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

Kobayashi, Masakazu

Chapter 5

Structure of the BRCT domain from RFC p140: A model Protein-DNA complex determined by NMR and mutagenesis data[1]

Abstract

Based on the chemical shift assignment of human RFC p140(375-480) in complex with double stranded DNA containing a 5'-recessed phosphate, 3D [^{15}N,^{1}H] NOESY-HSQC and [^{13}C,^{1}H] NOESY-HSQC spectra were assigned using the automated protocol CANDID. The resulting distance restraints and predicted dihedral angle restraints were used as input to CYANA 2.0, to produce an ensemble of 20 structures of the protein moiety, p140(375-480), bound to DNA. The protein consists of a well-defined core, corresponding to a consensus BRCT domain, and an N-terminal α helix, whose spatial orientation with respect to the core of the protein is less well defined. Due to a lack of sequence specific resonance assignments, the NMR data was insufficient to determine the structure of the DNA moiety of the complex. Therefore we used the HADDOCK protocol to dock the protein onto the DNA using ambiguous restraints derived from mutagenesis, amino acid residue conservation and ambiguously assigned intermolecular NOEs. In this model the 5' phosphate interacts with the positively charged surface of the BRCT domain while the N-terminal helix lies in the major groove of the DNA. The model supports previous observations that both the N-terminal sequence and the BRCT domain directly interact with the DNA and explains why no NOEs were observed between these two regions of the protein. Comparison of the 3D structure of the RFC BRCT domain with that of BRCA1 and the NAD^+ dependent DNA ligases reveals a remarkable structural conservation of the phospho-moeity binding residues. The shallow phospho-moiety binding surface may explain why the additional interactions with DNA made by the N-terminal α helix are essential for the stability of the complex.

[1] Parts of this chapter will be used to prepare a manuscript by M. Kobayashi, E. AB, A. M. J. J.Bonvin, and G. Siegal.

Introduction

Replication factor C (RFC) is a five subunit complex, which plays an important role in efficiently loading PCNA onto primer-template DNA during synthesis of the daughter strand in DNA replication (1). Human RFC consists of four subunits of 35-40 kDa and a fifth large subunit (p140) of 140 kDa. The C-terminus of p140 shares homology with the four small subunits, while the unique N-terminal sequence contains a single BRCT domain that is dispensable for its function in PCNA loading (2). The crystal structure of yeast RFC carrying a BRCT-truncated p140 indicated that the five subunits form a spiral complex that precisely matches that of B form DNA (3). Despite the lack of a role in DNA replication, the region including the BRCT domain (subsequently referred to as the BRCT region, residues 375-480) was shown to have binding activity specific for 5' phosphorylated dsDNA (4). Currently there is no structural information available regarding this type of structure-specific DNA recognition.

BRCT domains are small, consisting of roughly 90 amino acids, and are found in more than 900 proteins from all biological kingdoms (5). These proteins, which may bear more than a single copy of the BRCT domain, exhibit a broad range of functional activities in DNA replication, DNA repair and cell-cycle checkpoint regulation (5-7). Structural information is available for the BRCT domains of XRCC1 (8), BRCA1 (9), 53BP1 (10;11), DNA ligase III (12) and the bacterial NAD^+ dependent DNA ligase (13), all displaying a conserved fold. XRCC1 contains two copies of the BRCT domain, of which the C-terminal one forms a hetero-dimer with the BRCT domain of DNA ligase III through residues conserved between the two domains (14). BRCA1 also contains two BRCT domains, eliciting a very different function: they form an obligate paired structural unit that specifically binds to a phospho-serine containing sequence in the protein BACH1 (15) and CtIP (16). 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.

Although a limited number of BRCT-DNA interactions are known or have been implied from biochemical data, there is at present no structure of a BRCT-DNA complex. Deletion and mutagenesis data (Figures 2.2 and 3.5) suggest that the region spanning residues 375-480 in RFC p140 (hereafter called p140 (375-480)) is important for DNA binding. In order to investigate the structural basis of this recognition, we employed NMR

methods to determine the structure of p140 (375-480) bound to dsDNA. Despite numerous efforts, the data obtained were not sufficient to determine the solution structure of the DNA portion of the complex, therefore only the structure of the protein in the complex was determined from experimentally derived restraints. The resulting structure of p140(375-480) consists of a consensus BRCT fold preceded by an α-helix that connects to the core domain by a long loop. A model of the protein-DNA complex was generated using HADDOCK (17), an algorithm that docks two molecules using ambiguous interaction restraints based on a variety of experimental data including mutagenesis, ambiguously assigned intermolecular NOEs and amino acid conservation.

Materials and Methods

*Expression and purification of RFC p140 (*375-480)

The expression and purification of RFC p140 (375-480) were performed as described in the Materials and Methods in Chapter 4.

Preparation of DNA

The oligonucleotide of sequence pCTCGAGGTCGTCATCGACCTCGAGATCA was produced by standard solid state synthesis. The synthesized DNA was dissolved in 0.1 M NaOH and applied to a Q-sepharose column (Amersham Bioscience), also pre-equilibrated with 0.1 M NaOH. The DNA was eluted by increasing concentrations of NaCl in the same buffer. The volume of the collected peak fraction was reduced by rotary evaporator and the buffer was exchanged to 25 mM Tris-HCl pH 7.5, 50mM NaCl by PD10 desalting column (Amersham Bioscience). The purity of the DNA was analyzed by MS.

Protein-DNA complex preparations

Both the protein and the DNA solutions were diluted in 25mM Tris-HCl pH7.5, 50mM NaCl and 1mM DTT to 10 μM in order to prevent aggregation, mixed in the molar ratio of 1 to 1.2 and concentrated to approximately 0.5 mM by vacuum dialysis (Spectrum Labs) using a 10 kDa cut-off membrane. Subsequently, the buffer was exchanged to 25 mM D_{11}-Tris-HCl pH 7.5, 5mM NaCl in 95/5 H_2O/D_2O.

NMR spectroscopy

All NMR data were acquired at 298 K on a Bruker DMX600 spectrometer. Most of the sequential assignments for the backbone were obtained using 3D HNCACB,

CBCA(CO)NH and HBHA(CO)NH spectra. Aliphatic side-chain resonances were derived from 3D HCCH-TOCSY and CCH-TOCSY spectra. Additional data provided by 2D [^{1}H,^{1}H] NOESY (150 ms), 3D [^{15}N,^{1}H] NOESY-HSQC (150 ms) and [^{13}C,^{1}H] NOESY-HSQC experiments (150 ms) were used for further assignment of aromatic side-chain resonances as well as confirmation of the through-bond data. An additional 3D [^{15}N,^{1}H] NOESY-HSQC was recorded at 310 K for structure calculation. Spectral data were processed using NMRPipe (18).

The following half- and double- filtered experiments were acquired: a 2D NOESY (τ_m = 150 ms) recorded at 900 MHz with HMQC purge set to reject ^{13}C- and ^{15}N- coupled protons during t_1 and to accept ^{13}C- and ^{15}N- coupled protons during t_2, and a 2D NOESY (τ_m = 150 ms) run at 900M Hz with HMQC purge set to reject ^{13}C- and ^{15}N- coupled protons during both t_1 and t_2. Details of the pulse sequences used are given in the Supplementary Materials S2 and S3.

Resonance assignment

The assignment and the integration of NOE peaks was performed using the computer program CARA (19) available at http://www.nmr.ch) The majority of the chemical shift assignments of the protein bound to the DNA were obtained by comparing the data from the through-bond coupling experiment 3D CCH-TOCSY to the 3D [^{15}N,^{1}H] NOESY-HSQC. Approximately 83 % of the Cα-Hα and Cβ- Hα correlations were missing in the [^{13}C,^{1}H] NOESY-HSQC. The chemical shift assignments of the protein bound to DNA have been reported (20) and deposited (BMRB accession number 6353).

Spectra from through-bond coupling experiments contained substantially fewer peaks in the case of free protein than in that of the complex, at the same time approximately 43 % of the amide backbone correlations were missing in the [^{15}N,^{1}H] NOESY-HSQC in the free protein, presumably due to rapid exchange with water. Therefore no reliable chemical shift assignments were obtained for the free protein.

Structure Calculations

Distance restraints were derived from the automated NOE crosspeak assignment of the 3D [^{15}N,^{1}H] NOESY-HSQ (recorded at 310 K) and the [^{13}C,^{1}H] NOESY-HSQC (recorded at 298 k) using the algorithm CANDID implemented in the computer program CYANA 2.0 (21). The chemical shift tolerances used in the automated assignment were 0.02 ppm for protons and 0.1 ppm for heavy atoms. Structure calculations were performed

with CYANA 2.0 (22). The structures were calculated using the NOE derived distance restraints and the dihedral angle restraints calculated from the chemical shift values of Cα and Cβ by TALOS (23). One hundred structures were calculated starting from conformers with random dihedral angles and using simulated annealing and torsion angle dynamics (TAD) as implemented in CYANA 2.0. The distance restraints which resulted in the 20 structures with the lowest CYANA 2.0 target function were converted to CNS format. Since a water refinement protocol is not available for CYANA, the water-refinement protocol integrated in CNS 1.0 was used. The structures were then recalculated with the distance restraints and dihedral angle restraints using the standard simulated annealing protocol in the computer program CNS 1.0 (24). The 20 lowest-energy structures with no distance violations greater than 0.3 Å and no angle violations greater than 5° were subjected to water-refinement following the scheme described (25). The 20-lowest energy structures with no NOE violation greater than 0.3 Å and no angle violations greater than 5° were accepted as the final structures representing the solution conformation. The quality of the structures was assessed using the program PROCHECK (26).

Docking protocol

The docking was performed with HADDOCK (17) and the average relative solvent accessible surface for each residue was calculated using NACCESS (http://wolf.bms.umist.ac.uk/naccess). In HADDOCK, the docking is driven by ambiguous intermolecular restraints (AIRs) which are defined according to established criteria of the program (17), The "active" residues are either those that have been shown by mutations to abolish or significantly perturb complex formation with average solvent accessible surface above 50 %, or those with observable intermolecular NOE to the dsDNA. In this case, the counterpart “passive residues” correspond to the nucleotide residues of the dsDNA. As a result, 8 amino acids of p140(375-480) (Y382, Y385, R388, T415, G415, R423, K458, and K461) with > 50% solvent accessibility were defined as “active” residues. Ambiguous interaction restraints (AIRs) were then generated between the active residues of p140(375-480) and the passive residues of dsDNA (Table 5.4). A more specific AIR restraint was introduced between T415 and the 5’ phosphate group of dsDNA based on its conservation in the N-terminal BRCT domain of BRCA1 which is known to form hydrogen bonds to a phosphate. Side chain and backbone flexibility that allow local rearrangement during the semi-flexible simulated annealing step of HADDOCK, are confined to the segments around

the active and passive residues , that is residues 377-392 and 414-462 in p140(375-480) and the entire DNA molecule (Table 5.4).

The starting structures for docking were the 20 NMR structures of p140(375-480) and 3 models of dsDNA. Since no structure of the DNA portion of the complex is available, a model structure of 5' phosphorylated dsDNA with a 3' single stranded overhang in the standard B-form DNA with 3 conformations was generated, using the sequence of an oligonucleotide identical to that used for the NMR studies except for the fact that it did not contain a hairpin (5'pCTCGAGGTCG3'/5'CGACCTCGAGATCA3'). Docking of the p140(375-480)-dsDNA complex was performed following the protocol of HADDOCK1.3 (17). Inter- and intramolecular energies are evaluated using full electrostatic and van der Waal's energy terms with an distance cutoff using OPLS nonbonded parameters as defined in the default protocol (17). During the rigid body energy minimization, 2400 docking structures were generated (4 cycles of orientational optimization for each combination of starting structures were repeated 10 times). The best 200 structures in terms of intermolecular energies were then used for the semi-flexible simulated annealing, followed by explicit water-refinement. Finally the structures were clustered using a 5 Å RMSD as a cut-off based on the pairwise backbone RMSD.

Analysis of intermolecular contacts

Intermolecular contacts (hydrogen bonds and nonbonded contacts) in the ensemble of five best complex structures from the clusters with the lowest HADDOCK score were analyzed with the NBPLUS which is the part of NUCPLOT software (27). Used settings for hydrogen bonds were the distance cut-offs of < 2.7 Å and < 3.35 Å respectively for proton -acceptor (H-A) and proton donor –acceptor(D-A) provided that the D-H-A angle and H-A-AA are > 90°, where AA is the atom attached to the acceptor (for more details see (27)). 3.9 Å was the distance cut-off for nonbonded contacts.

Results

Preparation of the protein-DNA complex

The oligonucleotide used to form the protein-DNA complex was a 5' phosphorylated, 28-mer containing two self-annealing 10 nucleotide sequences separated by a 4 base hairpin turn. Upon annealing the oligonucleotide contains a single binding site for RFC p140 (375-480) with a 4 nucleotide long, 3' single stranded tail. This

oligonucleotide, has been shown by electrophoresis mobility shift assays (Figure 2.3) to bind RFC p140 (375-480) with K_D ~ 10 nM. Since mixing of protein and DNA at high concentrations (> 0.1 mM) resulted in severe protein precipitation, complex formation was performed under dilute conditions (range 5-10 μM of each constituent). The complex was then concentrated to 0.5 mM by vacuum dialysis using a 10 kDa cut-off membrane. The starting protein/DNA ratio was 1:1.2 to ensure the formation of a full complex with a 1:1 stoichiometry (Figure 2.2). Excess DNA eluted through the dialysis membrane. No signals from unbound protein could be detected in the NMR spectra.

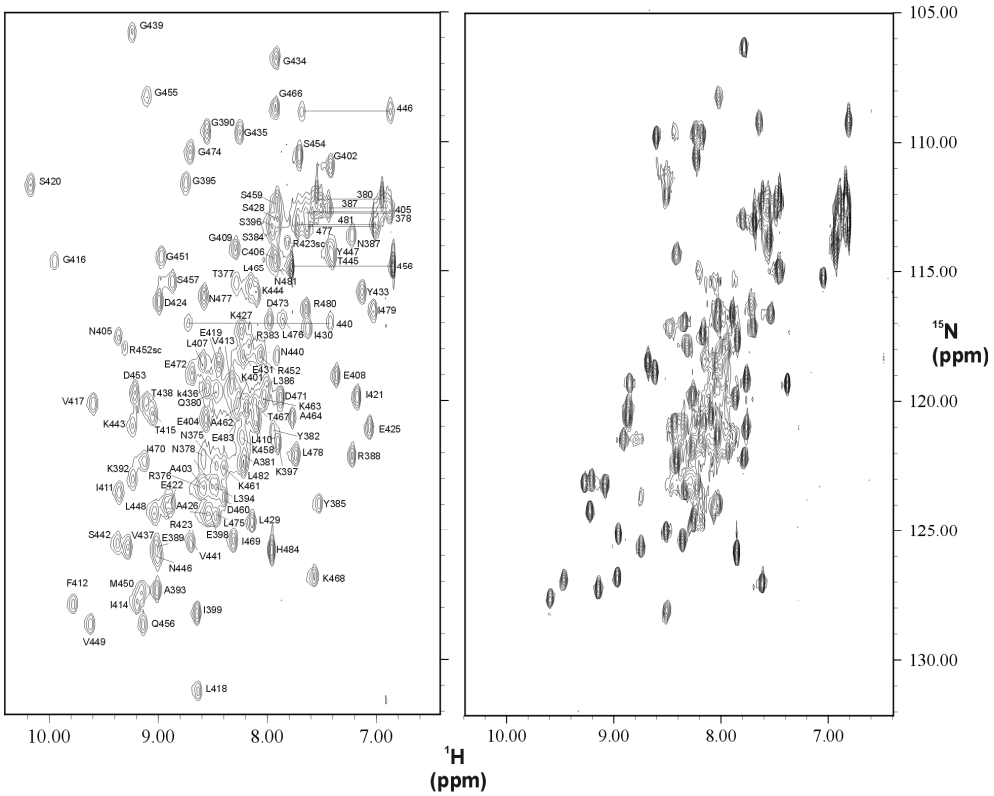
NMR spectroscopy and Resonance Assignment

Despite the moderate quality of the NMR data (Figure 4.1), we obtained over 90% of ^{1}H, ^{13}C and ^{15}N chemical shift assignments for the observable resonances of RFC p140(375-480) bound to the oligonucleotide (20). Approximately 83% of the Cα-Hα and β- Hα correlations were missing from the [^{13}C,^{1}H] NOESY-HSQC. A likely explanation is highly efficient relaxation due to dynamic behavior intermediate on the NMR time scale within the complex. The majority of the chemical shift assignments were obtained by analysis of data from the through-bond experiments HNCABC, CACBCONH, HCCH/CCH-TOCSY while the 3D [^{15}N,^{1}H] NOESY-HSQC supplied confirmatory and supplemental correlations. 99% of the backbone assignments were determined, with the only missing residues being P391, P400 and Y379. As typically is the case in NMR, the N-terminal residue, here methionine, was also not observed. The sequential assignment is available from the BMRB (http://www.bmrb.wisc.edu/) under the accession number 6353.

RFC p140(375-480) is rich in basic residues (11 lysines and 7 arginines). Many of the side chain carbon and proton resonances from the basic residues suffered from severe overlap in the 3D HCCH/CCH-TOCSY and 3D [^{13}C, ^{1}H] NOESY-HSQC spectra. The side chain Cε/Hε correlation of 9 lysines (residue numbers 375, 392, 401, 443, 444, 427, 436, 461 and 468) resonated very close to the random coil values, suggesting high mobility. The resonances of the side chains of K397 and K458 were mostly absent from the CCH-TOCSY and [^{13}C,^{1}H] NOESY-HSQC and only some of the resonances were assignable using the 3D [^{15}N,^{1}H] NOESY-HSQC and 2D NOESY spectra. Similarly, the side chain Cδ/Hδ correlation of R376 and R480 resonated close to their random coil values, while those of R383 and R432 had unique chemical shift values for Hδ. Much of the resonances

of R388, R423 and R452 were missing from both CCH-TOCSY and [^{13}C,^{1}H] NOESY-HSQC and were therefore assigned using [^{15}N,^{1}H] NOESY-HSQC and 2D NOESY spectra. The ^{1}H and ^{15}N resonances of the side chain HεNε correlation of R423 and R452 were assigned using the [^{15}N,^{1}H] NOESY-HSQC spectrum.

The [^{15}N,^{1}H] HSQC spectrum of the free protein was poorly dispersed (Figure 5.1) and exhibited heterogeneous linewidth and intensity, all characteristics of a flexible and dynamic molecule. A considerable number of the expected peaks were missing from the HNCACB spectrum and 40% of the backbone amide correlations were missing from the [^{15}N,^{1}H] NOESY-HSQC spectrum. Presumably the amide protons are in rapid exchange with those of water. The completed chemical shift assignment of the free protein is therefore not currently available. However, analysis of the partial sequential assignment and circular dichroism spectroscopy data (Figure 2.1) suggested the existence of secondary structure elements. The spectrum of the DNA-bound protein was clearly better than that of the free form (Figure 5.1). The spectrum of the complex contained 105 of the 106 expected amide correlations and exhibited good dispersion with more homogeneous linewidths. The observation of two distinct sets of resonances for the bound and the free protein is characteristic of slow dissociation of the Protein-DNA complex on the NMR timescale. As a consequence, it was not possible to deduce the DNA binding site on the protein by chemical shift perturbation analysis.



(Figure 5.1) [^{15}N,^{1}H] HSQC spectra of the protein-DNA complex (left) and of the free protein (right). The amide backbone assignments are plotted on the [^{15}N,^{1}H] HSQC spectra of the protein-DNA complex. Only 5% of the total amide correlations overlap in the complex and free-protein spectra. All the side chain amides Gln and Asn were identified and are connected by lines. The side chain guanidinium resonances (folded) are indicated as "R423sc" and "R452sc"

Structure calculation of p140(375-480) when bound to DNA

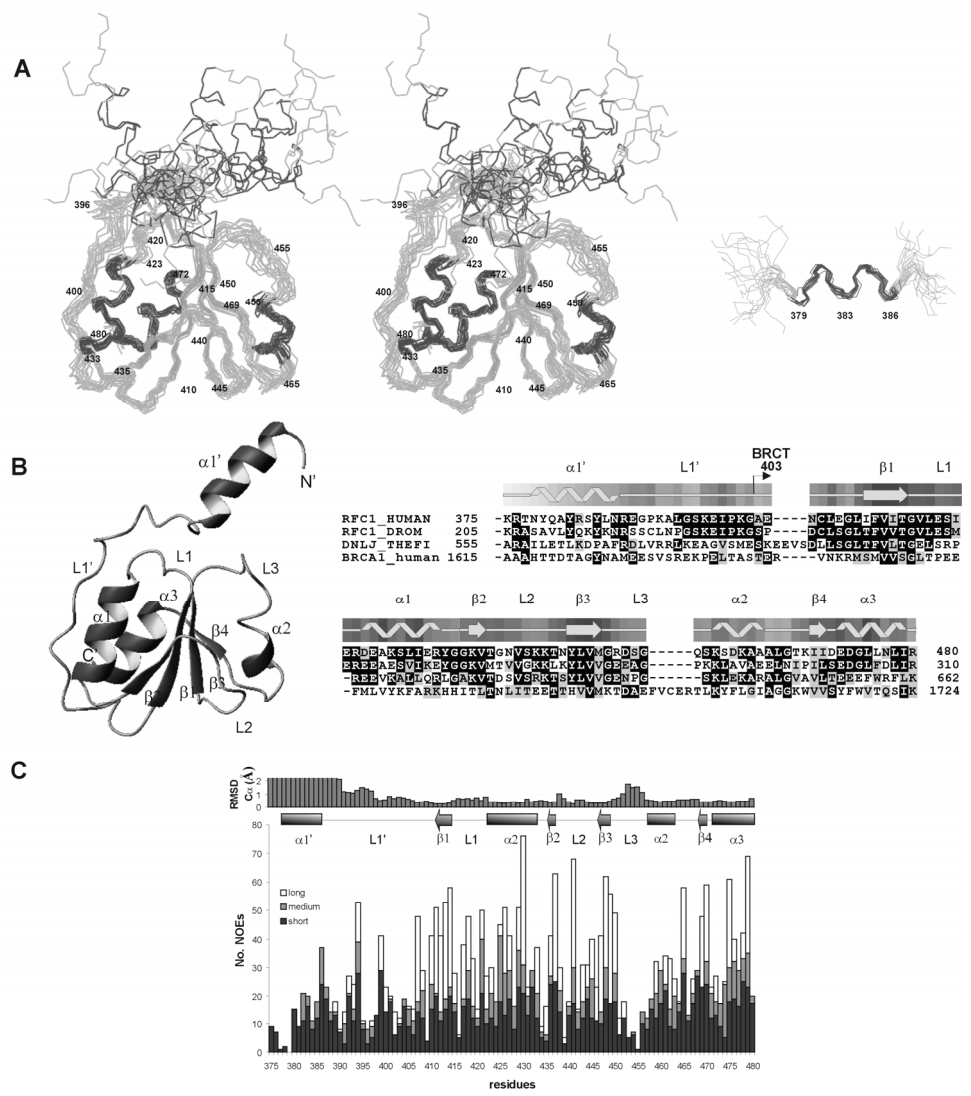
The structure of the protein moiety of the complex was determined primarily from distance restraints derived from NOEs in the 3D [^{15}N,^{1}H] NOESY-HSQC and 3D [^{13}C,^{1}H] NOESY-HSQC spectra as described in Materials and Methods. The best-fit superposition of the 20-lowest energy conformers is depicted in Figure 5.2A and the quality statistics of the structures are summarized in Table 5.1. With the exception of the N-terminal 16 residues (375-389), which were disordered with respect to the rest of the protein, and residues 451-457 in the loop 3 (L3), the backbone of p140 (375-480) is moderately well defined, with an average RMSD of 1.07 (± 0.17) Å for backbone atoms and 1.67 (± 0.23) Å for all heavy atoms (Table 5.1). The moderate definition of the structures reflects the low

number of long range distance restraints used in the structure calculation, which amounts to less than 5 restraints per residue (Table 5.1). None of the distances or dihedral angle violations was greater than 0.3 Å or 5.0º respectively in the final 20 structures. Analysis of the Ramachandran plot for all residues, using the computer program PROCHECK (26), showed that 83% of φ and ψ angles lie within the most favored and 14.8% lie in the additionally allowed regions, while only 2.4 % are in the generously allowed or disallowed regions (Table 5.1). The residues that fall into the latter regions are mostly found in loops. The 3D structures of a number of other BRCT domains have been published (8-13) and the fold is highly conserved and consists of the following secondary structure elements: β1-α1-β2-β3-α2-β4-α3. The various structural elements of p140(375-480) have been labeled as α(helix), β(sheet) and L(loop) for consistency with the reported structures of other BRCT domains and are depicted in Figure 5.2B.

(Table 5.1) Summary of restraints and structural statistics for the final RFCp140 (375-480) ensemble

Restraints used in the calculation		
total number of NOE upper distance limits :		1782
intraresidue and sequential ($\|i-j\| \leq 1$)		1003
medium-range ($1 < \|i-j\| < 5$)		283
long-range ($\|i-j\| \geq 5$)		496
total number of dihedral angle restraints		
predicted (TALOS)		36
CYANA 2.0 outputs (20 structures)		
target function value ($Å^2$)		0.85
number of distance restrain violations (> 0.2 Å)		1
number of dihedral angle constrain violations (> 5 °)		0
CNS 1.0/water-refinement (20 structures)		
number of distance restrain violations (> 0.3 Å)		0
number of dihedral angle constrain violations (> 5 °)		0
Structure statistics		
Final energies	(kcal/mol)	
total	(kcal/mol)	-3830
bonds	(kcal/mol)	42.5
angles	(kcal/mol)	151
improper	(kcal/mol)	54.4
dihedral	(kcal/mol)	485
van der Waals	(kcal/mol)	-430
electrostatic	(kcal/mol)	-4130
NOE	(kcal/mol)	0.351
RMS deviation from ideal values		
Bonds	(Å)	0.010
angles	(°)	1.25
PROCHECK Ramachandran plot analysis (375-480)		
Residues in most favored region (%)		83.0
Residues in additionally allowed region (%)		14.8
Residues in generously allowed region (%)		1.3
Residues in disallowed region (%)		0.9
RMSD to the averaged coordinates*		391-480
Backbone atoms		0.91 ± 0.16
Heavy atoms		1.44 ± 0.20

* RMSD to the average coordinates of the 20 structures calculated excluding residues 375-389 and 451-457.



(Figure 5.2) Structure of RFC p140(375-480) when bound to dsDNA. **(A)** Left**;** Stereoview of a superposition of the backbone (N, Cα and C') atoms for the 20 lowest-energy structures of RFC p140(375-480). α-helices and β-strands are colored in red and cyan respectively. Right; superposition of the backbone of residues 379-386 demonstrating the well defined helix α1' the N-terminus. **(B)** On the left, sequence alignment of the p140(375-480) with the homologous region of RFC p140 from *Drosophila melanogaster* (RFC1_DROM), NAD^+ dependent ligase of *Thermus filiformis* (DNLJ_THEFI) and human BRCA1 BRCT-n. The secondary structure of human RFC p140(375-480) was determined by PROCHECK-NMR(26) and the level of solvent accessibility of the residues is colored from light (accessible) to dark blue (not accessible). The loop regions are labeled as L1', L1, L2 and L3. On the left a ribbon diagram of the p140(375-480) structure with the lowest-energy is presented. The orientation is equivalent to (A). **(C)** The number of short, medium and long range distance restraints per residue are plotted along with the RMSD (root mean square deviation) values of the Cα atoms. The RMSD values were determined by superimposing the backbone atoms of each member of the ensemble of 20 conformers that form the BRCT domain itself (residues 403-480). Distant restraints are intra-residue NOEs (|i - j|< 1), inter-residue medium NOEs (1 <|i - j|< 5) and long-range NOEs (|i - j| > 5). Figures (A) and (B) (right) were generated using MOLMOL (28)

Structure description

Residues 403-480 of RFC p140 (375-480), which form the conserved BRCT domain, fold into a compact unit which consists of four parallel β-strands surrounded by helices α1 and α3 on one side and by helix α2 on the other (Figure 5.2A). With the β1α1β2β3α2β4α3 topology, the 3D fold of residues 403-480 (Figure 5.2B) closely resembles that of BRCT domains whose 3D structure has previously been determined. The packing of the parallel β-strands forming the central core of the BRCT domain is stabilized mainly by hydrophobic interactions between I411, F412, V413, I414 (β1), and Y447, L448, V449 (β3). These strands are further surrounded by V437 of β2 and I469-470 of β4. The packing of helices α1 and α3 against the central β-sheet occurs via hydrophobic interactions of A426, L430 (helix α1) and L475, L478 (helix α3) against the hydrophobic core. The BRCT domain of p140(375-480) belongs to a distinct subclass of the BRCT superfamily due to its amino acid sequence divergence from the rest of the family (5). One such difference from the rest of the superfamily is the substitution of the conserved tryptophan (α3) by G474 in RFC p140(375-480), even though the amino acid substitution of W74 (α3) to L in the XRCC1 BRCT revealed its significant contribution to the stability of the BRCT fold (14). The amino acid sequence identity among the BRCT homologs is very low (~14 %) (5), however the residues involved in hydrophobic interactions that make up the core of the protein are well conserved since substitution, when it occurs, is by a similar residue (Figure 3.1) (5). This pattern of conservation suggests that BRCTs from different species all adopt a similar, canonical three dimensional structure. Two of the most conserved residues in the BRCT superfamily are G434 and G435, which form a tight turn between α1 and β2 (5). The substitution of any of these glycines by a larger residue could potentially destabilize the 3D structure. In the case of BRCA1 for instance, the G1788V mutation renders the tandem BRCT repeat more sensitive to proteolytic digestion (29). Likewise, the G617I mutation reduces both DNA binding and nick-adenylation activity of the bacterial NAD^+ dependent DNA ligase(30) . Interestingly, the G193R mutation in the BRCT domain of Rev1 has been shown to interfere with the *in vivo* translesion bypass activity of REV1 in *S. cerevisiae* (31). This structural analysis suggests that the Rev1 G193R mutant results in decreased stability of the fold. The analogous G435R mutation in human RFC p140(375-480) resulted in a mutant prone to precipitation and with reduced DNA binding activity (Figure 2.2), both characteristics suggestive of a decrease in ΔG_{fold}.

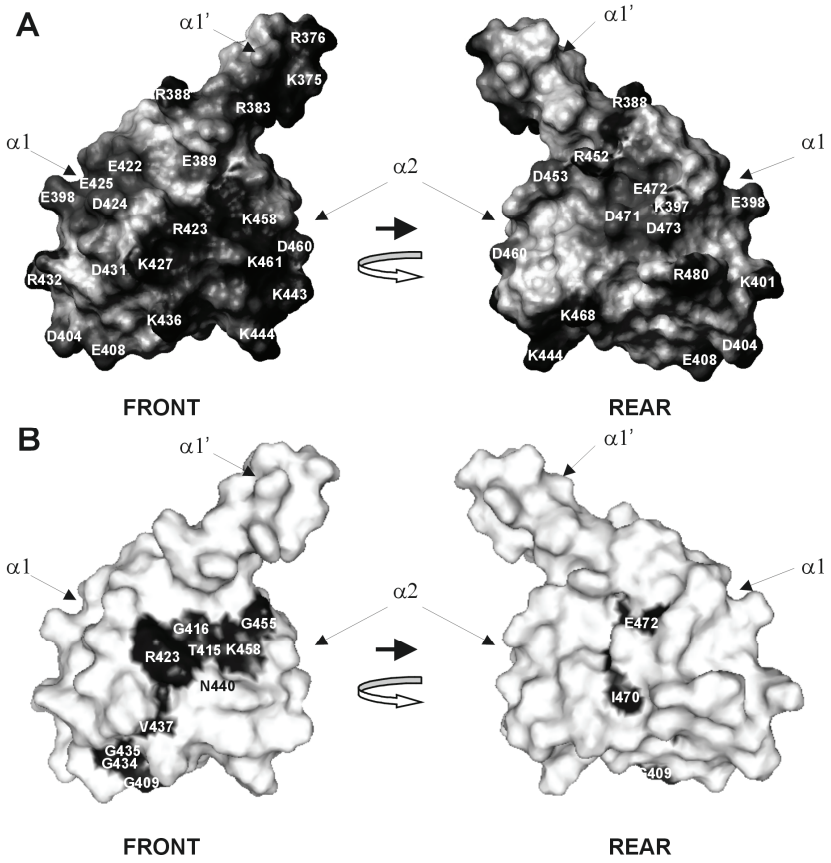
Helix α2 is formed by residues K458-A463 (Figure 5.2B), which together with the preceding loop L3 are the most variable in size and sequence in the BRCT family. For instance helix α2 is completely replaced by an extended L3 loop in the BRCT domain of human DNA ligase III (12). In the present structure, the L3 loop displays a high degree of disorder (Figure 5.1A) due to the limited number of restraints available within this region. It is not yet clear whether the disorder reflects actual dynamic motions within the L3 loop or simply a paucity of structural restraints (Figure 5.2C). To a lesser extent, loops L1, L2 and L4 display some conformational variation in the ensemble (Figure 5.2A and C). In most BRCT domains loop L1 appears to be more or less flexible as reflected by the high B-factors in X-ray crystal structures and poor definition in NMR structures (8;9;12;32). In relation to these other structures, the L1 loop is better defined and buried under loop L1' in the structure of RFC p140 (375-480) (Figure 5.2A and C).

The N-terminus of p140 (375-480), residues 375-403, forms an α-helix (α1') which is separated by a loop (L1') from the core of the protein. Loop L1' packs against helices α1 and α3 of the conserved BRCT domain. Helix α1' (residues 379-386) appears consistently in all 20 structures (Figure 2A, right), however it is poorly defined with respect to the rest of the protein (Figure 2A). This lack of definition certainly reflects the absence of observable long range NOEs between helix α1' and the core of the protein (Figure 5.2C). The loop L1' is anchored to helices α1 and α2 through burying of the side chains of residues L407, P400 and L399 between the two helices, and through potential salt-bridging between the side chains of K397 (L1') and E472 or D473 (α3), and of K392 (L1') and E419 (L1).

A potential DNA interaction site

The electrostatic potential at the surface of RFC p140(375-480) was calculated using the computer program MolMol (28) and is presented in Figure 5.3A. Casual inspection of Figure 5.3A reveals a large basic patch, composed of R423, K427, K436, K443, K444, K458 and K461 ("front" in Figure 5.3A), which might be a binding site for the negatively charged phosphate backbone of the DNA. Based on the amino acid sequence alignment (Figure 3.1), the surface exposed conserved residues of a number of orthologs of RFC p140 are mapped onto the surface of the p140(375-480) structure (Figure 5.3 B). The conserved residues are distributed mainly on the basic patch ("front" in Figure

5.3) of p140 (375-480) and localized within the BRCT domain rather than within the loop L1' or helix α1'. Comparison of Figures 3A and B indicates that the highly conserved residues R423 and K458 form part of the basic patch. Furthermore in the vicinity of this region, the highly conserved residues T415, G416 and G455 can also be found. Negatively charged surfaces, on the other hand, extend from the front to the "rear" of the molecule (Figure 5.3). A large patch of negative charge is located along helix α1 (423-432) on the front side while the conserved E472 is found in the negative charged patch on the rear. It is important to note that the location of helix α1' relative to the core of the protein in Figure 5.3A is arbitrary.



(Figure 5.3) Surface potential of p140(375-480). **(A)** Electrostatic potential of the accessible surface of p140(375-480) is shown . Negative potential is colored in red and positive potential in blue. The residues contributing to the surface charge are indicated on the surface. **(B)** Surface accessible residues that are conserved (defined in Figure 3.1) are colored with blue onto the structure of p140 (375-480). N440 is not a conserved residue. An intermolecular NOE transfer between N440 and unassigned DNA proton was observed (Table 5.3). The "front" and "rear" images where generated by 180° rotation around the Z-axis. The electrostatic potential surface was calculated using MolMol (28) and the residue conservation map (B) was created with Pymol (33).

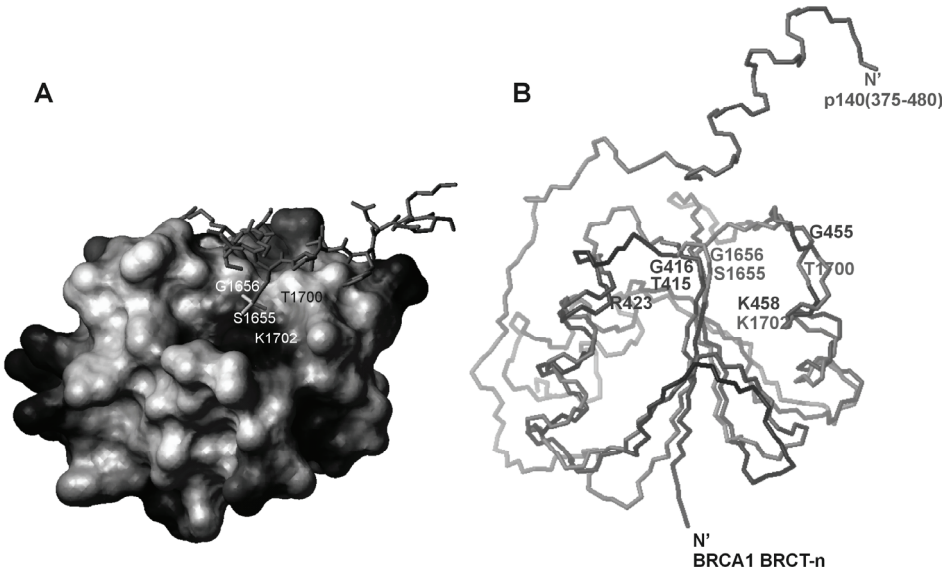
(Table 5.2) Comparison of the backbone fold of RFC p140(375-480) and various BRCT domains.

PDB	Description (reference)	Methods	RMSD* (Å)	Ref
1T29	BRCA1 BRCT-N	X-ray	2.5	(34)
1T29	BRCA1 BRCT-C	X-ray	3.0	(34)
1cdz	XRCC1 BRCT-N	X-ray	3.0	(8)
1l7b	NAD^+ dependent ligase BRCT	NMR	3.2	N/P
1gzh	p53BP BRCT-N	X-ray	1.9	(10)
1gzh	p53BP BRCT-C	X-ray	1.9	(10)
1wf6	RAD4+/CUT5+ PRODUCT	NMR	2.3	N/P
1in1	DNA ligase IIIα BRCT	NMR	2.7	(12)

N/P = Not published. *The backbone RMSD (Å) between the p140(375-480) and the PDB strcutures. The segments required for optimal structural alignment between the two structures were identified using DALI server (35).

Comparison with the structure of other BRCT domains

In order to compare the fold of RFC p140(375-480) to that of other BRCT domains, the backbone Cα atoms were superimposed (Table 5.2). Recently, the crystal structure of a complex of the BRCA1 tandem BRCT repeats bound to a phosphoserine peptide has been elucidated (34;36;37). Structural comparison of p140(375-480) with the N-terminal domain of the BRCA1 tandem BRCTs revealed a striking similarity between the binding site for the phosphate moiety of the phosphoserine on BRCA1(Figure 5.4A) and the conserved, basic patch of RFC p140(375-480) (Figure 5.4B). Residues of BRCA1 within direct hydrogen-bonding distance to the phosphate-moiety are identified as S1655 (β1), G1656 (L2) and K1702 (α2) (Figure 5.4A)(36-38). Despite the low amino acid conservation, the position of the phosphate binding residues, corresponds well with conserved residues T415, G416 and K458 respectively of p140(375-480) (Figure 5.2B and 4). The hydrogen bond donor S1655 of BRCA1 is conservatively substituted by T1898 in the related protein MDC1. The recent crystal structure of the tandem BRCT domains of MDC1 bound to a phosphoserine peptide (39) suggests that the phosphate binding pocket is highly conserved. Analogously, the positive patch created by the conserved residues of p140(375-480) is likely to be important for the interaction with the 5' phosphate of dsDNA.



(Figure 5.4) Structure comparison between the BRCT domains
(A). Electrostatic surface presentation of the N-terminal BRCT (BRCT-n) of BRCA1 (PDB:1T29) in complex with a phosphoserine peptide (in magenta). The C-terminal BRCT domain is not directly involved in the phosphate binding and therefore has been deleted from this figure for clarity. Positive potential is shown in blue and negative potential in red. The amino acid residues forming the pocket that binds the phosphate moiety (in yellow) of phosphoserine are indicated on the surface. Phosphate is directly hydrogen bonded by S1655, G1656 and K1702. T1700 is a conserved residue which forms a hydrogen bond to the side chain oxygen of S1655 stabilizing the S1655 side chain conformation. **(B)** Backbone superposition of p140 (375-480) (red) and the BRCT-n from BRCA1 (black). The orientation of the BRCT-n is identical to that of A. The backbone Cα, N and C' of the proteins are presented in red for p140(375-480) and in black for the BRCA1 BRCT-n. The conserved residues of p140(375-480) are presented in blue and the residues essential for the phosphate-moiety recognition of BRCA1 BRCT-n are presented in magenta (36-38). The overlay regions of p140 and (BRCA1) are 409-417(1649-1657), 421-452(1660-1691) and 453-479(1697-1723).

Detection of intermolecular NOEs

The conventional approach to structure elucidation of molecular complexes using NMR is based upon the assignment of intermolecular NOEs that can be used as structure restraints in the calculations. The combination of $^{13}C/^{15}N$ isotope labeling of RFC p140(375-480) and use of $^{13}C/^{15}N$ isotope editing NMR techniques enables selective observation of the components, isotope attached, non-attached or both of a complex (reviewed in ref (40)). The technique has been developed to distinguish between intra and intermolecular NOEs of a complex in which one component is uniformly isotope-labeled. In order to obtain intermolecular NOEs and structural information on p140(375-480) bound to DNA, an F_1F_2-double-half filtered 2D NOESY spectrum was recorded. Although this experiment has been successfully used to selectively observe intra- and intermolecular NOEs in the dimeric *arc* repressor (41), it did not yield high quality data on the RFC

p140(375-480)-dsDNA complex. The failure of the isotope-filtered experiment may possibly be due to magnetization loss due to T_2 relaxation during the long delay imposed by the refocused half-filter. We therefore tried an alternative approach based on purge pulses (42) which proved to be moderately successful. Two NOESY spectra were obtained by simultaneous suppression of $^{13}C/^{15}N$ attached protons in both F1 and F2, or only in F1. The resulting F_1F_2 double-filtered spectrum, which contains exclusively resonances from the unlabeled DNA, was different from the free oligonucleotide but was not sufficiently well resolved to perform a reliable sequential assignment. However, detailed comparison of the NOESY spectra listed in Table 5.3 allowed us to identify a few peaks arising from intermolecular magnetization transfer from DNA to protein (Figure 1S of supplementary materials). Due to the lack of sequence specific resonance assignments for the DNA, the identity of the source proton could not be ascertained.

(Table 5.3) Intermolecular NOEs observed between the p140(375-480) and DNA.

RFCp140(375-480)	DNA (possibilities) Ambiguously assigned	Intermolecular NOE (ppm)	Experiments
Y385 QD	CYT H5 or THY H1'	5.51	A,B
	CYT H6 or THY H6	7.72	
N440 HD21	CYT H5 or THY H1'	5.37	A,B,C
HB3	TCH_3	1.5	A,B
G416 HN	CYT H6, THY H6	7.7	C
	ADE H2, H8 or		
	NH_2 of ADE, CYT and GUA		
R423 HE	Ribose H2'' or H2'	2.19	B,C
	Ribose H4', H5'' or H5'	3.88	
G455 HN	Ribose H4', H5'' or H5'	3.94	A,B,C

A. 2D *F1*, [$^{13}C/^{15}N$]-filtered NOESY, B. 2D [$^1H,^1H$]-NOESY, C. 3D [$^{15}N,^1H$]-NOESY HSQC

As can be seen in Table 5.3, intermolecular NOEs between residue Y385 and the DNA were observed. The importance of this interaction was also demonstrated by mutagenesis (Figure 3.5). This observation supports the earlier speculation that the N-terminal helix (α1') is directly in contact with the DNA. The non-assignable NOE crosspeak correlations to the amide backbone of G416 and to the side chain guanidinium of R423 originate from DNA as well. The intermolecular NOEs to the amide side chain of N440 and the backbone of G455 were identified on the basis of the model p140(375-480)-dsDNA complexes generated by docking (described below). The majority of intermolecular NOEs that have been identified involve residues on one side of the protein surface within or in the vicinity of the conserved basic patch (Figure 5.3, Front side). However, the number of identifiable intermolecular NOEs is very small. This likely reflects

the mode of interaction between basic side chains of the protein and backbone phosphates of the DNA (although it is not possible to entirely rule out experimental limitations).

Protein-DNA docking by HADDOCK using mutagenesis and ambiguous intermolecular NOE data

Due to the lack of sequence specific resonance assignments for the DNA, it was not possible to calculate the structure of the bound dsDNA. In order to generate a model of the p140(375-480)-DNA complex despite the limited number of intermolecular structure restraints, an alternative approach using the docking program HADDOCK (17) was employed. HADDOCK can make use of a broader array of restraints including those derived from biochemical and biophysical data. This data was introduced as so-called Ambiguous Interacting Restraints (AIR) to drive the docking process. The mutagenesis (Figure 3.5), the intermolecular NOEs (Table 5.3) and the structural conservation (Figure 5.4) clearly indicate at least some of the residues that interact with the DNA. In the docking procedure, AIRs are therefore defined as ambiguous distances between these residues (called "active") and the 5' PO_4 or any/specific (if known) nucleotides on the DNA (called "passive") (Table 5.4). Although mutation of K379 resulted in the loss of DNA binding (Figure 3.5), it was not included as an AIR because its average solvent accessible surface is below 50% (see Materials and Methods).

An AIR was introduced between T415 and any nucleotide in the DNA and an additional AIR was specifically generated to the 5' phosphate of the DNA on the basis of the following three observations. (Table 5.4). 1) The resonance of the γ ^{1}H of T415 has been tentatively assigned indicating that it is in slow exchange with the solvent. Reduced exchange is a strong indication of involvement in hydrogen bonding while inspection of the protein structure indicates that there are no neighboring residues within sufficiently close distance to form a hydrogen bond. 2) Residue T415 is structurally conserved with S1655 of the BRCA1 and T1898 of the MDC1, which are crucial for binding the phosphate moiety of phosphoserine (Figure 5.3). 3) The T415A mutant (Figure 3.5) exhibited reduced DNA binding.

(Table 5.4) Active and passive residues used in the definition of the ambiguous distance restraints (AIRs) and the flexible fragments used in the HADDOCK of p140(375-480) and DNA.

Active residues of p140(375-480)	*Method determined*	*Passive residues of dsDNA*
Y382, Y385, R388, T415, R423 K458, K461 G416	Mutagenesis data (Figure 3.5) Residue conservation	Any DNA nucleotide (5'pCTCGAGGTCG3'/ 5'CGACCTCGAGATCA3')
Y385 Hδ T415 Hγ G416 HN R423 Hε	Intermolecular NOEs (Table 5.3)	Any H2, H6, H8 O1P, O2P, O3P of 5'pC19 Any H2, H6, H8 Any H4', H5', H5''

	Residue numbers of p140(375-480)	*dsDNA*
Flexible fragments	377-392, 414-464	All nucleodides

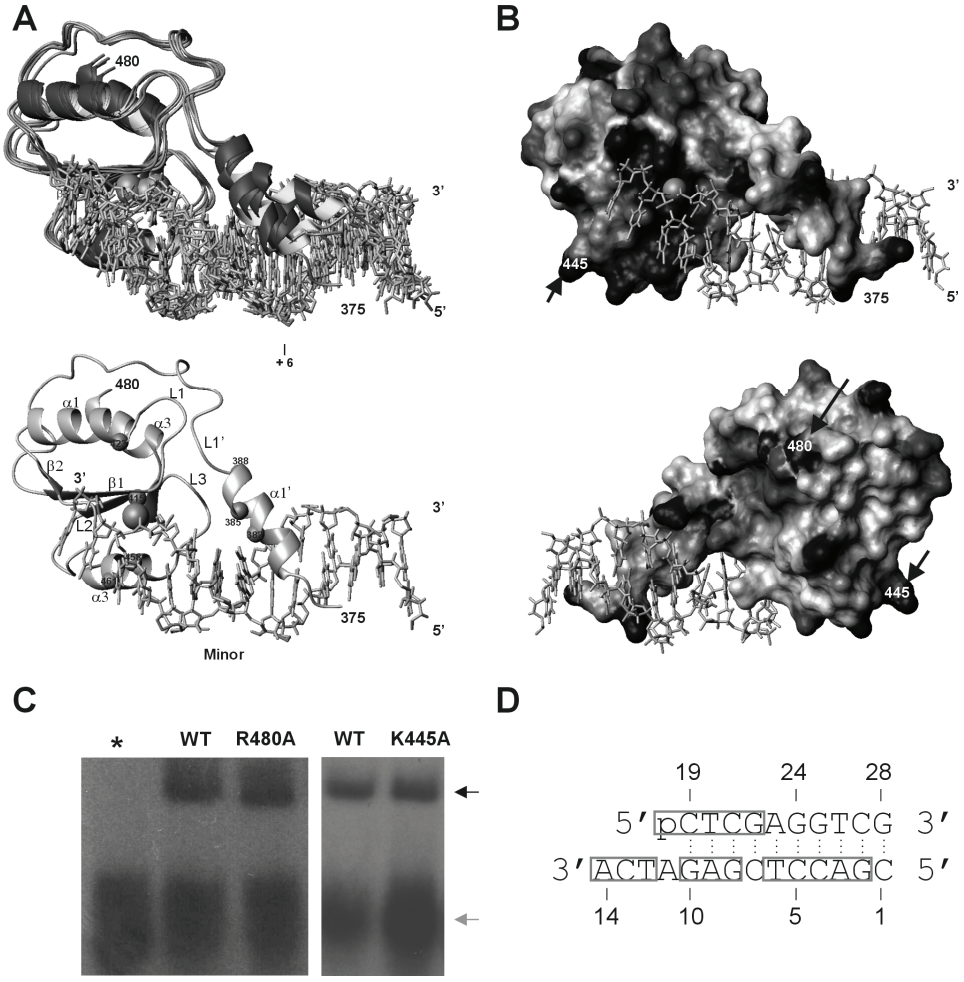
In HADDOCK, the docking protocol consists of three stages. Initially a protein and a DNA molecule are randomly rotated around the center of mass. Subsequently the orientation of each molecule is optimized in order to minimize the intermolecular energy functions, and then the two are docked using a rigid body energy minimization (EM). During this initial docking, proteins are treated as rigid bodies forbidding movement of side-chains or flexible loops. However, this lack of flexibility could be detrimental to determining the correct initial orientation of the interacting surfaces. In order to overcome this problem, the HADDOCK protocol was applied to each member of the ensemble of 20 NMR structures of p140(375-480). Since different side chain and/or backbone orientations allow better sampling of conformational possibilities, this method has been shown to produce better docking solutions that are closer to known complexes than use of a single structure (43). As no structure of the DNA portion of the complex is available, a model structure of 5' phosphorylated, dsDNA with a 3' single stranded overhang in the standard B-form conformation was generated using the sequence of an oligonucleotide used in the NMR studies but lacking the hairpin since it does not interact with p140(375-480). In the second stage of the docking protocol, HADDOCK accommodates conformational rearrangements of side chain and backbone of defined segments (Table 5.4) to optimize interaction at the protein-DNA interface. The N-terminal amino acid segment residues 377-392 of p140(375-480), which was less defined in the ensemble of structures, was allowed

to move freely. In the last stage of docking, the protein-DNA complexes are refined in an explicit water bath. As a result, the docking process generated 200 solutions that were sorted into clusters using a pairwise backbone RMSD of 5Å as a cutoff criterion (Table 5.5). This procedure resulted in 10 clusters, which were then ranked according to their HADDOCK scores calculated on the basis of the intermolecular energy (the sum of electrostatic, van der Waal's, and AIR energy terms) and their average buried surface area (Table 5.5). The five best structures from cluster 7, which had the lowest HADDOCK score of any cluster, were accepted as the best representative model of the complex (Figure 5.5A).

Table 5.5. Statistics of the five structures from each 8 clusters, which are ranked on the basis of their HADDOCK scores.

#cluster	RMSD[1]	Sd[2]	HADDOCK score[3]	Sd[4]	BSA[5]
clust7	1.3	0.8	-63	55	2081
clust1	4.8	1.9	-15.9	11.6	1924
clust4	3.3	1.5	17.4	32.5	2039
clust2	4.7	1.9	64.4	16.2	1419
clust6	5	2.4	117.9	43.9	1483
clust5	5	1.5	114.7	23.5	1346
clust3	4	1	135.2	23.3	1660
clust8	4.3	1.3	151.4	48.4	1502

[1] RMSD from the lowest energy structure in the cluster; [2] standard deviation of RMSD; HADDOCK score calculated in the basis of the intermolecular in interaction energy which is the sum of electrostatic, van der Waals and AIR energies at the interface. [4]Standard deviation of HADDOCK score. [5] BSA is a "buried surface area" in the complex ($Å^2$).



(Figure 5.5) Docking model of the p140 (375-480)-dsDNA complex generated by HADDOCK. **(A)** Ribbon representation of ensemble of the five lowest energy structures from the cluster 7. The 5'phosphopate is indicated by a magenta sphere. One of the ensemble structures in which the active residues defined in HADDCOK are presented in yellow spheres (bottom). **(B)** Electrostatic surface potential presentation of p140 (375-480) bound to dsDNA. The upper figure has the same orientation as in A while the lower has been rotated 180° around the Z-axis (bottom). The positive and negative charged surfaces are colored in blue and red respectively. Exposed hydrophobic residues are white or slightly colored **(C)** DNA binding activity of R480A and K445A. The mutations were designed as tests of the model of the Protein-dna complex. The wildtype DNA binding activity of the R480A and K445A mutants suggests that the "rear" side and the bottom of p140(375-480) (indicated with an arrow in **(B)**) do indeed not contact the DNA.**(D)** The sequence of the dsDNA used in the HADDOCK docking. The nucleotides with phosphates and bases in contact with the protein are boxed.

(Table 5.6) Intermolecular contacts identified over the ensemble of five best structure of cluster 7 representing p140(375-480)-dsDNA complex.

			Hydrogen bonds				*Nonbonded contacts*		
Residues	Secondary structure	DNA*	B-P	B-b	S-P	S-b	B-b	S-P	S-b
K375	α1'	G2			2				
		A3			1				
		C4			1				
Y382	α1'	C21			1			5	1
		C20			1				
		G22			1				
Y385	α1'	C21							3
		T6				1			1
		G22				1			
		C5							1
		T20							3
R388	L1'	T6			3			1	
		C5			1				
T415	β1	C19			4			3	3
G416	L1	C19	4					2	
R423	α1	C19			2			1	
		A14			2				
T438	L2	A14					3		
N440	L2	C13				3			2
R452	L3	C19			1			2	
		T20			1				
S454	L3	C19					4		
		T20					3		1
G455		C19					3		
Q456	L3	G8			2			1	1
		A9					1		
S457	L3	G10				3			4
		A9							4
K458	α2	C19			2		1	3	1
		T12				1			

* The numbering of the DNA refers to the Figure 5.5D.
Intermolecular contacts were analyzed by NUCPLOT (see Materials and Methods) and listed only when the interacting residues were present in three out of five structures. The identified hydrogen bonds and nonbonded contacts are classified into protein Backbone-DNA phosphate (B-P), Backbone-DNA base (B-b), side chain – DNA phosphate (S-P), side chain-DNA Base (S-b).

The residues interacting with the DNA in the five model complexes were identified using the HBPLUS (44) in the NUCPLOT (45) package. HBPLUS identified hydrogen bonds and non-bonded contacts made by pairs of atoms (Materials and Methods for the details), which were sorted according to the amino acid residue numbers. Residues which were identified as interacting with DNA in three out of five model complexes are listed in Table 5.6. In the model of the p140(375-480)-dsDNA complex, the 5' phosphate (C19, Figure 5.5D) of the dsDNA is accommodated by the positively charged surface (Figure 5.5B) mainly formed by the conserved residues T415, G416, R423 and K458 (Figure 5.3). The side chains of T415, R423 and K458, and the amide backbone of G416 are within the hydrogen bonding or salt-bridging distance to the oxygens of the 5' phosphate (Table 5.6). This is due to the AIR restraint introduced between T415 and the 5' phosphate (C19), while in the absence of this restraint; we did not obtain docking solutions in which the 5' phosphate is bound by the protein. Both mutants (R423A and T415A) exhibited reduced DNA binding but were not as severely affected as the K458E mutation (Figure 3.5). It is possible that the network of hydrogen bonds to the 5' phosphate by the remaining residues is sufficient to partially compensate for the loss of one hydrogen bond donor resulting from the electrically neutral alanine mutation. On the other hand, the introduction of negative charge in the K458E mutation, as might be expected, has a more drastic effect on DNA binding. Other than the conserved residues, R452 also interacts with the 5' phosphate via dominantly electrostatic interaction. Although no intermolecular NOE was observed between R452 and the DNA, the HεNε resonance of R452 is folded in [^{1}H,^{15}N] HSQC spectrum of the p140(375-480)-dsDNA complex, which suggests stabilization of the side chain conformation upon DNA binding often through a charge mediated, nonbonded interaction. In this model, K461 approaches the phosphates within coulombic distance (~ 6 Å) providing additional charge interaction. However in contrast to the K461E mutation data, which resulted in dramatic reduction of DNA binding (Figure 3.5), K461 does not seem to play a major role in DNA binding in the model structure.

In the model presented in Figure 5.5, the orientation of the helix α1' is better defined than in the free protein, and lies in the major groove (Figure 5.5A) clearly separated from the core of the protein, which explains the lack of long range NOEs in this region. The amino group of K375 (α1'), the hydroxyl group of Y382 (α1') and the guanidinium group of R388 (L1') participate in hydrogen bonds or electrostatic

interactions with the phosphate backbone of the DNA (Table 5.6). In addition, numerous interactions including hydrogen bonds and van der Waal's contacts occur between the aromatic side chain of Y385 (α1') and nucleotide bases (Table 5.6) of the major groove. Reduced DNA binding of p140(375-480) was observed when the size of the DNA duplex becomes less than 7 nucleotides long or when the +6 nucleotide position (G24) from the 5' phosphate end contains a non Watson-Crick basepair (T24:C5) (Figure 2.3). Accordingly the model complexes suggest that the shortening of the duplex and the non Watson-Crick base pair may interfere with the interaction of Y385 as well as with the interaction of R388 and K375 with dsDNA. No interaction between S384 and dsDNA was found in the model complexes, a result that is in concert with the mutagenesis study (Figure 3.5).

The model of the protein-DNA complex suggests that the 3' single stranded DNA tail (nucleotides C13 and A14) interacts via the backbone phosphates with the side chain of R423 and via the bases with the side chains of T438 and N440. Importantly, there are no restraints in the calculations that explicitly define this interaction. Therefore, the model explains the earlier observation that p140 (375-480) binds a 5' recessed dsDNA with higher affinity than blunt ended DNA (Figure 2.3). Based on the structure, we identified an intermolecular NOE between the side chain amide of N440 and the DNA (Table 5.3). The side chain amide resonance of N440 is shifted away from the random coil value (Figure 5.1) suggesting involvement in hydrogen bonding. Since there are no hydrogen bond acceptors on the protein in close vicinity of N440, the likely acceptor is the DNA. However, the present model does not include this restraint.

The backbone atoms of the S454, G455, Q456 and S457 in Loop 3 (L3) make a number of van der Waal's contacts with the bases of the DNA (Table 5.6). Although Loop 3 (L3) is disordered in the ensemble of 20 structures of the protein without DNA (Figure 5.2A), the loop orientation is better defined in the protein-DNA complexes (RMSD of the L3(451-457) is 1.36 ± 0.4 Å in the free protein in comparison to 0.39 ± 0.14 Å in the model complex). Similarly to N440, an intermolecular NOE between the backbone HN of G455 and an unassigned DNA proton was identified based on this model. Since the backbone HN of G455 resonates at the random coil position, it apparently does not take part in a hydrogen bond interaction with the DNA or protein, suggesting that it serves a structural role in the loop instead.

In order to test the veracity of the proposed model we sought to make further site directed mutations. Visual inspection of the structures presented in Figure 5.5 indicates that

residues R480 and K445, both of which have ≥ 60% amino acid identity among the 31 species analyzed (Figure 3.1), lie on the face of p140(375-480) that does not contact the DNA (Figure 5.5B arrows). We therefore generated R480A and K445A mutations and tested their effect on DNA binding. As shown in Figure 5.5C, neither of the mutations disrupts DNA binding, as was expected. This experiment serves as a further control that disruption of DNA binding observed in the earlier mutagenesis studies was not the result of a subtle global change in the protein fold that could not be detected in the 1D NMR spectra.

Discussion

Structure of p140(375-480) in the presence of DNA

Here the first structure of a BRCT domain bound to DNA is presented. Preliminary analysis of the NMR data suggests that the free protein is flexible in solution but becomes more rigid upon binding to DNA. Even in the presence of DNA, p140(375-480) still exhibits dynamic behaviour which caused substantial loss of NMR signals. The moderate resolution of the structures based on NOE derived restraints is a direct consequence of the poor quality NMR data.

The p140(375-480)-dsDNA complex generated by HADDOCK

The model complex of p140(375-480) and dsDNA was generated using the program HADDOCK which uses Ambiguous Interaction Restraints (AIRs) derived from mutagenesis, intermolecular NOEs and residue conservation to drive intermolecular docking. Although the model complex is in good agreement with all of the DNA binding properties characterized by biochemical analysis, it should nonetheless be considered as preliminary pending further experimental analysis.

Despite the fact that no specific restraints were included in the calculations, helix a1' lies in the major groove of the DNA helix in the best structures generated by HADOCK. The α-helix is a commonly used structural element for recognition of bases as well as backbone phosphates in sequence specific (reviewed in (46)) and non-sequence specific DNA binding (47). In the non-specific complex of DNA *–lac* headpiece-62, many of the side chains (Y7, Y17, Q18 and R22) that confer direct interactions with the base pairs in the major groove of the sequence specific complex, shift and participate in hydrogen bonds or electrostatic interactions with the backbone phosphates (47). Similarly in the p140(375-480)-dsDNA models, non-sequence specific interaction is mediated by

hydrogen bonds or salt bridges between K375, Y382 (α1') and R388 (L1') and backbone phosphates. In contrast to the sequence specific *lac* headpiece-DNA complex in which base specific contacts are made by several residues, in p140(375-480)-dsDNA Y385 is the only residue in helix α1' that points into the major groove. Amongst the five different structures in the cluster, Y385 contacts five different base pairs. This type of behavior is consistent with a dynamic, non-sequence specific complex.

Based upon analysis of the structure of the complex, 36 intermolecular NOEs would be expected. However, to date only intermolecular NOEs involving G455 and N440 have been definitively identified. This discrepancy can be explained in three ways. Cleary the model of the complex may be incorrect. However, since the model explains a considerable amount of biochemical data that was not used to generate it, this seems unlikely. Alternatively, the expected NOE correlations may be missing or overlapped in the NMR spectra. It is likely that at least some of the missing correlations are due to overlap in crowded regions of the spectra or to experimental limitations (Figure 1S). NMR methods commonly used to detect intermolecular NOE correlations failed on the p140-DNA complex despite the fact that the complex is not unusually large. An experiment based on purge pulses (42), which reduces the amount of time required to perform the magnetization filter, yielded moderate results. The limited amounts of intermolecular NOEs that have been assigned were derived from this experiment. This observation, in conjunction with the previously reported missing correlations in the 3D [^{13}C,^{1}H]-NOESY HSQC spectrum (Figure 4.1) and the below normal number of long range NOEs, reflects the nature of the complex in which the residues making contact with DNA likely undergo intermediate time scale exchange between conformations. This intermediate exchange rate will lead to loss of resonance signals due to efficient relaxation of transverse magnetization. In the case of the non specific DNA-*lac* headpiece-62 complex, residues located in the protein-DNA interface were clearly shown to undergo exchange dynamics on the μs-ms timescale indicating that they sample different base pair environments (47). Water molecules play important roles in filling the gaps in protein-DNA interfaces, which are otherwise energetically unfavorable, as well as in water-mediated hydrogen bond with bases and phosphate backbone at the protein-DNA interface of both specific (48) and non-specific complexes (49). In the interface of a non-specific protein-DNA complex of BamHI, all specific interactions are replaced by non-specific interactions mediated by hydrogen bonds

to the DNA backbone via highly mobile water molecules (49). Although water molecules were not included during the docking, one can not rule out their potential participation at the p140(375-480)-dsDNA interface, which maybe also contribute to the limited number of observable intermolecular NOEs.

The model of the complex shown in Figure 5.5 depends on the inclusion of a specific AIR between T415 and the 5' phosphate of the dsDNA. Justification of this assumption was previously provided (vida infra). While this assumption clearly demands further experimental support ("the reason for referring to the structure of the complex as preliminary"), the resulting structure explains a number of observations that were not included in the calculation and is therefore likely correct. This specific AIR was introduced, in part, based on the structural conservation with the functional residues of BRCA1 protein. Many of the tumor associated missense mutations are known to occur in the BRCT repeats of BRCA1 (50-52), which bind to phosphoserine containing peptide sequences (15;53;54). Peptide recognition by the BRCA BRCTs mainly involves the phosphoserine and a phenylalanine, that is located 3 residues carboxy terminal to the phosphoserine (34;37;38). The residues that bind the phosphoserine are found exclusively in the N-terminal domain of the BRCA1 BRCT repeat, while the phenylalanine is accommodated by a hydrophobic cleft created by the interface of the two BRCT repeats (Figure 5.4). Superposition of the N-terminal domain of the BRCA1 BRCT repeat with p140(375-480) showed a remarkable similarity in the positions of conserved residues of p140(375-480) and those residues involved in phosphoserine binding by BRCA1 (34;36;37). If these conserved residues, T415, G416, and K458 of the p140(375-480) bind specifically to the 5' phosphate, it would provide a possible explanation for the necessity of the N-terminal extension for high affinity DNA binding. In contrast to SH2 domains, which bind phosphotyrosine in a deep, positively charged pocket, the BRCT domains interact weakly with phosphoserine by a shallow basic site. High affinity binding requires a second interaction, such as the hydrophobic interaction involving the peptide phenylalanine in the BRCA1 tandem BRCTs. Analogously, p140(375-480) presents a shallow, positively charged pocket for 5' phosphate binding but in order to achieve high affinity DNA binding, additional interactions by the N-terminal α1 helix are necessary.

Other DNA binding BRCT domains

The BRCT domain of RFC p140 belongs to a distinct subgroup of the BRCT superfamily. Within the distinct subgroup, there is increasing evidence to suggest that the BRCT domain from the bacterial NAD^+ dependent ligase binds to DNA (30;55;56). This BRCT domain is located at the C-terminus of the multi-domain enzyme and recent studies (30) showed that the domain is responsible for stable association of protein and DNA (56). Amino acid sequence analysis of the distinct subgroup of BRCT domains indicates that the potential DNA-binding residues, including T415, G416, R423, G455 and K458, are absolutely conserved between the NAD^+ dependent DNA ligases and RFC p140 (Figure 3.6). Mutations in these residues severely affect the DNA binding or the adenylate-moiety transfer activities of this class of ligases (30). Similar effects on DNA binding observed upon mutations to the conserved residues shared between the bacterial ligases and RFC p140, implies that the 5' phosphate could also be the specific target for DNA binding by the BRCT domain of the DNA ligases. However, ligase activity of RFC p140 has not been reported and the cellular role of 5' phosphate binding remains unknown.

Acknowledgments

The author sincerely thanks Prof. R. Boelens for recording the isotope-filtered NOESY spectra, Marc van Dijk for initiating the early part of the HADDOCK calculation and Anneloes Blok for generation of the R480A and K445A mutants.

Reference list

1. Waga, S. and Stillman, B. (1994) *Nature* **369,** 207-212

2. Uhlmann, F., Gibbs, E., Cai, J., O'Donnell, M., and Hurwitz, J. (1997) *J.Biol.Chem.* **272,** 10065-10071

3. Bowman, G. D., O'Donnell, M., and Kuriyan, J. (2004) *Nature* **429,** 724-730

4. 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

5. Bork, P., Hofmann, K., Bucher, P., Neuwald, A. F., Altschul, S. F., and Koonin, E. V. (1997) *FASEB J.* **11,** 68-76

6. Caldecott, K. W. (2003) *Science* **302,** 579-580

7. Koonin, E. V., Altschul, S. F., and Bork, P. (1996) *Nat.Genet.* **13,** 266-268

8. 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

9. Williams, R. S., Green, R., and Glover, J. N. (2001) *Nat.Struct.Biol.* **8,** 838-842

10. 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

11. 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

12. Krishnan, V. V., Thornton, K. H., Thelen, M. P., and Cosman, M. (2001) *Biochemistry* **40,** 13158-13166

13. 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

14. Dulic, A., Bates, P. A., Zhang, X., Martin, S. R., Freemont, P. S., Lindahl, T., and Barnes, D. E. (2001) *Biochemistry* **40,** 5906-5913

15. Yu, X., Chini, C. C., He, M., Mer, G., and Chen, J. (2003) *Science* **302,** 639-642

16. Yu, X. and Chen, J. (2004) *Mol.Cell Biol.* **24,** 9478-9486

17. Dominguez, C., Boelens, R., and Bonvin, A. M. (2003) *J.Am.Chem.Soc.* **125,** 1731-1737

18. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J.Biomol.NMR* **6,** 277-293

19. Keller, R. (2004) *The Computer Aided Resonance Assignment Tutorial*, 1st Ed., CANTINA Verlag,

20. Kobayashi, M. and Siegal, G. (2005) *J.Biomol.NMR* **31,** 183-184

21. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J.Mol.Biol.* **319,** 209-227

22. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J.Mol.Biol.* **273,** 283-298

23. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J.Biomol.NMR* **13,** 289-302

24. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr.D.Biol.Crystallogr.* **54 (Pt 5),** 905-921

25. Nederveen, A. J., Doreleijers, J. F., Vranken, W., Miller, Z., Spronk, C. A., Nabuurs, S. B., Guntert, P., Livny, M., Markley, J. L., Nilges, M., Ulrich, E. L., Kaptein, R., and Bonvin, A. M. (2005) *Proteins* **59,** 662-672

26. Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) *J.Biomol.NMR* **8,** 477-486

27. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (1997) *Nucleic Acids Res.* **25,** 4940-4945

28. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J.Mol.Graph.* **14,** 51-32

29. Ekblad, C. M., Wilkinson, H. R., Schymkowitz, J. W., Rousseau, F., Freund, S. M., and Itzhaki, L. S. (2002) *J.Mol.Biol.* **320,** 431-442

30. Feng, H., Parker, J. M., Lu, J., and Cao, W. (2004) *Biochemistry* **43,** 12648-12659

31. Nelson, J. R., Gibbs, P. E., Nowicka, A. M., Hinkle, D. C., and Lawrence, C. W. (2000) *Mol.Microbiol.* **37,** 549-554

32. Gaiser, O. J., Ball, L. J., Schmieder, P., Leitner, D., Strauss, H., Wahl, M., Kuhne, R., Oschkinat, H., and Heinemann, U. (2004) *Biochemistry* **43,** 15983-15995

33. Warren, G. L. (2005) *The PyMOL Molecular Graphics System*, Delno Scientific LLC, San carlos, CA, USA http://www.pymol.org,

34. Shiozaki, E. N., Gu, L., Yan, N., and Shi, Y. (2004) *Mol.Cell* **14,** 405-412

35. Holm, L. and Sander, C. (1993) *J.Mol.Biol.* **233,** 123-138

36. 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

37. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. (2004) *Nat.Struct.Mol.Biol.* **11,** 519-525

38. Botuyan, M. V., Nomine, Y., Yu, X., Juranic, N., Macura, S., Chen, J., and Mer, G. (2004) *Structure.(Camb.)* **12,** 1137-1146

39. Lee, M. S., Edwards, R. A., Thede, G. L., and Glover, J. N. (2005) *J.Biol.Chem.*

40. Breeze, A. L. (2000) *Progress in Nuclear Magnetic Resonance Spectroscopy* **36,** 323-372

41. Burgering, M. J., Boelens, R., Caffrey, M., Breg, J. N., and Kaptein, R. (1993) *FEBS Lett.* **330,** 105-109

42. Ikura, M. and Bax, A. (2005) *J.Am.Chem.Soc.* **114,** 2433-2440

43. Dominguez, C., Bonvin, A. M., Winkler, G. S., van Schaik, F. M., Timmers, H. T., and Boelens, R. (2004) *Structure.(Camb.)* **12,** 633-644

44. McDonald, I. K. and Thornton, J. M. (1994) *J.Mol.Biol.* **238,** 777-793

45. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (1997) *Nucleic Acids Res.* **25,** 4940-4945

46. Garvie, C. W. and Wolberger, C. (2001) *Mol.Cell* **8,** 937-946

47. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) *Science* **305,** 386-389

48. Lawson, C. L. and Carey, J. (1993) *Nature* **366,** 178-182

49. Viadiu, H. and Aggarwal, A. K. (2000) *Mol.Cell* **5,** 889-895

50. Mirkovic, N., Marti-Renom, M. A., Weber, B. L., Sali, A., and Monteiro, A. N. (2004) *Cancer Res.* **64,** 3790-3797

51. Williams, R. S. and Glover, J. N. (2003) *J.Biol.Chem.* **278,** 2630-2635

52. Williams, R. S., Chasman, D. I., Hau, D. D., Hui, B., Lau, A. Y., and Glover, J. N. (2003) *J.Biol.Chem.* **278,** 53007-53016

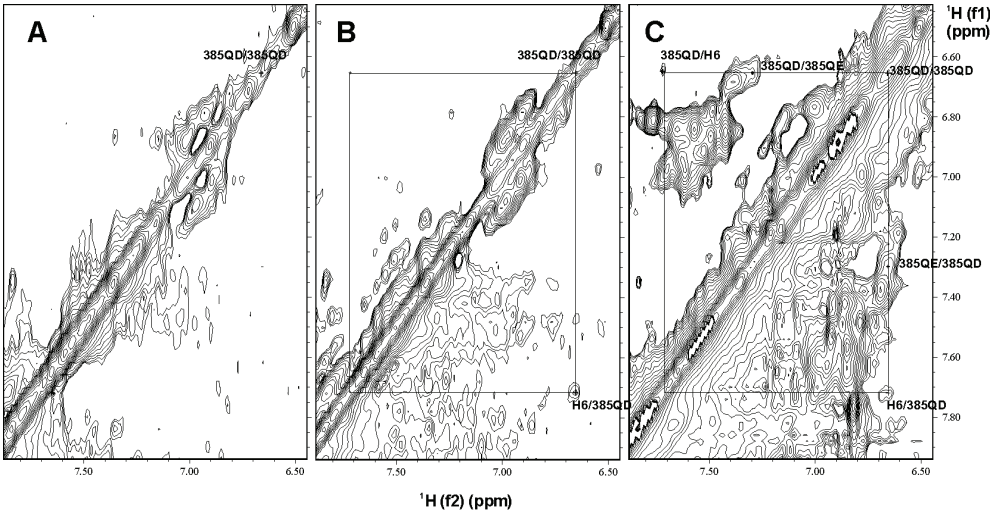
53. Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003) *Science* **302,** 636-639

54. Rodriguez, M., Yu, X., Chen, J., and Songyang, Z. (2003) *J.Biol.Chem.* **278,** 52914-52918

55. Jeon, H. J., Shin, H. J., Choi, J. J., Hoe, H. S., Kim, H. K., Suh, S. W., and Kwon, S. T. (2004) *FEMS Microbiol.Lett.* **237,** 111-118

56. Wilkinson, A., Smith, A., Bullard, D., Lavesa-Curto, M., Sayer, H., Bonner, A., Hemmings, A., and Bowater, R. (2005) *Biochim.Biophys.Acta* **1749,** 113-122

Supplementary materials



(Figure 1S) Identification of intermolecular NOE between DNA (H6) and Y385(QD) using spectra; **A)** *F1,F2*, [^{15}N/^{13}C]-filtered [^{1}H, ^{1}H]-NOESY contains only intra NOE peaks between protons of DNA, **B)** *F1* [^{15}N/^{13}C]-filtered [^{1}H, ^{1}H]-NOESY contains intra NOE peaks of DNA and intermolecular NOE peaks between DNA (f1) and protein (f2) and **C)** 2D [^{1}H, ^{1}H]-NOESY contains intramolecular NOE peaks of both protein and DNA, and intermolecular NOE peaks. In **B**), the magnetization of protons attached to the ^{15}N/^{13}C isotopes (protein) were suppressed during the first t_1 and the magnetization were allowed to transfer to protein from DNA during the mixing time, which results in the intermolecular NOE peaks observed only one side of the diagonal peaks (H6(f1)/385QD(F2) in **B**) in comparison to the symmetric peaks observed between intramolecular NOE transfers (Y385 QE in **C**).

S2. Bruker pulse program
2D *F1,F2*, [^{13}C/^{15}N]-filtered NOESY (Isotope filter based on HMQC purging)
Written by Rolf Boelens (Department of NMR spectroscopy, Bijvoet Center for Biomolecular research, University of Utrecht

```
# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
;qqnoesyphpr.rb2
```

```
;wgMQ4N1d.rb
;avance-version (00/02/07)
;2D homonuclear correlation via dipolar coupling
;dipolar coupling may be due to noe or chemical exchange.
;phase sensitive
;with presaturation during relaxation delay and mixing time
;;invieaV1_22.rb
;

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance2.incl
;   for 1
;
;avance-version (02/08/12)

;$Id: Avance2.incl,v 1.7.2.1 2002/08/12 13:19:57 ber Exp $
# 12 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for 1
;
;avance-version (02/05/31)

define list<gradient> EA=<EA>

;$Id: Grad2.incl,v 1.7 2002/06/12 09:04:22 ber Exp $
# 13 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl  -  include file for commonly used delays
;
;version 00/02/07


;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5


;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
```

```
define delay CEN_CP2


;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3



;$Id: Delay.incl,v 1.11 2002/06/12 09:04:22 ber Exp $
# 14 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2" 2


"p2=p1*2"
"p22=p21*2"

;"d24=2m"  ;wurst
"d5=5.55m"
"d11=30m"
"d12=20u"
"d13=4u"
"d0=4u"
"d18=d8-p16-d16-d12*5+d13*2"

"DELTA2=d24-p24*0.5+p2*0.5-d12-p16"
"DELTA6=d24-p24*0.5+p2*0.5-d12-p16-p1*2/3.1416"
"d15=p16-p3*2-d13"
"DELTA3=d24-p24*0.5+p2*0.5-d12-p16*2-d16"
"d25=p16+d16-p21-d0-p1*2/3.1416"
"d26=p16+d16-p21-d13-d12*2"
"DELTA4=d5-d24*2-d24+p24*0.5-p2*0.5"
"DELTA5=p21-p3*2-d13"
"CEN_HC2=(p24-p2)/2"
"CEN_HN2=(p24-p22)/2"

# 1 "mc_line 38 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding
       definition part of mc command before ze"
define delay MCWRK
define delay MCREST
define loopcounter ST1CNT
"ST1CNT = td1 / (2)"
"MCWRK = 0.500000*d11"
"MCREST = d11 - d11"
# 38 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
1 ze
# 1 "mc_line 38 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding
       definition of mc command after ze"
# 39 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
  3m pl12:f2
  3m pl16:f3
# 1 "mc_line 41 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding start
       label for mc command"
2 MCWRK  do:f2
LBLSTS1, MCWRK
LBLF1, MCREST
# 42 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"
  d12 do:f3
  d12 setnmr3|0 setnmr0|34|32|33
  p16:gp0
  d16
  d12 setnmr3^0 setnmr0^34^32^33
  d12 pl9:f1
  d1 cw:f1 ph29
  d13 do:f1
  d12 pl1:f1
3 (p1 ph1)
    d12
```

```
    p16:gp1
    DELTA6 pl0:f2 pl3:f3
  (CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (CEN_HN2 p22 ph0):f3
    d12
    DELTA2
    d15 pl2:f2
  (p3 ph20 d13 p3 ph10):f2
;
    d12
    p16:gp1
    DELTA3
    p16:gp2
    d16 pl0:f2
  (CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (DELTA4 p21 ph6):f3 (CEN_HN2 p22 ph0):f3
    d12
    DELTA3
    p16:gp2
    d25 pl2:f2
   (DELTA5 p3 ph16 d13 p3 ph10):f2 (p21 ph20):f3
;
   d0
  p1 ph2
  d13
  d12 pl9:f1 setnmr3|0 setnmr0|34|32|33
  p16:gp3
  d16
  d12 setnmr3^0 setnmr0^34^32^33
  d18 cw:f1
  d13 do:f1
  d12 pl1:f1
;
  (p1 ph3)
    d12
    p16:gp1
    DELTA6 pl0:f2 pl3:f3
  (CEN_HC2 p2 ph11) (p24:sp24 ph6):f2 (CEN_HN2 p22 ph0):f3
    d12
    DELTA2
    d15 pl2:f2
  (p3 ph21 d13 p3 ph10):f2
;
    d12
    p16:gp1
    DELTA3
    p16:gp2
    d16 pl0:f2
  (CEN_HC2 p2 ph11) (p24:sp24 ph6):f2 (DELTA4 p21 ph6):f3 (CEN_HN2 p22 ph0):f3
    d12
    DELTA3
    p16:gp2
    d26 pl2:f2
   (DELTA5 p3 ph17 d13 p3 ph10):f2 (p21 ph16):f3
   d13
   d12 pl12:f2
   d12 pl16:f3
;
  go=2 ph31 cpd2:f2 cpd3:f3
# 1 "mc_line 110 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2 expanding mc
       command in line"
  MCWRK  do:f2 do:f3 wr #0 if #0 zd dp2
  lo to LBLSTS1 times 2
  MCWRK id0
  lo to LBLF1 times ST1CNT
# 111 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qqnoesyphpr.rb2"

exit


ph0=0
ph10=1
```

```
ph6=0 2

ph20=0 0 2 2
ph16=0 0 0 0 2 2 2 2
ph21=0 0 0 0 0 0 0 0  2 2 2 2 2 2 2 2
ph17=0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  2 2 2 2 2 2 2 2  2 2 2 2 2 2 2 2

ph1=0 2
ph2=0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  2 2 2 2 2 2 2 2
        2 2 2 2 2 2 2 2  2 2 2 2 2 2 2 2  2 2 2 2 2 2 2 2
ph3=0 0 0 0 2 2 2 2  1 1 1 1 3 3 3 3
ph11=1 1 1 1 3 3 3 3  2 2 2 2 0 0 0 0

ph29=0
ph31=0 2 0 2 2 0 2 0  1 3 1 3 3 1 3 1  0 2 0 2 2 0 2 0  1 3 1 3 3 1 3 1  2 0 2 0 0 2 0
        2  3 1 3 1 1 3 1 3  2 0 2 0 0 2 0 2  3 1 3 1 1 3 1 3


;pl1 : f1 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;p1 : f1 channel -  90 degree high power pulse
;d0 : incremented delay (2D)
;d1 : relaxation delay; 1-5 * T1
;d8 : mixing time
;d11: delay for disk I/O                             [30 msec]
;d12: delay for power switching                      [20 usec]
;d13: short delay                                    [4 usec]
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;NS: 8 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC


;Processing

;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1


;$Id: noesyphpr,v 1.4 2002/06/12 09:05:10 ber Exp $
```

S3. 2D [*F1* filtered-$^{13}C/^{15}N$]] NOESY (Isotope filter based on HMQC purging)
Written by Rolf Boelens (Department of NMR spectroscopy, Bijvoet Center for biomolecular research, University of Utrecht

```
# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
;qnoesyphpr.rb1
;wgMQ4N1d.rb
;avance-version (00/02/07)
;2D homonuclear correlation via dipolar coupling
;dipolar coupling may be due to noe or chemical exchange.
;phase sensitive
;with presaturation during relaxation delay and mixing time
;;invieaV1_22.rb
;


# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Avance.incl" 1
;Avance2.incl
;   for 1
;
;avance-version (02/08/12)

;$Id: Avance2.incl,v 1.7.2.1 2002/08/12 13:19:57 ber Exp $
# 12 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Grad.incl" 1
;Grad2.incl  -  include file for Gradient Spectroscopy
;   for 1
```

```
;
;avance-version (02/05/31)


define list<gradient> EA=<EA>


;$Id: Grad2.incl,v 1.7 2002/06/12 09:04:22 ber Exp $
# 13 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2

# 1 "/opt/xwinnmr/exp/stan/nmr/lists/pp/Delay.incl" 1
;Delay.incl  -  include file for commonly used delays
;
;version 00/02/07


;general delays

define delay DELTA
define delay DELTA1
define delay DELTA2
define delay DELTA3
define delay DELTA4
define delay DELTA5
define delay DELTA6
define delay DELTA7
define delay DELTA8

define delay TAU
define delay TAU1
define delay TAU2
define delay TAU3
define delay TAU4
define delay TAU5


;delays for centering pulses

define delay CEN_HN1
define delay CEN_HN2
define delay CEN_HN3
define delay CEN_HC1
define delay CEN_HC2
define delay CEN_HC3
define delay CEN_HC4
define delay CEN_HP1
define delay CEN_HP2
define delay CEN_CN1
define delay CEN_CN2
define delay CEN_CN3
define delay CEN_CN4
define delay CEN_CP1
define delay CEN_CP2


;loop counters

define loopcounter COUNTER
define loopcounter SCALEF
define loopcounter FACTOR1
define loopcounter FACTOR2
define loopcounter FACTOR3



;$Id: Delay.incl,v 1.11 2002/06/12 09:04:22 ber Exp $
# 14 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2" 2


"p2=p1*2"
"p22=p21*2"
```

```
;"d24=2m"  ;wurst
"d5=5.55m"
"d11=30m"
"d12=20u"
"d13=4u"
"d0=4u"
"d18=d8-p16-d16-d12*5+d13*2"

"DELTA2=d24-p24*0.5+p2*0.5-d12-p16"
"DELTA6=d24-p24*0.5+p2*0.5-d12-p16-p1*2/3.1416"
"d15=p16-p3*2-d13"
"DELTA3=d24-p24*0.5+p2*0.5-d12-p16*2-d16"
"d25=p16+d16-p21-d0-p1*2/3.1416"
"DELTA4=d5-d24*2-d24+p24*0.5-p2*0.5"
"DELTA5=p21-p3*2-d13"
"CEN_HC2=(p24-p2)/2"
"CEN_HN2=(p24-p22)/2"

# 1 "mc_line 37 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding
definition part of mc command before ze"
define delay MCWRK
define delay MCREST
define loopcounter ST1CNT
"ST1CNT = td1 / (2)"
"MCWRK = 0.500000*d11"
"MCREST = d11 - d11"
# 37 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
1 ze
# 1 "mc_line 37 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding
definition of mc command after ze"
# 38 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
  3m pl12:f2
  3m pl16:f3
# 1 "mc_line 40 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding start
label for mc command"
2 MCWRK  do:f2
LBLSTS1, MCWRK
LBLF1, MCREST
# 41 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"
  d12 do:f3
  d12 setnmr3|0 setnmr0|34|32|33
  p16:gp0
  d16
  d12 setnmr3^0 setnmr0^34^32^33
  d12 pl9:f1
  d1 cw:f1 ph29
  d13 do:f1
  d12 pl1:f1
3 (p1 ph1)
    d12
    p16:gp1
    DELTA6 pl0:f2 pl3:f3
  (CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (CEN_HN2 p22 ph0):f3
    d12
    DELTA2
    d15 pl2:f2
  (p3 ph20 d13 p3 ph10):f2
;
    d12
    p16:gp1
    DELTA3
    p16:gp2
    d16 pl0:f2
  (CEN_HC2 p2 ph10) (p24:sp24 ph6):f2 (DELTA4 p21 ph6):f3 (CEN_HN2 p22 ph0):f3
    d12
    DELTA3
    p16:gp2
    d25 pl2:f2
   (DELTA5 p3 ph16 d13 p3 ph10):f2 (p21 ph16):f3
;
```

```
   d0
  p1 ph2
  d13
  d12 pl9:f1 setnmr3|0 setnmr0|34|32|33
  p16:gp3
  d16
  d12 setnmr3^0 setnmr0^34^32^33
  d18 cw:f1
  d13 do:f1
  d12 pl1:f1
  d12 pl12:f2
  d12 pl16:f3
  p1 ph3
  go=2 ph31 cpd2:f2 cpd3:f3
# 1 "mc_line 86 file /opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2 expanding mc
command in line"
  MCWRK  do:f2 do:f3 wr #0 if #0 zd dp2
  lo to LBLSTS1 times 2
  MCWRK id0
  lo to LBLF1 times ST1CNT
# 87 "/opt/xwinnmr/exp/stan/nmr/lists/pp/qnoesyphpr.rb2"

exit


ph0=0
ph1=0 2
ph20=0 0 2 2
ph6=0 2
ph16=0 0 0 0 2 2 2 2
ph10=1


ph2=0 0 0 0 0 0 0 0  0 0 0 0 0 0 0 0  2 2 2 2 2 2 2 2  2 2 2 2 2 2 2 2
ph3=0 0 0 0 2 2 2 2  1 1 1 1 3 3 3 3
ph29=0
ph31=0 2 0 2 2 0 2 0  1 3 1 3 3 1 3 1  2 0 2 0 0 2 0 2  3 1 3 1 1 3 1 3
;pl1 : f1 channel - power level for pulse (default)
;pl9 : f1 channel - power level for presaturation
;p1 : f1 channel -  90 degree high power pulse
;d0 : incremented delay (2D)
;d1 : relaxation delay; 1-5 * T1
;d8 : mixing time
;d11: delay for disk I/O                             [30 msec]
;d12: delay for power switching                      [20 usec]
;d13: short delay                                    [4 usec]
;in0: 1/(1 * SW) = 2 * DW
;nd0: 1
;NS: 8 * n
;DS: 16
;td1: number of experiments
;FnMODE: States-TPPI, TPPI, States or QSEC

;Processing

;PHC0(F1): 90
;PHC1(F1): -180
;FCOR(F1): 1

;$Id: noesyphpr,v 1.4 2002/06/12 09:05:10 ber Exp $
```