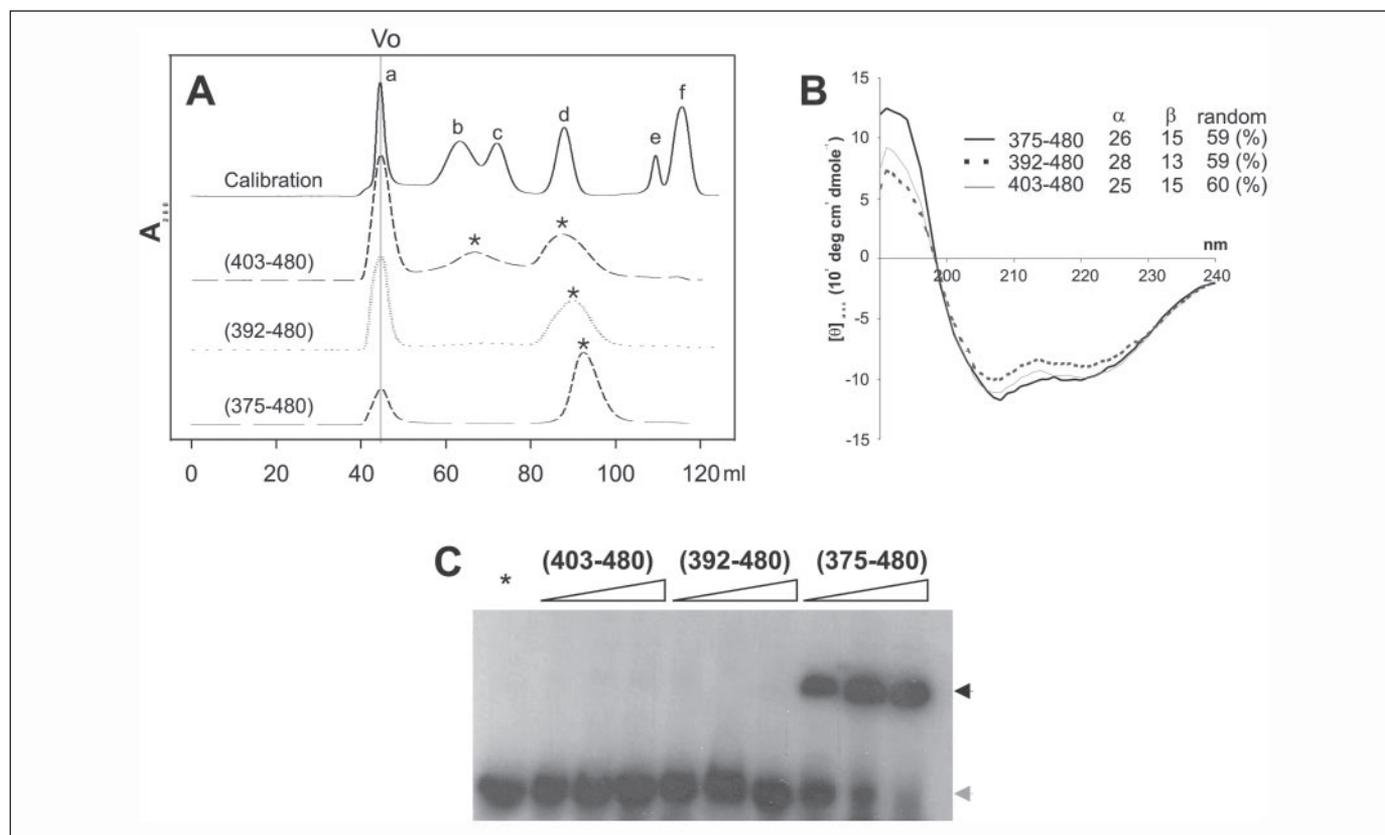
FIGURE 1. **Characterization of RFC p140 BRCT-containing proteins.** The predicted RFC p140 BRCT domain includes residues 403–480. Three different RFC proteins, amino acids 375–480, 392–480, and 403–480 were purified from *E. coli* cells and characterized as follows. *A*, gel-filtration chromatography using Superose 12 resin. The elution profile of each of the three different RFC proteins is shown with peaks containing the relevant protein indicated by an *asterisk*. The calibration set contained proteins of 670 (*a*), 158 (*b*), 44 (*c*), 17 (*d*), 6.5 (*e*) and 1.3 kDa (*f*), which were used to estimate the molecular weight of the RFC proteins (see text). *Vo* represents the void volume of the column. *B*, the far UV-CD spectra of all three proteins have characteristic minima at 208 and 220 nm, reflecting α-helical and β-strand content, respectively. The percentage of each secondary structure was predicted using the program K2d (34). *C*, a gel retardation assay was used to detect binding to the 5'-^{32}P-labeled oligodeoxynucleotide (oligo 1, Table 1) by each of the proteins. An increasing amount of (40, 200, or 400 fmol) of the indicated constructs of hRFC p140 was titrated into 40 fmol of DNA. The *asterisk* denotes a control in which no protein was added to the DNA. The *black arrow* indicates the position of the DNA-protein complex, and the *gray arrow* indicates that of the unbound DNA.

480) eluted as a broad peak corresponding to an estimated molecular mass of 18 kDa (Fig. 1*A*) as well as in an aggregate that elutes at the position expected of a 100-kDa protein. This data suggests that the core BRCT domain is in equilibrium between mono-, di-, and possibly trimeric states in solution. p140-(392–480) eluted slightly faster than expected, correlating with an approximate molecular mass of 15 kDa (Fig. 1*A*), which is somewhat larger than the expected value (10 kDa). The secondary structure of the three constructs was analyzed using CD spectroscopy (Fig. 1*B*). All three proteins gave rise to similar CD spectra suggesting a shared fold that remained intact even in the smallest protein. Prediction of secondary structure content from the CD spectra using the program K2d (34) shows that all of the peptide constructs contained a mixture of α and β structures, as would be expected of a BRCT domain.

Each of the three proteins was tested for DNA binding using a gel retardation assay in which purified proteins were titrated into a constant amount of a 5'-^{32}P-labeled oligo 1 (Fig. 1*C*; all DNA binding assays were performed in the presence of excess, unlabeled poly(dI-dC).). Oligo 1 (Table 1), which is used in all subsequent experiments, consists of a 10-bp double-stranded region connected by a 4 base hairpin turn and contains a 4 base recessed 5'-end. Only the protein containing 28 residues N-terminal to the predicted boundary of the BRCT domain (p140-(375–480)) demonstrated DNA binding activity. Given the CD data that indicates that the isolated BRCT domain remains folded, the DNA binding data suggest that sequences N-terminal to the conserved BRCT domain are directly required for DNA binding. We refer to this

TABLE 1

Sequences of oligonucleotides used in this study

For hairpin oligonucleotides, the hairpin sequence is highlighted with bold letters. The unpaired bases of oligo 2 are underlined. A "p" indicates 5'-PO$_4$, whereas "b" indicates a 5'-biotinylated.

Name	Duplex size	DNA sequence	Structure
Oligo 1	10 bp	pCTCGAGGTCG**TCAT**CGACCTCGAGATCA	
	9 bp	pCTCGAGTCG**TCAT**CGACTCGAGATCA	
	8 bp	pCTCAGTCG**TCAT**CGACTCAGATCA	
	7 bp	pCTAGTCG**TCAT**CGACTAGATCA	
	6 bp	pCTGTCG**TCAT**CGACAGATCA	
Oligo 2	9 bp	pCTCGA_T_GTCG**TCAT**CGAC_C_TCGAGATCA	
5'OH	10 bp	CTCGAGGTCG**TCAT**CGACCTCGAGATCA	
5'Pb	10 bp	pCTCGAGGTCG**TCAT**CGACCTCGAG	
5'B	20 bp	bTGCAGATTGCGCAATCTGCA	
Ss	Single stranded	TGGGGTGGGGT	
5'Ps	Single stranded	pTGGGGTGGGGT	

extended, minimal DNA-binding protein as the BRCT region to indicate the requirement for amino acid sequences not within the conserved BRCT domain.

Specificity of DNA Binding—We wished to further characterize the nature of the protein-DNA complex. Because the stoichiometry of the complex was not clear from previous reports, an increasing amount of p140-(375–480) was titrated into a constant amount of 10-bp oligo 1 (Table 1) until saturation of binding was clearly achieved. The protein-DNA complex was separated from unbound DNA using gel-filtration, and the amount of free DNA and DNA-protein complex was quantified to determine the fraction of DNA-bound protein. The fraction of bound DNA increased nearly linearly with increasing protein until a protein-DNA ratio of 1.25:1 at which binding begins to saturate. (Fig. 2A). Given the slope and saturation point, the binding curve strongly suggests a 1:1 protein-DNA ratio in the complex. To accurately measure the K_D of the complex, the titration was repeated using the gel retardation assay. The steps in the titration were more closely spaced about the estimated K_D for the complex. Fitting of the binding isotherm, assuming a 1:1 stoichiometry as suggested above, yielded an apparent K_D of 10 nM (not shown). To determine the length of duplex DNA required for optimal binding a series of oligonucleotides was synthesized that systematically reduced this region but left the overall base content intact (Table 1). The series of shorter oligonucleotides was titrated into ^{32}P-labeled 10-bp oligo 1, and binding was determined using gel retardation in the form of a competition assay. Fig. 2B demonstrates that DNA with a duplex region as short as 7 base pairs can effectively compete with the labeled oligo 1, but for maximal binding 9–10 base pairs are required. Duplex regions greater than 10 base pairs did not result increased binding affinity (not shown). In addition, fully base-paired DNA forms the best ligand because a mismatched Watson-Crick base pair 6 nucleotides away from the 5′-phosphate (oligo 2) reduced the affinity of the protein-DNA interaction (Fig. 2B).

The gel shift competition assay was subsequently used to assess the determinants for binding to the 5′-end of dsDNA (Fig. 2C). Using a protein concentration close to K_D, competition binding between two different oligonucleotide ligands was performed by mixing the 5′-^{32}P-labeled oligo 1 (10 base pairs) with oligonucleotides with various 5′-chemistries (Table 1) prior to the addition of p140-(375–480). As expected from previous studies (9, 25), binding to the 5′-phosphorylated hairpin oligonucleotide could not be competed with dsDNA containing either a 5′-OH or a 5′-biotin, nor a 5′-OH single-stranded oligodeoxynucleotide (Fig. 2C). Competition with a 5′-phosphorylated single-stranded DNA was just detectable, indicating a weak affinity of the protein for this ligand. Finally, competition was slightly more effective with dsDNA with a recessed 5′-phosphate end than with blunt ended DNA (Fig. 2C) correlating with the earlier report that the *Drosophila* RFC p140 binds 5′-end recessed dsDNA better than blunt ended DNA (9). DNA binding was also insensitive to salt up to 500 mM NaCl and not dependent on Mg^{2+} (data not shown).

Structural Characterization of the RFC p140 BRCT Region—Because the minimal DNA binding domain includes sequences N-terminal to the conserved BRCT domain, we wished to ascertain whether or not it contained elements of regular secondary structure. As a first approach we used the PSIPRED secondary structure prediction service (35). Sequences within the first 28 amino acids of the BRCT region were predicted to form an α-helix with a high degree of confidence (Fig. 3A). The secondary structure expected of a BRCT domain was also predicted between amino acids 403 and 480. As a control we performed a similar analysis on other BRCT domain sequences with known structures and obtained a close correlation between the predicted and experimentally
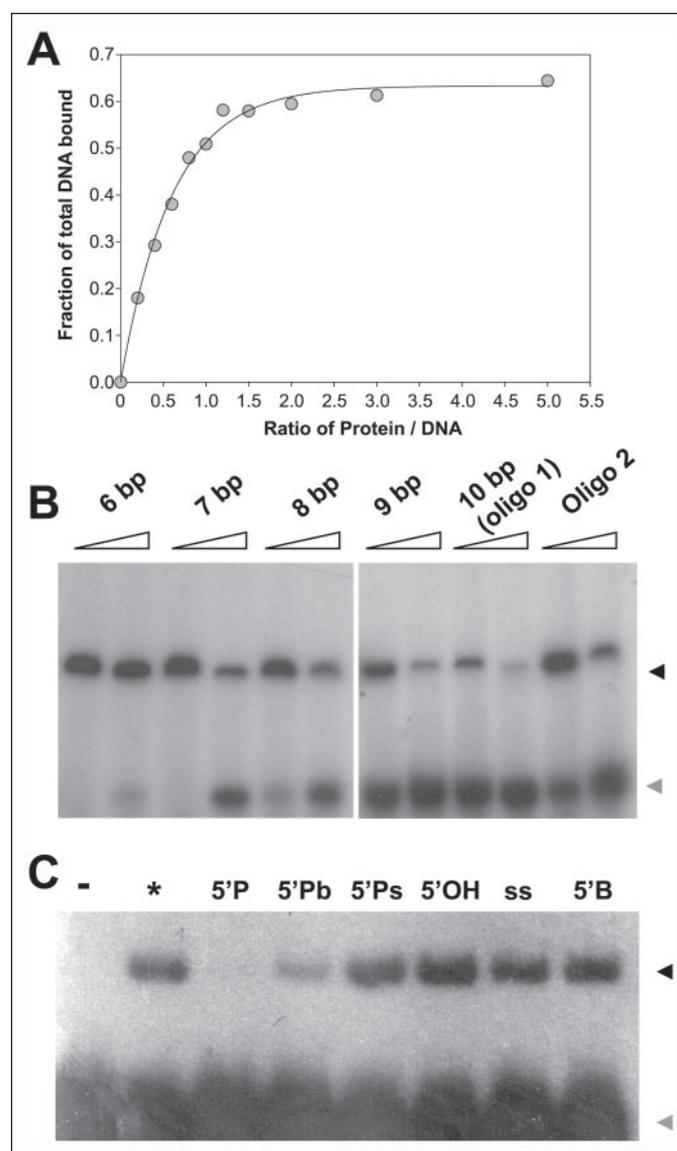

FIGURE 2. **Characterization of the requirements of RFC p140-(375–480) for DNA binding.** *A*, stoichiometric titration of the protein-DNA complex. A constant amount of oligo 1 (10 μM) was titrated with an increasing amount of RFC p140-(375–480), and the resulting DNA-protein complex was separated from the unbound DNA through an analytical gel-filtration column. *B*, competition gel retardation assay to characterize the extent of dsDNA required for binding. 40 fmol of ^{32}P-labeled oligo 1 was mixed with a 10- or 60-fold excess of the indicated unlabeled DNA (see Table 1) and subsequently allowed to bind to 200 fmol of protein. The *black arrow* indicates the position of the DNA-protein complex, and the *gray arrow* indicates that of the unbound DNA. *C*, the competition gel retardation assay was used to delineate the chemical features of DNA required for binding by RFC p140-(375–480). 40 fmol of 5′-^{32}P-labeled oligo 1 was mixed with a 50-fold excess of the competing DNA, and subsequently 200 fmol of protein was added. The competing DNA was a unlabeled oligo 1 (*5′P*), 5′-PO$_4$ blunt end (*5′Pb*) hairpin-oligonucleotide, 5′-PO$_4$ single-stranded (*5′Ps*) oligonucleotide, 5′-OH version of oligo 1 (*5′OH*), a single stranded (*ss*) and a 5′-biotinylated, 20-base-pair, blunt ended oligonucleotide (*5′B*).

determined elements of secondary structure. For comparison, the secondary structure as determined by NMR of the BRCT domain from the bacterial NAD^+-dependent DNA ligase, which is 53% homologous to RFC p140-(403–480), is shown in Fig. 3A. The PSIPRED-predicted secondary structure (not shown) is 99% identical. We also performed a secondary structure prediction on a number of other BRCT-containing proteins including the trans-lesion polymerase Rev1. Rev1 contains a consensus BRCT domain near its N terminus (Fig. 3A). Interestingly, PSIPRED predicts an N-terminal α-helix in a position analogous to that
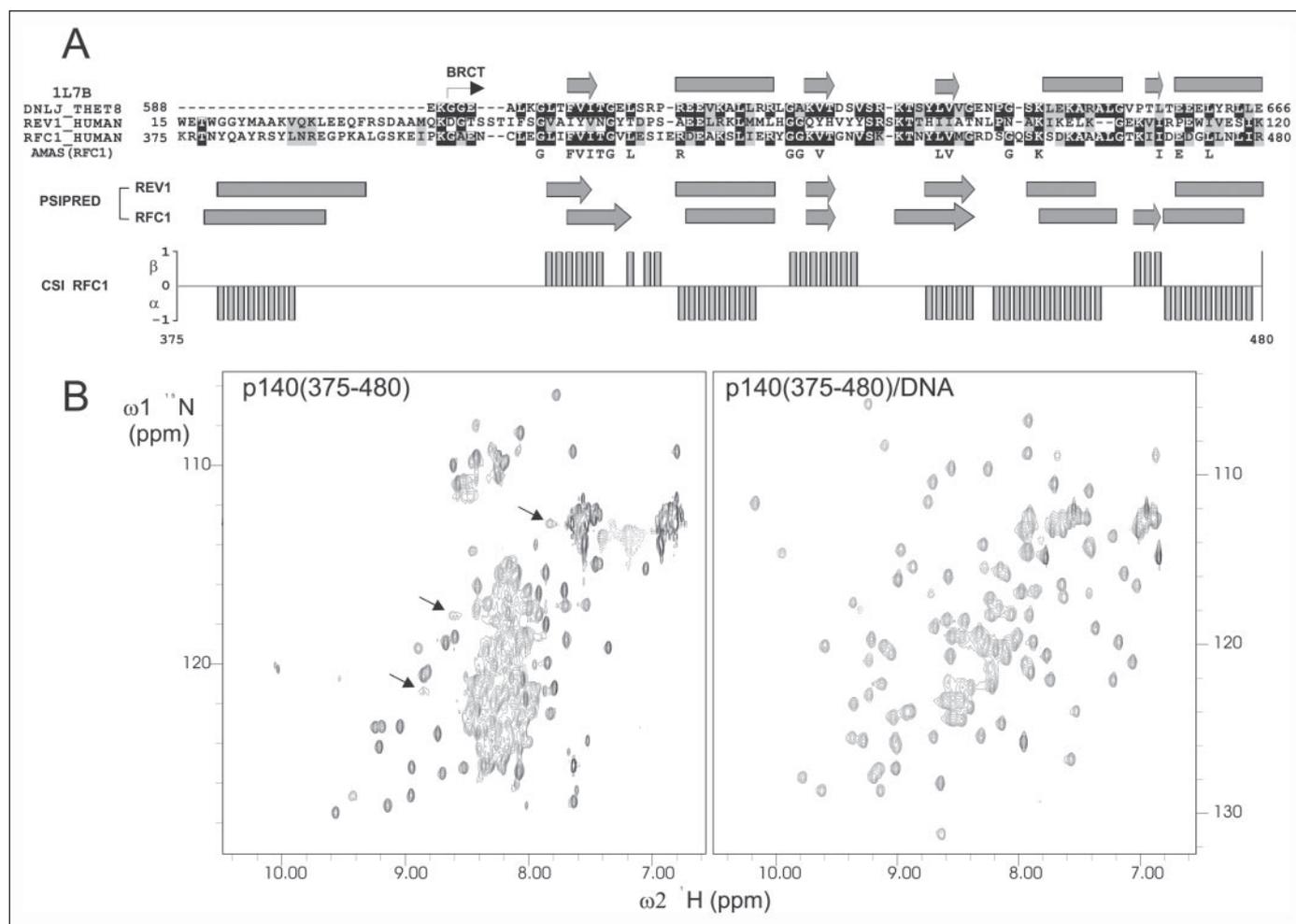
FIGURE 3. **Structural analysis of RFC p140-(375– 480).** *A*, sequence alignment of the BRCT domains of bacterial NAD$^+$-dependent ligase (DNLJ_THET8), human Rev1(REV1_HUMAN), and human RFC p140 (RFC1_HUMAN) was performed using ClustalW. In the alignment, residues with sequence identity >60% or similar properties are shaded *black* or *gray*, respectively. AMAS analysis of the sequence alignment of RFC p140 from 31 eukaryotic species identified only a few highly conserved residues, which are presented in *capital letters* below the alignment. The secondary structure elements, *arrows* for β-strands and *rectangles* for α-helices, are aligned with the corresponding amino acid sequences based on the NMR structure of the BRCT domain from the NAD$^+$-dependent ligase (PDB: 1L7B) or the PSIPRED (26) prediction of Rev1-(15–120) and RFCp140-(375– 480). The consensus secondary structure of hRFC p140-(375– 480) determined by the CSI method was generated from the NMR assignment using the Cα, Cβ, and Hα chemical shifts as input. *B*, two-dimensional [^{15}N,^{1}H] HSQC NMR spectra of RFC p140-(375– 480) without (*left*) and with DNA (oligo 1, *right*). NMR spectra were measured at 298 K at 600 MHz. Examples of some of the resonances with unusually broad lines, indicating intermediate exchange behavior, are indicated by *arrows*.

of Rfc1. Preliminary analysis indicates that the Rev1 BRCT region does indeed bind dsDNA.[5]

NMR is a powerful tool to characterize the solution structure of proteins (36). A significant advantage of NMR is that it can quickly provide information on the folded state of a protein by simple visual inspection of spectra. Further, it is possible to accurately determine the secondary structure of a protein based solely on backbone resonance assignments. Accordingly, we have used ^{13}C,^{15}N-labeled protein in conjunction with standard triple resonance NMR experiments, to determine the essentially complete sequential assignment of the p140-(375– 480)-DNA complex (32) and have obtained a preliminary backbone assignment of the free protein. The NMR data (Fig. 3*B*) suggest that the solution behavior of p140-(375– 480) is very different in the absence or presence of DNA. The two-dimensional [^{15}N,^{1}H] HSQC experiment, which detects the 1 bond correlation within an amide moiety, can be used to determine the folded state of a protein, to a first approximation, by analyzing the dispersion and line width of the peaks in the spectrum. Because one peak should be observed for every non-prolyl residue, the

chemical shift and line width are good indicators if that residue is well structured and/or undergoing conformational or chemical exchange. The spectra in Fig. 3*B* imply that large portions of p140-(375– 480) are conformationally flexible in the absence of DNA as indicated by the poor dispersion of resonances (see the overlapped peaks in the region from 7.8 to 8.6 ppm) and by the presence of very broad peaks indicative of exchange that is intermediate on the NMR time scale (*e.g.* $k_{ex} \sim \Delta\delta$(Hz), indicated by *arrows* in Fig. 3*B*). It is for this reason that the sequential assignment of the isolated protein is incomplete. In contrast, the spectrum of the DNA-bound protein in Fig. 3*B* shows good peak dispersion and more uniform line width, despite the fact that it is part of a complex that is nearly twice as large. These data suggest that the DNA-bound protein experiences reduced dynamic behavior and that likely, a number of amide ^{1}Hs are protected from exchange with the solvent.

Once the sequential chemical shift assignment of a protein is known, the CSI is a useful means of correlating the so-called secondary chemical shift (the difference between the observed chemical shift and that expected for the same residue in random coil) with the secondary structure of a protein (33). We used CSI analysis of the NMR data to deter-

---

[5] M. Kobayashi, F. Figaroa, and G. Siegal, manuscript in preparation.

FIGURE 4. **Three-dimensional model of the sequence conservation in the BRCT domain of RFC p140.** *A*, ribbon presentation of the model BRCT domain. The model was based on the NMR structure of the BRCT domain from the bacterial NAD[+]-dependent DNA ligase (PDB code: 1L7B), which has 53% sequence identity to RFC p140 and was generated using the program 3D-JIGSAW (38). The predicted N-terminal α-helix is schematically shown as a *rectangle*, and the connecting loop (387–402) is presented as a *dotted line*, as there is no available homologous structure. Point mutations created in this study are represented by *blue balls* at the appropriate Cα position. *B*, surface presentation of the homology modeled RFC p140 BRCT domain. The *blue-shaded* areas represent the accessible surface occupied by the conserved amino acids (see Fig. 3) defined from the amino acid sequence alignment using AMAS. The position of other residue mentioned in the text is also indicated.
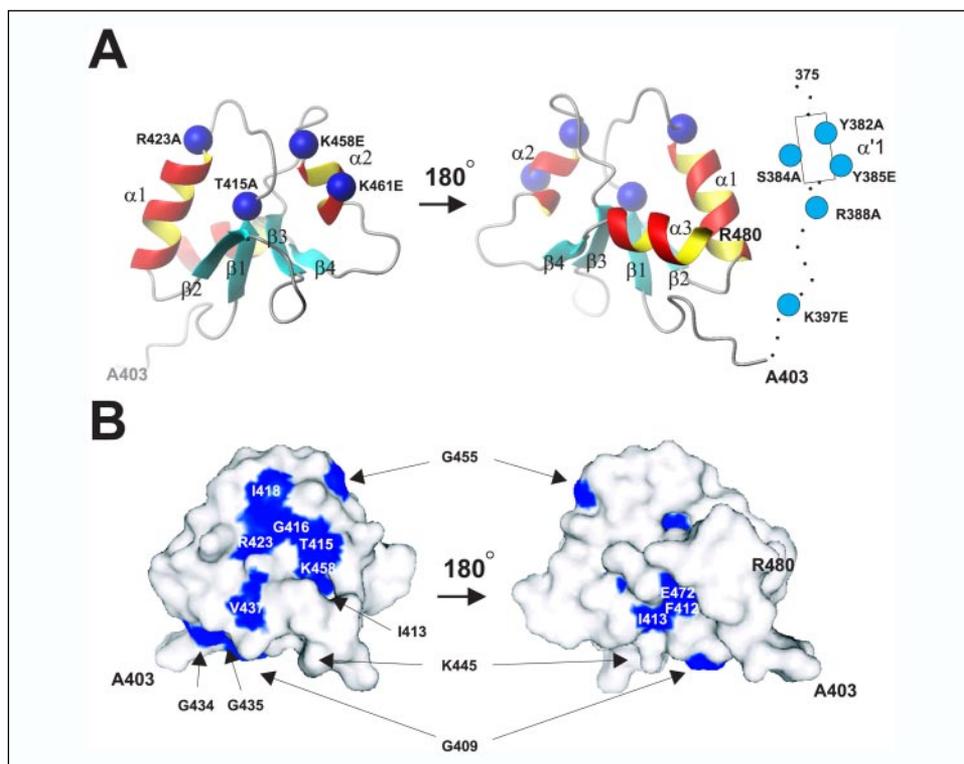
mine the secondary structure of the DNA-bound RFC p140-(375–480) (Fig. 3*A*). The secondary structure of residues 403–480 is consistent with that of other BRCT domains whose three-dimensional structure has been determined and closely matches that predicted by PSIPRED. Importantly, the α-helix between residues 379 and 386 that is predicted by PSIPRED is experimentally confirmed by the CSI analysis of the backbone chemical shifts. Our preliminary backbone resonance assignment of RFC p140-(375–480) in the absence of DNA indicates that the secondary structural elements are retained but they are shorter. This analysis is complicated by the remaining gaps in the sequential assignment of the isolated p140-(375–480). However, the reduced secondary structure content in the free protein suggested by the NMR analysis is supported by the secondary structure prediction based on the CD spectrum of the free p140-(375–480) (Fig. 1*C*), from which we estimate that 26 and 15% of all amino acids are in α-helices or β-strands, respectively. The CD-derived values of the free protein compare with the CSI-predicted values of 31 and of 21% for α-helices and β-strands, respectively, in the DNA complex (Fig. 3).

Sequence Conservation Analysis—To find potential residues involved in the recognition of DNA, a sequence alignment of the BRCT region in the p140 subunit of 31 eukaryotic RFC complexes was analyzed using the program AMAS (37), allowing the identification of a few highly conserved residues within the BRCT domain (Figs. 3*A* and 4). In contrast, the N-terminal sequence of the BRCT region lacked any absolute conservation among all 31 species despite its indispensable nature in DNA binding. A homology model of the RFC BRCT domain (residues 403–480, Fig. 4) was generated by the program 3D-JIGSAW (38) using the available structure of the BRCT domain from the bacterial NAD[+]-dependent DNA ligase (Fig. 3*A*). The resulting model consists of four parallel β-strands (β1–β4) forming the core, which are surrounded by three α-helices (α1–α3) (Fig. 4*A*). As with PSIPRED, the secondary structure predicted by 3D-JIGSAW closely matched that determined by the CSI analysis. The N-terminal sequence between residues 375 and 402 is not included in the model because of the absence of sequence

homology; hence, the N-terminal helix (379–386) and the loop (387–402) are schematically presented as a *rectangle* (α1′) and a *dotted line*, respectively (Fig. 4*A*, *right*). Most of the conserved hydrophobic residues are involved in stabilizing the core created by residues within the parallel β-strands and the α-helices packed against it (12). The structural importance of these conserved hydrophobic residues in the BRCT domain has been demonstrated by the reduced stability of the folded protein when mutated (18). Mapping of the solvent-exposed, conserved residues onto the model structure reveals an interesting pattern where the majority cluster on one molecular surface (Fig. 4*B*). Furthermore, a number of these residues (Thr-415, Arg-423, Lys-458) provide the requisite hydrogen bonding potential to interact with charged backbone phosphate groups on DNA as would be expected for a non-sequence-specific complex (40). The model therefore provides an interesting starting point from which to base experimental analysis of the protein-DNA interface.

The most straightforward approach to determining the involvement of a given amino acid residue in DNA binding is mutagenesis. However, it is important to independently determine that mutagenesis has not severely disrupted the tertiary structure of the protein, merely the desired local properties. Accordingly, based on sequence alignment and the structural model, amino acid residues identified as conserved or with an identity of at least 50% were selected for mutagenesis. Conserved hydrophobic residues contributing purely to the stability of folding were excluded from mutagenesis. The surface-exposed GG repeat (amino acids 434 and 435) (Fig. 4*B*), highly conserved throughout the BRCT family, forms a tight turn between α1 and β2, which can be disrupted by substitution with bulky residues (41). Therefore this sequence was also excluded from mutagenesis. The assumption was made that most residues involved in DNA binding would do so via salt bridge or hydrogen bond interactions, and we therefore selected basic or polar residues for mutagenesis. Based on these considerations four residues (Thr-415, Arg-423, Lys-458, and Lys-461) within the BRCT domain itself were selected for mutagenesis. To probe the DNA binding

FIGURE 5. **Spectroscopic characterization of the folding of mutant RFC p140-(375–480) proteins.** A, one-dimensional ^{1}H NMR spectra were recorded of each purified mutant protein (indicated). Resonances that either disappear or are different from the wt spectrum are indicated with *asterisks*. The spectra are ordered in decreasing similarity to the wt protein (descending top to bottom left and then right). The similarity of the spectrum of each mutant to that of the wt protein indicates a nearly native fold. B, CD spectra of the mutants T415A, R423A, K458E, and K397E compared with wild type (*WT*) RFC p140-(375–480).
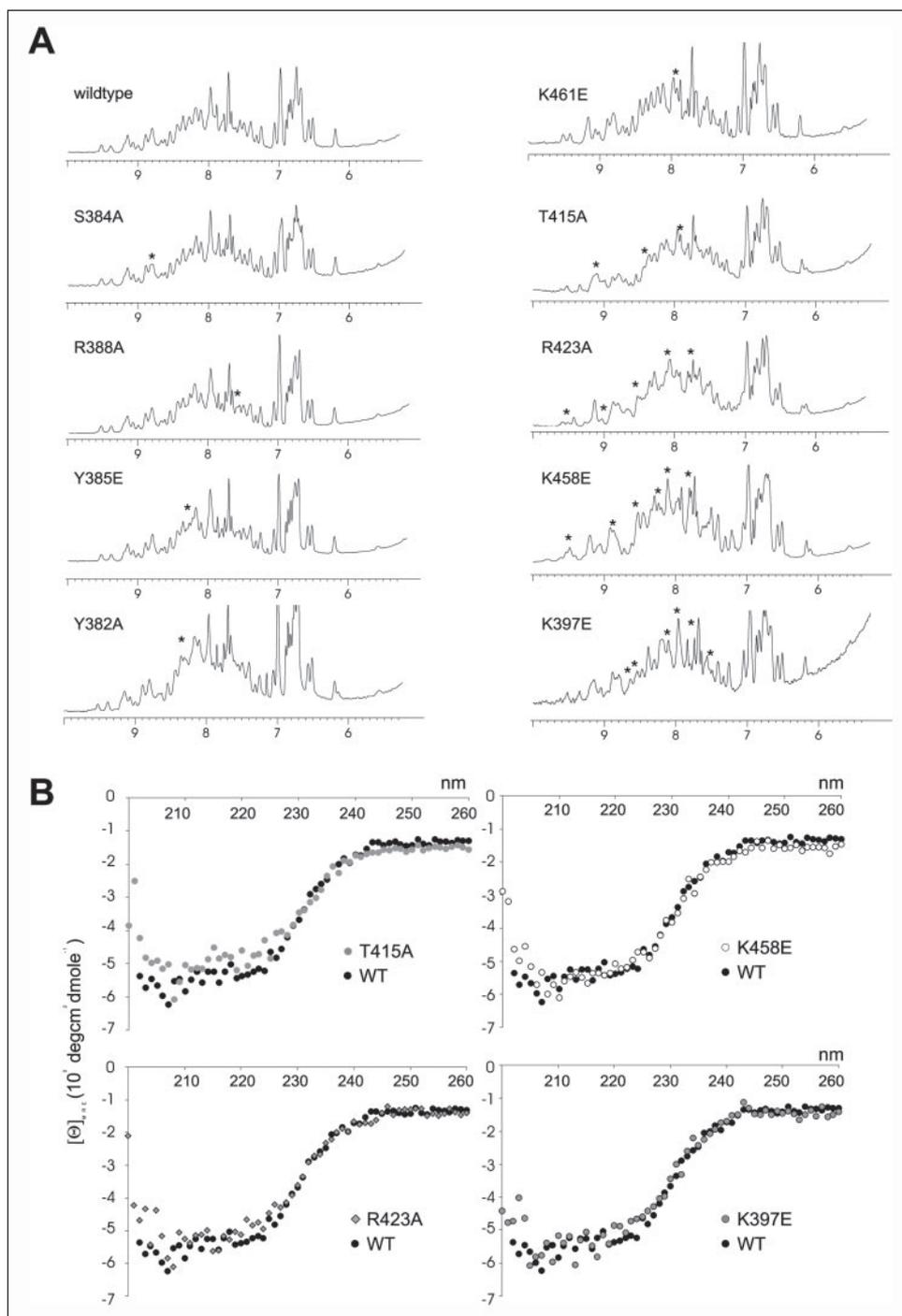
interface of the N-terminal $\alpha1'$-helix, which is clearly important for DNA binding, we also selected three residues (Tyr-382, Ser-384, Tyr-385) as well as two residue within the connecting loop (Arg-388, Lys-397). These residues were preferably substituted with the negatively charged glutamic acid, if the substitution was not expected to disturb the secondary structure as predicted by PSIPRED, otherwise with an alanine. As a result, four residues, Tyr-385, Lys-397, Lys-458, and Lys-461, were substituted with glutamic acid, and five residues, Tyr-382, Ser-384, Arg-388, Thr-415, and Arg-423 were substituted with alanine (indicated as *blue balls* on the structure model in Fig. 4A).

Physical Characterization of Mutated Proteins—The mutated proteins each contained a 6-histidine tag at the C terminus and were puri-

fied using the identical protocol as for the wild type protein. Each mutant protein eluted at the same position as the wild type (wt) protein during gel filtration (not shown) suggesting that the mutations did not result in aggregation or complete unfolding of the protein. The existence of natively folded protein in the purified fraction was further confirmed by one-dimensional ^{1}H NMR spectroscopy comparing the spectral region containing resonances from the backbone amide [1]Hs (7–10 ppm) and aliphatic [1]Hs (0–4 ppm). Fig. 5A presents the downfield portion of the spectrum of each mutant protein in order of decreasing similarity to the wt protein. All spectra are indicative of folded proteins, exhibiting a similar peak dispersion and pattern to the wt protein. Mutants T415A and R423A exhibit increased line width in comparison

to the wt, indicating possible weak aggregation in solution or dynamic behavior that is different from that of wt. There were no significant differences in the resonances arising from aliphatic ^{1}Hs (not shown). The mutants (Thr-415, Arg-423, K397E, and K458E) that displayed spectra with some deviations from the wt protein were further checked using CD spectroscopy to determine the secondary structure content (Fig. 5B). Because of the poor signal to noise ratio of the CD spectra, a quantitative determination of secondary structure was not possible. However, qualitative inspection does not indicate significant differences between wt and mutant proteins. Taken together, the chromatographic, NMR, and CD data all suggest that the mutants are natively folded with only minor structural changes.

Effect of Mutations on DNA Binding—The effect of each mutation on dsDNA binding was assessed using the gel shift assay. To maximize the sensitivity we titrated a constant amount of 5′-^{32}P-labeled, dsDNA with amounts of each mutant protein below and above the previously determined K_D of the wt protein (Fig. 6A). Of all the mutants generated, only S384A exhibited DNA binding equivalent to wt. The mutations, Y385E, R388A, K458E, and K461E, essentially abrogated DNA binding activity. Significant reductions in DNA binding were also observed for the mutations Y382A, T415A, and R423A. As might be expected, the substitution of residues with the negatively charged glutamic acid side chain generally affected DNA binding more drastically than substitution by alanine (Fig. 6B). This observation likely reflects the greater change in surface electrostatic potential caused by incorporation of a charged residue than a neutral residue. The mutagenesis results suggest that one face of the BRCT domain as well as the N-terminal helix and connecting loop directly contact the DNA. To confirm these findings we made two further mutants that were expected to have no impact on DNA binding. Residue Lys-445 is in the "lower" face of the BRCT domain (Fig. 4B), whereas Arg-480 is on the "back" face with respect to the conserved cluster. Because both residues are highly surface-exposed we mutated them to alanine and tested for DNA binding. As demonstrated in Fig. 6C, both mutants bound DNA with affinity similar to the wt protein.

DISCUSSION

Protein Requirements for DNA Binding—Our data clearly show that despite the sequence divergence, amino acids 403–480 of the p140 subunit of human RFC form a consensus BRCT fold. Further, the folded BRCT domain by itself is insufficient for DNA binding. Additional sequences N-terminal to the BRCT are required for DNA binding. Although these additional sequences form in part, an α-helix, it is not clear whether this helix is an integral part of the folded BRCT domain or whether it is isolated from the core domain. It is also clear from the data that the BRCT region is only moderately well structured in the absence of its DNA ligand, but that it becomes more rigid, with a better defined structure, upon DNA binding. Because the protein studied here is removed from its normal context of the complete five-subunit RFC complex it is not possible to say whether the behavior observed *in vitro* is a reflection of the *in vivo* situation. Both biochemical and structural studies, however, suggests that the entire N-terminal half of RFC p140 is only loosely connected to the remainder of the clamp loading complex (28, 29, 42). If so, the *in vitro* observations could well be relevant to the *in vivo* function and may represent some type of regulation. This type of behavior, structural rigidification upon ligand binding, is rather common (43).

Specificity of DNA Binding by p140-(375–480)—Because there appears to be no restriction on the sequences that form the duplex region of bound DNA, it is likely that the majority of protein contacts are made to the backbone and 5′-phosphate of the DNA as observed recently in the structure of a non-sequence-specific protein-DNA complex (40). Further, the protein appears to directly contact both strands of the duplex because 5′-phosphorylated single-stranded DNA bound only very weakly, but more subtly, an unpaired base within the duplex had a significant negative effect on binding. Whether or not the BRCT region actually contacts all 10 base pairs of DNA awaits the elucidation of the structure of the complex. Alternatively, it is possible that stability of the duplex became limiting in our gel shift experiments as the number of base pairs was reduced. This does not seem likely however because the entire assay was performed at 4 °C, a temperature at which even the short 6 bases duplex should be stable in the context of a hairpin oligonucleotide.

Model for the Protein-DNA Interaction—The face of the homology model that contains surface-exposed, conserved residues, also bears a large patch of positive electrostatic potential formed primarily from the side chains of Arg-423, Lys-458, and the backbone amide of Gly-416. Mutagenesis of the conserved, potential hydrogen bonding residues reduces or abrogates DNA binding. Despite the lack of amino acid conservation in the N-terminal sequence of the BRCT region, mutagenesis of one face of helix $\alpha1'$ and charged residues in the loop that connect this helix to the BRCT domain also disrupts DNA binding. The combination of sequence conservation and mutagenesis data suggests a possible mode of DNA binding by the BRCT region. DNA binding, both sequence- and non-sequence-specific, is commonly achieved through interaction of an α-helix with the major groove. In nonspecific interactions, side chains of helical residues interact primarily with the backbone phosphate of the DNA and weakly, typically via bridging waters, to the bases (40). Because there are any number of ways in which such interactions can be achieved, there is little evolutionary force to maintain sequence conservation in a helix that is meant only for nonspecific interaction with the DNA. Thus we propose that helix $\alpha1'$ lies within the major groove of the DNA and provides the majority of the protein-DNA contacts. This model would be consistent with reduced affinity binding to duplex DNA containing unpaired bases, because this would be expected to distort the major groove reducing the ability of the helix to productively interact with both phosphate backbones simultaneously. It seems likely that the binding to the conserved feature on the DNA, the 5′-phosphate, occurs through a conserved motif on the protein. The conserved, positively charged patch containing Thr-415 is clearly the likely site of this interaction. Even if these suppositions are correct, the currently available data cannot provide any insight into the relative orientation of helix $\alpha1'$ and the BRCT domain. Thus a three-dimensional structure of the protein-DNA complex remains an important goal. However, given the quality of the NMR data of the complex (see Ref. 32 for details), a high resolution structure determination may prove to be beyond reach.

Potential Conserved Mechanism—An increasing number of BRCT domains have been reported to bind DNA. Early reports suggested that BRCT domains could bind the termini of DNA (10). It seems likely that this interaction is an artifact of the phosphoserine peptide binding activities more recently attributed to these proteins (6). However, recent reports clearly identify DNA binding by the BRCT domain of bacterial ligases (44), the most closely related sequences to RFC p140. Not only are the residues of the charged patch conserved in these proteins, but mutations of these residues disrupt DNA binding or related functions in the context of the full-length ligase (39). In our secondary structure prediction exercise we found that the trans-lesion polymerase Rev1 contains sequences N-terminal to the BRCT domain that likely form an α-helix in addition to the conserved, positively charged patch. We have cloned and purified a Rev1 protein consisting of the N-terminal α-helix
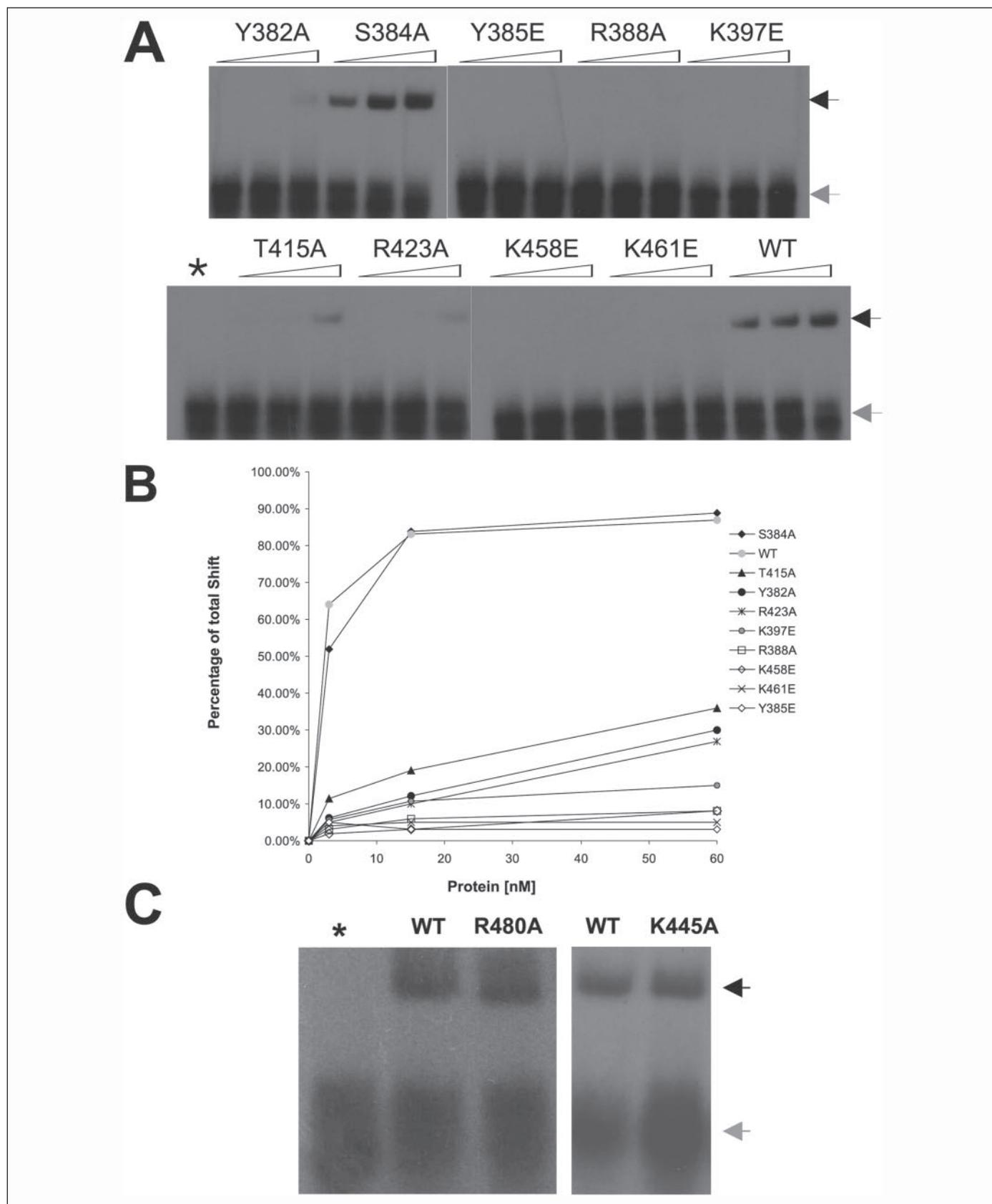
FIGURE 6. **The effect of point mutations on DNA binding by RFC p140-(375– 480).** *A*, the gel retardation assay was used to detect protein-DNA complex formation. 20, 100, or 200 fmol of the indicated protein was added to 20 fmol of ^{32}P-labeled oligo 1. The lane marked with an *asterisk* contained no protein. *B*, quantification of DNA-protein complex formation. The intensity of the bands in *A* was determined using a phosphor imager. The percentage of the total shift (DNA-protein complex, *black arrow* in *A*) was calculated as described under Materials and Methods." *C*, to test the model of DNA binding, two further mutants in p140-(375– 480), R480A and K445A, were generated. 20 fmol of ^{32}P-labeled oligo 1 was added to 100 fmol R480A, K445A, or wt protein and analyzed by gel retardation.

and the BRCT domain. Preliminary experiments indicate that the Rev1 BRCT region does indeed bind DNA. Whether or not the DNA binding is relevant in light of the known phosphoserine peptide binding is not yet known however (8). In any case, it does seem likely that despite the overall poor sequence conservation with the BRCT domain family, the conserved positively charged patch surrounding Thr-415 in human RFC, serves a dual role enabling DNA binding for some BRCT domains and phosphoserine-dependent peptide binding for others.

REFERENCES

1. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) *FASEB J.* **11,** 68–76
2. Callebaut, I., and Mornon, J. P. (1997) *FEBS Lett.* **400,** 25–30
3. Nash, R. A., Caldecott, K. W., Barnes, D. E., and Lindahl, T. (1997) *Biochemistry* **36,** 5207–5211
4. Masson, M., Niedergang, C., Schreiber, V., Muller, S., Menissier-de Murcia, J., and de Murcia, G. (1998) *Mol. Cell Biol.* **18,** 3563–3571
5. Soulier, J., and Lowndes, N. F. (1999) *Curr. Biol.* **9,** 551–554
6. Liu, K., Lin, F. T., Ruppert, J. M., and Lin, W. C. (2003) *Mol. Cell Biol.* **23,** 3287–3304
7. Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) *Science* **302,** 636–639
8. Yu, X., Chini, C. C. S., He, M., Mer, G., and Chen, J. J. (2003) *Science* **302,** 639–642
9. Allen, B. L., Uhlmann, F., Gaur, L. K., Mulder, B. A., Posey, K. L., Jones, L. B., and Hardin, S. H. (1998) *Nucleic Acids Res.* **26,** 3877–3882
10. Yamane, K., and Tsuruo, T. (1999) *Oncogene* **18,** 5194–5203
11. Ma, Y. M., Lu, H. H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K., and Lieber, M. R. (2004) *Mol. Cell* **16,** 701–713
12. Zhang, X., Morera, S., Bates, P. A., Whitehead, P. C., Coffer, A. I., Hainbucher, K., Nash, R. A., Sternberg, M. J., Lindahl, T., and Freemont, P. S. (1998) *EMBO J.* **17,** 6404–6411
13. Williams, R. S., Green, R., and Glover, J. N. (2001) *Nat. Struct. Biol.* **8,** 838–842
14. Derbyshire, D. J., Basu, B. P., Serpell, L. C., Joo, W. S., Date, T., Iwabuchi, K., and Doherty, A. J. (2002) *EMBO J.* **21,** 3863–3872
15. Joo, W. S., Jeffrey, P. D., Cantor, S. B., Finnin, M. S., Livingston, D. M., and Pavletich, N. P. (2002) *Genes Dev.* **16,** 583–593
16. Krishnan, V. V., Thornton, K. H., Thelen, M. P., and Cosman, M. (2001) *Biochemistry* **40,** 13158–13166
17. Lee, J. Y., Chang, C., Song, H. K., Moon, J., Yang, J. K., Kim, H. K., Kwon, S. T., and Suh, S. W. (2000) *EMBO J.* **19,** 1119–1129
18. Dulic, A., Bates, P. A., Zhang, X. D., Martin, S. R., Freemont, P. S., Lindahl, T., and Barnes, D. E. (2001) *Biochemistry* **40,** 5906–5913
19. Clapperton, J. A., Manke, I. A., Lowery, D. M., Ho, T., Haire, L. F., Yaffe, M. B., and Smerdon, S. J. (2004) *Nat. Struct. Mol. Biol.* **11,** 512–518
20. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. M. (2004) *Nat. Struct. Mol. Biol.* **11,** 519–525
21. Shiozaki, E. N., Gu, L. C., Yan, N., and Shi, Y. G. (2004) *Mol. Cell* **14,** 405–412
22. Waga, S., and Stillman, B. (1998) *Annu. Rev. Biochem.* **67,** 721–751
23. Mossi, R., and Hubscher, U. (1998) *Eur. J. Biochem.* **254,** 209–216
24. Reynolds, N., Fantes, P. A., and MacNeill, S. A. (1999) *Nucleic Acids Res.* **27,** 462–469
25. Fotedar, R., Mossi, R., Fitzgerald, P., Rousselle, T., Maga, G., Brickner, H., Messier, H., Kasibhatla, S., Hubscher, U., and Fotedar, A. (1996) *EMBO J.* **15,** 4423–4433
26. Burbelo, P. D., Utani, A., Pan, Z. Q., and Yamada, Y. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90,** 11543–11547
27. Tsurimoto, T., and Stillman, B. (1991) *J. Biol. Chem.* **266,** 1950–1960
28. Gomes, X. V., Gary, S. L., and Burgers, P. M. J. (2000) *J. Biol. Chem.* **275,** 14541–14549
29. Uhlmann, F., Cai, J. S., Gibbs, E., ODonnell, M., and Hurwitz, J. (1997) *J. Biol. Chem.* **272,** 10058–10064
30. Horton, R. M., Cai, Z. L., Ho, S. N., and Pease, L. R. (1990) *BioTechniques* **8,** 528–535
31. Rippe, K. (1997) *B.I.F.Futura* **12,** 20–26
32. Kobayashi, M., and Siegal, G. (2005) *J. Biomol. NMR* **31,** 183–184
33. Wishart, D. S., and Sykes, B. D. (1994) *J. Biomol. NMR* **4,** 171–180
34. Andrade, M. A., Chacon, P., Merelo, J. J., and Moran, F. (1993) *Protein Eng.* **6,** 383–390
35. McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) *Bioinformatics* **16,** 404–405
36. Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, pp. 1–10, John Wiley & Sons, New York
37. Livingstone, C. D., and Barton, G. J. (1993) *Comp. Appl. Biosci.* **9,** 745–756
38. Bates, P. A., Kelley, L. A., MacCallum, R. M., and Sternberg, M. J. (2001) *Proteins* **5,** (suppl.) 39–46
39. Feng, H., Parker, J. M., Lu, J., and Cao, W. G. (2004) *Biochemistry* **43,** 12648–12659
40. Kalodimos, C. G., Biris, N., Bonvin, A. M. J. J., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) *Science* **305,** 386–389
41. Ekblad, C. M. S., Wilkinson, H. R., Schymkowitz, J. W. H., Rousseau, F., Freund, S. M. V., and Itzhaki, L. S. (2002) *J. Mol. Biol.* **320,** 431–442
42. Bowman, G. D., O'Donnell, M., and Kuriyan, J. (2004) *Nature* **429,** 724–730
43. Dyson, H. J., and Wright, P. E. (2002) *Curr. Opin. Struct. Biol.* **12,** 54–60
44. Wilkinson, A., Smith, A., Bullard, D., Lavesa-Curto, M., Sayer, H., Bonner, A., Hemmings, A., and Bowater, R. (2005) *Biochim. Biophys. Acta Proteins Proteomics* **1749,** 113–122 .0