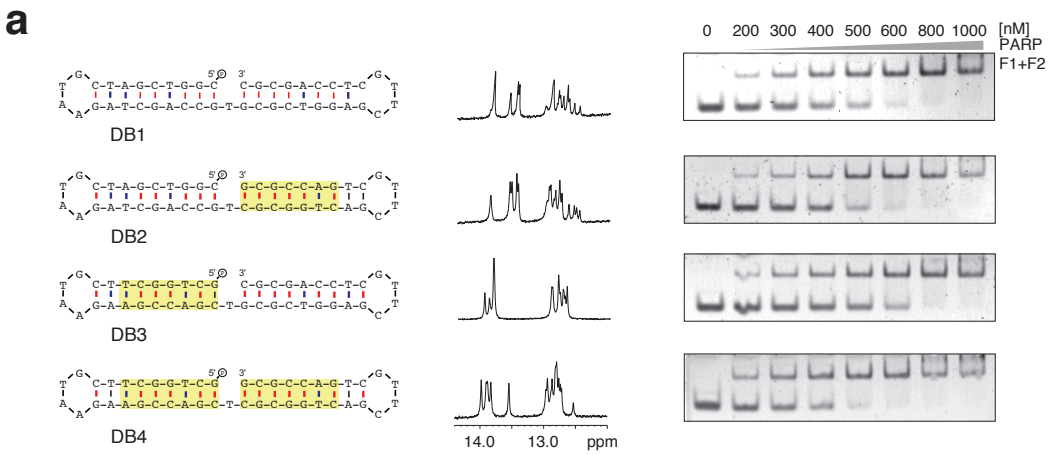
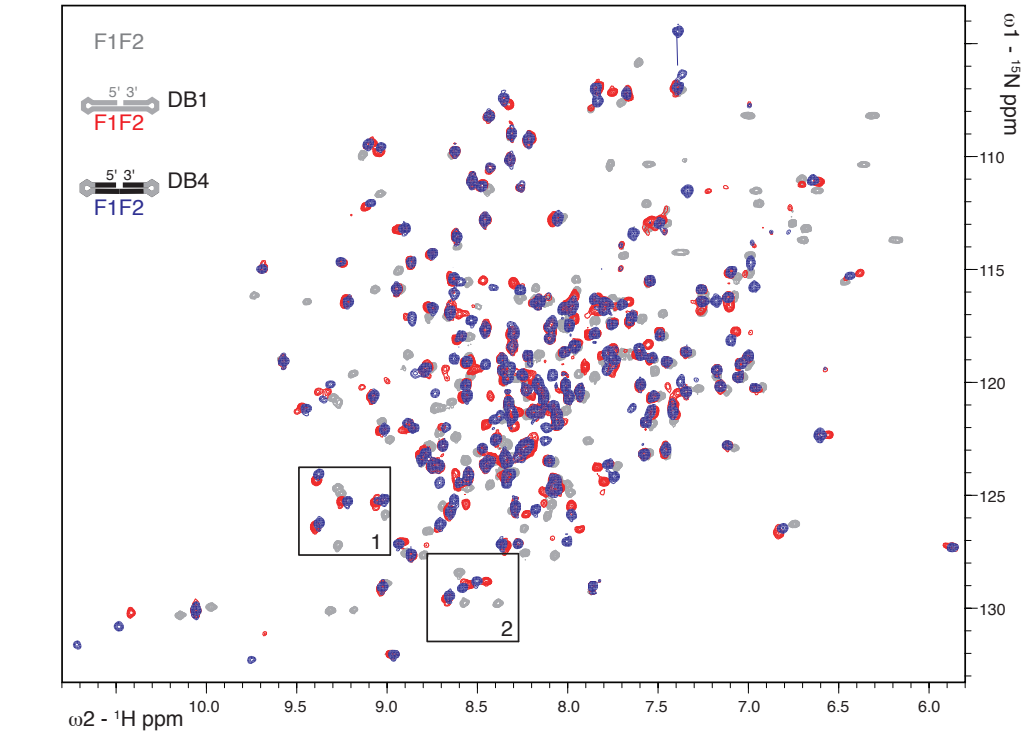
Molecular Cell, Volume *60*

Supplemental Information

Structural Basis of Detection and Signaling of DNA Single-Strand Breaks by Human PARP-1

Sebastian Eustermann, Wing-Fung Wu, Marie-France Langelier, Ji-Chun Yang, Laura E. Easton, Amanda A. Riccio, John M. Pascal, and David Neuhaus


a
0 200 300 400 500 600 800 1000 [nM]
PARP
F1+F2
DB1
DB2
DB3
DB4
14.0 13.0 ppm


b
F1F2
DB1
F1F2
DB4
F1F2
1
2
ω1 - 15N ppm
110
115
120
125
130
ω2 - 1H ppm
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0

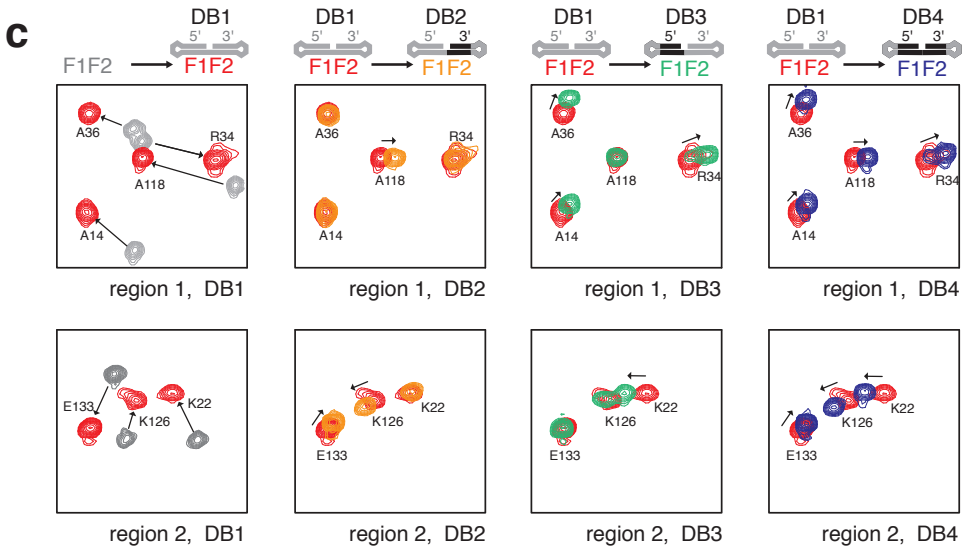

c
DB1
F1F2 → F1F2
DB1 DB2
F1F2 → F1F2
DB1 DB3
F1F2 → F1F2
DB1 DB4
F1F2 → F1F2
A36
R34
A118
A14
region 1, DB1
region 1, DB2
region 1, DB3
region 1, DB4
E133
K22
K126
region 2, DB1
region 2, DB2
region 2, DB3
region 2, DB4

Figure S1, Related to Figure 1:

Comparison of CSPs induced by different DNA dumbbell sequences reveals the orientation of F1F2 on the SSB ligand. a) DNA dumbbell sequences. DB1 was used in Eustermann *et al.* (2011), DB2 differs only in the 3' stem, DB3 differs only in the 5' stem, while DB4 (as used elsewhere in the present work, since complex F1F2-DB4 gave slightly higher quality spectra than did F1F2- DB1) combines both sets of variations. In each case, the imino region of the ^{1}H 1D spectrum confirms that basepairing in the stems is maintained, and an EMSA assay confirms that F1F2 binding occurs with comparable affinity. b) [^{15}N,^{1}H] correlation spectra of F1F2 in the free state (gray; HSQC), bound to DNA dumbbell ligand DB1 (red; TROSY) or bound to DNA dumbbell ligand DB4 (blue; TROSY); TROSY spectra were displaced by $^{1}J(^{15}N,^{1}H)/2$ in each dimension to make their peak positions comparable to those in the HSQC spectra. c) Expansions of regions 1 and 2 (defined in panel b) for complexes with dumbbells DB1-DB4; protein signal assignments are shown on the plots. The first box in each series shows CSPs of free protein signals on binding dumbbell DB1, while subsequent boxes compare complexes with different DNA dumbbell sequences. When the sequence of the 3' stem of the DNA dumbbell is varied, it is exclusively signals from F2 that are perturbed (A118, K126 and E133), whereas when the sequence of the 5' stem is varied only signals from F1 are perturbed (A14, K22, R34 and A36). When both are varied, protein perturbations are cumulative.

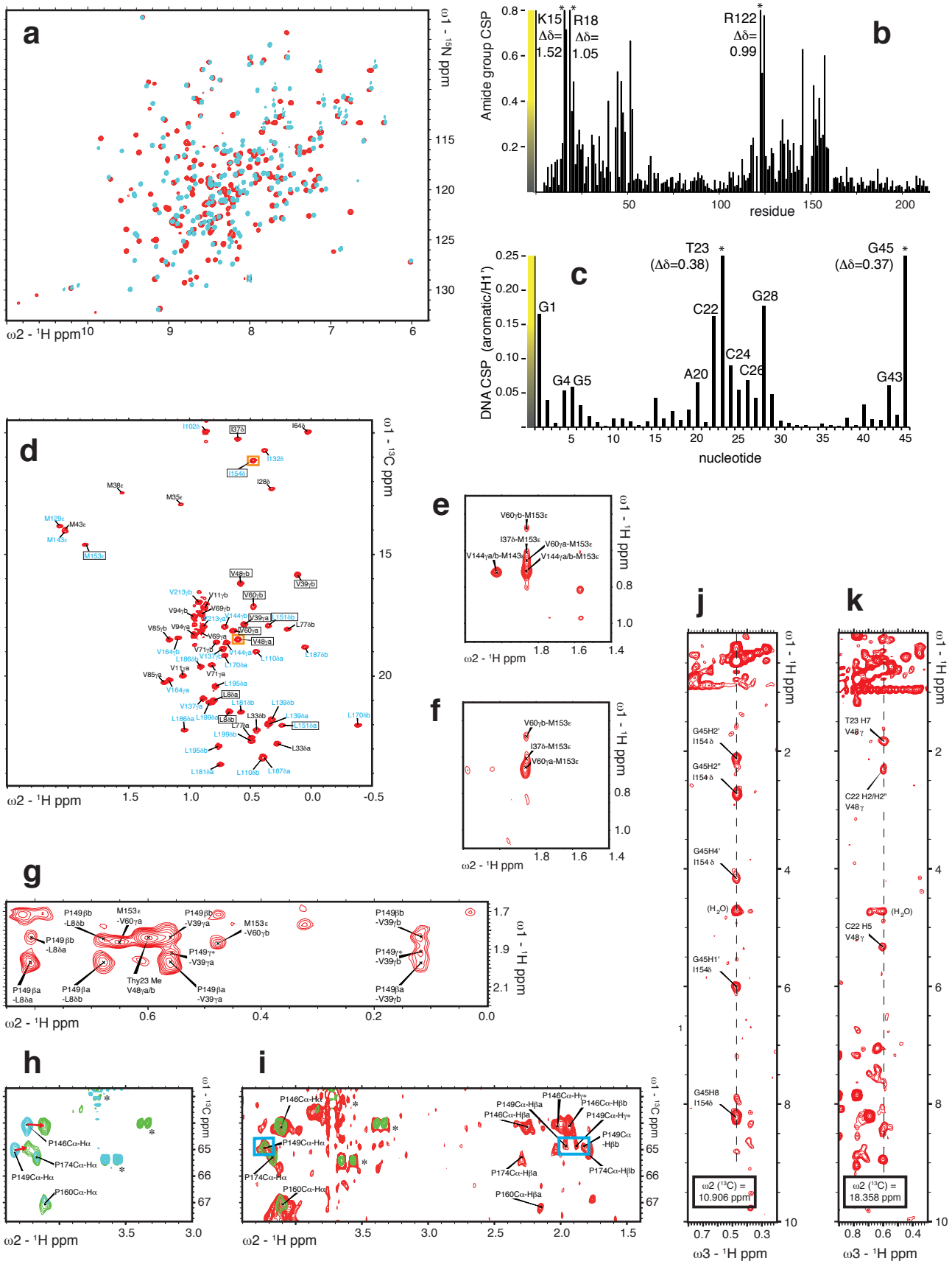

a
ω1 - ¹⁵N ppm
ω2 - ¹H ppm
b
Amide group CSP
K15 Δδ= 1.52
R18 Δδ= 1.05
R122 Δδ= 0.99
residue
c
DNA CSP (aromatic/H1')
G1
G4 G5
A20
C22
T23 (Δδ=0.38)
C24
C26
G28
G43
G45 (Δδ=0.37)
nucleotide
d
ω1 - ¹³C ppm
ω2 - ¹H ppm
e
V60γb-M153ε
I37δ-M153ε
V60γa-M153ε
V144γa/b-M143ε
V144γa/b-M153ε
f
g
h
i
P146Cα-Hα
P174Cα-Hα
P149Cα-Hα
P160Cα-Hα
j
k
G45H2'
G45H2"
G45H4'
(H2O)
G45H1'
G45H8
ω2 (¹³C) = 10.906 ppm
T23 H7
C22 H2/H2"
C22 H5
ω2 (¹³C) = 18.358 ppm
ω3 - ¹H ppm

Figure S2, Related to Figure 1:

Representative NMR data used during the signal assignment and structure determination of the F1F2-DNA dumbbell complex; all labeling schemes referred to below are defined in Supplemental Experimental Procedures. a) Overlay of [^{15}N,^{1}H] HSQC spectra of F1F2 in the free (cyan) and DNA-bound (red) states (for clarity, assignments are not marked), recorded at 800 MHz and 37°C. These data were used to calculate the CSP values shown in b), which in turn were used to generate the color ramp mapped on the molecular structure in Fig. 2; the color ramp runs from gray ($\Delta\delta = 0$) to yellow ($\Delta\delta = 0.4$; $\Delta\delta > 0.4$ shown yellow). c) Corresponding CSP values measured for the DNA, using the formula $((\Delta\delta(H1'))^2 + (\Delta\delta(H^{arom}))^2)^{1/2}$ (where H^{arom} is H8 for Ade, H6 for Cyt, H8 for Gua and H6 for Thy); in this case the color ramp runs from gray ($\Delta\delta = 0$) to yellow ($\Delta\delta = 0.15$; $\Delta\delta > 0.15$ shown yellow). d) Methyl region of a 2D [^{13}C,^{1}H] HMQC spectrum of the F1F2-DNA dumbbell complex; only methyl signals of Ile, Leu, Val and Met residues appear since just these are "reverse-labeled" in an otherwise deuterated background (labeling scheme IVa). Here methyls appear from both finger F1 (assignments labeled in black; all methyls assigned) and from finger F2 (assignments labeled in cyan; only Val202 unassigned), but these contributions were separated in other samples produced using sortase ligation, in which residues of only one finger were reverse-labeled. Methyls for which interdomain or intermolecular NOE interactions were used in the structure calculations appear boxed in black. Panels e) and f) show 2D NOESY experiments designed to confirm assignments of cross peaks linking Val60 and Met153. In e), the sample is reverse-labeled (Ile, Leu, Val and Met residues) in both fingers F1 and F2 (labeling scheme V), and the spectrum includes crosspeaks at positions corresponding to both intradomain (V144/M143) and interdomain (V60/M153 and I37/M153) interactions. Repeating the experiment using a sortase ligated sample in which only finger F2 was reverse-labeled, in this case only with Met (and Arg) residues, while finger F1 uniformly contained ^{1}H and ^{12}C (labeling scheme VIII) eliminates the crosspeaks due to Val/Met interactions within finger F2, leaving only the interdomain crosspeaks to Val60. These assignments (and others) were all confirmed by observation of corresponding crosspeaks in both 3D [^{1}H,^{13}C,^{1}H] NOESY-HMQC and 3D [^{13}C,^{13}C,^{1}H] HMQC-NOESY-HMQC spectra (Table S2). g) Part of a filtered 2D [^{1}H, ^{1}H] NOESY spectrum of F1F2-DNA dumbbell complex. Here F1F2 was produced by sortase ligation such that finger F1 was reverse-labeled ([^{13}C, ^{1}H] methyl groups for Ile, Leu, Val and Met in a deuterated background), while finger F2 uniformly contained ^{1}H and ^{12}C (labeling scheme VII), and in this filtered NOESY experiment only ^{13}C-bound protons were selected in ω2, while no selection was made in ω1. NOE crosspeaks in this spectrum therefore arise from interactions between ^{13}C-labeled methyls in finger F1 (appearing with shifts in ω2) and protons in either finger F2 or the DNA (appearing with shifts in ω1). The region shown contains, *inter alia*, crosspeaks linking methyls of Val39 and Leu8 to multiplets at 1.8-2.0 ppm assigned to Pro149. Panels h) and i) show spectra that were used to confirm the assignment of these multiplets to the Hβ and Hγ protons of Pro149 (asterisks indicate small molecule impurities). Back-labeling the prolines of (only) finger F2 with [^{13}C,^{1}H] in a deuterated context (labeling scheme XI) allowed unambiguous identification of the bound-state Hα and Cα signals of Pro149 (by comparing

free-state (green) and bound-state (cyan) HSQC spectra (panel h) in combination with knowledge of Cα assignments from backbone experiments in both free and bound states). This assignment was then extended to the Hβ and Hγ bound-state signals using a 2D [^{13}C,^{1}H] HSQC-NOESY spectrum of the complex (panel i); the Pro149 signals are shown boxed, and correspond closely to those seen for Pro149 in panel g). Panels j) and k) show parts of (ω1, ω3) planes from a 3D [^{1}H,^{13}C,^{1}H] NOESY-HMQC spectrum of the same sample as in panel d), showing intermolecular NOE cross peaks linking j) Ile154δ to protons of nucleotide G45, and k) both γ methyls of Val48 (too close together to be resolved in this spectrum) to protons of nucleotide C22; the positions of Ile154δ and Val48γ methyl signals are indicated with orange boxes in panel d).

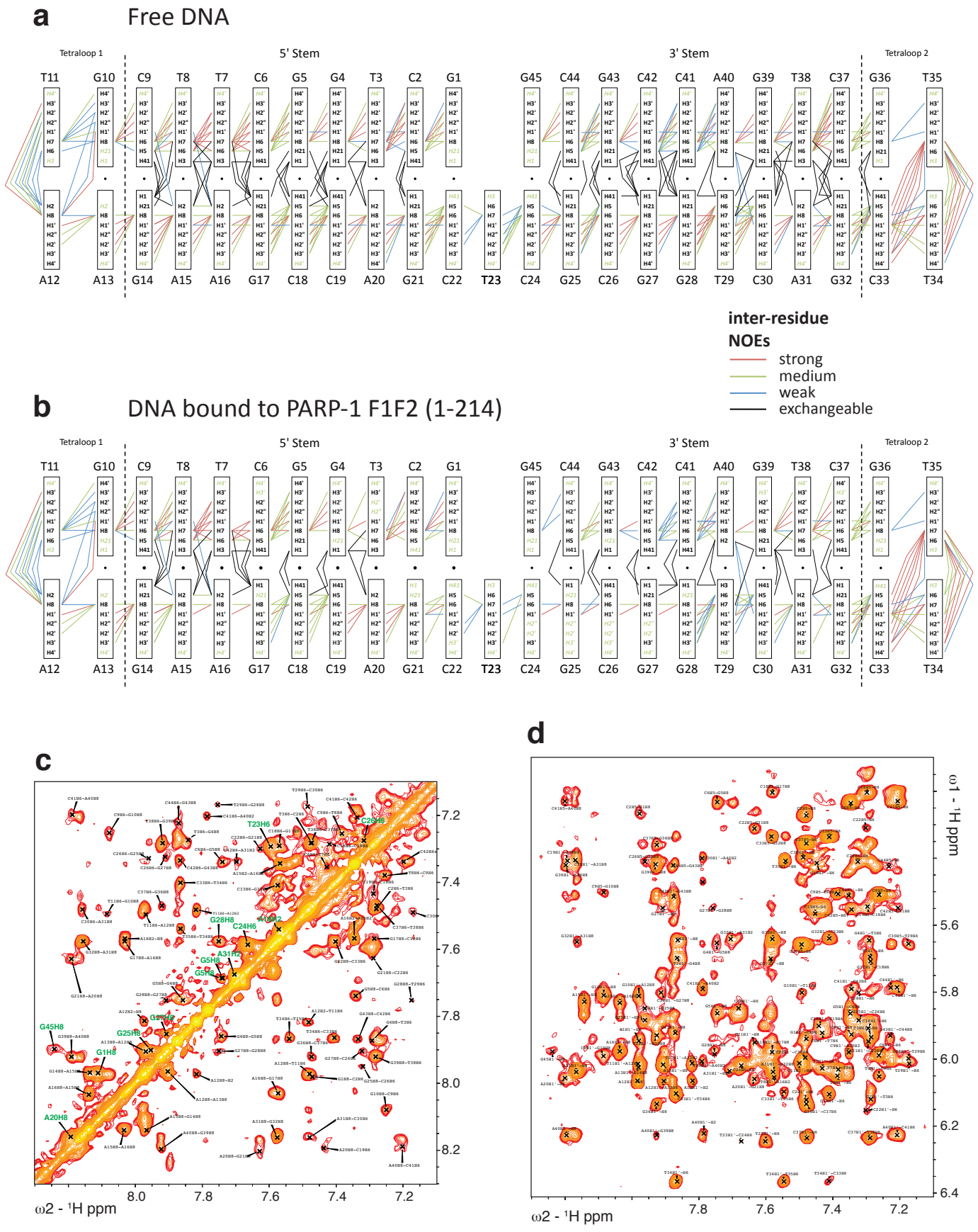

a
Free DNA
Tetraloop 1
5' Stem
3' Stem
Tetraloop 2
T11 G10 C9 T8 T7 C6 G5 G4 T3 C2 G1 G45 C44 G43 C42 C41 A40 G39 T38 C37 G36 T35
A12 A13 G14 A15 A16 G17 C18 C19 A20 G21 C22 T23 C24 G25 C26 G27 G28 T29 C30 A31 G32 C33 T34
inter-residue
NOEs
strong
medium
weak
exchangeable
b
DNA bound to PARP-1 F1F2 (1-214)
Tetraloop 1
5' Stem
3' Stem
Tetraloop 2
T11 G10 C9 T8 T7 C6 G5 G4 T3 C2 G1 G45 C44 G43 C42 C41 A40 G39 T38 C37 G36 T35
A12 A13 G14 A15 A16 G17 C18 C19 A20 G21 C22 T23 C24 G25 C26 G27 G28 T29 C30 A31 G32 C33 T34
c
T23H6
C22H6
G28H8
C24H6
A31H2
G5H8
G5H8
G45H8
G25H8
G1H8
A20H8
7.2
7.4
7.6
7.8
8.0
8.2
8.0
7.8
7.6
7.4
7.2
ω2 - ¹H ppm
d
ω1 - ¹H ppm
5.6
5.8
6.0
6.2
6.4
7.8
7.6
7.4
7.2
ω2 - ¹H ppm

Figure S3, Related to Figure 1:

Comparison of free- and bound-state DNA signal assignments. Charts summarizing all the sequential NOE connectivities found during assignment of a) free and b) bound states of the 45nt gapped DNA dumbbell ligand. The similarity of the patterns (allowing for the reduced sensitivity and increased complexity of spectra of the complex) shows that the secondary structure of the dumbbell is maintained in the complex. Non-assigned protons are indicated in green italics (for simplicity, H5' and H5" signals are omitted from the schemes; in the free state, these were assigned for G1, G4, C6, T7, T8, G10 and T35, while for the bound state they were assigned only for T8). No attempt was made to quantitate crosspeaks involving exchangeable signals as the intensity of these varies greatly due to their different exchange rates.

Panels c) and d) show regions of a 2D [^{1}H,^{1}H] NOESY spectrum of the F1F2-DNA complex, showing data quality and extent of assignment. This sample contained a slight excess of DNA, which led to clear exchange crosspeaks linking corresponding free and bound signals (assignments shown in green on panel c); these were very useful in making assignments for those signals that shifted substantially upon complex formation.

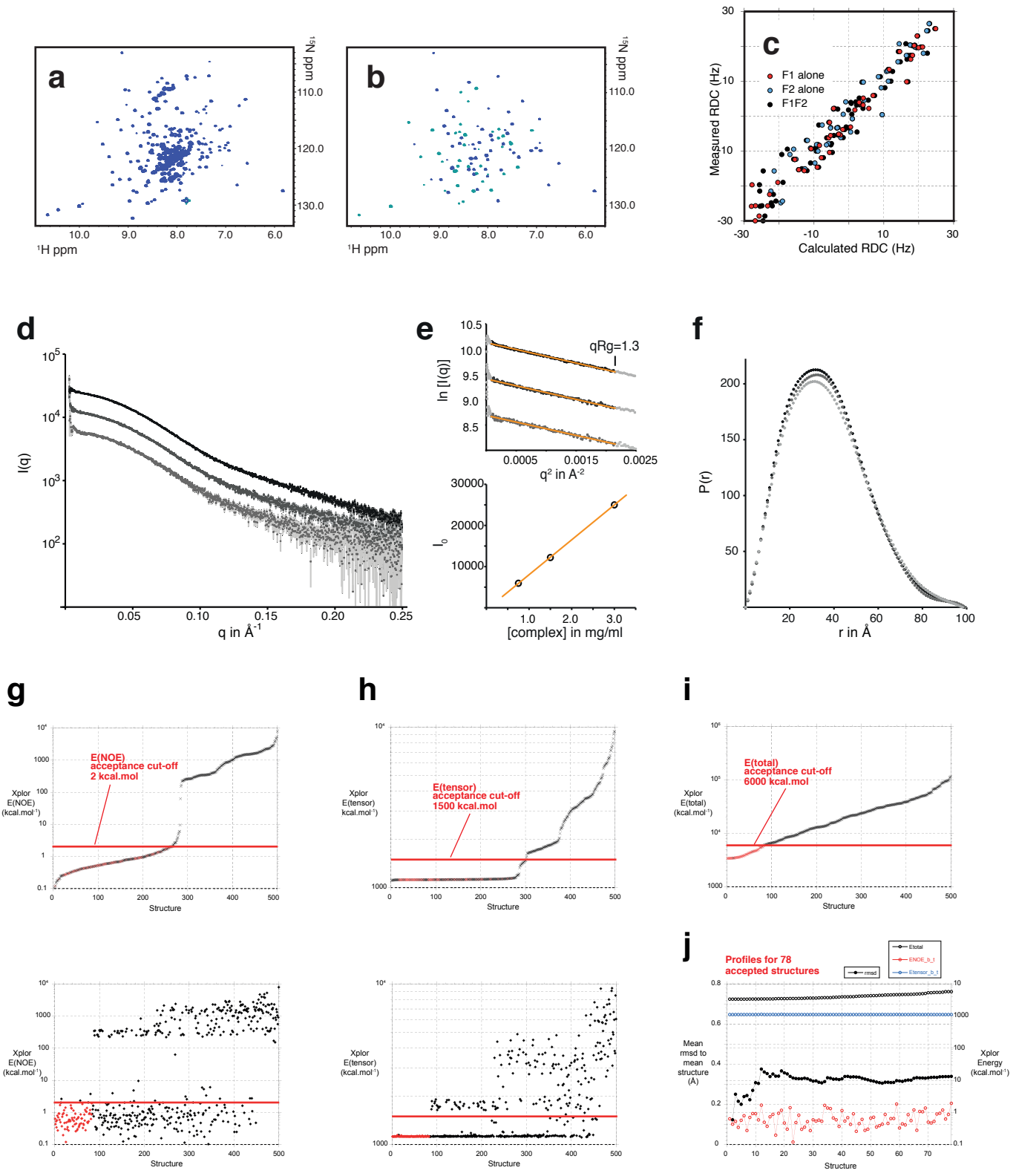

a
b
c
^{15}N ppm
^{1}H ppm
Measured RDC (Hz)
Calculated RDC (Hz)
F1 alone
F2 alone
F1F2
d
I(q)
q in Å^{-1}
e
ln [I(q)]
qRg=1.3
q^2 in A^{-2}
I$_0$
[complex] in mg/ml
f
P(r)
r in Å
g
h
i
E(NOE) acceptance cut-off 2 kcal.mol
E(tensor) acceptance cut-off 1500 kcal.mol
E(total) acceptance cut-off 6000 kcal.mol
Xplor E(NOE) (kcal.mol^{-1})
Xplor E(tensor) kcal.mol^{-1})
Xplor E(total) (kcal.mol^{-1})
Structure
j
Profiles for 78 accepted structures
Etotal
ENOE_b_t
Etensor_b_t
rmsd
Mean rmsd to mean structure (Å)
Xplor Energy (kcal.mol^{-1})

Figure S4, Related to Figure 1:

a) and b) ARTSY spectra used to measure RDC values for the F1F2-DNA dumbbell complex. In the presence of pf1 phage, the sample becomes aligned and ^{15}N-^{1}H splittings become a composite of $^{1}J(^{15}N,^{1}H)$ and RDC values, which are measured indirectly using the intensity difference for a given peak between the "reference" (panel a) and "attenuated" (panel b) versions of the ARTSY spectrum (Fitzkee et al., 2010). c) Amide group RDC values measured for F1F2-DNA complex aligned by pf1 phage (see Supplemental Experimental Procedures) are plotted against corresponding values calculated using co-ordinates derived from crystal structures of F1-DNA complex (pdb 3ODA) and F2-DNA complex (pdb 3ODC); the RDC data used were restricted to residues shown to be rigid by the ^{15}N relaxation data and the conformations of which were invariant amongst different copies of the protein in the crystal structures. When F1 and F2 are treated separately (i.e. when each has its own independent alignment tensor), the Q values (18.9% for F1 and 29.9% for F2) show that in each case the structures were preserved to the degree expected for X-ray structures in this resolution range (2.6-2.8Å) (Bax et al., 2005), thereby demonstrating that the internal structures of F1 and F2 are similar in solution and in the crystals. When a single, combined tensor is used during the structure calculation protocol, this causes F1 and F2 to adopt a specific mutual orientation that optimises agreement with the RDC data. It is clear both from the plot and from the Q value (24.5%) that the fit obtained using a combined tensor is of comparable quality to that obtained using independent tensors, showing that the mutual orientation of the fingers in the calculated structure of the complex has converged to an optimal fit to the RDC data.

Panels d)-f) show measures of data quality for the Small Angle X-ray Scattering (SAXS) experiments. d) SAXS profiles of PARP-1 F1+F2 in complex with gapped 5'-phosphorylated DNA (DB4) at three different concentrations (3mg/ml black, 1.5mg/ml grey, 0.75 mg/ml light grey). Errors are indicated by light grey bars. e) Particle interference was shown to be absent using a Guinier analysis within the $qRg<1.3$ limit, as well as by demonstrating a direct proportionality between I_0 (derived from P(r)) and the sample concentrations used. f) Pair-distance distribution functions of the three data sets are largely consistent.

Panels g)-j) illustrate structure selection for the calculated structures of the F1F2-DNA dumbbell complex. Structures were included in the accepted ensemble if they simultaneously met the three criteria E(total) <= 6000 kcal.mol^{-1}, E(tensor) <=1500 kcal.mol^{-1} and E(NOE) <= 2 kcal.mol^{-1}; here E(total) represents the total XPLOR-NIH energy term (Schwieters et al., 2003), E(NOE) describes the quality of the fit to the NOE restraints and E(tensor) describes the quality of the fit to the RDC restraints. g)-i) Profiles showing the variation in energy terms g) E(NOE), h) E(tensor), and i) E(total) across all 500 calculated structures; accepted structures are shown in red, rejected in black, and the acceptance criteria for each energy term is shown with a horizontal red line. In g) the upper panel shows points ordered by increasing E(NOE) values, while the lower panel shows the same data points but reordered according to increasing E(total) values; similarly, in h) the upper panel shows points ordered by increasing E(tensor), the lower panel shows the same points re-ordered by increasing

E(total). These plots show how the acceptance criteria were chosen; for E(NOE) and E(tensor), it is clear that the accepted structures all lie within well-defined minimum plateau regions, but that these plateau regions also contain many structures with high E(total) values. E(total) varies over a wide range across the ensemble; this is not unexpected, since structures were calculated starting from conformers in which the 10 rotatable bonds of the DNA linker had been completely randomized and the ability of the calculation to converge on the lowest energy solution is strongly dependent on starting conformation. The acceptance criterion of E(total)<=6000 kcal.mol^{-1} was chosen since this corresponds to roughly the highest value of E(total) for which structures with higher E(NOE) and E(tensor) values (outside the plateau regions) are largely excluded. However, as this single criterion was not sufficient to exclude a minority of structures with E(total)<=6000 kcal.mol^{-1} but poor E(NOE) and/or E(tensor) values, we employed selection based on the three simultaneous criteria for E(total), E(NOE) and E(tensor) as described above. j) Profiles for the 78 accepted structures, showing values for E(total), E(NOE) and E(tensor); in all cases structures in this plot are arranged in order of increasing E(total). For the rmsd profile, values were independently calculated for each ensemble size using the program CLUSTERPOSE (Diamond, 1992, 1995), adding successive structures in order of increasing E(total).

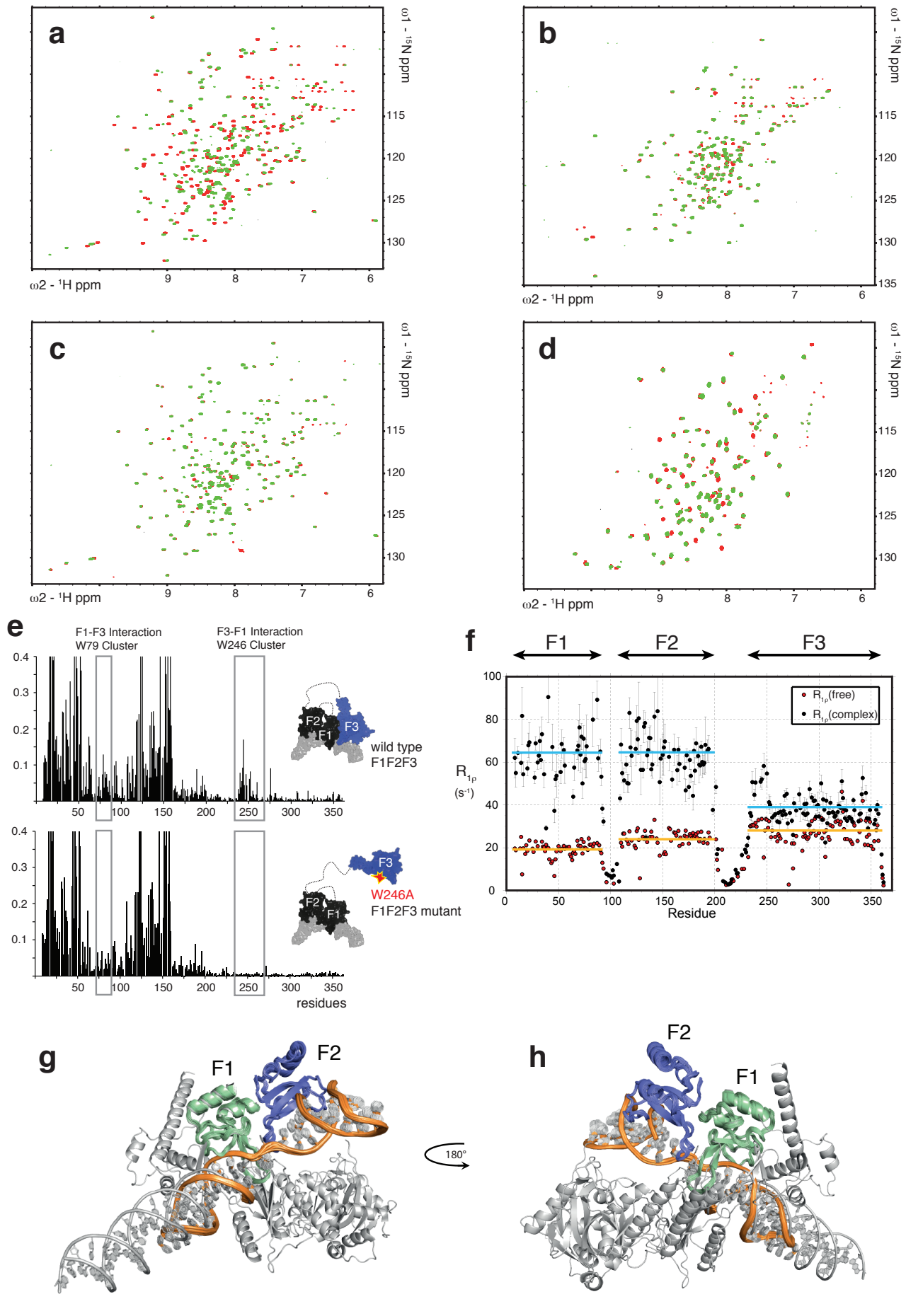

a
b
c
d
ω1 - 15N ppm
ω2 - 1H ppm
e
F1-F3 Interaction
W79 Cluster
F3-F1 Interaction
W246 Cluster
wild type
F1F2F3
W246A
F1F2F3 mutant
residues
f
F1
F2
F3
R1ρ(free)
R1ρ(complex)
R1ρ
(s-1)
Residue
g
h
180°

Figure S5, Related to Figure 4:

Interactions of the F1F2-DNA dumbbell complex with the F3 and WGR domains of PARP-1. Panels a)–d) show the NMR data (all acquired at 800 MHz and 30°C) that were used to derive the chemical shift perturbations plotted in Figure 3. a) Overlay of TROSY spectra for the free (red) and DNA-bound (green) states of a sortase–ligated sample of F1F2F3 in which only F1F2 contributes signals (F1F2F3 labeling scheme II, see Supplemental Experimental Procedures). b) Overlay of TROSY spectra for the free (red) and DNA-bound (green) states of a sortase–ligated sample of F1F2F3 in which only F3 contributes signals. c) Overlay of TROSY spectra for the bound states of F1F2 (red) and F1F2F3 (green), signals from the latter being restricted to those from F1F2 by sortase ligation; these data were used to identify interaction surfaces between F3 and F1F2-DNA in the complex. d) Overlay of HSQC spectra for the free (red) and F1F2F3-DNA-bound (green) states of the WGR domain. e) CSP values observed upon DNA binding of the F1F2F3 W246A mutant (lower panel) plotted against corresponding CSP values for wild-type F1F2F3 (upper panel). The W246A mutation clearly abrogates the DNA-dependent interaction of F3 with F1 (CSPs in the W246 cluster are abolished in the mutant). The CSPs due to F3 interaction observed on F1 are also reduced in the mutant, though here interpretation is more difficult due to the large background of CSPs due to the DNA-F1 interaction. (For the wild-type protein, the data shown here were combined from measurements using the same two sortase ligated samples as used to construct the plots in Figure 3 of the main paper). f) ^{15}N $R_1\rho$ data ($1/T_1\rho$) for free and DNA-bound F1F2F3; average values are indicated for each domain in the free protein (orange) and the DNA-bound protein (cyan). The F1 and F2 domains show a strong enhancement of relaxation in the bound state, consistent with tumbling as a single entity in combination with the DNA, whereas the F3 signals show a more modest increase in relaxation rate, suggesting the interaction of F3 with the rest of the complex is relatively weak. The F1F2 and F2F3 linkers clearly remain flexible in the complex, as relaxation rates for linker residues are low.

g) and h) Superposition of the F1F2-DNA dumbbell structure with the crystal structure of F1, F3 and WGR-CAT bound to a DNA blunt end (pdb 4DQY); the structures were fitted using the backbone atoms of F1 (N, Cα and C' atoms of residues 2-7) and its associated DNA stem (C1', C2', C3', C4', C5', O3', O4' O5' and P of first 7 basepairs only). It is clear that this fit positions F2 and its associated DNA stem into a region of free space, without steric clashes.

Construct	Template used for (sub)cloning	Vector	Restriction sites used for cloning	Sequence boundaries	Expression/purification tag and protease site (N- and *C*-terminal)	Vector derived N- and *C*-terminal residues of purified construct
F1 (SL)	MGC clone IMAGE 5193735	pET28a	BamHI/NcoI	1-102	LPXTG sortase motif, His6	**None** *LPETGGGRRHHHHHH*
F2 (SL)	MGC clone IMAGE 5193735	pET28a	BamHI/NcoI	105-214	**His6 TEV site** *none*	**GGG** *none*
F1F2	MGC clone IMAGE 5193735	pET13	BamHI/NcoI	1-214	**none** *none*	**none** *none*
F1F2 (SL)	codon-optimized PARP-1 gene (pQE7 plasmid; Qiagen)	pET28a-lip	NdeI/XhoI	1-214	**none** *LPXTG sortase motif, His6*	**none** *LPETGGGRRHHHHHH*
F3	MGC clone IMAGE 5193735	pGEX-6P1	BamHI/EcoRI	215-362	**GST-domain** *none*	**GPGS** *none*
F3 (SL)	codon-optimized PARP-1 gene (pQE7 plasmid; Qiagen)	pET28a-lip	NdeI/XhoI	215-362	**His6, lipoyl domain, TEV** *none*	**GG** *none*
F1F2F3	codon-optimized PARP-1 gene (pQE7 plasmid; Qiagen)	pET28a-lip	NdeI/XhoI	1-362	**His6, lipoyl domain, TEV site** *none*	**G** *none*
WGR	codon-optimized PARP-1 gene (pQE7 plasmid; Qiagen)	pET28a-lip	NdeI/XhoI	523-649	**His6, lipoyl domain, TEV site** *none*	**G** *none*
Full length PARP-1	(Langelier et al. 2008)	pET28a	NdeI/XhoI	1-1014	**His6, Thrombin site** *none*	**MGSSHHHHHHSSGLVPRGSH** *none*
Full length PARP-1	(Langelier et al. 2008)	pET28a-SMT	NdeI/XhoI	1-1014	**His6, SUMO domain** *none*	**His6, SUMO domain** *none*
GFP-PARP(F1)	(Langelier et al. 2011)	pEGFP-N1	NdeI/XhoI	1-234 Δ97-216	**eGFP domain** *none*	Construct not purified; used for *in vivo* localization study
GFP-PARP(F2)	(Langelier et al. 2011)	pEGFP-N1	NdeI/XhoI	97-234	**eGFP domain** *none*	Construct not purified; used for *in vivo* localization study
GFP-PARP(F1F2)	(Langelier et al. 2011)	pEGFP-N1	NdeI/XhoI	1-234	**eGFP domain** *none*	Construct not purified; used for *in vivo* localization study
GFP-PARP(ΔF1F2)	(Langelier et al. 2011)	pEGFP-N1	NdeI/XhoI	201-1014	**eGFP domain** *none*	Construct not purified; used for *in vivo* localization study
GFP-PARP(wt)	(Langelier et al. 2011)	pEGFP-N1	NdeI/XhoI	1-1014	**eGFP domain** *none*	Construct not purified; used for *in vivo* localization study
wt Sortase Δ59 (Srt)	pGBMCS containing Srt Δ59 (gift Inagaki Lab)	pGBMCS (pet21b)	BamHI/XhoI	60-	**GB1 domain** *His6*	**GB1 domain** *HHHHHH*
Mutant Sortase Δ59 (SrtM) P94S, D160N, D165A, K196T	Synthesized gene (IDT) mutations according to (Liu et al. reference)	pGBMCS (pet21b)	BamHI/XhoI		**GB1 domain** *His6*	**GB1 domain** *HHHHHH*

Table S1, Related to Figure 1: PARP-1 protein constructs used in this work.

Intermolecular NOE	Expt./Sample		Interdomain NOE	Expt./Sample
Ile154 Hδ – G45 H1'	1, 2, 3, 4		Val60 Hγa – Met153 Hε	4, 5, 6, 7, 8, 9
Ile154 Hδ – G45 H2'	1, 2, 3		Val60 Hγb – Met153 Hε	4, 5, 6, 7, 8, 9
Ile154 Hδ – G45 H2"	1, 2, 3, 4		Ile37 Hδ – Val144 Hγa	8
Ile154 Hδ – G45 H3'	1, 2, 4		Ile37 Hδ – Val144 Hγb	8
Ile154 Hδ – G45 H4'	1, 2, 3, 4		Ile37 Hδ – Met153 Hε	5, 6, 7, 8
Ile154 Hδ – G45 H8	1, 2, 3, 4		Val39 Hγa – Pro149 Hβa	4
Leu151 Hδa - G45 H1'	1, 2, 3, 4		Val39 Hγa – Pro149 Hβb	4
Leu151 Hδa - G45 H2"	3		Val39 Hγb – Pro149 Hβa	4
Leu151 Hδb - G45 H1'	4		Val39 Hγb – Pro149 Hβb	4
Val48 Hγa – T23 M7	3, 4		Val39 Hγa – Pro149 Hγa/b	4
Val48 Hγb - T23 M7	3, 4		Val39 Hγb – Pro149 Hγa/b	4
Val48 Hγa – C22 H5	3, 4		Leu8 Hδa – Pro149 Hβa	4
Val48 Hγb – C22 H5	3, 4		Leu8 Hδa – Pro149 Hβb	4
Val48 Hγa/b – T23 H6	4		Leu8 Hδb – Pro149 Hβa	4
Val48 Hγa/b – C22 H1'	4		Leu8 Hδb – Pro149 Hβb	4
V48 Hγa – C22 H2'/H2"	3, 4			
V48 Hγb – C22 H2'/H2"	3, 4			

Table S2, Related to Figure 1: Intermolecular (left) and interdomain (right) NOE interactions observed for the F1F2-DNA dumbbell complex. For each of these assigned NOE interactions, the list of spectra in which the corresponding cross peaks were seen is not comprehensive, but lists cases where assignment was unambiguous. For definitions of isotope labelling schemes, see Supplemental Experimental Procedures.

Expt./Sample:

1) Experiment: 2D [^{1}H, ^{1}H] NOESY (τ_m = 200ms), filtered to accept only ^{13}C-bound ^{1}H signals in ω_2 (unfiltered in ω_1);
Sample: F1F2 isotope labelled according to scheme VII, in complex with 45nt gapped dumbbell DNA in $^{2}H_2O$.

2) Experiment: 2D [^{1}H, ^{1}H] NOESY (τ_m = 200ms), filtered to accept only ^{13}C-bound ^{1}H signals in ω_2 (unfiltered in ω_1);
Sample: F1F2 isotope labelled according to scheme VII, in complex with 45nt gapped dumbbell DNA in $^{2}H_2O$, no ^{13}C-decoupling applied in F_1 (only artefacts should differ between experiments 1 and 2).

3) Experiment: 3D [^{1}H, ^{13}C, ^{1}H] NOESY-HMQC (τ_m = 200ms);
Sample: F1F2 isotope labelled according to scheme IVa, in complex with 45nt gapped dumbbell DNA in H_2O.

4) Experiment: 2D [^{1}H, ^{1}H] NOESY (τ_m = 200ms), filtered to accept only ^{13}C-bound ^{1}H signals in ω_2 (unfiltered in ω_1);
Sample: F1F2 isotope labelled according to scheme IX, in complex with 45nt gapped dumbbell DNA in H_2O. Interdomain NOE crosspeaks listed for this experiment were all missing in a corresponding control spectrum acquired in the absence of DNA, showing that the interdomain interactions are DNA-dependent.

5) Experiment: 2D [^{1}H, ^{1}H] NOESY (τ_m = 200ms), filtered to accept only ^{13}C-bound ^{1}H signals in ω_2 (unfiltered in ω_1);
Sample: F1F2 isotope labelled according to scheme VIII, in complex with 45nt gapped dumbbell DNA in H_2O.

6) Experiment: 2D [^{1}H, ^{1}H] NOESY (τ_m = 200ms), filtered to accept only ^{13}C-bound ^{1}H signals in ω_2 (unfiltered in ω_1);
Sample: F1F2 isotope labelled according to scheme VIII, in complex with 45nt gapped dumbbell DNA in H_2O, no ^{13}C-decoupling applied in F_1 (only artefacts should differ between experiment 6 and 7).

7) Experiment: 2D [^{1}H, ^{1}H] NOESY (τ_m = 200ms), filtered to accept only ^{13}C-bound ^{1}H signals in ω_2 (unfiltered in ω_1);
Sample: F1F2 isotope labelled according to scheme V, in complex with 45nt gapped dumbbell DNA in $^{2}H_2O$.

8) Experiment: 3D [^{13}C, ^{13}C, ^{1}H] HMQC-NOESY-HMQC (τ_m = 200ms);
Sample: F1F2 isotope labelled according to scheme V, in complex with 45nt gapped dumbbell DNA in $^{2}H_2O$.

9) Experiment: 3D [^{1}H, ^{13}C, ^{1}H] NOESY-HMQC (τ_m = 200ms);
Sample: F1F2 isotope labelled according to scheme V, in complex with 45nt gapped dumbbell DNA in $^{2}H_2O$.

Supplemental Experimental Procedures

Protein constructs:

Plasmids coding for human PARP-1 constructs were (sub)cloned from a human cDNA library, a MGC cDNA clone [IMAGE 5193735 (Geneservice,UK)] or from a PQE7 plasmid containing the human PARP-1 sequence codon-optimized for expression in *E. coli* (Qiagen); for details see Table S1. Full-length PARP-1 expression constructs were cloned in the pET28 vector as described (Langelier et al., 2008; Langelier et al., 2011), or into a pET28 vector modified to include the coding sequence for the SUMO polypeptide tag (SMT) in-frame at the N-terminus of PARP-1 (SMT-PARP-1). PARP-1 fusions with GFP (WT: 1-1014, ΔF1F2: 201-1014, F1F2: 1-234, F1:1-234 with 97-216 deleted, and F2: 97-234) were cloned in the pEGFP mammalian expression vector as described (Steffen et al., 2014). The plasmid for bacterial expression of wild-type Staphylococcus aureus Sortase Δ59 (SrtΔ59M) was a generous gift of Fuyuhiko Inagaki (Hokaido University, Japan). The vector (pGBMCS; http://www.addgene.org/21931/) was based on pET21b and contains a non-cleavable N-terminal GB1 fusion tag and a non-cleavable C-terminal His6 tag. The sortase gene for SrtΔ59M containing the mutations P94S, D160N, D165A, K196T (Chen et al., 2011) was synthesized by IDT (Integrated DNA Technologies), and cloned into pGBMCS using BamHI and XhoI restriction sites. Vectors (based on pET28a) containing DNA sequences encoding Bacillus stearothermophilus dihydrolipoamide acetyltransferase lipoyl domain (pet28a-lip) and Streptococcal protein GB1 (pET28a-GB1) as N-terminal expression tags with TEVpro cleavage sites were a generous gift from Peter Lukavsky (CEITEC, Masaryk University, Brno). Site-directed mutagenesis was achieved by PCR using QuikChange II methodology (Stratagene) using KOD Hot Start DNA Polymerase (Novagen).

Recombinant expression and isotope labeling of PARP-1 constructs:

Full-length PARP-1 WT and mutants, and SMT-PARP-1 WT and E988K mutant were expressed and purified using a published protocol (Langelier et al., 2010; Langelier et al., 2011) that uses Ni^{2+} and heparin affinity followed by size exclusion chromatography. Constructs of human PARP-1 derived from codon-optimized sequences were recombinantly expressed in *E.coli* BL21-DE3, while BL21-CodonPlus(DE3)-RP cells (Stratagene) were used for all other PARP-1 constructs. Unless otherwise stated all isotopically labeled compounds were obtained from from Sigma Isotec or Cambridge Isotope Laboratory.

Uniform ^{13}C and ^{15}N isotope labeling of protein constructs was essentially as described in Eustermann *et al.* (2011). PARP-1 fragments F1, F2 and F1F2 were "reverse" labeled using adapted protocols from (Marley et al., 2001; Takeuchi et al., 2007; Tugarinov et al., 2006). In each case a colony of freshly transformed cells was cultured in LB medium at 37°C to A_{600}=0.6, then cells were pelleted at 4°C, washed in M9-salt solution lacking all nitrogen and carbon sources, then pelleted as before and resuspended in M9 minimal medium supplemented with $^{15}NH_4Cl$ (0.5 g/L), [$^{2}H_7{}^{13}C_6$]-glucose

(8 g/L) and 98% $^{2}H_2O$ (all Sigma Aldrich Isotec). The volume of this culture was approx. one quarter that of the initial culture in LB, yielding A_{600} after resuspension of approximately 2.4. After incubation for 1h at 37°C, the medium was supplemented with sodium α-ketoisovalerate and sodium α-ketobutyrate (see schemes IVa and IVb in the NMR spectroscopy section for the labeling patterns used for these precursors) and/or [^{1}H,^{13}C,^{15}N] uniformly labeled amino acids (methionine [250 mg/liter], proline [156 mg/liter], arginine [120 mg/liter], lysine [120 mg/liter], tyrosine [120 mg/liter], phenylalanine [120 mg/liter] and/or ^{1}H,^{15}N uniformly labeled tryptophan [150 mg/liter]) (all compounds from Sigma Isotec or Cambridge Isotope Laboratories). After further incubation for 1h at 22°C, cultures were supplemented with 0.5 mM $ZnSO_4$ and protein expression induced by adding 0.5 mM IPTG, followed by incubation for 8-10h at 22°C, then cells were harvested and proteins purified as described. Metabolic scrambling during isotope labeling by residue type was minimized by using a protocol from Takeuchi *et al.* (2007) and supplementing, when required, the respective M9 medium with 8.5 g/l deuterated ^{13}C ^{15}N Celtone (Cambridge Isotope Laboratories).

Uniform ^{2}H, ^{13}C and ^{15}N labeling of F1, F2. F1F2, F1F2F3, F3 and WGR was achieved using similar protocols as described above for reverse labeling. Cells for expression of F1F2F3, F3 and WGR were grown in 2xTY medium until an A_{600} of 2 was reached, then exchanged into 98% $^{2}H_2O$ M9 medium supplemented with [$^{2}H_7{}^{13}C_6$]-glucose (4 g/L) and $^{15}NH_4Cl$ (0.5 g/L) as the sole carbon and nitrogen sources. After additional incubation for 45 min at 37°C followed by 45min at 22°C, expression was induced as described above.

These protocols yielded protein-deuteration levels of >95% (assayed by NMR), and selective ^{13}C labeling and protonation were sufficiently efficient for the NMR experiments described to succeed. Amide protons were essentially fully back-exchanged by overnight incubation in aqueous buffer. Segmental labeling was achieved by using Sortase-mediated protein ligation as described below.

Protein purification:

All purification steps were carried out at 4°C or on ice. Proteins carrying N- or C-terminal residues for sortase ligation (see below) are designated (SL). F1 (SL), F2 (SL) and F1F2 were purified as described in Eustermann *et al.* (2011). Harvested cells were resuspended in lysis buffer containing 50 mM Tris pH 7.4, 150 μM $ZnSO_4$, 4 mM DTT, 25% (w/v) sucrose and protease inhibitor mix (Roche Complete Protease Inhibitor Cocktail EDTA free; 1 tablet per 50 ml). After sonication, the cell lysate was cleared by centrifugation and filtered using 0.22 μM PVDF Stericup filter (Millipore). Initial protein purification employed ion-exchange chromatography using SP-Sepharose (GE-Healthcare), eluting with a linear NaCl gradient in 50 mM Tris pH 7.4, 150 μM $ZnSO_4$, 4 mM DTT and protease inhibitor mix (Roche Complete Protease Inhibitor Cocktail EDTA free; 1 tablet per 1 L). Eluted protein was exchanged to the same buffer without NaCl and purified further using a HiTrap Heparin HP column (GE-Healthcare), again eluting with a linear NaCl gradient. F2 (SL) was subjected to TEV cleavage and NiNTA affinity purification, as described below, to produce the N-terminal triple glycine motif required

for sortase ligation. Finally, all proteins were purified to homogeneity by size exclusion chromatography using a Superdex-S75 column (GE Healthcare) equilibrated with 50 mM Tris pH 7.4, 200 mM NaCl, 150 μM $ZnSO_4$, 4 mM DTT.

For purification of GST-tagged F3, cells were harvested after expression and resuspended in glutathione binding buffer (50 mM Tris pH 7.4, 1M NaCl, 150 μM $ZnSO_4$ and 1 mM DTT) containing EDTA free Complete Protease Inhibitor Cocktail (Roche) and lysed by sonication. The lysate was cleared by centrifugation and the supernatant incubated with Glutathione-Sepharose 4B (GE Healthcare), equilibrated in gluthathione binding buffer. Protein-bound beads were washed thoroughly with binding buffer and equilibrated into cleavage buffer (50 mM Tris pH 7.4, 150 mM NaCl, 100 μM $ZnSO_4$ and 1 mM DTT). The GST fusion protein was cleaved on the Glutathione-Sepharose beads by GST-tagged Precission Protease (GE Healthcare), leaving N-terminal vector-derived residues Gly-Pro-Gly-Ser N-terminal. Protein in supernatant was cleared off from gluthione sepharose 4B, equilibrated to gelfiltration buffer (50 mM Tris pH 7.4 , 250 mM NaCl, 100 μM $ZnSO_4$ and 1mM DTT) and concentrated using Vivaspin 20 MWCO 3000 (Sartorius-Stedim Biotech). Finally, F3 was purified to homogeneity by gelfiltration using Superdex S-75 equilibrated in the respective gelfiltration buffer.

For purification of F1F2F3, F1F2 (SL), F3 (SL) and WGR, harvested cells were resuspended in NiNTA binding buffer (50 mM Na-HEPES pH 7.5, 1 M NaCl, 4 mM β-mercaptoethanol, 20 mM imidazole, 5% glycerol) containing EDTA free Complete Protease Inhibitor Cocktail (Roche) as well as additional protease inhibitors (4 μg/ml aprotinin, 10 μg/ml leupeptin, 4 μg/ml pepstatin A, 4 μg/ml E64 and 1 mM PMSF). Cells were lysed by sonication, the lysate cleared off by centrifugation and the filtered supernatant loaded on a HisTrap HP Ni-affinity column (GE Healthcare). The His6-tagged proteins were eluted using a linear gradient of imidazole and dialysed overnight in 50 mM Na-HEPES pH 7.0, 0.2 M NaCl, 1 mM DTT, 5% glycerol. If required, the N-terminal fusion tag and His6 Tag were cleaved off by addition of His-tagged TEVpro during dialysis and the sample passed through a HisTrap HP Ni-affinity column to remove the protease and the cleaved tags. Following this, F1F2F3, F1F2 (SL), F3 (SL) were loaded on a HiTrap Heparin HP column (GE-Healthcare) and eluted with a linear NaCl gradient. Finally, all proteins were purified to homogeneity by size exclusion chromatography using a Superdex-S75 column (GE Healthcare) equilibrated with 50 mM Na-HEPES pH 7.5, 0.2 M NaCl, 2 mM DTT, 5% glycerol.

Sortase ligation:

Segmental labeling of PARP-1 constructs was achieved by sortase mediated protein ligation. *Staphylococcus aureus* Sortase Δ59 as well as a mutant variant Sortase Δ59 M (P94S, D160N, D165A, K196T), that has been evolved by yeast display to be catalytically more active (Chen et al., 2011), were purified as described previously (Kobashigawa et al., 2009). For each ligation, two protein starting materials were purified: one carrying a C-terminal LPXTG sortase recognition cleavage site as well as a His6 Tag, the other an N-terminal triple glycine motif, which has been previously shown to be well suited for efficient ligation(Huang et al., 2003) (see Table S1). The latter was produced by TEV

digestion of a N-terminal GENLYFQGGG motif. These constructs were mixed at a ratio of 1:1.25-2 at 100 μM in ligation buffer (50 mM Na-HEPES pH 7.5, 500 mM NaCl, 10 mM $CaCl_2$, 50 μM $ZnSO_4$, 2 mM DTT) and incubated in the presence of 2 μM Sortase Δ59 or Sortase Δ59 M for 20h at either 20C or 4C, respectively. The C-terminal cleavage by-product (GGGRRHHHHHH) was removed by dialysis against ligation buffer by using Slide-A-Lyzer cassettes with a MWCO of 3.500 or 10.000 (Pierce), as it would otherwise be a substrate for the back-reaction. The ligation mixture was then loaded onto a 5 ml HisTrap HP column (GE Healthcare) equilibrated in the ligation buffer without $CaCl_2$. The ligation product as well as the unligated C-terminal by-product were collected in the flow-through, while the other components were bound to the column and could be used for another round of ligation after elution with imidazol. Finally, the ligation product was purified to homogeneity by ion exchange chromatography using a MonoS 16/60 column.

Preparation of DNA dumbbell ligands:

5'-phosphorylated DNA dumbbell ligands were obtained from the in-house DNA synthesis facility of the MRC Laboratory of Molecular Biology or from Integrated DNA Technologies. The sequences of the DNA dumbbell ligands are:

DB1 5'-P-CGGTCGATCGTAAGATCGACCGTGCGCTGGAGCTTGCTCCAGCGC-3'

DB2 5'-P-CGGTCGATCGTAAGATCGACCGTCGCGGTCAGCTTGCTGACCGCG-3'

DB3 5'-P-GCTGGCTTCGTAAGAAGCCAGCTGCGCTGGAGCTTGCTCCAGCGC-3'

DB4 5'-P-GCTGGCTTCGTAAGAAGCCAGCTCGCGGTCAGCTTGCTGACCGCG-3'

Unless otherwise stated all experiments were performed with DB4. For AUC as well as fluorescence anisotropy experiments, a DNA dumbbell was synthesized that contained a fluoresceine derivatized thymidine at sequence position 8. DNA ligands have been further purified using denaturing polyacrylamide gel electrophoresis according to the protocol of Price *et al.* (1998). Correct folding of DNA ligands into a monomeric dumbbell conformation was verified using NMR as well as native PAGE.

Isothermal titration calorimetry:

ITC experiments employed a MicroCal iTC200 calorimeter (GE Healthcare). PARP-1 F1F2 solutions of 100-200 μM were titrated into 5-10 μM gapped DNA dumbbell ligand at 25° C. Typically, one initial injection of 0.5 μl was followed by 19 injections of 2 μl with 120 seconds between each step. Before each experiment, protein and DNA solutions were dialyzed overnight at 4 °C into ITC buffer (50 mM Tris pH 7.4, 50 μM $ZnSO_4$, 1 mM DTT and 200 mM NaCl) using Slide-A-Lyzer cassettes

(Pierce) with a MWCO of 3500Da. Protein and DNA concentrations were determined photometrically and adjusted by dilution. Excess heat upon each injection was integrated using an automatically adjusted base line and derived values corrected by heat of dilution. The latter was derived from the endpoints of protein-DNA titrations under saturating conditions. Data analysis and curve fitting was performed using the Origin 7 software package provided with the Microcal iTC200 calorimeter. The iTC200 was calibrated with EDTA-Ca^{2+} titrations prior to use according to the manufacturer's recommendations.

Fluorescence anisotropy:

Fluorescence polarization experiments were performed as described (Langelier et al., 2010) in 12 mM HEPES (pH 8.0), 60 mM KCl, 0.05 mg/ml BSA, 8 mM $MgCl_2$, and 4 % glycerol and using fluorescently labeled dumbbell DNA DB4.

Analytical ultracentrifugation:

Analytical ultra-centrifugation (AUC) experiments were performed and analyzed in SEDFIT(Schuck, 2005) as described(Langelier et al., 2012) in 25 mM Tris (pH 8.0), 150 mM NaCl, 1 mM EDTA, and 0.1 mM TCEP using full-length PARP-1 (4 μM) and fluorescently labeled dumbbell DNA DB4 (2 μM).

Activity assays:

The SDS-PAGE automodification assay was performed as described(Langelier et al., 2012) in 20 mM Tris (pH 7.5), 50 mM NaCl, 5 mM $MgCl_2$, and 0.1 mM TCEP using 1 μM protein, 5mM NAD^+ and 0.5 or 1 μM DNA depending on the template. When two proteins were combined in the same reaction 1 μM of each protein was used. Reactions containing plasmid DNA used the pUC19 vector modified to include a single nicking enzyme site (Nb.BsmI). The colorimetric assay of PARP-1 activity was performed as described (Langelier et al., 2010; Langelier et al., 2011) in 18 mM HEPES (pH 8.0), 150 mM NaCl, 0.5 mM TCEP, and 10 μg/ml BSA using 40 nM of dumbbell DNA DB4 and 20 nM PARP-1 WT or mutants, and 500 μM NAD^+.

Live cell imaging and laser irradiation:

HeLa cell transfection, sensitization and microirrradiation were performed essentially as described (Langelier et al., 2010; Steffen et al., 2014). Live cell images (512 x 512 pixels) were collected at 1 sec intervals on a Zeiss LSM-510 Meta Confocal laser scanning microscope by excitation with a 488 nm argon laser (8% power) focused through a 40x oil-immersion lens. The nuclei selected

for analysis had a starting GFP intensity ranging from 70 to 160 arbitrary units under this setup. A defined region of interest (ROI) in the nucleus was scanned with 5 iterations of a 405 nm diode laser (100% transmission) to create localized DNA damage. Images were processed and quantified using LSM Image (Zeiss). For quantification of fluorescence intensity changes, the ROI intensity was compared to the intensity of a non-irradiated region of the nucleus, providing a ratio of intensities (ROI/non-irradiated) that was averaged over ≥ 4 cells from a representative experiment.

Small-angle X-ray scattering:

SAXS data was acquired at 20°C at a wavelength of 1.24Å at the EMBL P12 beamline (Petra III, DESY, Hamburg) equipped with a photon counting Pilatus 2M pixel X-ray detector (Blanchet et al., 2015). F1F2-DNA complex was reconstituted as described for NMR spectroscopy using natural abundance protein and DNA (DB4 sequence). Prior to the SAXS measurement F1F2-DNA dumbbell complex was reconstituted as described for NMR spectroscopy and then subjected to another size exclusion chromatography step using a S75 column (GE Healthcare) equilibrated in 50 mM Tris pH 7.2, 200 mM NaCl, 1 mM DTT, 100 μM $ZnSO_4$ to yield monodisperse 1:1 complex. A series of concentrations was measured (adjusted to 3 mg/ml, 1.5 mg/ml and 0.75 mg/ml); the gelfiltration buffer also served as reference sample for the SAXS measurements. Data processing and analysis was performed by using the ATSAS program package: After averaging, data reduction and buffer subtraction, the program PRIMUS (Konarev et al., 2003) was used for evaluation of data quality by Guinier analysis (Fig. S4k). Radius of Gyration Rg as well as forward scattering I(0) were calculated using the entire scattering data using the program GNOM (Svergun, 1992), providing also pair distribution functions P(r) (Fig. S4m). Theoretical X-ray scattering profiles of individual NMR ensemble structures were calculated using the program Crysol (Svergun et al., 1995). The average profile shown in Fig. 1h is a linear combination of all 78 theoretical profiles calculated for the whole ensemble.

NMR spectroscopy:

NMR data were acquired using Bruker Avance I 800, Avance II+ 700, DMX600 and DRX500 spectrometers (Bruker BioSpin GmbH), each equipped with a cryogenically cooled triple resonance ($^{1}H/^{15}N/^{13}C$) 5mm probe. Solutions contained protein, DNA or protein-DNA complexes at 100-300 μM in either (A) 50 mM [$^{2}H_{11}$]-tris, 1 mM [$^{2}H_{10}$] DTT and 100 μM $ZnSO_4$ at pH7.2, $H_2O/^{2}H_2O$ 95:5 or $^{2}H_2O$ 98%, or (B) the same buffer with addition of 200 mM NaCl. Conditions (A) were used for NMR experiments involving F1F2-DNA complex or free DNA. Since constructs including the F3 domain were prone to precipitation at low ionic strength, conditions (B) were used for all experiments involving F1F2F3 free protein, F1F2F3-DNA complex, and for WGR and F1F2F3-DNA titrations with WGR; in addition, some assignment experiments for the F1F2-DNA complex were repeated under conditions (B) to allow chemical shift differences to the F1F2F3-DNA complex to be assessed under identical

conditions. When necessary, restricted volume NMR tubes were used to optimize sensitivity; for samples under conditions (A) these were Shigemi microtubes (Shigemi), while for samples under conditions (B) shaped NMR tubes (Bruker BioSpin GmbH) were used. Water suppression was achieved using the WATERGATE 3-9-19 pulse-sequence element (both for samples in H_2O and in ^2H_2O) (Piotto et al., 1992). Spectra were recorded at 25°C, 30°C or 37°C; ^1H chemical shifts were calibrated using sodium 3,3,3-trimethylsilylpropionate (TSP) as an external ^1H reference; ^{15}N and ^{13}C chemical shifts were indirectly referenced to the ^1H shifts using the ratio of gyromagnetic ratios. Spectra were processed using the program Topspin (Bruker BioSpin GmbH) and the NMRpipe package (Delaglio et al., 1995), while spectral analysis was performed in Sparky (Goddard) or the CCPN analysis package (Vranken et al., 2005).

Protein-DNA complexes were reconstituted using chemically synthesized DNA at natural abundance and recombinant protein isotopically labeled according to one of the following schemes:

F1F2 scheme I: Uniform [^2H,^{15}N,^{13}C];

F1F2 scheme II: Uniform [^2H (approx. 70%),^{15}N,^{13}C];

F1F2 scheme III: Uniform [^2H,^{15}N];

F1F2 scheme IVa: Uniform [^2H,^{15}N,^{13}C], back-labeled with [^1H,^{13}C] in the δ-methyl groups of Ile and all methyl groups of Leu and Val residues, using sodium [4-^{13}C, 3,3-$^2H_2$]α-ketobutyrate and sodium [3-^2H, 4,4'-$^{13}C_2$]α-ketoisovalerate as precursors to maximize protonation of methyl groups, for use in NOE experiments; sodium [3-^2H, 4,4'-$^{13}C_2$]α-ketoisovalerate was prepared from sodium [4,4'-$^{13}C_2$]α-ketoisovalerate by exchange with ^2H_2O at pH 12.5 and 45°C for 3 hrs (Goto et al., 1999);

F1F2 scheme IVb: Uniform [^2H,^{15}N,^{13}C], back-labeled with [^1H,^{13}C] in the δ-methyl groups of Ile and all methyl groups of Leu and Val residues, using sodium [3,3-$^2H_2$,$^{13}C_4$]α-ketobutyrate and sodium [3-^2H,$^{13}C_5$]α-ketoisovalerate as precursors to produce linear chains of ^{13}C in the sidechains of Val and Leu, for use in assignment experiments to link methyl signals to Cα signals;

F1F2 scheme V: Uniform [^2H,^{15}N,^{13}C], back-labeled with [^1H,^{13}C] in the methyl groups of Met residues in addition to Ile, Leu and Val methyl groups as in (IVa);

F1F2 scheme VI: Uniform [^2H,^{15}N,^{13}C]; back-labeled with [^1H,^{13}C] in the methyl groups of Ile, Leu and Val methyl groups as in (IVa) and [^1H,^{15}N,^{13}C] Phe residues;

F1F2 scheme VII: [^1H,^{15}N,^{12}C] F1 sortase ligated to [^2H,^{15}N,^{13}C] F2 back labeled with [^1H,^{13}C] Ile, Leu Val and Met methyl groups labeled as in (V);

F1F2 scheme VIII: [^1H,^{15}N,^{12}C] F1 sortase ligated to [^2H,^{15}N,^{13}C] F2 back labeled with [^1H,^{13}C] Met methyl groups as in (V) and [^1H,^{15}N,^{13}C] Arg residues;

F1F2 scheme IX: [^2H,^{15}N,^{13}C] F1 back-labeled with Ile, Leu and Val methyl groups labeled as in (IVa) and [^1H,^{15}N,^{13}C] Arg residues sortase ligated to [^1H,^{15}N,^{12}C] F2;

F1F2 scheme X: [^{2}H,^{15}N,^{13}C] F1 back-labeled with Ile, Leu and Val methyl groups labeled as in (IVa) and [^{1}H,^{15}N,^{13}C] Phe residues sortase ligated to [^{1}H,^{15}N,^{12}C] F2;

F1F2 scheme XI: [^{2}H,^{15}N,^{13}C] F1 back-labeled with Ile, Leu and Val methyl groups labeled as in (IVa) sortase ligated to [^{2}H,^{15}N,^{12}C] F2 back-labeled with [^{1}H,^{15}N,^{13}C] Pro;

F1F2F3 scheme I: Uniform [^{2}H,^{15}N,^{13}C];

F1F2F3 scheme II: F1F2 [^{2}H,^{15}N,^{13}C] sortase ligated to F3 at natural abundance;

F1F2F3 scheme III: F1F2 at natural abundance sortase ligated to F3 [^{2}H, ^{15}N, ^{13}C].

NMR assignment:

Essentially complete amide group assignments of PARP-1 constructs, both free and in complex with dumbbell DNA ligands, were obtained using the following strategy. First, full sets of assignments were obtained for the isolated domains; in the cases of F1 and F2, backbone and sidechain assignments were transferred from those reported in Eustermann et al., (2011), and in the cases of the isolated F3 and WGR domains, assignments were made using [^{15}N,^{13}C] labeled protein and a standard suite of NMR experiments ([^{15}N,^{1}H] HSQC, [^{13}C,^{1}H] HSQC, CBCANH, CBCA(CO)NH, HBHANH and HBHA(CO)NH). These assignments were then used as a starting point for making sequential assignments of the amide signals of multidomain PARP-1 fragments, in both their free and DNA-bound states. Assignments were made using [^{2}H,^{15}N,^{13}C] labeled protein in conjunction with TROSY-based experiments ([^{15}N,^{1}H] TROSY, [^{13}C-^{1}H] HMQC, TROSY-NHCACB (optimized for CB), TROSY-HNCA and TROSY-HN(CO)CA); chemical shifts of amide signals recorded using TROSY-based experiments were corrected by subtraction of 47Hz (=^{1}J(^{15}N,^{1}H)/2) in each dimension. Severe spectral crowding was overcome by using sortase-mediated protein-protein ligation as described above to prepare samples in which the observed signals were limited to those from particular domains; thus, in different samples, NMR signals were restricted to the F1, F2, F1F2 or F3 domains and their respective linker regions, while in each case the other parts of the same polypeptide chain were kept NMR silent (this corresponds to labeling schemes F1F2 VII-XI and F1F2F3 II and III as listed above). Segmental labeling of F1F2F3 was similarly used to relieve overlap and facilitate protein backbone assignments for ternary F1F2F3-WGR-DNA complexes; different samples were made that comprised F1F2F3 labeled according to schemes I, II or III in complex with DNA and WGR domain, or alternatively [^{2}H,^{15}N,^{13}C] labeled WGR domain in complex with DNA and (natural abundance) F1F2F3. Using this strategy, amide signals in the various fragments and complexes were assigned to the following extents: F1F2-DNA, 200/204, 98.0%%; F1F2F3, 331/346, 95.7%%; F1F2F3-DNA, 330/346, 95.4%; F1F2F3-WGR-DNA, 324/346, 93.6%; F1F2F3(W246A), 327/346 94.5%; F1F2F3(W246A)-DNA, 325/346. 93.9%; F1F2F3(W246A)-DNA-WGR, 320/346, 92.5%.

All δ-methyl groups of Ile, all Leu and Val methyl groups (with the sole exception of the methyls of Val202) as well as all Met methyl groups of DNA-bound F1F2 were assigned (see Figure S2). The

majority of the methyls of Ile, Val and Leu were assigned using HMCM(CG)CBCA experiments (Tugarinov and Kay, 2003), to link to the corresponding Cα signals (that had themselves been previously assigned using the backbone experiments), recorded using DNA-bound F1F2 labeled according to scheme IVb and in $^{2}H_2O$ buffer. These methyl assignments were confirmed and extended using NOE-based data, particularly 3D [^{1}H,^{13}C,^{1}H] NOESY-HMQC and 3D [^{13}C,^{13}C,^{1}H] HMQC-NOESY-HMQC experiments (τ_c=200ms) recorded using DNA-bound F1F2 labeled according to scheme IVa and in $^{2}H_2O$ buffer; these experiments, together with 2D [^{1}H,^{1}H] NOESY and ^{15}N NOESY-HSQC experiments recorded in H_2O buffer, established an extensive network of NOE connectivities, all of which were consistent with the known structures of the DNA-unbound proteins. A further check on the assignments was provided by segmental labeling of F1 and F2 (F1F2 schemes VII-XI), which established unambiguously from which finger any given signal originated. Assignments for the methionine methyl signals were made using NOE connectivities to Ile, Val and Leu methyl signals (in experiments recorded in $^{2}H_2O$) and to backbone amide signals (in experiments recorded in H_2O), again using sortase ligated samples to distinguish from which finger given signals originated. In addition, assignments were obtained for the aromatic signals of phenylalanine residues in F1 and for sidechain signals of Pro residues in F2. The Phe44 aromatic signals were identified using unambiguous NOEs to a methyl of Val48 and to backbone amide signals of Met43 and Asp45 in experiments using DNA-bound F1F2 labeled according to scheme X and in $^{2}H_2O$ buffer. Sidechain signals of prolines in F2 were assigned starting from Cα assignments previously established in backbone experiments, using samples of DNA-bound F1F2 labeled according to scheme XI together with 2D [^{13}C,^{1}H] HSQC experiments recorded in $^{2}H_2O$ to extend these to the Hα signals and then using 2D [^{13}C,^{1}H] HSQC-NOESY (τ_c=100ms) recorded on the same samples to further extend the assignments onto corresponding Hβ and Hγ signals (Figure S2); this process did not allow a full assignment of all proline sidechain protons, but unambiguous assignments were obtained for most Hβ and Hγ signals, including those of Pro149 (Figure S2). Samples containing back-labeled Arg, Lys, Trp and Tyr residues were also made in an effort to measure further intermolecular protein-DNA and inter-domain NOEs (see below). However, any signals from these residues in the NOE-based experiments were below the detection threshold, presumably because they have more complicated multiplet structures and suffer faster relaxation than do the methyl signals; consequently, we did not pursue their assignment, except for sidechain NH signals of Trp.

DNA assignment:

Non-exchangeable signals of the 45 nucleotide free DNA ligand (sequence DB4) were assigned following standard protocols for B-form DNA (Wüthrich, 1986) using 2D (^{1}H,^{1}H) NOESY, TOCSY and DQF-COSY spectra recorded in $^{2}H_2O$ (recorded at 800 MHz and 37°C, τ_m=200ms), and these assignments were extended to base-paired imino- and amino- signals using cross-peaks in NOESY spectra recorded in H_2O. A complete assignment was obtained for the H1', H2', H2" and H3' and aromatic protons (excepting only A13 H2), while for the H4' protons, and particularly for the H5' and

H5" protons, signal overlap limited the extent of assignment; imino and H-bonded amino proton assignments were obtained for all of the A:T and G:C basepairs except for the G:C pairs on either side of the gap. Signals in the two tetraloops were assigned unambiguously using connectivity patterns observed in the ^2H_2O spectra; in addition, those of the CTTG loop were very similar to assignments previously published for a small circular DNA dumbbell containing a CTTG loop (Ippel et al., 1992, 1995). DNA signals in the complex were assigned using patterns of connectivities in 2D (^{1}H,^{1}H) NOESY spectra filtered to remove ^{13}C-bound signals in ω2, where possible making sequential walks characteristic for B-form DNA, and comparing these with corresponding patterns and assignments in the free DNA. During this process, a number of factors were helpful in overcoming the severe crowding in the 2D NOESY spectra of the complex; comparison of spectra at different temperatures and in either H_2O or ^2H_2O helped in resolving many ambiguities due to overlap, while signals arising from the tetraloops and adjacent basepairs, away from the protein-binding region, occurred at very similar chemical shifts to those for the free DNA. A particularly powerful approach for signals that changed shift substantially upon protein-binding, and would thus otherwise have been the most difficult to assign, was to use exchange cross-peaks observed in 2D NOESY spectra from samples containing a slight excess of DNA (protein:DNA ratio, 0.95:1). Under the conditions used, the kinetics of the complex were such that exchange cross-peaks linking corresponding DNA signals in the free and bound states were of comparable intensity to strong NOE cross-peaks (see Figure S3); for this to occur k_{off} must presumably be comparable to or faster than $1/T_1$ (Combrisson et al., 1971; Neuhaus and Williamson, 2000). Assignments made using such cross-peaks could additionally be confirmed by the further presence of exchange-relayed NOE cross-peaks in some cases, as well as by comparison with spectra from samples lacking excess DNA where cross-peaks caused by free/bound exchange were missing. The extent of the DNA assignment free and bound to PARP-1 F1F2 is summarized in Figure S3.

Chemical Shift Perturbation analysis:

Using the assignments described above, chemical shift perturbations upon protein-DNA interactions were calculated according to the formula $((\Delta\delta(^1H))^2 + (\Delta\delta(^{15}N))^2/5)^{1/2}$ for protein amide groups, $((\Delta\delta(^1H))^2 + (\Delta\delta(^{13}C))^2/10)^{1/2}$ for protein methyl groups and $((\Delta\delta(H1'))^2 + (\Delta\delta(H^{arom}))^2)^{1/2}$ for DNA chemical shift perturbations (where H^{arom} is H8 for Ade, H6 for Cyt, H8 for Gua and H6 for Thy).

Residual Dipolar Couplings:

RDCs were measured for DNA-bound [^{2}H,^{15}N] labeled PARP-1 F1F2 protein using the ARTSY pulse sequence (Fitzkee and Bax, 2010) at 600 MHz, recorded in an interleaved manner with ^{1}H dephasing durations of 5.375 ms and 10.75 ms. ^{1}JNH were extracted from measured ARTSY peak intensity ratios (Fitzkee and Bax, 2010) of an isotropic sample adjusted to buffer condition A, while

^{1}JNH+^{1}DNH values were determined for a weakly aligned sample made by addition of filamentous phage Pf1 (ASLA Biotech) to a final concentration of approximately 8mg/ml. The RDC data used in the structure calculations were restricted to values for residues shown to be rigid by the ^{15}N relaxation data and for which the conformations were invariant amongst different copies of the protein in the crystal structures. For the RDC analysis of the crystal structures of F1 and F2 we used the ISAC protocol (Sass et al., 2001) as implemented in XPLOR-NIH (Schwieters et al., 2003). In order to judge the quality of correlation between the RDCs back-calculated for respective structures and the experimentally determined values, we calculated Q values ($Q=RMS(D_i^{obs}-D_i^{calc})/RMS(D_i^{obs})$) as described by Bax and Grishaev (2005).

Relaxation analysis:

For the DNA-bound F1F2 complex (labeling scheme V), per-residue effective correlation times ($\tau_{c(effective)}$) at 37°C were determined at 800 MHz by using ^{15}N TRACT experiments (Lee et al., 2006) that measure the rate difference between the sharp and broad components of TROSY multiplets. Experiments were carried out in an interleaved manner with relaxation periods of 2, 2, 4, 8, 14, 14, 20, 22, 32, 48, 66, 96, 96, 140 and 240ms. Relaxation rates were determined using peak analysis scripts within the NMRpipe software package (Delaglio et al., 1995). Associated errors were estimated using data from the duplicated delays, and per-residue effective correlation times were then calculated according to (Lee et al., 2006). For F1F2F3 complexes, the broad TROSY component decayed too quickly for TRACT experiments to be feasible; instead ^{15}N T_1, $T_{1\rho}$ and steady-state $\{^{1}H\}^{15}N$ NOE data were acquired at 800 MHz and 30°C, using the optimized pulse sequences of Lakomek et al. (2012). The relaxation delays used were as follows: T_1 (free protein): 0, 40, 40, 100, 180, 320, 320, 500, 800, 1200ms; T_1 (complex): 0, 60, 60, 140, 240, 400, 700, 1200, 2000ms; $T_{1\rho}$ (free protein): 0.2, 5, 5, 9, 17, 25, 40, 40, 70, 100ms; $T_{1\rho}$ (complex): 0.2, 1, 3, 3, 6, 10, 16, 25, 40ms; steady-state $\{^{1}H\}^{15}N$ NOE saturation/relaxation period, 6.6s. $T_{1\rho}$ values were corrected for resonance offset effects (spin-lock tilt-angle) as described by Lakomek et al. (2012). Curve-fitting of these data was carried out using routines within the CCPN analysis software package (Vranken et al., 2005), and errors were estimated using data from the duplicated delays.

NOE-based restraints:

NOE restraints provided key atom-specific information required to characterize the architecture of the F1F2-DNA complex. Initial attempts to measure intermolecular NOE interactions between uniformly [^{1}H,^{13}C,^{15}N] labeled protein and natural abundance DNA by using conventional filtered NOE experiments all failed; not unexpectedly, given the size of the system, rapid transverse relaxation rendered the signals undetectable. To overcome this we turned to reverse labeling schemes in which various combinations of selected amino acid types (Phe, Tyr, Trp, Lys, Arg, Pro and Met) as

well as methyl groups of Ile, Val and Leu were protonated and ^{13}C-labeled in an otherwise deuterated protein background (see isotope labeling schemes F1F2 I–XI described above), thereby reducing simultaneously spectral complexity and transverse relaxation. Although it necessarily reduced the number available, this approach made it possible to detect NOE interactions at high resolution using 2D [^{1}H-^{1}H] NOESY experiments having only a single isotope half-filter, set to accept only ^{13}C-bound ^{1}H signals in ω2, and thereby restricting DNA signals to appear only in ω1. Spectral complexity was further reduced by segmental labeling of F1 and F2 (see below), and assignment of intermolecular NOE interactions was further aided by using 3D [^{1}H,^{13}C,^{1}H] NOESY-HMQC and 3D [^{13}C,^{13}C,^{1}H] HMQC-NOESY-HMQC experiments. Intriguingly, exchange-relayed intermolecular NOE crosspeaks were also observed in samples that contained an excess of DNA (see DNA assignment); once correctly identified, these were useful in confirming assignments. Identification of artifacts in the 2D spectra was aided by recording data separately with and without ^{13}C decoupling in ω1; unwanted intramolecular protein crosspeaks due to breakthrough gain a splitting in ω1 in the undecoupled dataset, whereas genuine intermolecular cross-peaks remain un-split. Intermolecular NOE crosspeaks that were detected using this strategy were largely restricted to interactions involving methyl groups; attempts to measure intermolecular NOE interactions involving sidechain $^{13}C^{1}H_{2}$ or $^{15}N^{1}H$ groups of Arg or Lys residues, that could have helped characterize protein interactions with the DNA phosphate backbone, were all unsuccessful, as were attempts to detect NOE interactions of the aromatic rings of [^{1}H^{13}C]-labeled Phe residues intended to detect intermolecular interactions of Phe44. The assigned intermolecular NOE crosspeaks and the experiments and samples that were used to detect them are summarized in Table S2.

A similar strategy was employed for identification of the DNA-dependent interdomain NOE interactions linking F1 and F2. In order to isolate these from others, samples of the F1F2-DNA complex were made that contained sortase-ligated protein in which one domain was reverse-labeled and deuterated while the other was [^{1}H,^{12}C,^{15}N] labeled, according to one of the schemes VII-XI. Such samples allowed detection of interdomain NOE crosspeaks at high resolution when used in similar experiments to those described for detecting the intermolecular NOE interactions. 2D [^{1}H-^{1}H] NOESY experiments were recorded with and without heteronuclear decoupling in ω1, as well as with a single half-filter set to accept only ^{13}C-bound ^{1}H signals in ω2, so as to restrict signals from ^{12}C attached protons (from the [^{1}H,^{12}C,^{15}N] labeled finger and the DNA) to the ω1 dimension. As an additional control, some spectra were recorded for the protein in the free state, in order to show that the observed interdomain contacts were only present in the DNA-bound state of the protein. Unambiguous assignment of interdomain NOE interactions between methyl groups was further aided by using 3D [^{1}H,^{13}C,^{1}H] NOESY-HMQC and 3D [^{13}C,^{13}C,^{1}H] HMQC-NOESY-HMQC experiments recorded using DNA-bound F1F2 labeled according to scheme IVa and in $^{2}H_{2}O$ buffer. The assigned interdomain NOE crosspeaks and the experiments and samples that were used to detect them are summarized in Table S2.

In order to use these assigned NOE interactions as restraints in the structure calculations, their associated upper distance bounds were uniformly set to 6.0Å. A true, experiment-based calibration of these NOE intensities would be extremely difficult to envisage, as the nature of the various interacting pairs varies considerably; in most cases the interaction spans between a deuterated and a non-deuterated domain or the DNA, involves a methyl group sometimes at one site and sometimes at both, and in addition many of the protein signal intensities are likely to be strongly modulated by differing label incorporation efficiencies at the different sites. Setting the upper bounds to 6Å results in negligible violations in the best fitting structures, and reducing the upper-bound distances uniformly to 5Å has a negligible impact on the calculated co-ordinates and the violations (data not shown), from which we conclude that use of 6Å upper bounds represents a conservative approach. Lower bounds were uniformly set to zero, and r^{-6} averaging was employed for equivalent and non-stereoassigned groups. Methyl signals of Leu and Val residues were not stereoassigned..

Structure calculations; i) template structures:

The starting point for calculations were the deposited co-ordinates of PARP-1 F1 and F2 domains in complex with DNA double-strand breaks; for the F1 complex the co-ordinates of chains B, I and J from pdb 3ODA were used, and for the F2 complex those of chains B, E and F from pdb 3ODC. The program XPLOR-NIH was used for all calculations. Hydrogen atoms were first added according to standard geometries, and the structures then subjected to 3000 cycles of Powell minimization while rigidly fixing the co-ordinates of the N, C' and O atoms of every peptide bond in the protein as well as all atoms of the DNA. This allowed the protein structures largely to equilibrate with XPLOR-NIH force field while exactly preserving the relative orientations of the individual amide bond vectors, as well as the relationship of the protein to the DNA. Only very small movements of the protein backbone occurred during this minimization; for F1 (residues 6-91) the backbone co-ordinate shift (rmsd for N, Cα, C') was 0.081Å, while for F2 (residues 109-200) the corresponding shift was 0.077Å. We refer to the resulting structures as F1-DNA_{min} and F2-DNA_{min} respectively.

To model the DNA dumbbell, initially a simulated annealing protocol was used to calculate an ensemble of 50 structures from torsion-angle randomized starting conformations, restraining the stems to reproduce ideal B-form geometry and the tetraloops, in the absence of corresponding deposited DNA structures, to reproduce the RNA tetraloop structures 1MSY (for G10-A13) and 1RNG (for C33-G36; the lowest energy structure of 1RNG was used). Experimental NOE connectivities measured for the DNA dumbbell in both free and bound states (summarized in Figure S3) are largely consistent with these RNA tetraloop structures and suggest they are an adequate approximation to the true solution conformations. One difference concerns the base of T35, which is clearly in the anti conformation in the DNA case. This is not unexpected, as the syn conformation of the corresponding uridine in 1RNG is maintained by an H-bond to the 2'OH group of the preceding ribose, which clearly cannot form in the DNA case; consequently the χ angle restraint used for T35 in the calculations was

altered from -155° to +25°. Overall, the restraints used comprised 478 dihedral angle restraints (set to ranges of ±5°; of these, 383 were in the stems and 95 in the tetraloops), 55 H-bonding distance restraints (49 in the stems set to ranges of ±0.1Å, and 6 in the tetraloops set to ±0.2Å), 18 weak basepair planarity restraints (force constant 50 kcal.mol^{-1}.Å^{-2}, except for the stem-closing basepairs G1:C22 and C24:G45 where the force constant was 200 kcal.mol^{-1}.Å^{-2}) for basepairs in the stems as well as for G10:A13 and C33:G36 in the tetraloops, and in the stems additional 1-3 phosphorus-phosphorus lower limit constraints (distance > 10.5Å) to aid convergence (Cerdan et al., 2001).

The lowest energy structure from this ensemble was then subjected to further annealing using a set of constraints designed to reproduce almost exactly, in the context of the dumbbell, the backbone conformation of the DNA in the complexes 3ODA and 3ODC within the region of the protein footprints. We refer here to the stem of the dumbbell which binds to F1 and contains the 5' terminus as stem 1, and that which binds F2 and contains the 3' terminus as stem 2. The restraints applied to stem 1 were based on target values measured for chains I and J in 3ODA, and comprised 193 dihedral angles in nucleotides 1-9 and 14-22, in addition to 48 long-range intra- and inter-strand distances in nucleotides 1-7 and 16-22 (these spanned between 2 and 7 basepairs, and comprised 24 intra-strand O3'-O5', 12 inter-strand O3'-O3' and 12 inter-strand O5'-O5' distances), while those applied to stem 2 were based on target values measured for chains E and F in 3ODC, and comprised 168 dihedral angles in nucleotides 24-31 and 38-45, in addition to 48 long-range intra- and inter-strand distances in nucleotides 24-30 and 45-39; dihedral angle restraints were applied with a range of ±5° and force constant increasing during the protocol to reach 100 kcal.mol^{-1}, and the long-range distance restraints were exact (±0Å), applied with a force constant of 50 kcal.mol^{-1}. Dihedral angle restraints in the tetraloops as well as basepair H-bond distance restraints throughout were applied as before (GC H-bond restraints: G(N1)-C(N3), 2.95 ± 0.1 Å; G(N2)-C(O2), 2.86 ± 0.1 Å; G(O6) – C(N4), 2.91 ± 0.1 Å. AT H-bond restraints: A(N1)-T(N3), 2.82 ± 0.1 Å; A(N6)-T(O4), 2.95 ± 0.1 Å; all force constants 50 kcal.mol^{-1}). Basepair planarity restraints were applied very weakly (force constant 20 kcal.mol^{-1}.Å^{-2} in the stems, and 10 kcal.mol^{-1}.Å^{-2} for basepairs G10:A13 and C33:G36 in the tetraloops). We refer to the resulting structure as DNA-dumbbell$_{anneal}$.

Structure calculations; ii) ensemble calculations:

To calculate an ensemble of models for the F1F2-DNA complex, 500 different random starting conformations were first created by randomizing the ten rotatable bonds in the continuous strand of the DNA linker between C22 O3' and C24 O5' within DNA-dumbbell$_{anneal}$. Next, the common DNA backbone atoms in the 7-basepair region of each protein domain's DNA-binding footprint were used to fit F1-DNA$_{min}$ and F2-DNA$_{min}$ onto the corresponding DNA backbone atoms of the stems of DNA-dumbbell$_{anneal}$, thereby carrying each protein domain accurately into the same spatial relationship to the appropriate stem of the dumbbell as it had with blunt-ended DNA in complex 3ODA (for F1) or 3ODC (for F2); the fits obtained are shown below:

<table>
<tr><th colspan="3">Stem 1</th></tr>
<tr><th>3ODA, chain I</th><th>DNA-dumbbell$_{anneal}$ stem 1</th><th>rmsd</th></tr>
<tr><td>nucleotide 1:
C1', C2', C3', C4', C5', O3', O4' and P
nucleotides 2-7:
C1', C2', C3', C4', C5', O3', O4', O5', P</td><td>nucleotide 1:
C1', C2', C3', C4', C5', O3', O4' and P
nucleotides 2-7:
C1', C2', C3', C4', C5', O3', O4', O5', P</td><td rowspan="3">0.234Å</td></tr>
<tr><th>3ODA, chain J</th><th>DNA dumbbell$_{anneal}$ stem 1</th></tr>
<tr><td>nucleotide 4:
C1', C2', C3', C4', C5', O3', O4'
nucleotides 5-10:
C1', C2', C3', C4', C5', O3', O4', O5', P</td><td>nucleotide 16:
C1', C2', C3', C4', C5', O3', O4'
nucleotides 17-22:
C1', C2', C3', C4', C5', O3', O4', O5', P</td></tr>
<tr><td colspan="3"></td></tr>
<tr><th colspan="3">Stem 2</th></tr>
<tr><th>3ODA, chain F</th><th>DNA-dumbbell$_{anneal}$ stem 2</th><th>rmsd</th></tr>
<tr><td>nucleotide 1:
C1', C2', C3', C4', C5', O3', O4' and P
nucleotides 2-7:
C1', C2', C3', C4', C5', O3', O4', O5', P</td><td>nucleotide 24:
C1', C2', C3', C4', C5', O3', O4' and P
nucleotides 25-30:
C1', C2', C3', C4', C5', O3', O4', O5', P</td><td rowspan="3">0.202Å</td></tr>
<tr><th>3ODA, chain E</th><th>DNA dumbbell$_{anneal}$ stem 2</th></tr>
<tr><td>nucleotide 2:
C1', C2', C3', C4', C5', O3', O4'
nucleotides 3-8:
C1', C2', C3', C4', C5', O3', O4', O5', P</td><td>nucleotide 39:
C1', C2', C3', C4', C5', O3', O4'
nucleotides 40-45:
C1', C2', C3', C4', C5', O3', O4', O5', P</td></tr>
</table>

Once these fits had been carried out, the original DNA atoms derived from 3ODA and 3ODC were deleted, leaving only the F1 and F2 domains in their bound conformations on the dumbbell. The structures were then subjected to 50 cycles of Powell minimization while holding the positions of the peptide bond atoms (N, C', O and HN) rigidly fixed to preserve relative amide bond orientations.

These starting structures of the dumbbell with bound F1 and F2 domains were then subjected to a simulated annealing protocol to generate an ensemble of conformers consistent with the RDC, intermolecular and interdomain NOE restraints. The protocol comprised 4000 cycles of Powell

minimization, followed by 40,000 steps of Langevin dynamics at 500K with reduced timestep of 0.0001ps, then 40,000 steps of cooling to 300K with timestep 0.001ps, and 5000 steps of final Powell minimization. For calculating the RDC energy terms a single alignment tensor was used for the whole complex, and the tensor parameters were allowed to float for best fit during the calculation, as previously described (Sass et al., 2001). The force constant used for the NOE restraints was 50 kcal.mol^{-1}. During these calculations strong non-crystallographic symmetry (NCS) constraints were used to maintain the internal structure of the DNA stems, the F1 and F2 domains, and the relationship of each finger to the stem to which it is bound. To achieve this the entire structure was duplicated, the co-ordinates of one copy rigidly fixed and groups of NCS constraints defined between the fixed and unfixed copies so as to maintain these elements of the structure in the evolving co-ordinates of the unfixed copy. Once preliminary rounds of calculation had established likely regions where contacts between F1 and F2 might occur, the NCS terms for just these regions were restricted to backbone atoms and were applied with a much reduced force constant; the NCS constraints used in the final rounds of calculations are shown below:

NCS constraint group	Atoms constrained	Force constant (kcal.mol^{-1})
F1/stem1	All atoms of protein residues 9-36, 49-59 and 66-89; All atoms of DNA nucleotides 1-22.	10,000
	Backbone heavy atoms (N, Cα, C') of protein residues 6-8, 37-48, 60-65 and 90-91; All carbon atoms of protein residue 44.	50
F2/stem2	All atoms of protein residues 114-140 and 155-199; All atoms of DNA nucleotides 25-45. All atoms except O5', P, O1P and O2P of DNA nucleotide 24.	10,000
	Backbone heavy atoms (N, Cα, C') of residues 109-113, 141-154 and 200-201.	50

Structures were included in the accepted ensemble if they met simultaneously all of the following criteria: E(total) <= 6000 kcal.mol^{-1}, E(tensor) <=1500 kcal.mol^{-1}, and E(NOE) <= 2 kcal.mol^{-1}.

Finally, the remaining atoms of the protein N- and C-terminal tails (residues 1-5 and 202-214) and the F1-F2 linker (residues 92-108) were added in a separate simulated annealing protocol. Initially

all atoms of the full structure (residues 1-214 of the protein and 1-45 of the DNA, including the 5' terminal phosphate group) were placed at fully randomized positions within a 200Å cube, then for those residues included in the previous stages of the calculations (i.e. residues 6-91 and 109-201 of the protein and nucleotides 1-45 of the DNA, not including the 5' terminal phosphate group) the randomized co-ordinates were replaced by the previously calculated values and rigidly fixed in place. The structures were then subjected to 200 cycles of Powell minimization, followed by 20,000 steps of Langevin dynamics at 500K with timestep 0.001ps and reduced Van der Waals radii and force constant, followed by 152,000 steps during which the Van der Waals radii and force constant were gradually increased and 102,000 steps during which the temperature was lowered to 300K, and a final 1000 steps of Powell minimization. This protocol for adding the tails and linker was repeated independently 5 times for each input structure using a different randomization seed each time, and the structure with the lowest value of E(total) retained.

As a control, an ensemble of 100 calculations was run in which the DNA was omitted and only the RDC constraints applied, to verify that the under these circumstances the protocol resulted in a statistical distribution of the four possible mutual orientations of F1 and F2 consistent with the RDC data. Analysis of the results (data not shown) demonstrated that there was essentially complete convergence to the four possible relative orientations, the four groups having populations of 23, 25, 25 and 27.

Parallel calculations were also run for alternative models:

1) Ensembles in which the template structure for F1 and its associated DNA stem were extracted from pdb 4DQY (chain A, with DNA backbones of chains M 1-10 and N 13-22). This was done in order to refine the fitting of the F3, WGR and CAT domains from 4DQY to the hybrid model; because the duplex DNA conformations in 3ODA and 4DQY differ somewhat, the quality of the fit is degraded when the template used for calculation of the hybrid structure is derived from 3ODA (Figure S5g,h). Alternatively, if 4DQY is itself used as the source of the template structures for F1 and its associated DNA stem, then a very precise fit is obtained, allowing more reliable modeling of the relationship between the domains. The lowest energy structure from an ensemble calculated in this way was used to prepare Figures 2, 3 and 4 of the main paper.

2) Ensembles in which the linker was shortened by deletion of residues 94-102 as described by (Ali et al., 2012), showing that the shorter linker is easily able to connect the F1 and F2 domains in the model (Figure S4j).

Structure analysis:

Structural superpositions optimized for simultaneous best fits across each ensemble were calculated using the program CLUSTERPOSE (Diamond, 1992, 1995). F1 fits refer to protein residues 6-91, F2 fits to protein residues 109-201, DNA stem 1 fits to nucleotides 1-22 and DNA stem 2 fits to residues 24-45 (excluding C24 O5' and P). For backbone fits the N, Cα and C' atoms of protein residues and the C1', C2', C3', C4', C5', O3', O4', O5' and P atoms of DNA were used. Ramachandran statistics for the protein were calculated using the program Procheck-NMR (Laskowski et al., 1996); these statistics demonstrated that the backbone geometry of the crystal structures 3ODA and 3ODC was maintained through the calculation protocol. DNA bend angles were calculated using the program 3DNA (Lu and Olson, 2003), using co-ordinates of basepairs 1:22-5-18 to define the 5' stem and 24:45-28-41 to define the 3' stem, and interface areas were calculated using the program PISA (Krissinel and Henrick, 2007). Structures were visualized using the program PyMol (DeLano, 2002).

Supplementary References

Blanchet, C.E., Spilotros, A., Schwemmer, F., Graewert, M.A., Kikhney, A., Jeffries, C.M., Franke, D., Mark, D., Zengerle, R., Cipriani, F.*, et al.* (2015). Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). Journal of applied crystallography *48*, 431-443.

Cerdan, R., Payet, D., Yang, J.C., Travers, A.A., and Neuhaus, D. (2001). HMG-D complexed to a bulge DNA: an NMR model. Protein Sci *10*, 504-518.

Chen, I., Dorr, B.M., and Liu, D.R. (2011). A general strategy for the evolution of bond-forming enzymes using yeast display. Proc Natl Acad Sci U S A *108*, 11399-11404.

Combrisson, S., Roques, B., Rigny, P., and Basselier, J.J. (1971). Description de l'effet Overhauser dans les systèmes en échange: Application à l'analyse conformationnelle d'aldéhydes hétérocycliques. Canadian Journal of Chemistry *49*, 904-913.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR *6*, 277-293.

DeLano, W.L. (2002). The PyMOL Molecular Graphics System (Palo Alto, CA: DeLano Scientific).

Diamond, R. (1992). On the multiple simultaneous superposition of molecular structures by rigid body transformation. *Protein Sci* ***1***, 1279-1287.

Diamond, R. (1995). Coordinate-based cluster analysis. Acta Cryst D ***51***, 127-135.

Fitzkee, N.C., and Bax, A. (2010). Facile measurement of (1)H-(1)5N residual dipolar couplings in larger perdeuterated proteins. J Biomol NMR *48*, 65-70.

Goddard, T.D., Kneller, D. G. SPARKY 3 (San Francisco: University of California).

Goto, N.K., Gardner, K.H., Mueller, G.A., Willis, R.C., and Kay, L.E. (1999). A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. J Biomol NMR *13*, 369-374.

Huang, X., Aulabaugh, A., Ding, W., Kapoor, B., Alksne, L., Tabei, K., and Ellestad, G. (2003). Kinetic mechanism of Staphylococcus aureus sortase SrtA. Biochem *42*, 11307-11315.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., van den Boogaart, J.E., Pikkemaat, J.A., and Altona, C. (1992). An NMR study of the conformation and thermodynamics of the circular dumbbell d [formula: see text] Slow exchange between two- and four-membered hairpin loops. Journal of Biomolecular Structure & Dynamics *9*, 821-836.

Ippel, J.H., Lanzotti, V., Galeone, A., Mayol, L., Van den Boogaart, J.E., Pikkemaat, J.A., and Altona, C. (1995). Slow conformational exchange in DNA minihairpin loops: a conformational study of the circular dumbbell d<pCGC-TT-GCG-TT>. Biopolymers *36*, 681-694.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. Journal of applied crystallography *36*, 1277-1282.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. J Mol Biol *372*, 774-797.

Lakomek, N.A., Ying, J., and Bax, A. (2012). Measurement of (1)(5)N relaxation rates in perdeuterated proteins by TROSY-based methods. J Biomol NMR *53*, 209-221.

Langelier, M.F., Planck, J.L., Servent, K.M., and Pascal, J.M. (2011). Purification of human PARP-1 and PARP-1 domains from Escherichia coli for structural and biochemical analysis. Methods in molecular biology *780*, 209-226.

Langelier, M.F., Ruhl, D.D., Planck, J.L., Kraus, W.L., and Pascal, J.M. (2010). The Zn3 domain of human poly(ADP-ribose) polymerase-1 (PARP-1) functions in both DNA-dependent poly(ADP-ribose) synthesis activity and chromatin compaction. J Biol Chem *285*, 18877-18887.

Langelier, M.F., Servent, K.M., Rogers, E.E., and Pascal, J.M. (2008). A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation. J Biol Chem *283*, 4105-4114.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR *8*, 477-486.

Lee, D., Hilty, C., Wider, G., and Wuthrich, K. (2006). Effective rotational correlation times of proteins from NMR relaxation interference. J Magn Reson *178*, 72-76.

Lu, X.J., and Olson, W.K. (2003). 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. Nucleic Acids Res *31*, 5108-5121.

Marley, J., Lu, M., and Bracken, C. (2001). A method for efficient isotopic labeling of recombinant proteins. J Biomol NMR *20*, 71-75.

Neuhaus, D., and Williamson, M.P. (2000). The nuclear Overhauser effect in structural and conformational analysis; Second edition (New York: Wiley-VCH), pp. 136-143.

Piotto, M., Saudek, V., and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. J Biomol NMR *2*, 661-665.

Price, S., Oubridge, C., Varani, G., and Nagai, K. (1998). Preparation of RNA:protein complexes for X-ray crystallography and NMR. In RNA:protein interactions : a practical approach, C.W.J. Smith, ed. (Oxford ; New York: Oxford University Press), pp. 37-72.

Schuck, P. (2005). In “Modern Analytical Ultracentrifuge: Techniques and Methods”, D.J. Scott, S.E. Harding, and A.J. Rowe, eds. (Cambridge: The Royal Society of Chemistry), pp. 26-60.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. J Magn Reson *160*, 65-73.

Steffen, J.D., Tholey, R.M., Langelier, M.F., Planck, J.L., Schiewer, M.J., Lal, S., Bildzukewicz, N.A., Yeo, C.J., Knudsen, K.E., Brody, J.R*., et al.* (2014). Targeting PARP-1 allosteric regulation offers therapeutic potential against cancer. Cancer Res *74*, 31-37.

Svergun, D. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. Journal of applied crystallography *25*, 495-503.

Svergun, D., Barberato, C., and Koch, M.H.J. (1995). CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. Journal of applied crystallography *28*, 768-773.

Takeuchi, K., Ng, E., Malia, T.J., and Wagner, G. (2007). 1-C-13 amino acid selective labeling in a (HN)-H-2-N-15 background for NMR studies of large proteins. J Biomol NMR *38*, 89-98.

Tugarinov, V., Kanelis, V., and Kay, L.E. (2006). Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. Nature protocols *1*, 749-754.

Tugarinov, V., and Kay, L.E. (2003). Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. J Am Chem Soc *125*, 13868-13878.

Vranken, W.F., Boucher, W., Stevens, T.J., Fogh, R.H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J., and Laue, E.D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins *59*, 687-696.

Wüthrich, K. (1986). NMR of Proteins and Nucleic Acids (New York: John Wiley & Sons).