Supplementary Information

Title:

Solution structures of the two PBZ domains from human APLF and their interaction with poly(ADP-ribose)

Sebastian Eustermann,[1] Christoph Brockmann,[1] Pawan Vinod Mehrotra,[2] Ji-Chun Yang,[1] David Loakes,[1] Stephen C. West,[3] Ivan Ahel[2] & David Neuhaus.[1]

[1]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

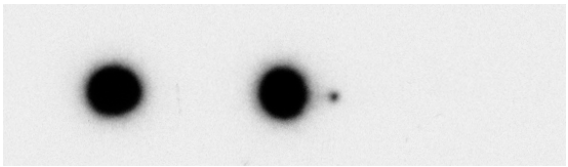
[2]Paterson Institute for Cancer Research, University of Manchester, Wilmslow Road, Manchester M20 4BX, UK.

[3]Cancer Research UK, London Research Institute, Clare Hall Laboratories, South Mimms, Herts EN6 3LD., U.K.

Supplementary Information comprises:

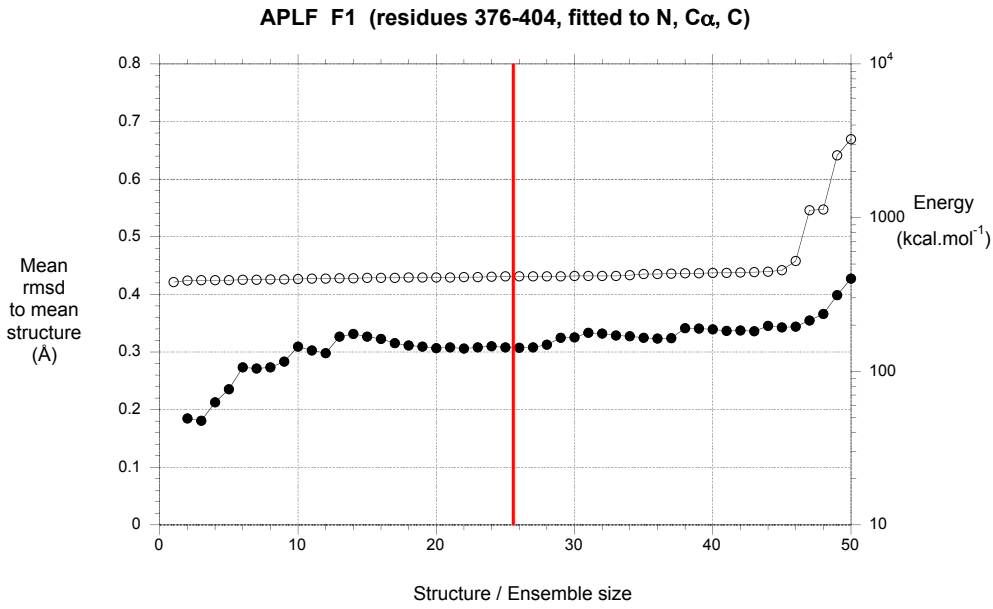
Supplementary Figures 1-11
Supplementary Tables 1 and 2
Supplementary Methods

GST-APLF GST-PBZ GST

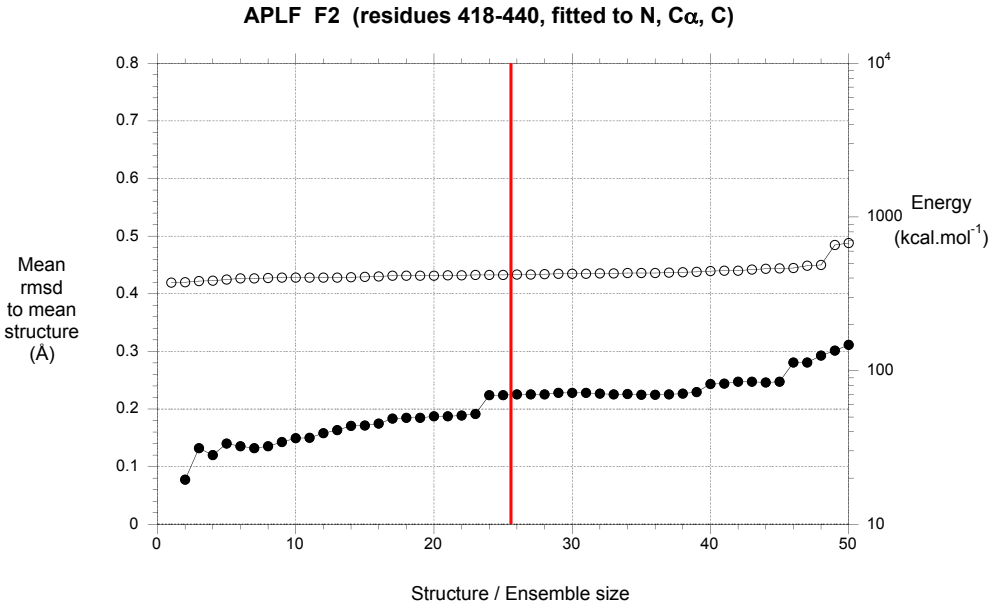


Supplementary Figure 1. Dot blot assay showing that GST-fused APLF 368-451 shows equivalent PAR-binding to that observed for GST-fused wild type APLF; the control shows that GST alone does not bind PAR.

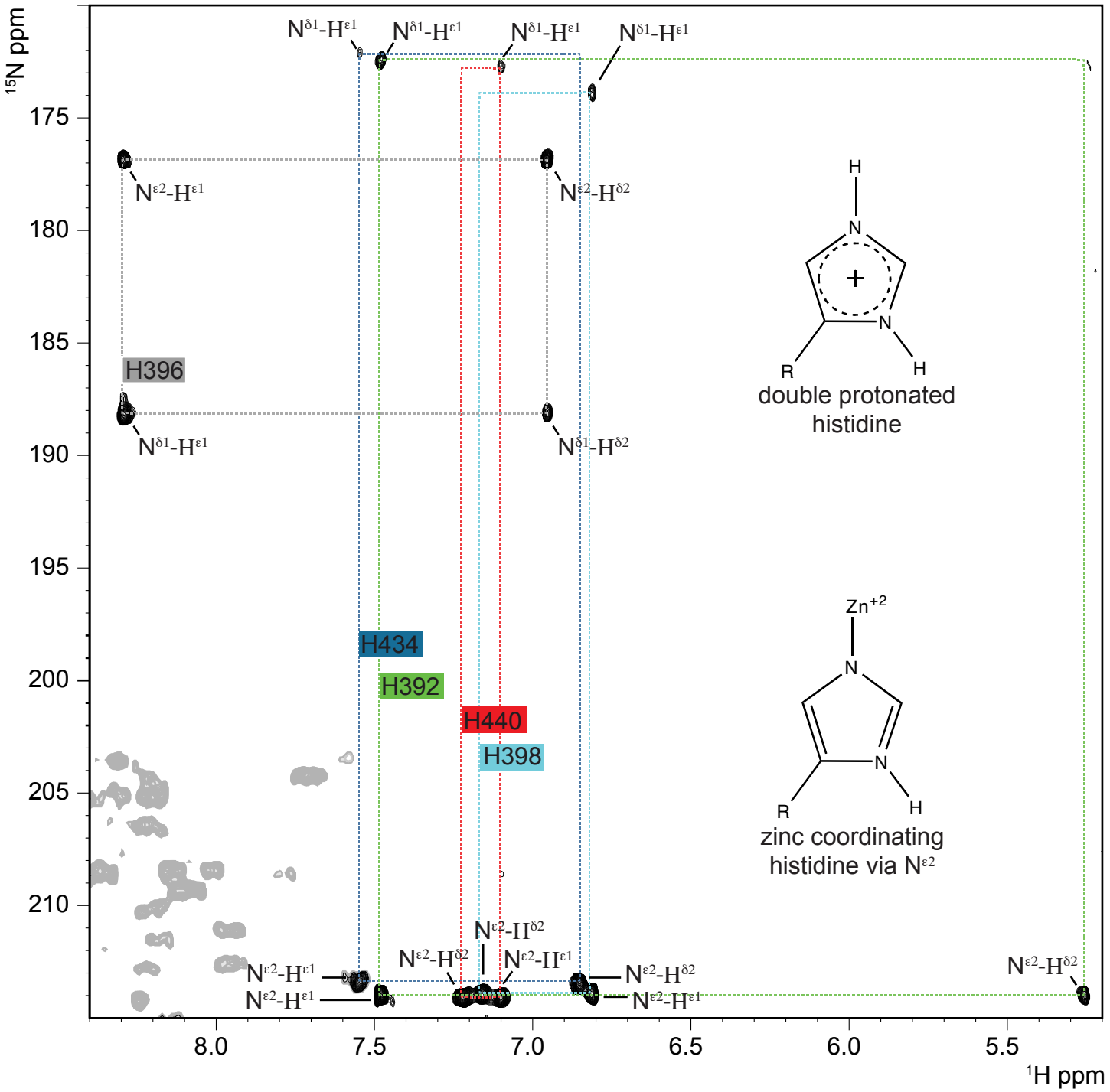
a



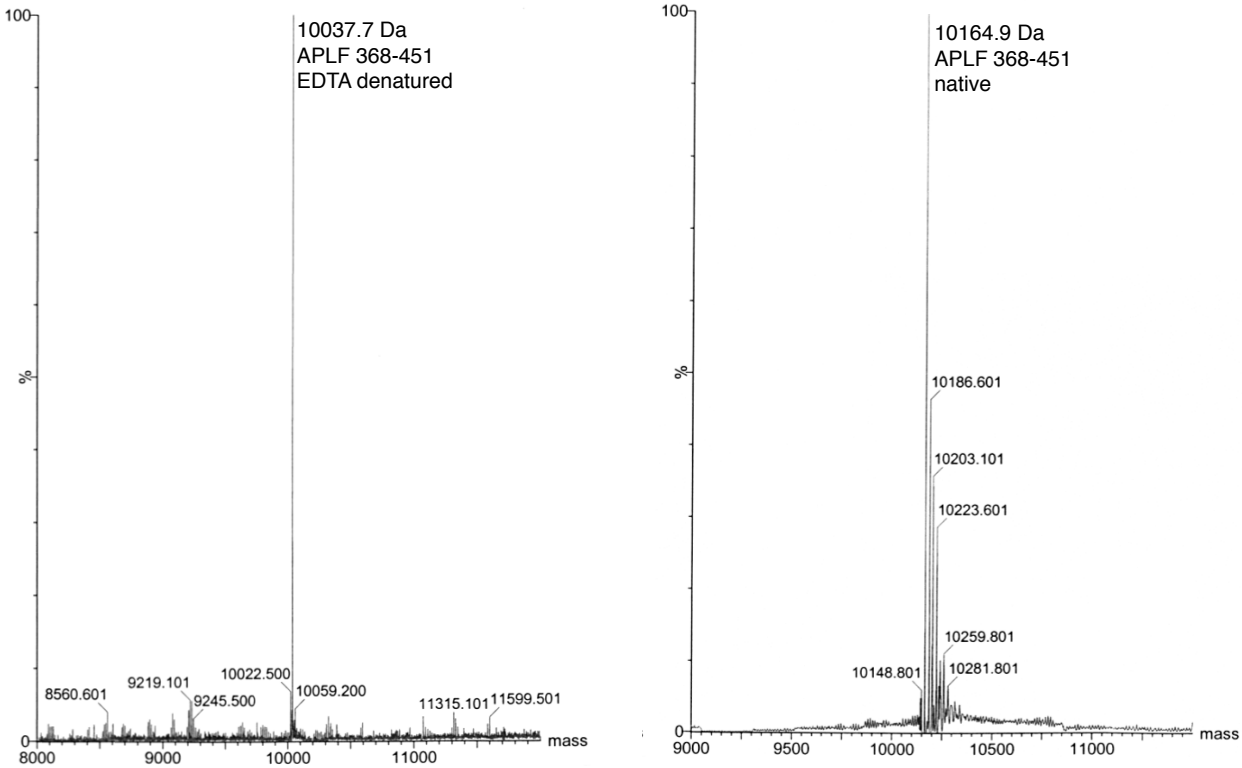
b



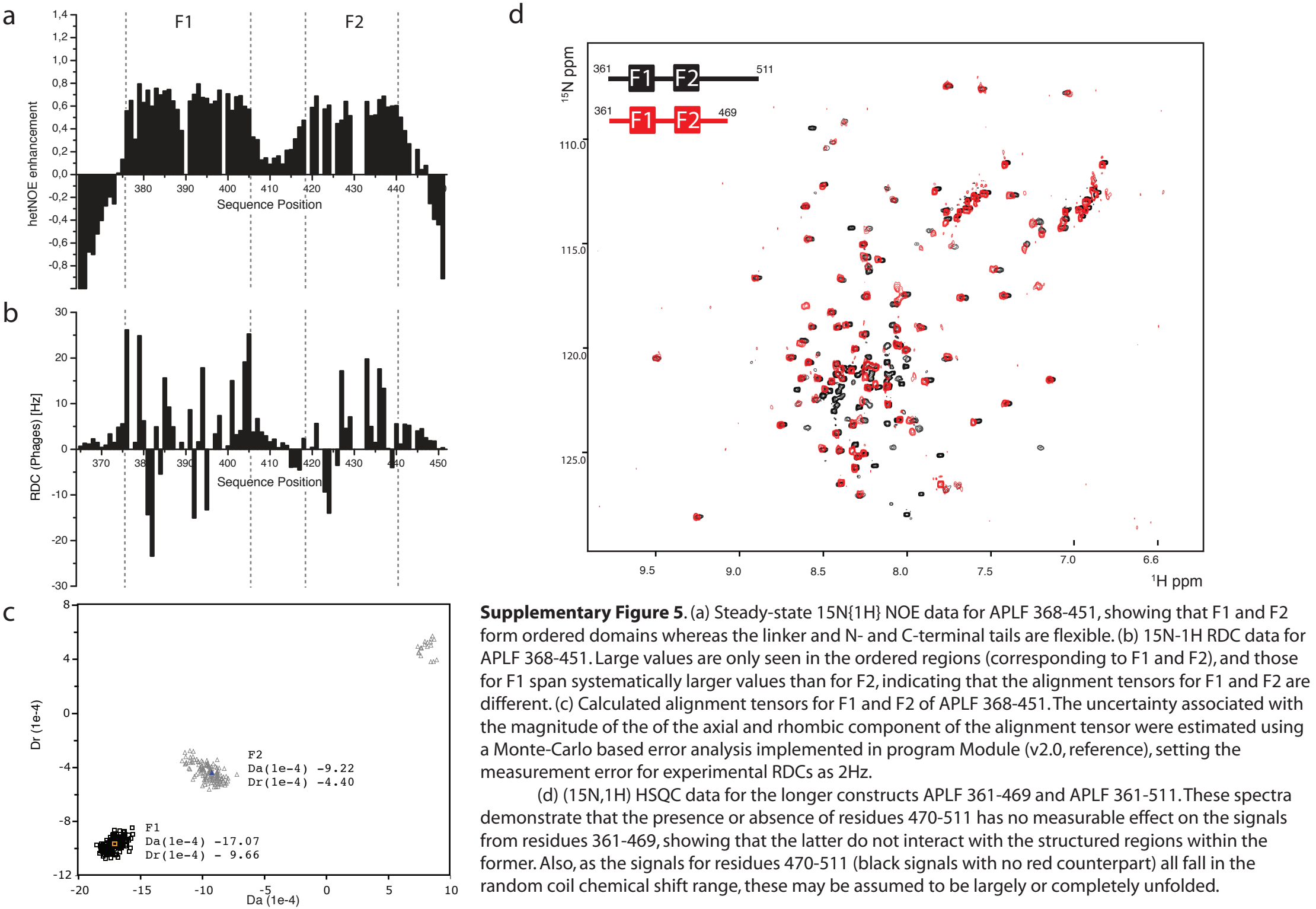
Supplementary Figure 2. Rmsd profiles for the deposited calculated ensembles of structures for (a) APLF F1 (b) APLF F2. Rmsd values (filled circles) are independently calculated using each ensemble size, adding successive structures in order of increasing XPLOR total energy term. Open circles represent the XPLOR total energy terms. Only structures to the left of the vertical red line were included when calculating the structural statistics. Rmsd calculations employed the program CLUSTERPOSE (Diamond, Acta Cryst. D (1995) 51, 127-135).



Supplementary Figure 3. Long-range HSQC spectrum for APLF 368-451. For all four metal-binding histidines, the pattern of cross peaks is uniquely compatible with co-ordination through the Nε2 atom (c.f. Legge et al., J.M.B. (2004) 343, 1081–1093). Histidine H396 shows a pattern typical for a non-ligating histidine in a partially protonated state.

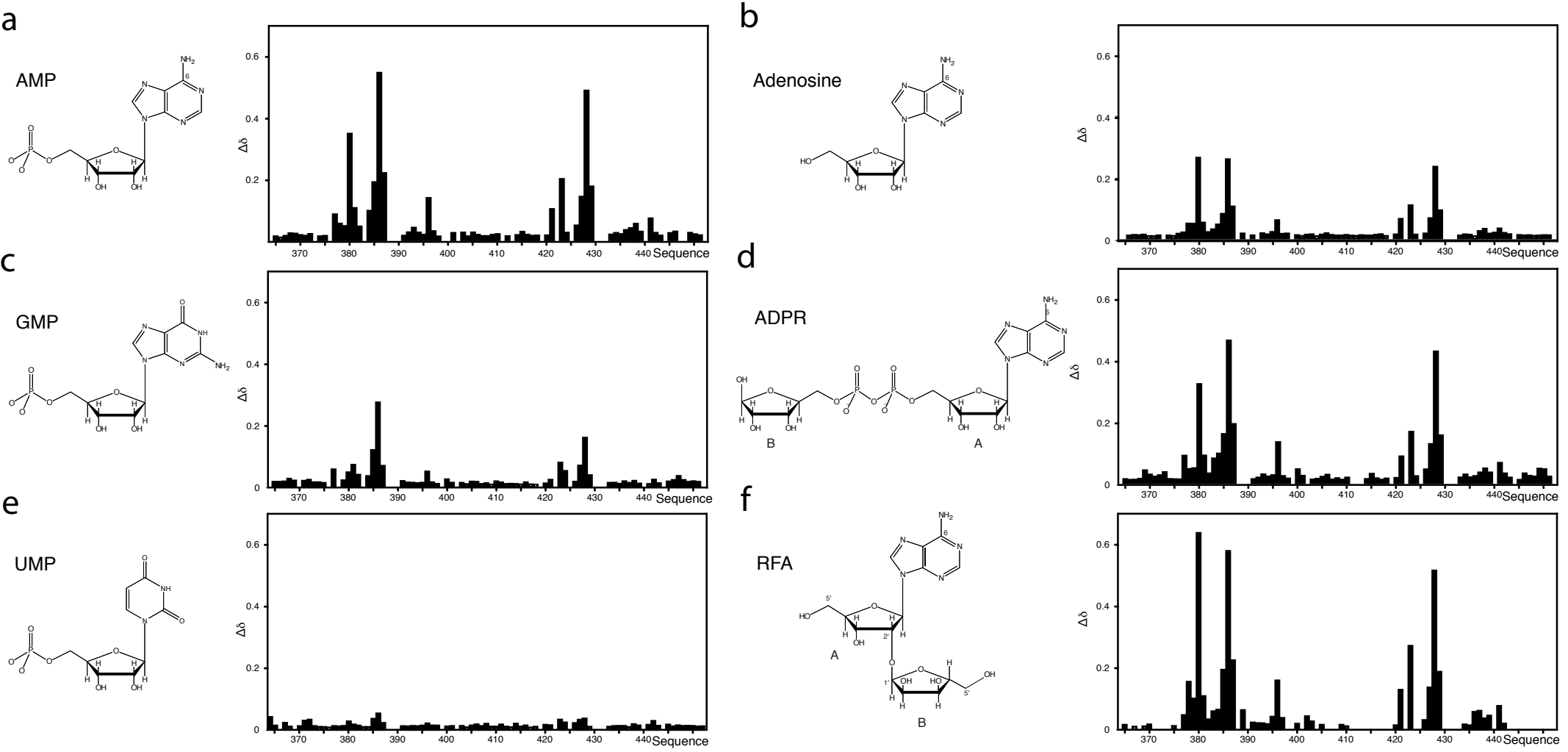


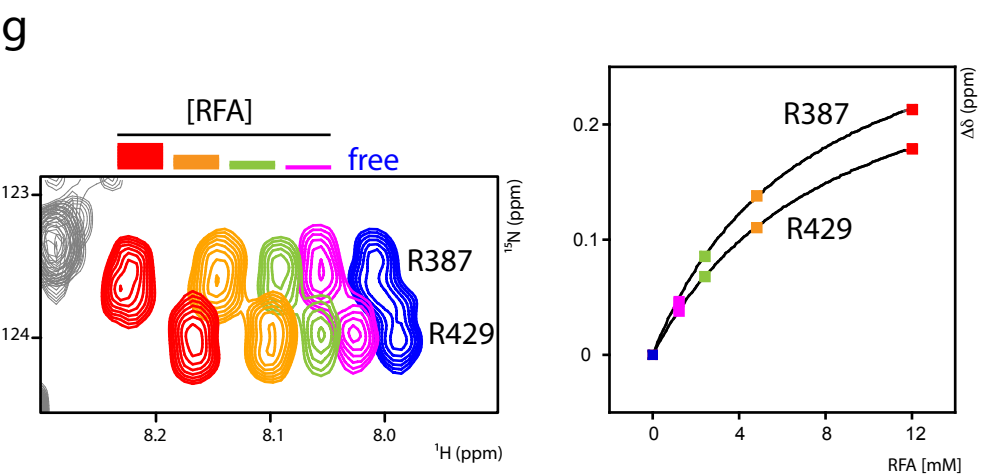
Supplementary Figure 4. Mass spectra showing that native APLF 368-451 binds two zinc ions (10164.9 – 10037.7 = 127.2Da; 2x Zn = 130.76).

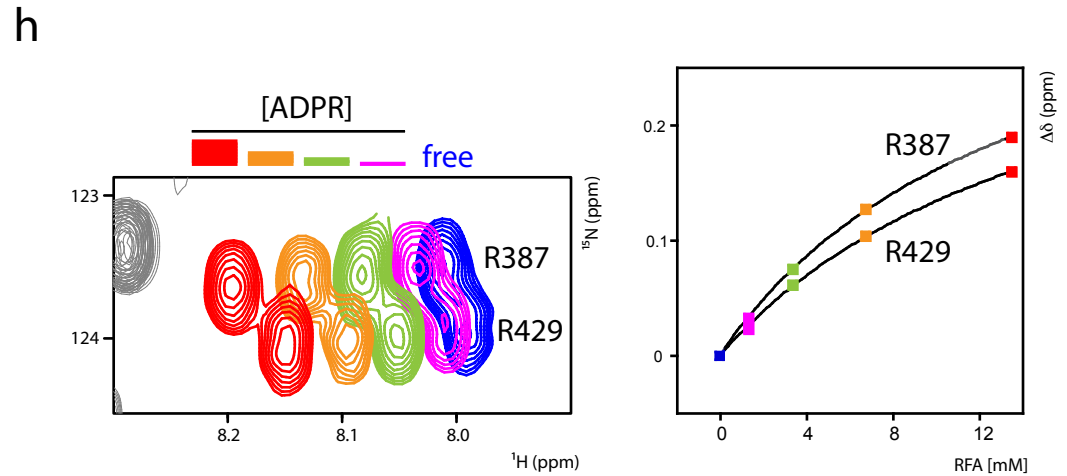


Supplementary Figure 5. (a) Steady-state 15N{1H} NOE data for APLF 368-451, showing that F1 and F2 form ordered domains whereas the linker and N- and C-terminal tails are flexible. (b) 15N-1H RDC data for APLF 368-451. Large values are only seen in the ordered regions (corresponding to F1 and F2), and those for F1 span systematically larger values than for F2, indicating that the alignment tensors for F1 and F2 are different. (c) Calculated alignment tensors for F1 and F2 of APLF 368-451. The uncertainty associated with the magnitude of the of the axial and rhombic component of the alignment tensor were estimated using a Monte-Carlo based error analysis implemented in program Module (v2.0, reference), setting the measurement error for experimental RDCs as 2Hz.

(d) (15N,1H) HSQC data for the longer constructs APLF 361-469 and APLF 361-511. These spectra demonstrate that the presence or absence of residues 470-511 has no measurable effect on the signals from residues 361-469, showing that the latter do not interact with the structured regions within the former. Also, as the signals for residues 470-511 (black signals with no red counterpart) all fall in the random coil chemical shift range, these may be assumed to be largely or completely unfolded.


a
AMP
b
Adenosine
c
GMP
d
ADPR
e
UMP
f
RFA
Δδ
Sequence
370
380
390
400
410
420
430
440

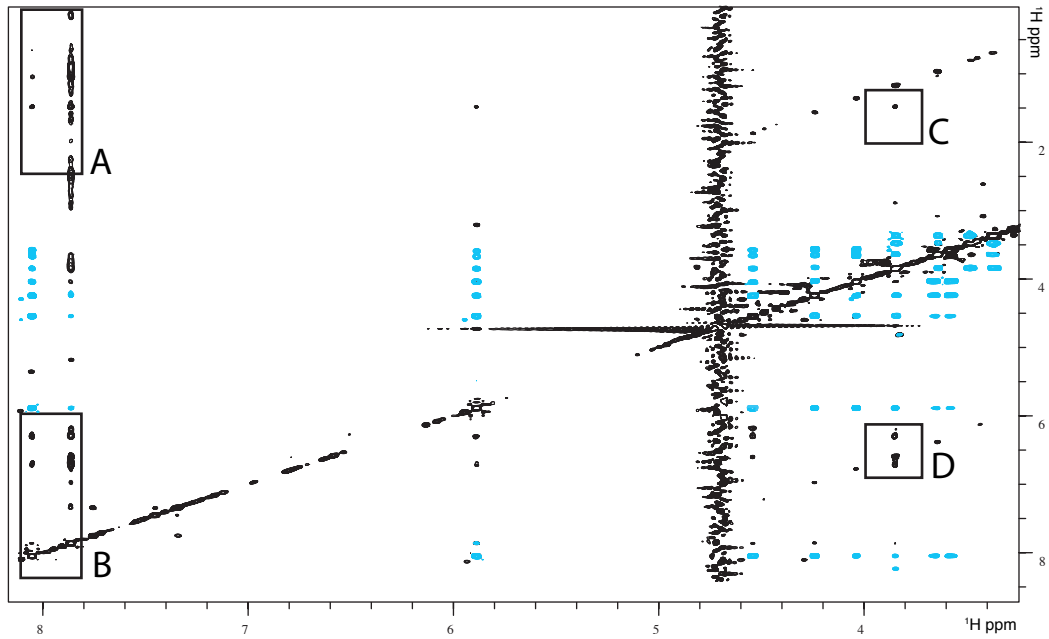

g
[RFA]
free
R387
R429
^{15}N (ppm)
^{1}H (ppm)
Δδ (ppm)
RFA [mM]

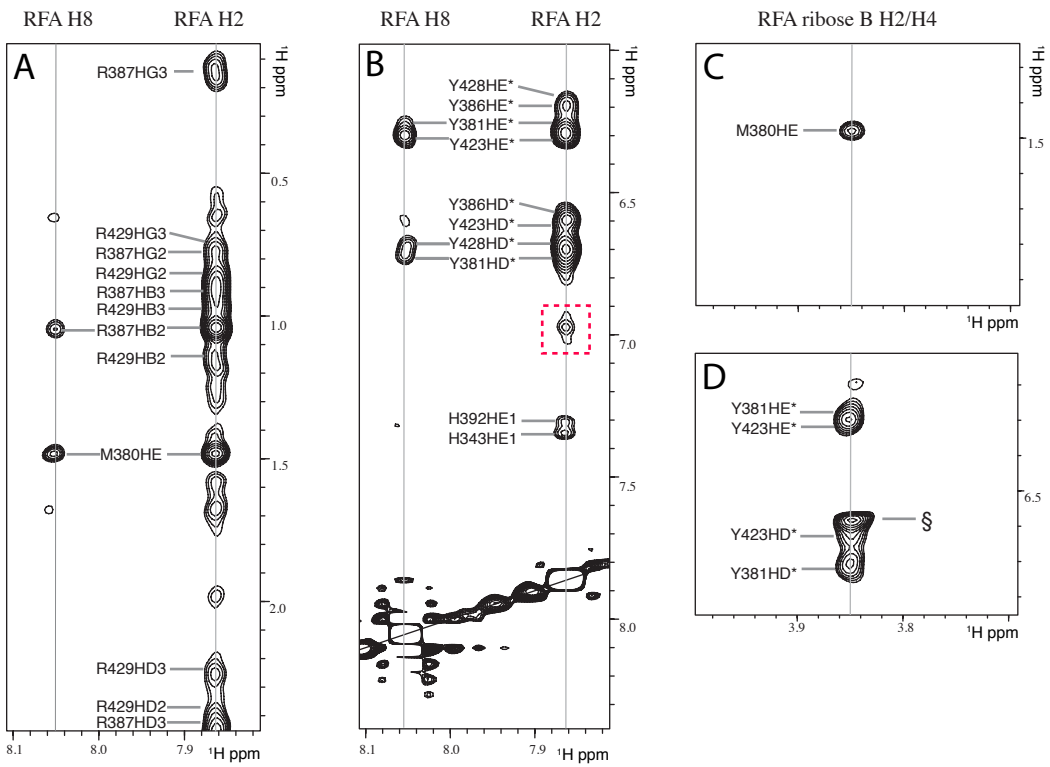

h
[ADPR]
free
R387
R429
^{15}N (ppm)
^{1}H (ppm)
Δδ (ppm)
RFA [mM]

Supplementary Figure 6. Amide group chemical shift perturbation data for addition of various ligands to APLF 368-451, calculating Δδ as √((δ1H)2 + (δ15N ÷ 5)2). Concentrations: APLF 100μM, ligand 8mM. Addition of AMP (a) caused substantial and specific chemical shift perturbations involving the absolutely conserved surface-exposed Tyr residues (Y381 and Y386 in F1, Y423 and Y428 in F2) and neighbouring residues, suggesting that these tyrosines recognise the base through stacking interactions. Addition of adenosine (b) or GMP (c) caused a similar pattern of changes, though more weakly, whereas addition of UMP (e) caused only minor perturbations, showing the PBZ modules do not bind purine bases. Addition of ADPR (d) caused very similar shift perturbations to those caused by RFA (e), showing that the pyrophosphate of ADPR is not recognized.

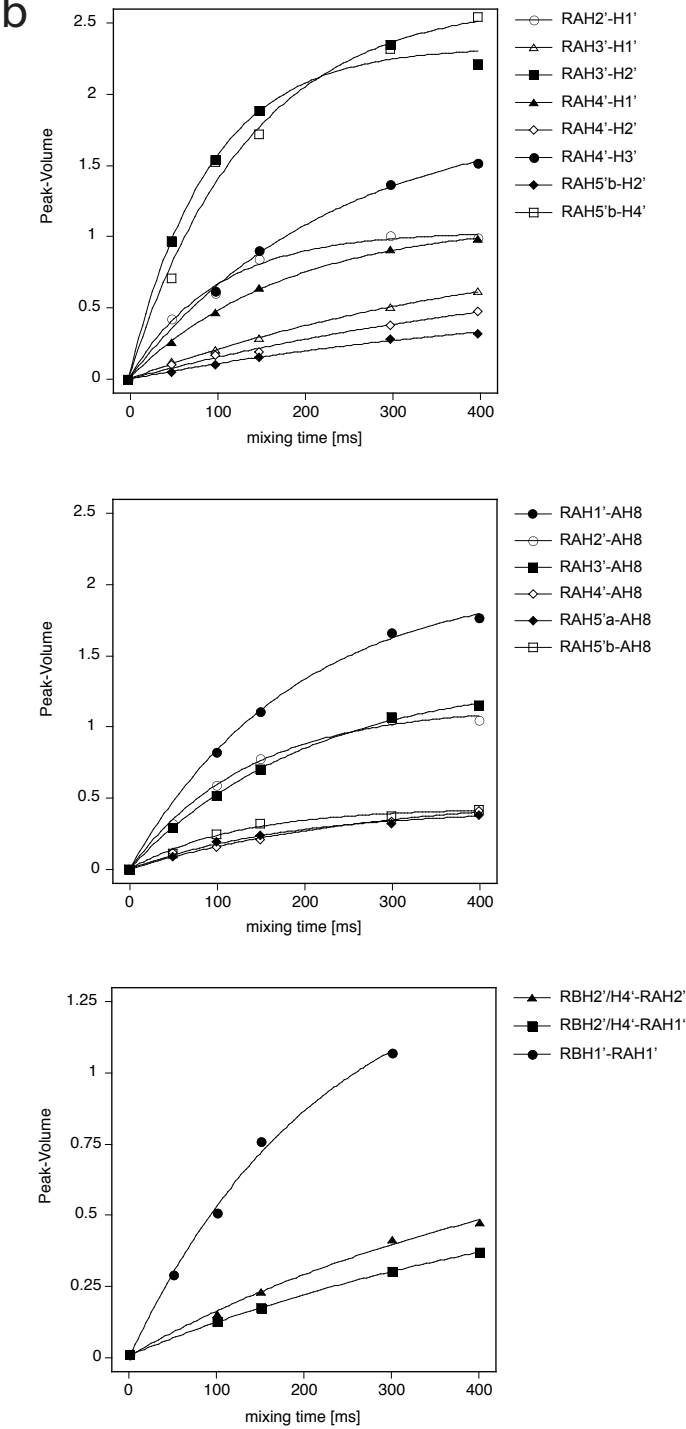
Binding curves for addition of (g) RFA and (h) ADPR to APLF 368-451, showing the chemical shift perturbations for the arginine residue of the conserved CYR motif in each finger, i.e. R387 in F1 and R429 in F2. The ligand concentrations used (see color code on figure) were 1.2mM, 2.4mM, 4.8mM and 12mM for RFA and 1.35mM, 3.37mM, 6.75mM and 13.5mM for ADPR, and the approximate Kd values derived from these curves are 7.2±0.2 mM (RFA binding to F1), 8.3±0.05 mM (RFA binding to F1), 14.0±1.5 mM (ADPR binding to F1) 16.5±1.7 mM (ADPR binding to F1); data were fit to the equation Δδ= Δδmax x [ligand]/(Kd + [ligand]) with Δδmax and Kd as the fitting variables.

a


A
B
C
D
¹H ppm


RFA H8
RFA H2
A
R387HG3
R429HG3
R387HG2
R429HG2
R387HB3
R429HB3
R387HB2
R429HB2
M380HE
R429HD3
R429HD2
R387HD3
B
Y428HE*
Y386HE*
Y381HE*
Y423HE*
Y386HD*
Y423HD*
Y428HD*
Y381HD*
H392HE1
H343HE1
RFA ribose B H2/H4
C
M380HE
D
Y381HE*
Y423HE*
Y423HD*
Y381HD*
§
¹H ppm

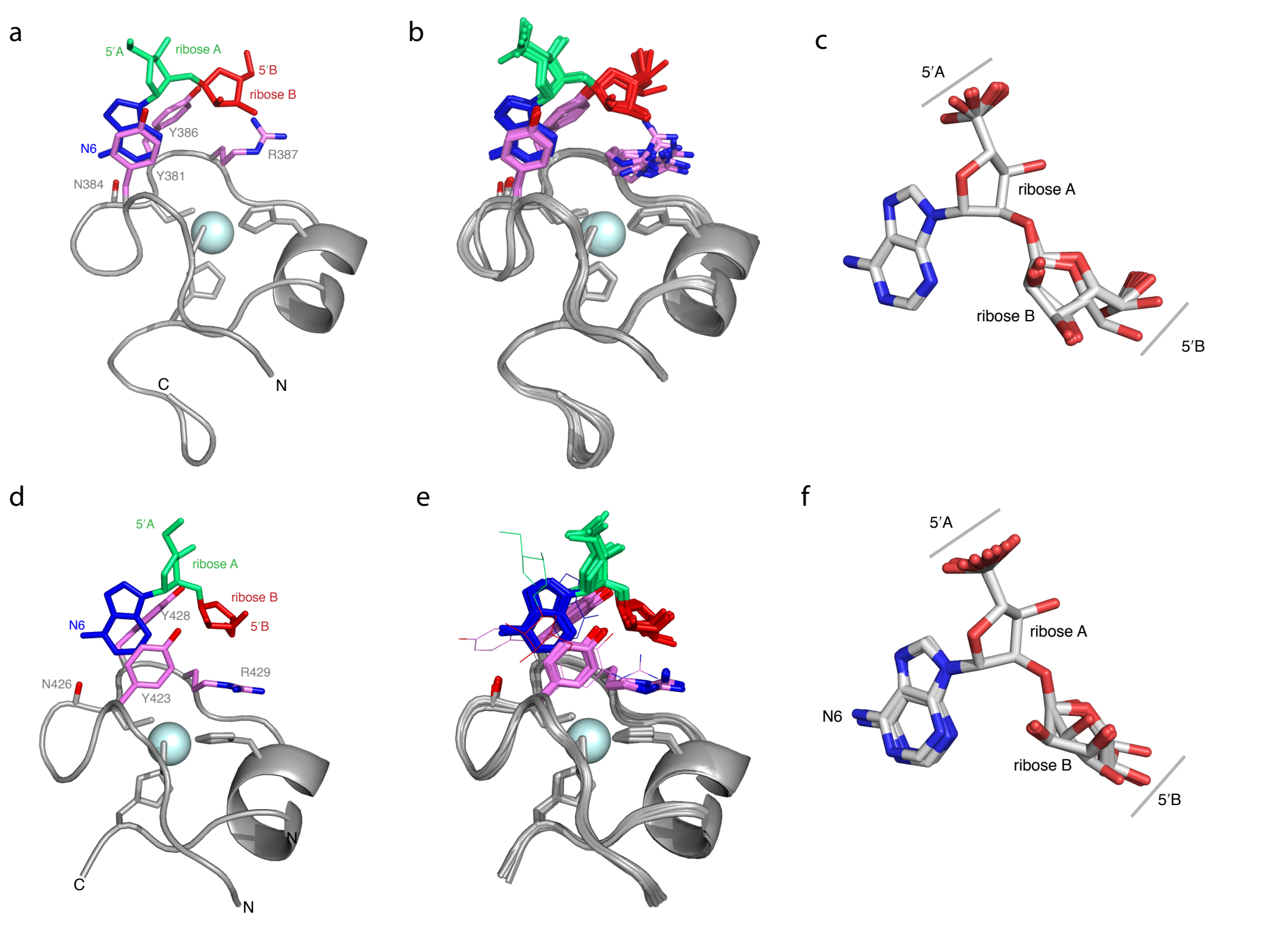
b


Peak-Volume
mixing time [ms]
RAH2'-H1'
RAH3'-H1'
RAH3'-H2'
RAH4'-H1'
RAH4'-H2'
RAH4'-H3'
RAH5'b-H2'
RAH5'b-H4'
RAH1'-AH8
RAH2'-AH8
RAH3'-AH8
RAH4'-AH8
RAH5'a-AH8
RAH5'b-AH8
RBH2'/H4'-RAH2'
RBH2'/H4'-RAH1'
RBH1'-RAH1'

Supplementary Figure 7. (a) NOESY spectrum of the complex of APLF 368-451 with RFA, showing both intermolecular NOE cross-peaks and, in light blue, transferred NOE cross peaks. The latter are cross-peaks connecting two ligand signals and reflecting the bound-state conformation (all peaks are positive). Cross-peak assignments are indicated on the expansions, which in turn are boxed on the full spectrum; HD* and HE* refer to rotationally averaged signals from both HD1 and HD2, or from HE1 and HE2, respectively. The peak labelled § is an artefact. (b) Buildup curves for selected transferred NOE cross peaks.

For ADPR, no intermolecular NOEs were observed to ribose B, which in the case of ADPR lies beyond the pyrophosphate, and no transferred NOEs are seen for ribose B, suggesting that it has relatively high mobility (data not shown). In contrast, ribose B of RFA, which is linked to ribose A through the glycosidic bond, shows several intermolecular contacts to the APLF PBZ modules and numerous transferred NOEs. In combination these observations show that both ribose A and B of RFA are recognised by the PBZ modules of APLF, whereas for APLF only ribose A is recognised.

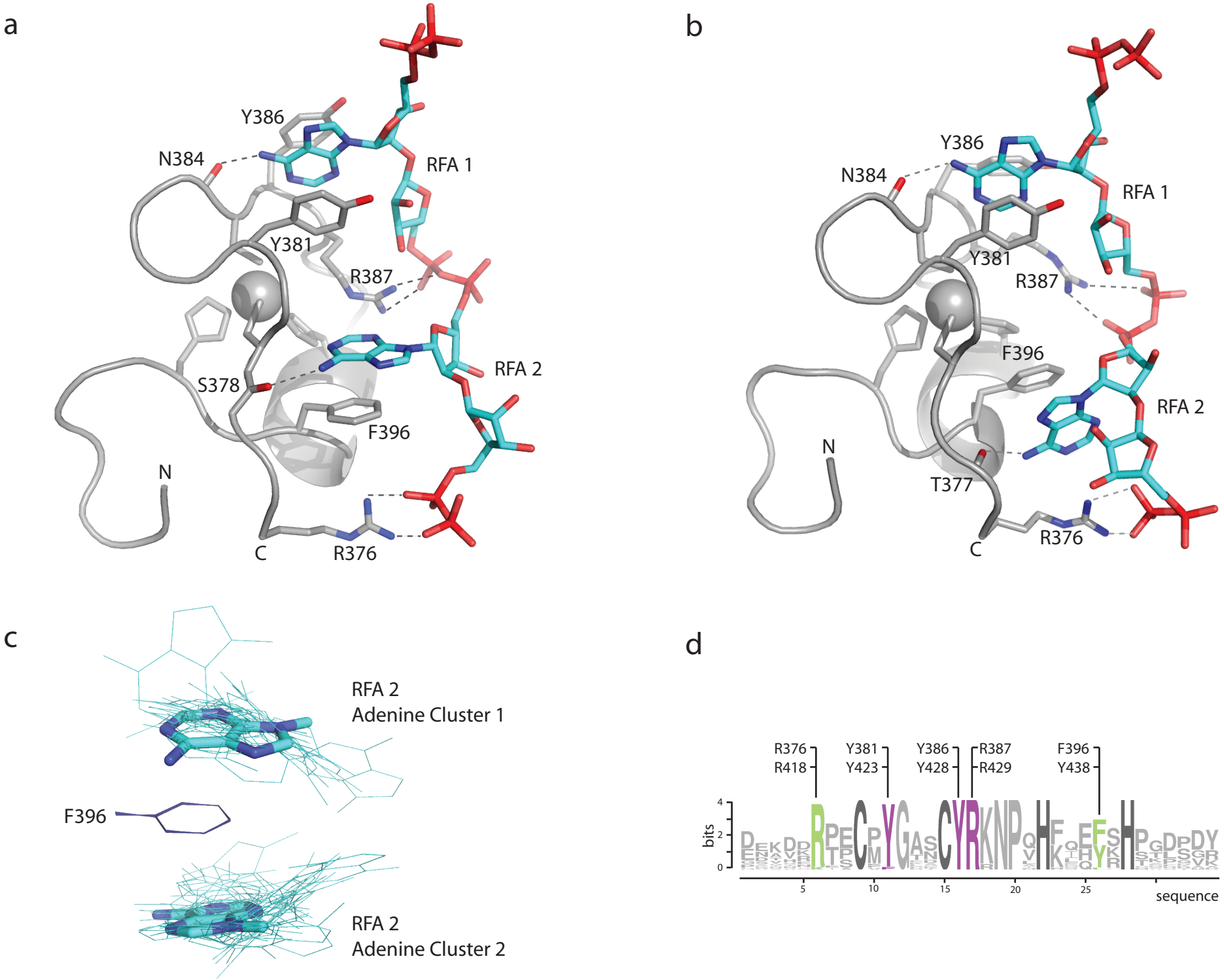
The weak NOE cross peak boxed in red in box B of panel a) links H2 of the adenine of RFA to the Hζ proton of F396. This NOE cannot arise from adenine bound in the primary binding site as the corresponding distance would be least 13Å; we assign it to binding of RFA in a second site (see Supplementary Figure 9). Due to overlap in the protein spectrum, no other intermolecular NOE can be unequivocally assigned to RFA binding in the second site (either in F1 or in F2), but it is likely that there are other weak contributions to the NOESY spectrum arising from second site binding. However, under the experimental conditions used for the NOESY experiments, occupancy of this second site appears to be low, so we do not expect such contributions to interfere with the experimental determination of the structure of the complex with RFA bound at the first binding site. Note also that, since binding of RFA is in the fast exchange regime on the chemical shift timescale, only one set of averaged signals is seen for both protein and ligand; intermolecular NOEs from either binding site therefore all involve the same, single, set of chemical shifts.


a
5′A
ribose A
5′B
ribose B
Y386
N6
R387
Y381
N384
C
N
b
c
5′A
ribose A
ribose B
5′B
d
5′A
ribose A
ribose B
N6
Y428
5′B
R429
N426
Y423
N
C
N
e
f
5′A
ribose A
N6
ribose B
5′B

Supplementary Figure 8. Structures calculated for the complex of APLF 368-451 with RFA; the binding sites of both F1 and F2 are simultaneously and independently occupied under the experimental conditions. Panels (a), (b) and (c) show results for the F1-RFA complex, (d), (e) and (f) show results for the F2-RFA complex. Panels (a) and (d) show the lowest energy structure, panels (b) and (e) show the ensemble of 10 lowest energy structures (of 50 calculated) superposed on the protein backbone, and panels (c) and (f) show the ligand only for the same ensemble, superposed on the heavy atoms of ribose A. In (a), (b), (d) and (e), the adenine ring is blue, ribose A is green, ribose B is red and interacting sidechains of APLF are purple with red or blue heteroatoms.

These ensembles clearly show the predominant binding mode of RFA to both F1 and F2. In each case, the adenine ring is stacked with the two conserved tyrosines (Y381 and Y386 in F1, Y423 and Y428 in F2), though for two (Y386 and Y428) the interaction may be edge-face. The N6 amino group of the RFA adenine ring is positioned to form a hydrogen bond to a backbone carbonyl of the protein (N384 in F1, S426 in F2), and the C5'OH group of ribose B of RFA is positioned close to the arginine sidechain of the highly conserved CYR sequence motif (R387 in F1, R429 in F2).

Since these ligands bind quite weakly to the PBZ modules (each binding site is approximately 30-40% saturated under the experimental conditions), it is to be expected that there may be some exchange involving different binding configurations, and since the system is in fast exchange all of the measured NMR data will reflect an average over any such exchange. Consistent with this, we see a small number of calculated structures in which the adenine ring is reversed in the binding site, resulting in a substantially different binding arrangement; one such outlier exists for F2 in the ensemble shown but none for F1 (in the figure, the outlier is shown using thinner lines for the ligand and interacting protein sidechains). It is possible that such minor forms are indeed present in solution, but our data cannot determine this uniquely.


a
Y386
N384
RFA 1
Y381
R387
RFA 2
S378
F396
N
C
R376
b
Y386
N384
RFA 1
Y381
R387
F396
RFA 2
N
T377
C
R376
c
RFA 2
Adenine Cluster 1
F396
RFA 2
Adenine Cluster 2
d
R376
R418
Y381
Y423
Y386
Y428
R387
R429
F396
Y438
bits
4
2
0
5
10
15
20
25
sequence

Supplementary Figure 9. Models of possible structures for a putative complex of APLF F1 with a longer PAR fragment, pop-RFA-pop-RFA-pop.

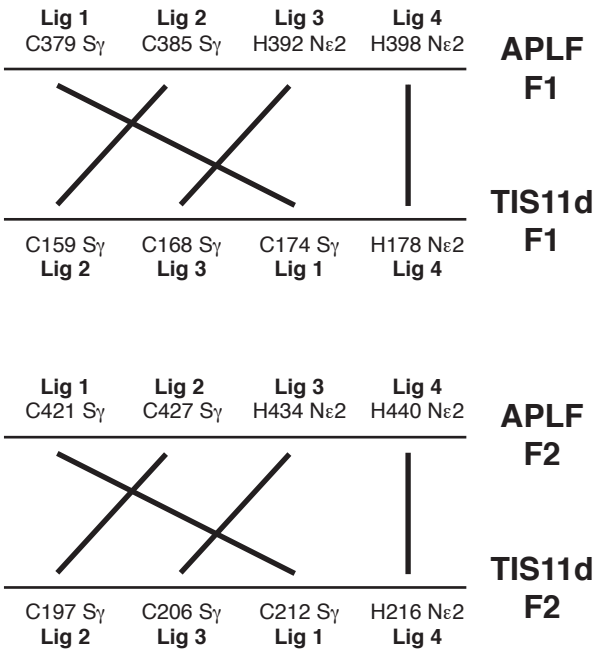
The starting point for this modelling was the observation of a weak intermolecular NOE from H2 of the adenine ring of RFA to Hζ of residue F396 in F1 (highlighted in Supplementary Figure 7). When bound in its primary binding site, the adenine ring of RFA is too remote from F396 to explain the occurrence of this intermolecular NOE, so we propose that these data probably indicate binding of the adenine to a secondary site; the NOE is weak presumably due to partial occupancy of the second binding site under the experimental conditions. Consistent with the existence of such a second binding site, the ring protons of F396 were strongly perturbed during titration of APLF PBZ domain with ADPR to excess, strongly suggesting that the ring of F396 stacks with the adenine of RFA when bound in the second site.* Although we have only modelled a second site interaction for F1, a similar interaction is equally possible for F2 (involving mainly Y438), but in the case of F2 spectroscopic overlap obscures the regions where corresponding evidence would occur.

Inspection of the structures of manually constructed models quickly showed that the most plausible mode of interaction for a second adenine ring with F396 would be for the two rings to stack together, and further that such an interaction could bring the third pyrophosphate into proximity to the sidechain of the absolutely conserved arginine R376. Such an arrangement would recapitulate a number of features of the first binding site. To test whether such an arrangement was indeed plausible, we used a three-stage protocol to model the interaction as described in Supplementary Methods. Initial XPLOR-NIH calculations, similar to those used for the calculation of the APLF-RFA complexes, were used to generate a set of starting structures compatible with the experimental data and starting assumptions, and in which the second RFA moiety was distributed over a wide range of spatial locations relative to the protein. In a second step, simulated annealing calculations in a full force field using AMBER 9 were used to test whether these starting conformations could converge on energetically plausible bound state structures. Finally, the program DOCK-6 was used to rank the resulting conformations.

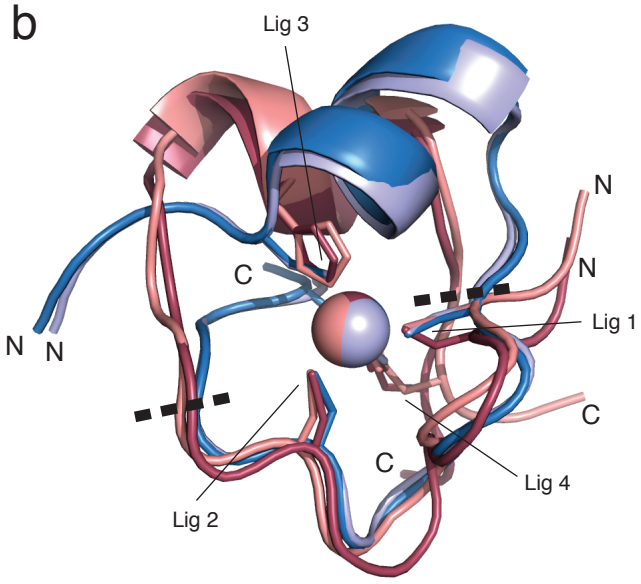
The resulting ensemble of 50 conformers includes energetically reasonable binding arrangements having the second adenine ring on either side of the ring of F396; all the DOCK-6 scores for these structures fall in the range between -62 and -41 kcal.mol-1. Representative single examples of each binding arrangement are shown in panels a) and b) (these are the second and sixth in order of binding energy, respectively). These conformers demonstrate that a two-unit fragment of PAR can occupy both the primary and putative secondary binding sites simultaneously without causing significant steric strain, and while making a combination of favorable stacking, electrostatic and hydrogen bonding interactions that are quite similar between the two sites (note however that the hydrogen bonding of the adenine N6 amino group, though entirely plausible as shown here, is not particularly well conserved across the ensemble). The available data do not allow us to distinguish which of the two possible arrangements a) or b) is more likely to be correct. Panel c) illustrates how the ensemble of 50 calculated conformers partitions into two sets according to which side of the F396 ring stacks against the second adenine; in this view, just the rings of the second adenine and F396 are shown, superposed on the ring carbon atoms of F396, and the individual conformations shown in panels a) and b) are shown using sticks. Panel d) shows a sequence logo of conservation amongst PBZ fingers F1 and F2 in APLF orthologues (the same set as is used in Supplementary Figure 11), showing that the key interacting residues in the putative second binding site are comparably well conserved to those in the primary binding site. In this view, ligand interacting residues in the primary ligand-binding site are colored violet, potential ligand-interacting residues in the second ligand-interacting site are colored light green and metal-binding residues are colored dark grey. The sequence logo was generated using the program WebLogo 3.0 (Crooks G.E., Hon G., Chandonia J.M. and Brenner S.E. (2004) "WebLogo: A sequence logo generator" Genome Research, 14, 1188-1190).

**Note that, as far as the adenine/F396 interaction is concerned, we attach no significance to the different ligands used in these two experiments; they differ only because the ADPR titration experiment was the only one to employ 13C-labelled protein, required for monitoring of the aromatic signals.*

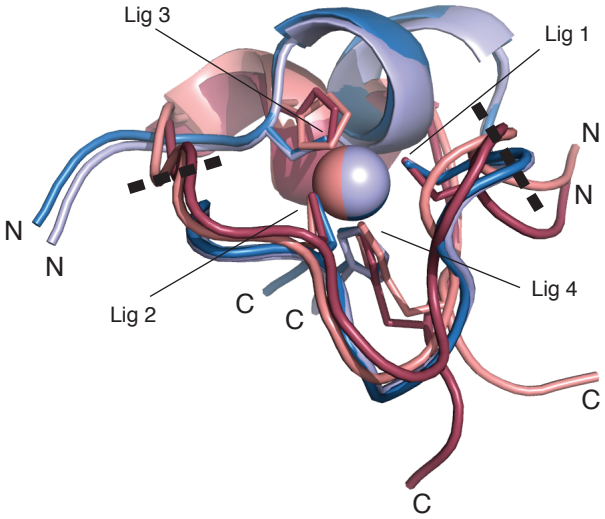
a


Lig 1
C379 Sγ
Lig 2
C385 Sγ
Lig 3
H392 Nε2
Lig 4
H398 Nε2
APLF
F1
TIS11d
F1
C159 Sγ
Lig 2
C168 Sγ
Lig 3
C174 Sγ
Lig 1
H178 Nε2
Lig 4
Lig 1
C421 Sγ
Lig 2
C427 Sγ
Lig 3
H434 Nε2
Lig 4
H440 Nε2
APLF
F2
TIS11d
F2
C197 Sγ
Lig 2
C206 Sγ
Lig 3
C212 Sγ
Lig 1
H216 Nε2
Lig 4

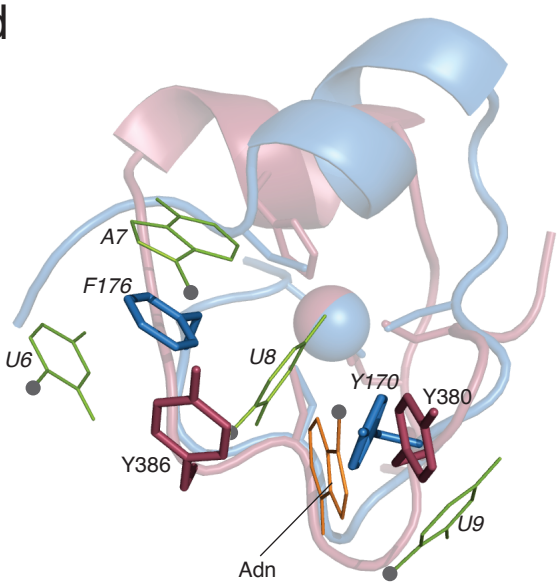
b


Lig 3
N
N
C
Lig 1
N
N
C
C
Lig 4
Lig 2

c


Lig 3
Lig 1
N
N
N
N
Lig 4
Lig 2
C
C
C

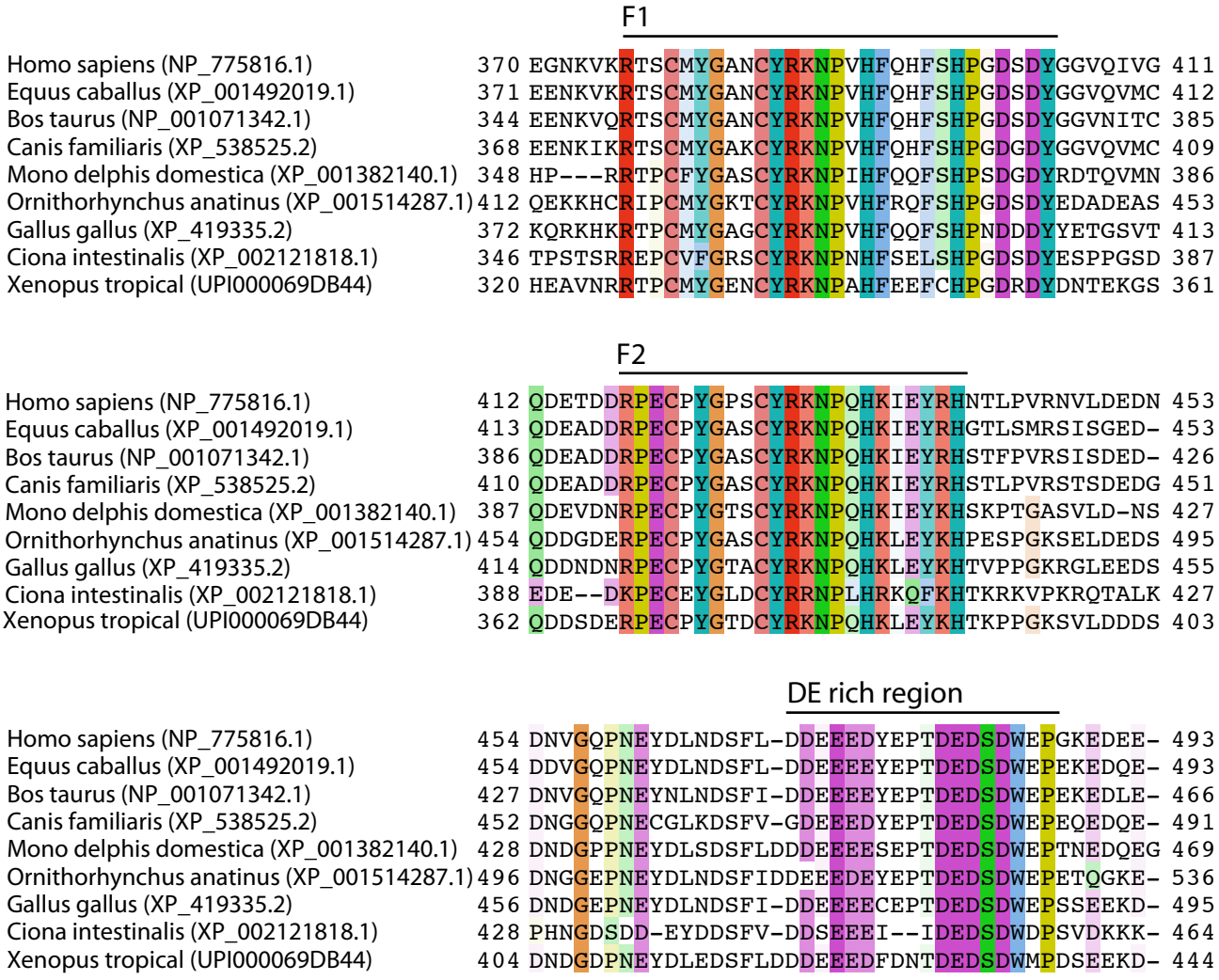
d


A7
F176
U6
U8
Y170
Y380
Y386
U9
Adn

Supplementary Figure 10. Similarity of the zinc-finger domains from APLF (dark red, F1, and pink, F2) and TIS11d (dark blue, F1, and light blue, F2). Conventional sequence-based structural alignment does not reveal particularly strong similarities between the zinc fingers of these two proteins. However, superimposition using just the four atoms directly bonded to zinc in each case (labelled Lig1-4 in the figure) allows different permutations to be explored, and in this way a close fit was found using a non-sequential superposition in which pairs of zinc-bonded atoms were superposed as shown in a). This arrangement is obtained from the sequential alignment by rotating one of the molecules through 120° about the axis of the bond connecting the zinc to ligand 4. Note that, because this alignment is non-sequential, the N- and C- termini of the fingers are in different spatial locations between the APLF and TIS11d fingers.

When the molecules are superposed in this way (panels b and c, views related by a 45° rotation about x), there is a striking similarity in the backbone conformation of the fingers in the common regions (that is, between the dashed lines in panel b and c). For the two F1 structures, the backbone rmsd (N, Cα, C') between APLF F1 379-386 and TIS11d F1 168-175 is 0.72Å, while for the two F2 structures the corresponding rmsd between APLF F2 421-428 and TIS11d F2 206-213 is 0.46Å. Also, while the axes of the helices do not align closely in this superposition, the helices do occupy a common region of space.

Still more strikingly, in this arrangement the residues used for ligand binding by the different fingers are in closely similar locations. Panel d shows the key aromatic side chains that stack with an adenosine ring in the case of APLF F1, or with RNA bases in the case of TIS11d F1 (results are similar for the two F2 domains; not shown). However, because of the different nature of the ligands, in particular the greater distance between successive adenosines in PAR than between successive bases in RNA, the nature of the stacking is different between APLF and TIS11d. Thus the aromatic rings of residues Y380 and Y386 in finger 1 of APLF stack with a single adenosyl ring of PAR (mimicked by RFA in our work), whereas the rings of residues Y170 and F176 of finger 1 TIS11d are each sandwiched between two RNA bases. The adenine ring (of RFA) is shown in orange, and RNA bases are shown in light green, and for all bases the position of the C1' atom is shown by a grey dot. Both APLF and TIS11d are shown in their bound conformations in all panels.



Supplementary Figure 11. Sequence alignment for APLF. Sequences were retrieved from the NCBI (www.ncbi.nlm.nih.gov) and the UniProt database (www.uniprot.org) and multiple sequence alignment performed using the program ClustalW2 (Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. (2007) "ClustalW and ClustalX version 2." Bioinformatics, 23, 2947-2948). The figure was produced using the program Jalview 2 (Waterhouse, A.M., Procter, J.B., Martin, D.M.A, Clamp, M. and Barton, G. J. (2009) "Jalview Version 2 - a multiple sequence alignment editor and analysis workbench" Bioinformatics doi: 10.1093/bioinformatics/btp033).

	APLF F1	APLF F2
NMR distance and dihedral restraints		
Distance restraints		
Total NOE	715	595
Intra-residue	151	145
Inter-residue		
Sequential ($\|i - j\| = 1$)	269	228
Medium-range ($2 \leq \|i-j\| \leq 4$)	163	138
Long-range ($\|i - j\| > 4$)	132	84
Total dihedral angle restraints	36	18
ϕ	12	5
ψ	13	5
$\chi 1$	11	8
Total RDC restraints	27	21
Structure statistics		
Violations		
Distance viols. > 0.2Å (average per structure)	2.0 ± 1.1	2.6 ± 1.3
Dihedral angle viols. > 5° (average per structure)	1.3 ± 0.5	6.6 ± 1.1
R.m.s. deviation to distance restraints (Å)	0.026	0.031
R.m.s. deviation to angle restraints (°)	3.43	3.10
Max. distance restraint violation (Å)	0.57	0.52
Max. dihedral angle restraint violation (º)	6.5	8.7
Deviations from idealized geometry		
Bond lengths (Å)	0.0093	0.0102
Bond angles (º)	0.78	0.93
Impropers (º)	0.50	0.70
Mean pairwise r.m.s. deviation to mean structure	**Residues**	**Residues**
for 25 lowest energy structures of 50 calculated (Å)	**376-404**	**418-440**
Heavy	0.86 ± 0.23	0.74 ± 0.13
Backbone	0.31 ± 0.09	0.22 ± 0.14
Mean pairwise r.m.s. deviation amongst all	**Residues**	**Residues**
25 lowest energy structures of 50 calculated (Å)	**376-404**	**418-440**
Heavy	1.26 ± 0.35	1.09 ± 0.18
Backbone	0.44 ± 0.13	0.33 ± 0.20

Supplementary Table 1 NMR and refinement statistics for APLF F1 and F2 structures. Ramachandran statistics for F1: most favoured, 80.3%; additionally allowed 19.5%; generously allowed, 0.2%; disallowed 0.0%. Ramachandran statistics for F2: most favoured, 70.7%; additionally allowed 27.3%; generously allowed, 2.0%; disallowed 0.0%.

	APLF F1 + RFA	APLF F2 + RFA
NMR distance and dihedral restraints		
Distance restraints		
Total NOE	715	595
Intra-residue	151	145
Inter-residue		
Sequential ($\|i - j\| = 1$)	269	228
Medium-range ($2 \leq \|i-j\| \leq 4$)	163	138
Long-range ($\|i - j\| > 4$)	132	84
Intraligand (transferred NOEs)	24	24
Intermolecular	19	21
Total dihedral angle restraints (protein)	36	18
ϕ	12	5
ψ	13	5
$\chi 1$	11	8
Total RDC restraints (protein)	27	21
Structure statistics (protein component)		
Violations		
Distance viols. > 0.2Å (average per structure)	4.7 ± 0.9	7.7 ± 2.1
Dihedral angle viols. > 5° (average per structure)	1.3 ± 0.47	5.7 ± 1.4
R.m.s. deviation to distance restraints (Å)	0.034	0.043
R.m.s. deviation to angle restraints (°)	4.21	2.83
Max. distance restraint violation (Å)	0.59	0.71
Max. dihedral angle restraint violation (º)	5.7	9.6
Deviations from idealized geometry		
Bond lengths (Å)	0.0090	0.0102
Bond angles (º)	0.80	0.95
Impropers (º)	0.58	0.68
Mean pairwise r.m.s. deviation to mean structure for 10 lowest energy structures of 50 calculated (Å)	**Residues 376-404**	**Residues 418-440**
Heavy	0.73 ± 0.10	0.78 ± 0.26
Backbone	0.23 ± 0.07	0.19 ± 0.08
Mean pairwise r.m.s. deviation amongst all 10 lowest energy structures of 50 calculated (Å)	**Residues 376-404**	**Residues 418-440**
Heavy	1.46 ± 0.25	1.52 ± 0.40
Backbone	0.35 ± 0.09	0.29 ± 0.11

Supplementary Table 2 NMR and refinement statistics for APLF F1 and F2 co-structures with RFA. Ramachandran statistics for F1: most favoured, 76.8%; additionally allowed 23.2%; generously allowed, 0.0%; disallowed 0.0%. Ramachandran statistics for F2: most favoured, 76.1%; additionally allowed 22.8%; generously allowed, 0.6%; disallowed 0.6%.

Materials and Methods

Cloning, expression and purification of APLF fragments. DNA coding for different fragments (residues 361-511, 361-469 and 368-451) of human APLF protein containing the C-terminal zinc finger domains was amplified from an MGC cDNA clone (IMAGE:6042653, MRC geneservice) by PCR and subcloned into a pGEX-6P-1 vector (GE Healthcare) using BamH1 and XhoI restriction sites. The resulting plasmids were transformed into *E.Coli* BL21-CodonPlus(DE3)-RP cells (Stratagene). Cells were cultured either in LB medium or in M9 minimal medium, supplemented either with $^{15}NH_4Cl$ or with $^{15}NH_4Cl$ and [$^{13}C_6$]-glucose (Sigma-Aldrich Isotec) as a sole nitrogen- or carbon-source, respectively. The GST-fusion protein was expressed overnight at 20°C after induction with 0.5 mM IPTG. Upon induction, the medium was supplemented with 0.5 mM $ZnSO_4$. Protein purification was carried out at 4ºC. Harvested cells were resuspended in glutathione binding buffer (50mM Tris pH 7.4, 1M NaCl, 150uM $ZnSO_4$ and 1mM 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 to cleavage buffer (50mM Tris pH 7.4, 150 mM NaCl, 100uM $ZnSO_4$ and 1mM DTT). The GST fusion protein was cleaved on the Glutathione-Sepharose beads by GST-tagged Precission Protease (GE Healthcare), leaving vector-derived residues Gly-Pro-Gly-Ser N-terminal (applies for all APLF constructs described in this work). Protein in supernatant was cleared off from gluthione sepharose 4B, equilibrated to gelfiltration buffer (20mM Pyrophosphate (APLF 368-451) or 50 mM Tris pH 7.0 (all other APLF constructs), 200mM NaCl, 100uM $ZnSO_4$ and 1mM DTT) and concentrated using Vivaspin 20 MWCO 3000 (Sartorius-Stedim Biotech). Protein was purified to homogeneity by gelfiltration using Superdex S-75 equilibrated in the respective gelfiltration buffer. GST-fused APLF for the in vitro PAR binding assays was similarly purified, except that the cleavage step was omitted, the fusion protein being eluted directly from the Glutathione-Sepharose beads using 20mM glutathione.

Mass Spectrometry. Samples were prepared from stock solutions of 1.6 mM APLF (368-451) in 20mM pyrophosphate pH 6.0, 200mM NaCl, 2mM DTT and 100uM $ZnSO_4$ by dilution with 200 mM ammonium acetate (pH 6.0) to give final solution concentrations of 10-25 µM APLF (368-451). Mass spectra were acquired using LCT instrument (Micromass) configured for nanoflow electrospray ionisation (ESI) in positive ion mode. All spectra were acquired with capillary voltage of 2500 V, a cone voltage of 45 V and a desolvation temperature of 80ºC. Mass spectra were acquired also after addition of a solution of EDTA (pH 6.0) in 80 fold molar excess to APLF (368-451). Data was analyzed using Mass Lynx NT software v4.0 (Micromass).

NMR Experiments. All data was acquired on Bruker Avance 800, DMX600 and DRX500 spectrometers, each equipped with a triple resonance ($^1H/^{15}N/^{13}C$) cryoprobe. ^1H, ^{15}N and ^{13}C chemical shifts were calibrated using sodium 3,3,3-trimethylsilylpropionate (TSP) as an external ^1H reference.[1] Unless otherwise stated, all NMR experiments for the free protein were performed at 27°C using ^{15}N- or ^{15}N, ^{13}C-labelled protein samples adjusted to 20mM pyrophosphate pH 6.0 (APLF 368-451) or 50 mM Tris pH 7.0 (all other APLF constructs), 200mM NaCl, 100uM $ZnSO_4$, 2mM [$^2H_6$] DTT and either 5% or 100% [$^2H_2$] H_2O (v/v). Resonance assignments were made using a standard suite of triple resonance NMR experiments. For experiments used to derive structural constraints the samples

comprised 0.5-0.6 mM ^{15}N, ^{13}C-labelled solutions of APLF 368-451. The following spectra were acquired: 2D: [^{15}N-^{1}H] HSQC, long-range-optimised [^{15}N-^{1}H] HMQC to correlate histidine ring ^{1}H and ^{15}N signals,[2] [^{13}C-^{1}H] HSQC covering the full ^{13}C spectral width, constant-time [^{13}C-^{1}H] HSQC covering only the aliphatic ^{13}C region, constant-time [^{13}C-^{1}H] HSQC covering only the aromatic ^{13}C region, [^{1}H-^{1}H] NOESY experiments (without heteronuclear filtering; τ_m = 150 ms), [^{1}H-^{1}H] NOESY experiments filtered to remove ^{15}N-coupled signals in F_2 only (τ_m = 150 ms); 3D data sets: CBCANH, CBCACONH, HBHANH, HBHACONH, [^{1}H-^{13}C-^{1}H] HCCH-TOCSY, [^{13}C-^{13}C-^{1}H] HCCH-TOCSY, HNHB, ^{15}N NOESY-HSQC (τ_m = 150 ms and τ_m = 50 ms), ^{13}C NOESY-HSQC (τ_m = 150 ms), separate datasets acquired for ^{13}C aliphatic and aromatic spectral regions. All of the NOESY datasets used for structure calculations (see below) were acquired using pulse sequences modified to ensure equal RF heating in each case, e.g. for ^{13}C experiments, a period of 15N decoupling equal in length to the acquisition period was applied at the beginning of the interscan delay, and for ^{15}N experiments an equivalent period of ^{13}C decoupling was similarly applied. Residual dipolar couplings were measured using a 0.3 mM ^{15}N, ^{13}C-labelled solution of APLF 368-451 adjusted to 20 mM pyrophosphate pH 6.0, 300mM NaCl, 100uM $ZnSO_4$, 2mM [$^{2}H_6$] DTT and 5% D2O (v/v), to which Pf1 phage (ASLA Biotech) was added to a final concentration of 12mg/ml; splittings were measured in F_1 cross-sections of [^{15}N-^{1}H] HSQC IPAP spectra.[3] Titrations were carried out by adding successive aliquots of ligands (ADP Ribose (Sigma), UMP (Sigma), GMP (Sigma), AMP (Sigma) or 2'-*O*-D-ribofuranosyladenosine (RFA) to a solution of 0.15 mM ^{15}N, ^{13}C-labelled APLF 368-451, up to a maximum ligand:protein molar ratio of 80:1. Chemical shift perturbation upon ligand addition was monitored by acquisition of HSQC spectra.

For experiments used to derive structural constraints of APLF in complex with RFA, samples contained ^{15}N, ^{13}C-labelled APLF 368-451 (0.8mM) and RFA (2mM), adjusted to 20mM pyrophosphate pH 6.0, 200mM NaCl, 100uM $ZnSO_4$, 2mM [$^{2}H_6$] DTT and 99% % [$^{2}H_2$] H_2O (v/v). This protein:ligand ratio was selected to give the best compromise between saturation of protein binding sites (approximately 30-40% under these conditions at 278K, as judged from the extent of chemical shift perturbation) and artefacts arising from intense ligand signals. Intermolecular NOE and transferred NOE cross peaks were measured in 2D NOESY spectra (τ_m = 50, 100,150, 300 and 400 ms) recorded at 278K and 286K and with half filters set to reject ^{13}C coupled protons during t_1 and to accept ^{13}C coupled protons in t_2, and with heteronuclear decoupling applied during t_2 but not during t_1. Spectral overlap within the half-filtered 2D NOESY spectrum was partially overcome by analysing an additional 3D ^{13}C NOESY-HSQC spectrum (τ_m = 250 ms) obtained at 286K.

Ligand-bound protein signals were assigned by following changes in a set of [^{13}C-^{1}H] HSQC spectra on successive addition of ligand up to the final protein:ligand ratio, followed by a stepwise decrease in temperature (300K to 278K). In cases of ambiguity, a ^{13}C NOESY-HSQC spectrum (τ_m = 250 ms) obtained at 286K was used to verify resonance assignments. Signals of the free ligands were assigned using 2D [^{1}H, ^{1}H] TOCSY, NOESY and [^{13}C-^{1}H] HSQC spectra of 20 mM solutions adjusted to 20mM pyrophosphate pH 6.0, 200mM NaCl, 100uM $ZnSO_4$, 2mM [$^{2}H_6$] DTT and 99% D2O (v/v). Intermolecular NOE and transferred NOE cross peaks were measured in 2D NOESY spectra (τ_m = 50, 100,150, 300 and 400 ms) recorded at 278K and 286K and with half filter set to reject ^{13}C coupled protons during t_2 and with heteronuclear decoupling applied

during t_1 but not during t_2. Essentially the same procedure was followed for the complex of APLF 368-451 and ADPR.

Spectra were processed using the program TOPSPIN (Bruker GmbH, Karlsruhe) and analysed using either the program SPARKY [4] or CCPN analysis (http://www.ccpn.ac.uk/ccpn).

Free Protein Structure Calculations. Initial structures for the free protein fragments were calculated using the semi-automatic program ATNOS-CANDID,[5,6] for which the input comprises the protein sequence (residues 363-451), the full resonance assignment and the following 3D NOESY datasets: ^{15}N NOESY-HSQC (τ_m = 150 ms), ^{13}C aliphatic region NOESY-HSQC (τ_m = 150 ms) and ^{13}C aromatic region NOESY-HSQC (τ_m = 150 ms). Dihedral restraints for the backbone were obtained using the program TALOS,[7] but restraints were only applied in the region of the helix (confirmed by TALOS). Within ATNOS-CANDID, the internal generation of backbone dihedral restraints was suppressed and the TALOS restraints were specified as external input. Dihedral restraints for $\chi 1$ angles were derived from measurements of $^3J_{\alpha\beta}$ and $^3J_{N\beta}$ using the HACAHB-COSY and HNHB datasets respectively. During the ATNOS/CANDID calculations no metal was represented explicitly, but the effect of metal binding was approximated by including inter-ligand distance constraints as follows: S^γ to S^γ, 3.7-4.0Å; S^γ to histidyl-N, 3.4-3.8Å; histidyl-N to histidyl-N, 3.1-3.5Å; at this stage all histidyl N atoms were assigned ambiguously as either N^δ or N^ε atoms. Some early rounds of ATNOS-CANDID calculations showed ambiguities in the $\chi 2$ rotamer of one of the metal-binding histidines (H392). Careful manual analysis of 2D NOESY spectra (filtered to remove ^{15}N coupled signals in the indirect dimension), making use of similarities in the cross peak patterns seen for this histidine and its counterpart in finger 2 (H434), allowed assignment of a small number of key NOE cross-peaks that resolved this ambiguity in H392 $\chi 2$. These cross peak assignments were used in later rounds of ATNOS-CANDID calculations, and comprised the following: H392 $H^{\varepsilon 1}$ – F396 $H^{\delta 1/2}$ < 6.0Å; H392 $H^{\varepsilon 1}$ – F396 $H^{\beta 2}$ < 6.0Å; H434 $H^{\varepsilon 1}$ – Y423 $H^{\delta 1/2}$ < 6.0Å; H434 $H^{\varepsilon 1}$ – C421 $H^{\beta 2}$ < 6.0Å; H434 $H^{\varepsilon 1}$ – P422 $H^{\delta 2}$ < 6.0Å. Separate calculations were carried out for F1 and for F2, which differed only in that the rmsd range (used by ATNOS-CANDID in the first round to help select structures) was set to 379-404 in the F1 calculations and to 420-440 in the F2 calculations.

In order to be able to employ explicit zinc bonding and geometry terms in the force-field for the calculations (including bond-angle and, for the histidines, in-plane constraints), as well as constraints based on RDC measurements, we next calculated structures using XPLOR-NIH.[8] As input, these calculations used the set of NOE restraints generated by the final (seventh) cycle of ATNOS/CANDID, re-formatted for use in XPLOR-NIH. Since the XPLOR-NIH calculations employed r^{-6} summation for all groups of equivalent protons and non-stereospecifically assigned prochiral groups, and since no stereoassignments were made (and the assignment-swapping protocol within XPLOR-NIH for deriving stereoassignments indirectly during the structure calculation itself was not applied), the constraints for all such groups were converted to group constraints (i.e. such groups were specified using wildcards such as HB*). All lower bounds were set to zero. The pattern of zinc connectivities to the histidine residues was established using long-range ^{15}N-HMQC experiments as described by Legge *et al*. [2] which showed unambiguously that the $N^{\varepsilon 2}$ atom binds the zinc in all four cases (see supplementary material). Structures were calculated from polypeptide chains with randomized ϕ and ψ torsion angles using a two-stage simulated annealing protocol within the program XPLOR-NIH, essentially as described elsewhere[9] but employing larger numbers of cycles as follows: First stage calculations comprised Powell energy minimization (500 steps), dynamics at 1000K

(50000 steps), increase of the van der Waals force constant and tilting of the NOE potential function asymptote (10000 steps), switching to a square-well NOE function then cooling to 300K in 2000 step cycles, and final Powell minimization (1000 steps). Second stage calculations used Powell minimization (500 steps), increasing dihedral force constant during 4000 step cycles of dynamics at 1000K (with a strong van der Waals force constant and square-well NOE potential function), cooling to 300K in 6000 step cycles, and 2000 steps of final Powell minimisation. A final stage of refinement against measured values of amide group ^{15}N-^{1}H residual dipolar couplings was employed using the ISAC protocol of Sass *et al*.[10] For the whole protocol, completely separate calculations were run for F1 and for F2, using residues 363-417 for F1 and 405-451 for F2, and dividing the constraints into separate files as appropriate.

The program CLUSTERPOSE was used to calculate the mean rmsd of ensembles to their mean structure.[11,12] Structures were visualized using the program pymol (http://www.pymol.org).

Complex Structure Calculations. Co-structures of the complexes of APLF F1 and F2 with ADPR and RFA were calculated using XPLOR-NIH. Intra-ligand (i.e. transferred) NOE cross-peak volumes were measured using the program SPARKY in 2D filtered NOESY spectra, mainly those recorded at 278K but in few cases where signals at 278K are overlapped with the water signal, also the spectra recorded at 286K. For each temperature, distance calibration was achieved by fitting the buildup curves to an exponential function $f(\tau) = a(1-e^{-b\tau})$, where a and b are adjustable and τ is the NOESY mixing time; initial slopes of the curves were used to calculate distances, using the H1' H4' distance (3.4 Å) as reference. Restraints were set as a range $0 < r < D*1.2$, where D is the distance calculated from the buildup curves. Intermolecular NOEs were also measured in the 2D filtered NOESY spectra, taking intensities from the 300ms mixing time spectra; these were classified as either medium ($0 < r < 3.0$ Å) or weak ($0 < r < 4.5$Å). Constraints for the free protein were used unchanged for the complex calculations also; inspection of the NOESY spectra for the complex showed no changes for any of the intraprotein NOE cross peaks that were compared.

Calculations for the complexes used essentially the same protocol as the XPLOR-NIH part of the calculations for the free protein, but with the ligands added to the starting structures and the intra-ligand and intermolecular constraints added to the input. The numbers of cycles used were similar to those for the free protein, except that in the first stage calculations the 50000 steps of dynamics at 1000K were divided into 25000 steps with the ligand translated and fixed 15Å in all three dimensions from an atom near the protein centre (NE2 of His392 for F1, NE2 of His434 for F2), followed by 25000 steps with the ligand released. Starting structures were calculated by randomizing all flexible torsion angles in both the protein and the ligand, then subjecting the structures to energy minimization followed by dynamics as in the main protocol except that all experimental constraints (except the intra-ligand NOE constraints) were excluded, and finally introducing 15Å translations between the x, y and z co-ordinates of the centers of masses of the protein and the ligand. Parameter and topology files for the ligands were obtained by initially creating pdb files manually within pymol, using the program XPLO2D (http://xray.bmc.uu.se/usf/xplo2d_man.html) to generate initial parameter and topology files, and then hand-editing these to substitute standard regularized geometries. Ribose parameters were taken from the file dna-rna-allatom.param in the XPLOR-NIH distribution, using the values for a 3'-endo conformation; the transferred NOEs from adenine H8 to RAH2' and from adenine H8 to RAH3' are approximately equal, consistent with a 3' endo conformation, whereas a 2'endo conformation would be expected to yield a significantly larger transferred NOE H8 to RAH2' (J-coupling data that would normally be used to characterize sugar pucker cannot easily be applied here as the ligand is >60% in the

unbound conformation and, unlike transferred NOEs, J-couplings average linearly between free and bound states). For ribose B our data do not allow us to deduce the ring pucker in the complex.

Modeling calculations for a complex of APLF F1 with pop-RFA-pop-RFA-pop. Initial structures were calculated in XPLOR-NIH, following an identical protocol to that described for the APLF RFA complexes. Identical experimental data were used for the input as before, with the sole exception that the additional intermolecular NOE assigned to the second adenosine ring binding site (H2 of the adenosine to Hζ of F396) was now included. In addition, a small number on non-experimental constraints were added to improve stacking geometry in the second site and to bring the third pyrophosphate into proximity with R376; these comprised an ambiguous distance constraint from C4 or C5 of the adenine ring to either Cδ atom of F396 (set to 0Å < *r* < 2.5Å) and a distance constraint from Arg 376 Cζ to the central oxygen of the third pyrophosphate (set to 0Å < *r* < 6Å). An ensemble of 200 such structures was calculated, and of these the 100 having the lowest total XPLOR-NIH energy term were selected for further refinement in AMBER 9.

Preparatory to running in AMBER 9,[13] mol2 files were generated for the pop-RFA-pop-RFA-pop ligand (where "pop" stands for pyrophosphate) using the "antechamber" program, specifying the AM1-BCC charge model and a net ligand charge of -8. Charges calculated for the lowest energy ligand conformation were used for all starting structures and gaff or ff99SB forcefields were used for the ligand and protein, respectively. For the protein, charges were neutralized by addition of sodium ions and the zinc-binding arrangement was explicitly specified. Calculations in "sander" comprised an initial minimization (250 steps steepest descent then 250 steps conjugate gradient), then 20ps of molecular dynamics using a simulated annealing protocol (5000 fs-steps heating from 0K to 300K; 13000 fs-steps cooling to 4K; 1000 fs-steps cooling to 1K; 1000 fs-steps of final cooling to 0.1K) and a final minimization (250 steps steepest descent then 250 steps conjugate gradient). Implicit solvent representation using the generalized Born method was employed throughout (igb=1), and Langvin temperature control was used (ntt=3; gamma_ln=5). The experimental intermolecular NOE constraints (including the one constraint to F396 in the second binding site) were applied throughout, but the artificial constraints to enforce stacking and the R376 - pyrophosphate interaction were omitted. Resulting structures were then submitted to the program DOCK 6.3.[14] Structures were ranked using calculated DOCK 6.3 "amber score" interaction energies after an additional 200 steps of amber 9 minimization.

Synthesis of RFA

RFA was synthesized as described in reference 7 of main paper.

Immunoprecipitation

Human embryonic kidney 293T cells were transiently transfected using Polyfect transfection reagent (Qiagen), according to the manufacturer's specifications, with FLAG-tagged wild-type and mutant APLF constructs. Following transfection (48 h), cells were resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% Triton X-100, 1 mM dithiothreitol) supplemented with 50 U/μL benzonase nuclease (Sigma) and protease inhibitor cocktail tablet (Roche). Whole cell extracts were clarified by centrifugation and incubated with anti-FLAG M2 agarose (Sigma) for 20 minutes at 4°C. Following extensive washing with lysis buffer, the immunoprecipitates were boiled in SDS-PAGE loading buffer and analysed by immunoblotting. In the PARP inhibition experiments, 293T cells were pre-treated with 0.5 μM KU-0058948 or PBS for 12 h.

PAR-binding assay

Proteins (0.5 pmol) were spotted onto a nitrocellulose membrane, and analyzed as described previously.[15]

References:

1. Wishart, D.S., Bigam, C. G., Yao, J., Abildaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., Sykes, B. D. ^{1}H, ^{15}N and ^{13}C chemical shift referencing in biomolecular NMR. *J. Biomol. NMR* **6**, 135-140 (1995).
2. Legge, G.B. et al. ZZ domain of CBP: an unusual zinc finger fold in a protein interaction module. *J. Mol. Biol.* **343**, 1081-93 (2004).
3. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373-378 (1998).
4. Goddard, T.D., Kneller, D. G. SPARKY 3. (University of California, San Francisco).
5. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. *J. Biomol. NMR* **24**, 171-89 (2002).
6. Herrmann, T., Guntert, P. & Wuthrich, K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol* **319**, 209-27 (2002).
7. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* **13**, 289-302 (1999).
8. Schwieters, C.D., Kuszewski, J.J., Tjandra, N. & Clore, G.M. The Xplor-NIH NMR molecular structure determination package. *J. Magn. Reson.* **160**, 65-73 (2003).
9. Muto, Y. et al. The structure and biochemical properties of the human spliceosomal protein U1C. *J Mol Biol* **341**, 185-98 (2004).
10. Sass, H.J., Musco, G., Stahl, S.J., Wingfield, P.T. & Grzesiek, S. An easy way to include weak alignment constraints into NMR structure calculations. *J. Biomol. NMR* **21**, 275-80 (2001).
11. Diamond, R. On the multiple simultaneous superposition of molecular structures by rigid body transformation. *Protein Sci.* **1**, 1279-1287 (1992).
12. Diamond, R. Coordinate-based cluster analysis. *Acta Cryst. D* **51**, 127-135 (1995).
13. Case, D.A. et al. AMBER 9. (University of California, San Francisco, 2006).
14. Lang, P.T. et al. DOCK 6: Combining Techniques to Model RNA-Small Molecule Complexes. *RNA* **15**, 1219-1230 (2009).
15. Ahel, I. et al. Poly(ADP-ribose)-binding zinc finger motifs in DNA repair/checkpoint proteins. *Nature* **451**, 81-5 (2008).