Supplementary Materials for

The large family of PC4-like domains – similar folds and functions throughout all kingdoms of life

Robert Janowski[1], and Dierk Niessing[1,2,] *

[1] Institute of Structural Biology, Helmholtz Zentrum München – German Research Center for Environmental Health, Neuherberg, 85764 Germany

[2] Institute of Pharmaceutical Biotechnology, Ulm University, Ulm, 89081 Germany

* To whom correspondence should be addressed. Tel: +49-731-50-23161; Fax: +49-731-50-23169; Email: niessing@helmholtz-muenchen.de

A number of functionally uncharacterized protein domains adopt PC4-like fold, albeit it is unclear whether they also bind nucleic acids.

Type I

SsgA-like proteins (SALPs) are a family of homologous cell division-related proteins that occur exclusively in morphologically complex actinomycetes. The crystal structure of its family member SsgB from *Thermobifida fusca* (residues 1–137*,* SsgA-like sporulation-specific cell division protein), represents the first structural example for this family (PDB ID: 3cm1) (1). It revealed a PC4-like fold (Figure S1A) and displays similarities to the Whirly family. The structural similarity to our initial Pur-α search model from *D. melanogaster* (fragment 41-185; chain A of PDB ID: 3k44) (2) is remarkable with a Z-score of 6.9. However in contrast to Pur-α, the electro-negative surface of the SALPs suggests that neither SsgB nor any of the other SALPs are likely to interact with nucleic acids (1). In the crystal lattice, SsgB forms a bell-shaped trimer, which is assembled through α3 interaction with α1 and α2 of a neighboring protomer. Interestingly, multimerization of PBF-2 and MRP1/2 also involves mostly helices, suggesting that the same strategy is followed for oligomer formation for other members of the PC4-like family (1).

Hydroquinone dioxygenase from *Pseudomonas* is the key enzyme in the hydroquinone pathway of *para*-nitrophenol degradation. It catalyzes the ring cleavage of hydroquinone to γ-hydroxymuconic semialdehyde (3). This enzyme is another example of a type I PC4-like fold (PDB ID: 4zxa). It is composed of small and big subunit where the latter one, spanning amino acids 16-140, exhibits a β-β-β-β-β-α-linker-β-β-β-β-α topology (Figure S1B). Of note, this protein has an additional β-strand at the N-terminal part of the first repeat, thereby expanding the β-sheet. It has not been reported that hydroquinone dioxygenase interacts with nucleic acids. Although the potential nucleic acid interacting surfaces of this domain are solvent exposed, assessment of their electrostatic potentials indicates that an interaction with RNA/DNA is rather unlikely.

CdiI immunity protein from *Enterobacter cloacae* (ECL) is a representative of type I PC4-like fold with other function than nucleic acid binding. CdiI functions as a neutralizing protein (antitoxin) for CdiA-CT toxins and protects toxin-producing cells from auto inhibition (PDB ID: 4ntq) (4). Apart from lacking its first β-strand (topology β-β-β-α-linker-β-β-β-β-α) the scaffold highly resembles the PC4-like fold with the Z-score of 4.6 (Table S1). Also the CdiI homologue from *Burkholderia dolosa* (PDB ID: 5ffp; unpublished) functions as an antitoxin and is more similar to PC4-like fold as it poses all four β-strands in the first repeat (Figure S1C). Both proteins contain an extra helix, long insertions between α1 and β5 for CdiI from *B. dolosa* (PDB ID: 5ffp) and between β6 and β7 for *E. cloacae* (PDB ID: 4ntq). Analysis of the surface electrostatic potential rather precludes nucleic acid binding properties of CdiI proteins.

Proteins of unknown function: Although their functions are not known so far, the proteins described below are worth mentioning due to their high similarity to the PC4-like fold. The crystal structure of a protein with unknown function from the DUF155 family (YP_292156.1) from *Prochlorococcus marinus* sp. NATL2A has been solved to 1.80 Å resolution (Joint Center for Structural Genomics; PDB ID: 2it9). It has a β-β-β-β-α-linker-β-β-β-β-α topology with Z-score of 9.5 to our search model (Figure S1D). Even higher similarity (Z-score 10.8) is found for the protein YP_400729.1 with unknown function from *Synechococcus* (Figure S1E). It has been solved at 2.50 Å resolution (Joint Center for Structural Genomics; PDB ID: 2nvn). Both structures show high symmetry of the fold of both repeats.

Non-PC4-like proteins with similarity to PC4-like family

We noted that some proteins do not fulfill the definition of a PC4-like domain but nevertheless show considerable similarities with this class of domains. The example given here is the core of the Alu domain of the human signal recognition particle SPR (5). Heterodimer of proteins SRP9 and SRP14 bound to the 5' and 3' terminal sequences of SRP RNA (PDB ID: 1e8o) resemble the PC4 type II fold, however without helix swapping. Also unlike all the nucleic-acids interactions of PC4-like proteins, the SPR RNA is bound not only by its extended β-sheet but also on the other side by its helices (Figure S4E).

TABLES AND FIGURES

Figure S1. Comparison of different protein structures exhibiting type I PC4-like fold. Each structure is depicted in two different orientations. Electrostatic potential surface is shown for every structure in the same orientation as in the middle panel with the β-sheet surface directed to the front. The coordinates are taken from: (**A**) *Thermobifida fusca* SsgB (PDB ID: 3cm1) (1), (**B**) *Pseudomonas* Hydroquinone dioxygenase (PDB ID: 4zxa) (3), (**C**) CdiI homologue from *Burkholderia dolosa* (PDB ID: 5ffp; unpublished), (**D**) *Prochlorococcus marinus* DUF155 (PDB ID: 2it9; unpublished), (**E**) *Synechococcus elongates* YP_400729.1 (PDB ID: 2nvn; unpublished). The structures of type I PC4-like fold have been represented as cyan (repeat I) and navy blue (repeat II). The figure has been prepared in *PyMol* v 1.3 (pymol.org).

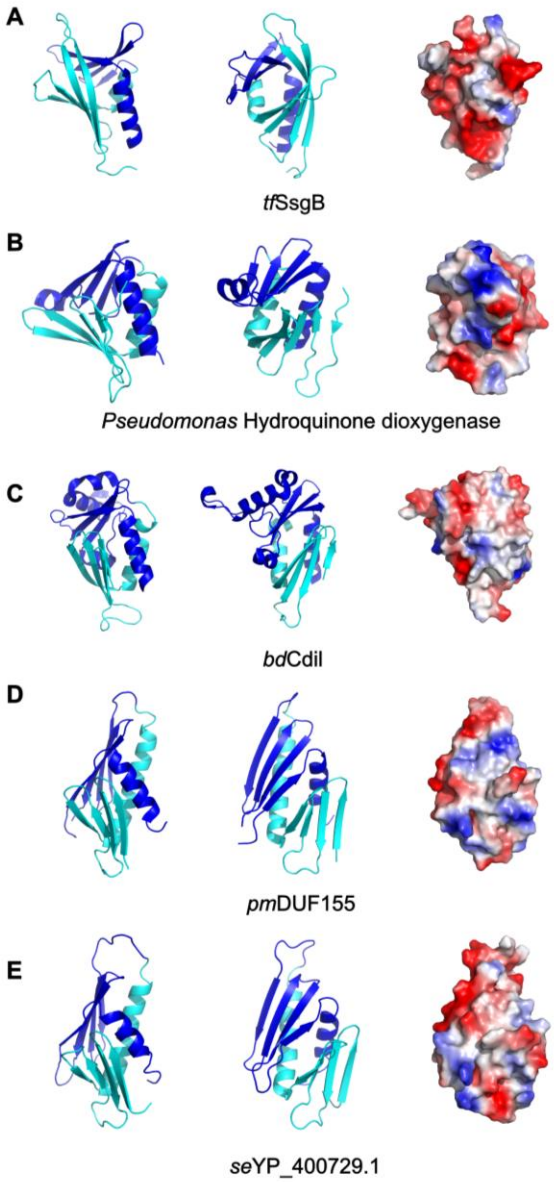

Figure S2. Phylogenetic tree for the selected members of PC4-like family.
DIST = percentage divergence (/100).

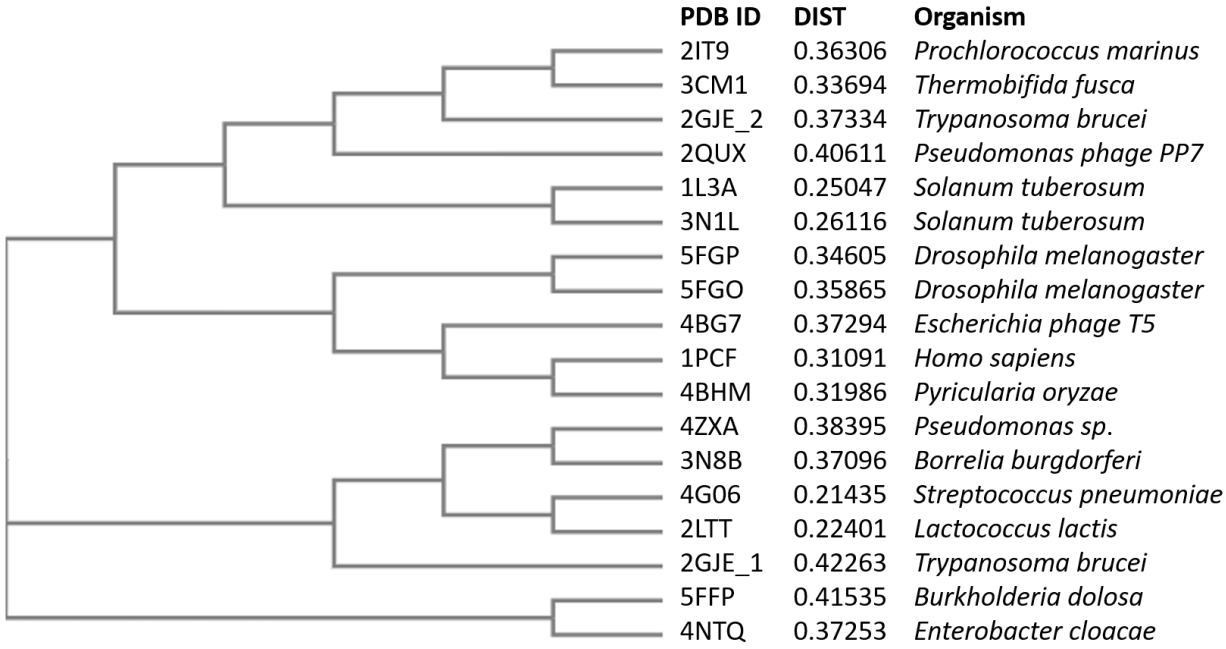

Figure S3. Comparison of the electrostatic potential surface of selected protein structures exhibiting type I and II PC4-like fold. Every structure is presented in the same orientation with the β-sheet surface directed to the front. The coordinates are taken from: (**A**) *Solanum tuberosum* P24 subunit of PBF-2 (PDB ID: 1l3a) (6), (**B**) *Trypanosoma brucei* MRP1 (PDB ID: 2gje) (7), (**C**) The human replication and transcription cofactor PC4 (PDB ID: 1pcf) (8), (**D**) Bacteriophage T5 homologue of PC4 (PDB ID: 4bg7) (9), (**E**) *Lactococcus lactis* DUF2128 family member YdbC (PDB ID: 2ltt) (10), (**F**) Pur-α from *Borrelia burgdorferi* (PDB ID: 3n8b) (11), (**G**) *Streptococcus pneumonia* SP_0782 protein (PDB ID: 5zkl) (12), (**H**) The bacteriophage coat protein PP7 (PDB ID: 2qux) (13). The figure has been prepared in *PyMol* v 1.3 (pymol.org).

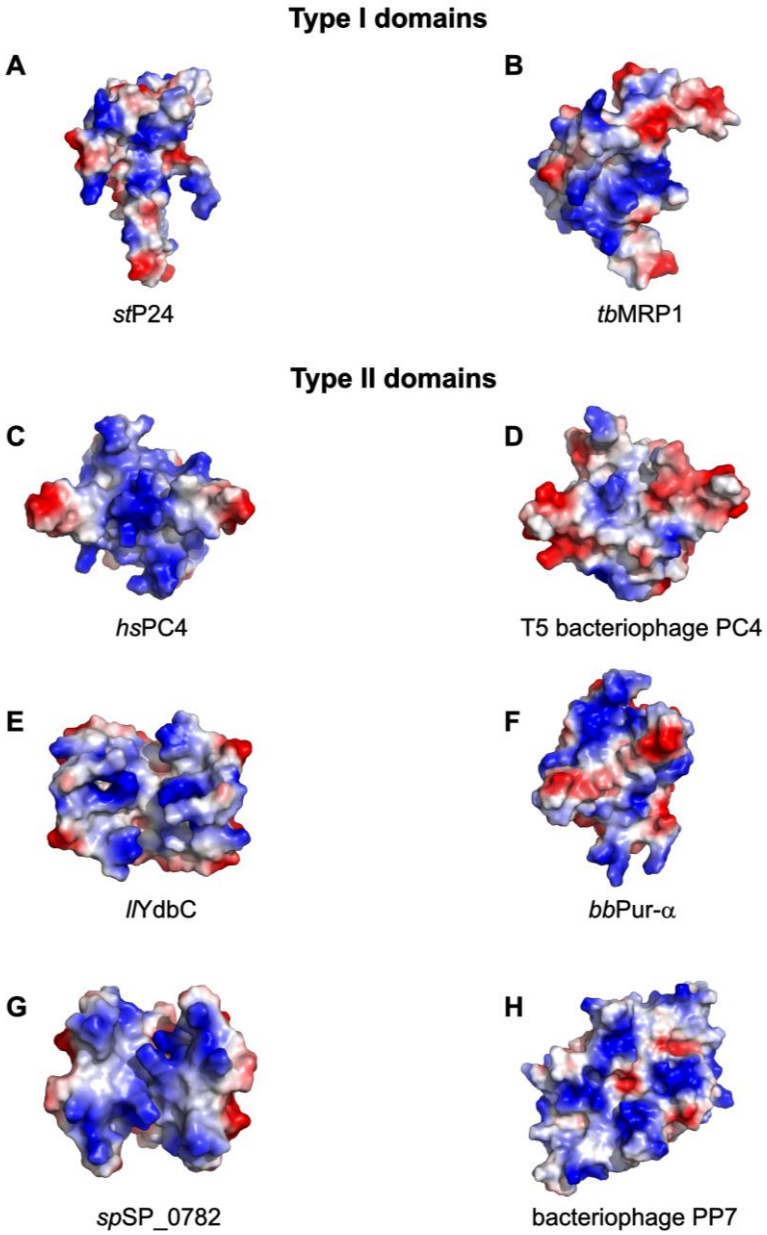

Figure S4. (**A**) Tetrameric arrangement of P24 of PBF-2 from *Solanum tuberosum* (PDB ID: 1l3a) (6), (**B**) Schematic presentation of unwindase activity of Pur-α. (**C**) Oligomerization of *Pyricularia oryzae* MoSub1 in complex with ssDNA (PDB ID: 4bhm) (14), (**D**) Oligomerization of *Magnaporthe oryzae* MoSub1 in complex with ssDNA (PDB ID: 5zg9) (15), (**E**) The core of the Alu domain of the human signal recognition particle SPR. The binding mode of the heterodimer of SRP9 and SRP14 proteins and the SRP RNA (PDB ID: 1e8o) (5). The figure has been prepared in *PyMol* v 1.3 (pymol.org).

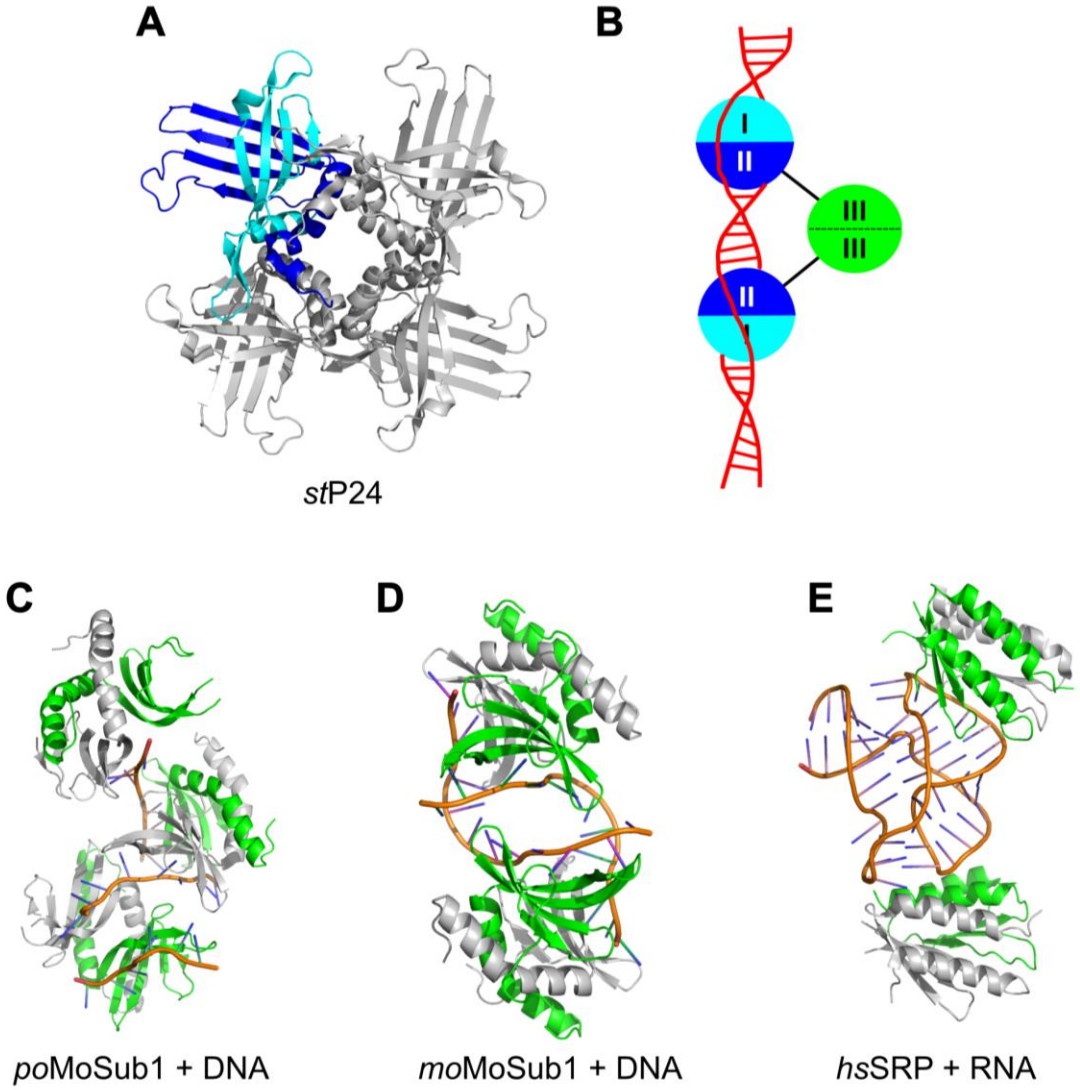

Figure S5. Comparison of the electrostatic potential surface of *Drosophila melanogaster* Pur-α protein exhibiting type I (**A**) (PUR repeat I+II; PDB ID: 5fgp) (16) and type II (**B**) (PUR repeat III; PDB ID: 5fgo) (16) PC4-like fold. Both structures are presented in the same orientation with the β-sheet surface directed to the front. The figure has been prepared in *PyMol* v 1.3 (pymol.org).

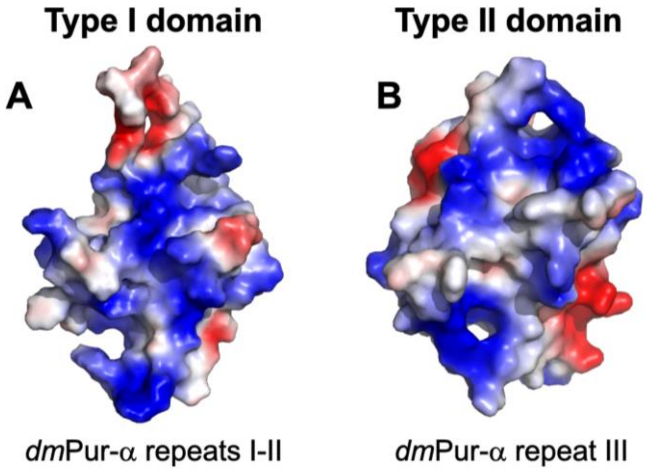

Figure S6. Electrostatic potential surface of the selected protein structures with DNA/RNA bound (shown in green) exhibiting type I PC4-like domain fold; as well as the close-up of nucleic acid (shown as balls and sticks) interaction with a protein (shown as ribbon; amino acid residues important for binding are shown as ball and stick and labelled). The coordinates taken from: (**A** and **B**) *Solanum tuberosum* WHY2 in complex with ssDNA (PDB ID: 3n1l) (17), (**C** and **D**) *Trypanosoma brucei* MRP1/MRP2 complex with RNA (PDB ID: 2gje) (7), (**E** and **F**) *Drosophila melanogaster* Pur-α repeat I-II in complex with ssDNA (PDB ID: 5fgp) (16). The figure has been prepared in *PyMol* v 1.3 (pymol.org).

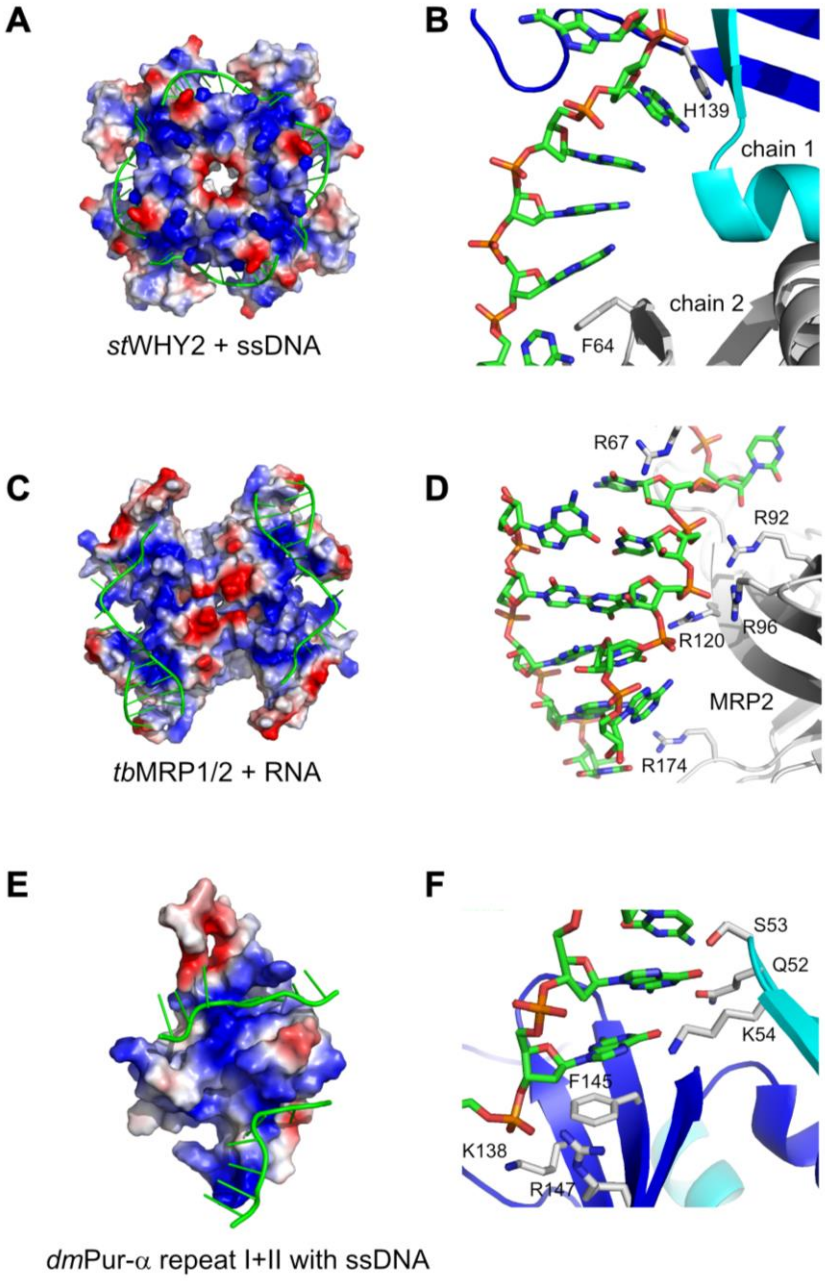

Figure S7. Electrostatic potential surface of selected protein structures with DNA/RNA bound (shown in green) exhibiting type II PC4-like domain fold; as well as the close-up of DNA (shown as balls and sticks) interaction with a protein (shown as ribbon; amino acid residues important for binding are shown as ball and stick and labelled). The coordinates are taken from: (**A** and **B**) Human PC4 CTD in complex with ssDNA (U mode; PDB ID: 2c62) (18), (**C** and **D**) *Pyricularia oryzae* MoSub1 in complex with ssDNA (Straight mode; PDB ID: 4bhm) (14), (**E** and **F**) *Magnaporthe oryzae* MoSub1 in complex with ssDNA (L mode; PDB ID: 5zg9) (15). The figure has been prepared in *PyMol* v 1.3 (pymol.org).

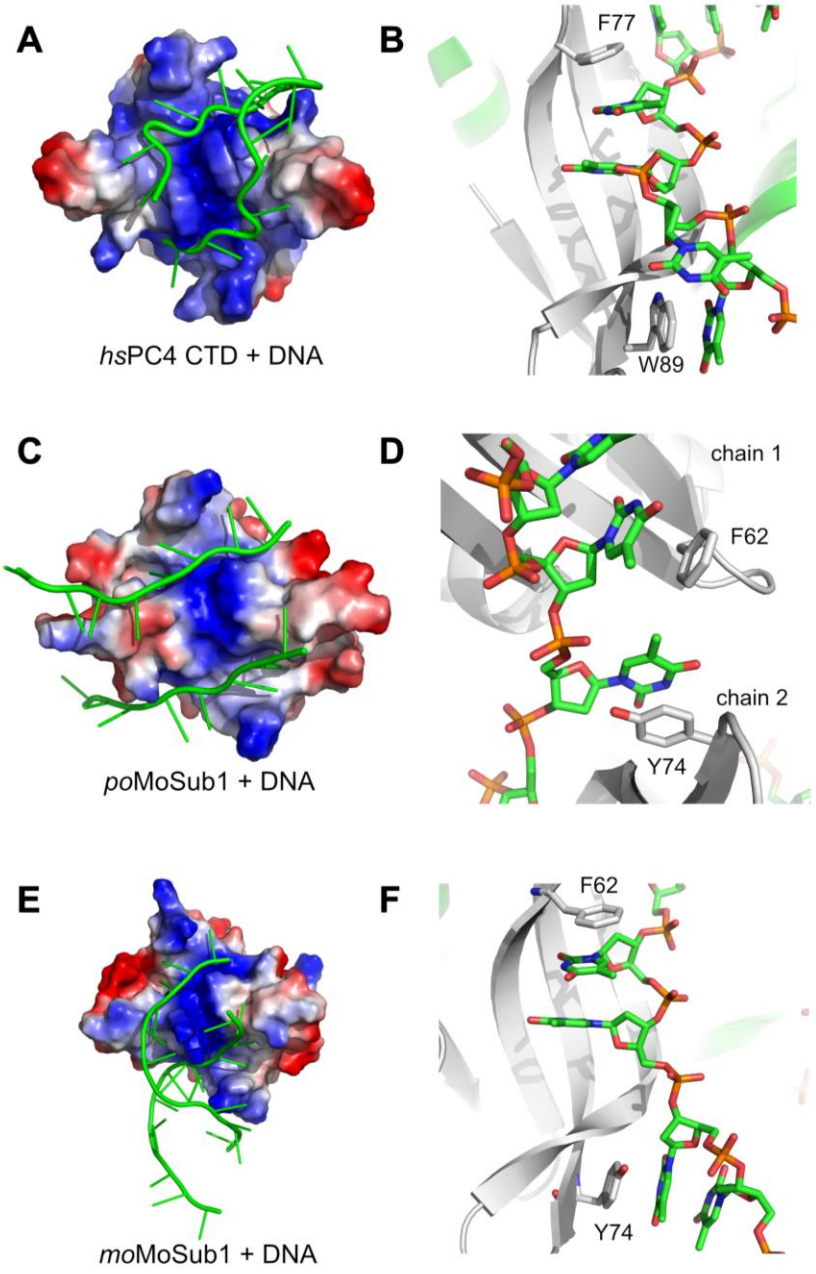

Figure S8. Electrostatic potential surface of selected protein structures with DNA/RNA bound (shown in green) exhibiting type II PC4-like domain fold; as well as the close-up of nucleic acid (shown as balls and sticks) interaction with a protein (shown as ribbon; amino acid residues important for binding are shown as ball and stick and labelled). The coordinates are taken from: (**A** and **B**) *Lactococcus lactis* Ydbc in complex with ssDNA (PDB ID: 2ltt) (10), (**C** and **D**) *Streptococcus pneumonia* SP_0782 protein in complex with ssDNA dT12 (PDB ID: 5zkl) (12), (**E** and **F**) *Pseudomonas* phage protein PP7 in complex with RNA (PDB ID: 2qux) (13). The figure has been prepared in *PyMol* v 1.3 (pymol.org).

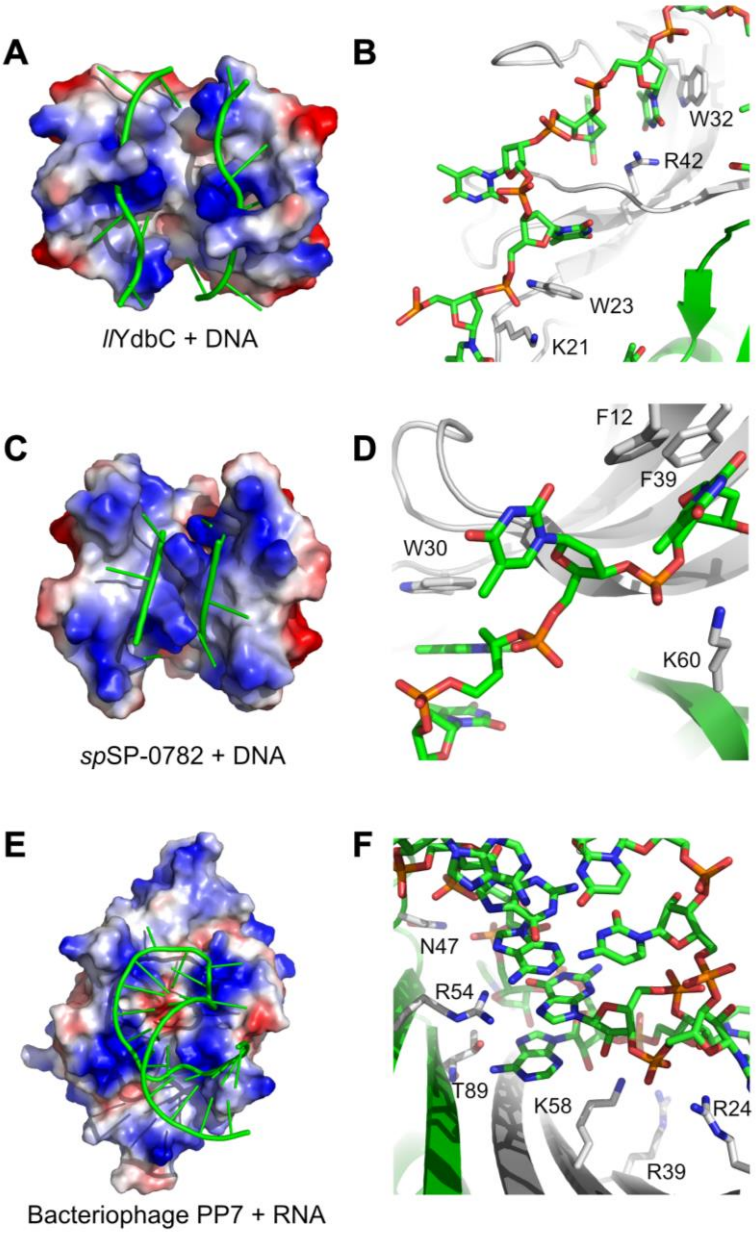

Table S1. The structures described in this paper ordered according to their Z-score compared to the search model the structure of *dm*Pur-α (fragment 41-185, chain A, PDB ID: 3k44) (2).

PDB-ID	Name	RNA/DNA binding	Type	Z-score	r.m.s.d.	n-align	Seq ID
5fgp	Pur-α repeat I and II from fly	ssDNA/RNA	I	24	1.6	141	97
2nvn	YP_400729.1	No info	I	10.8	3.9	115	15
2it9	DUF155 unknown function	No info	I	9.5	4.0	116	12
2gje	MRP1	RNA	I	8.2	4.0	109	11
2gje	MRP2	RNA	I	8.2	4.0	109	11
3cm1	SSGA-like sporulation-specific cell division protein	No info	I	6.9	4.6	100	9
4zxa	Hydroquinone 1,2-dioxygenase	No info	I	6.6	4.1	104	6
1l3a	transcription factor PBF-2 (P24, WHY1)	ssDNA	I	6.4	4.2	105	7
5fgo	Pur-α repeat III	no	II	6.4	2.1	62	23
3n8b	Pur-α from *B. burgdorferi*	ssDNA/RNA	II	5.6	2.2	68	18
3n1l	stWhy2	DNA	I	5.2	3.0	87	9
5ffp	CdiI immunity protein	No info	I	4.7	3.8	102	3
1pcf	Replication and transcription cofactor PC4 CTD	ssDNA	II	4.7	3.2	60	17
5zkl	SP_0782	ssDNA	II	4.6	3.9	66	7
4ntq	CdiI immunity protein	No info	I	4.6	3.4	92	9
4bhm	MoSub1-DNA PC4	ssDNA	II	4.0	3.5	60	15
2ltt	YdbC (DUF2128 family member)	ssDNA	II	3.8	4.3	61	11
4bg7	PC4 putative transcriptional coactivator p15	ssDNA	II	3.7	6.3	70	9
2qux	PP7 coat protein	RNA stemloop	II	3.2	3.6	53	9

Table S2. Sequence similarity comparison between selected members of PC4-like family

1: 2IT9	100.00	25.58	30.00	10.14	21.25	6.56	4.44	6.82	3.33	7.41	3.85	18.00	5.56	4.55	15.38	6.98	8.33	6.98
2: 2GJE_2	25.58	100.00	24.00	15.88	15.98	16.90	11.70	9.71	8.27	5.56	9.71	15.56	9.09	10.34	9.76	4.81	9.18	5.75
3: 3CM1	30.00	24.00	100.00	20.69	17.07	19.44	18.03	8.57	8.51	7.35	4.94	10.87	5.00	10.67	12.90	10.53	7.94	7.55
4: 2QUX	10.14	15.88	20.69	100.00	19.90	17.69	13.93	8.59	11.56	6.73	6.84	14.29	14.71	11.76	13.70	14.50	12.82	10.91
5: 1L3A	21.25	15.98	17.07	19.90	100.00	48.84	17.76	10.71	13.33	13.54	7.77	14.04	9.57	14.00	17.14	15.32	14.29	11.88
6: 3N1L	6.56	16.90	19.44	17.69	48.84	100.00	18.07	17.05	12.00	10.13	6.33	5.88	7.59	14.44	18.00	19.61	16.00	13.48
7: 5FGP	4.44	11.70	18.03	13.93	17.76	18.07	100.00	29.53	15.79	17.28	13.33	13.95	13.10	11.43	13.24	18.75	16.33	14.29
8: 5FGO	6.82	9.71	8.57	8.59	10.71	17.05	29.53	100.00	12.26	14.77	12.24	13.95	7.53	12.82	23.19	17.48	14.85	13.98
9: 4BG7	3.33	8.27	8.51	11.56	13.33	12.00	15.79	12.26	100.00	25.00	20.89	12.68	12.39	12.28	20.63	9.23	16.83	14.58
10: 1PCF	7.41	5.56	7.35	6.73	13.54	10.13	17.28	14.77	25.00	100.00	36.92	6.25	8.54	8.00	9.62	17.71	16.00	11.59
11: 4BHM	3.85	9.71	4.94	6.84	7.77	6.33	13.33	12.24	20.89	36.92	100.00	7.41	10.68	6.90	19.23	12.96	12.20	10.39
12: 4ZXA	18.00	15.56	10.87	14.29	14.04	5.88	13.95	13.95	12.68	6.25	7.41	100.00	12.33	15.58	17.65	24.51	21.43	21.57
13: 5FFP	5.56	9.09	5.00	14.71	9.57	7.59	13.10	7.53	12.39	8.54	10.68	12.33	100.00	13.76	21.21	17.89	23.86	18.52
14: 2GJE_1	4.55	10.34	10.67	11.76	14.00	14.44	11.43	12.82	12.28	8.00	6.90	15.58	13.76	100.00	17.95	15.32	15.53	20.62
15: 4NTQ	15.38	9.76	12.90	13.70	17.14	18.00	13.24	23.19	20.63	9.62	19.23	17.65	21.21	17.95	100.00	13.95	13.58	16.87
16: 3N8B	6.98	4.81	10.53	14.50	15.32	19.61	18.75	17.48	9.23	17.71	12.96	24.51	17.89	15.32	13.95	100.00	23.01	27.78
17: 5ZKL	8.33	9.18	7.94	12.82	14.29	16.00	16.33	14.85	16.83	16.00	12.20	21.43	23.86	15.53	13.58	23.01	100.00	56.16
18: 2LTT	6.98	5.75	7.55	10.91	11.88	13.48	14.29	13.98	14.58	11.59	10.39	21.57	18.52	20.62	16.87	27.78	56.16	100.00

REFERENCES

1. Xu, Q., Traag, B.A., Willemse, J., McMullan, D., Miller, M.D., Elsliger, M.A., Abdubek, P., Astakhova, T., Axelrod, H.L., Bakolitsa, C. *et al.* (2009) Structural and functional characterizations of SsgB, a conserved activator of developmental cell division in morphologically complex actinomycetes. *J Biol Chem*, **284**, 25268-25279.
2. Graebsch, A., Roche, S. and Niessing, D. (2009) X-ray structure of Pur-alpha reveals a Whirly-like fold and an unusual nucleic-acid binding surface. *Proc Natl Acad Sci USA*, **106**, 18521-18526.
3. Liu, S., Su, T., Zhang, C., Zhang, W.M., Zhu, D., Su, J., Wei, T., Wang, K., Huang, Y., Guo, L. *et al.* (2015) Crystal structure of PnpCD, a two-subunit hydroquinone 1,2-dioxygenase, reveals a novel structural class of Fe2+-dependent dioxygenases. *J Biol Chem*, **290**, 24547-24560.
4. Beck, C.M., Morse, R.P., Cunningham, D.A., Iniguez, A., Low, D.A., Goulding, C.W. and Hayes, C.S. (2014) CdiA from Enterobacter cloacae delivers a toxic ribosomal RNase into target bacteria. *Structure*, **22**, 707-718.
5. Weichenrieder, O., Wild, K., Strub, K. and Cusack, S. (2000) Structure and assembly of the Alu domain of the mammalian signal recognition particle. *Nature*, **408**, 167-173.
6. Desveaux, D., Allard, J., Brisson, N. and Sygusch, J. (2002) A new family of plant transcription factors displays a novel ssDNA-binding surface. *Nat Struct Biol*, **9**, 512-517.
7. Schumacher, M.A., Karamooz, E., Zikova, A., Trantirek, L. and Lukes, J. (2006) Crystal structures of T. brucei MRP1/MRP2 guide-RNA binding complex reveal RNA matchmaking mechanism. *Cell*, **126**, 701-711.
8. Brandsen, J., Werten, S., van der Vliet, P.C., Meisterernst, M., Kroon, J. and Gros, P. (1997) C-terminal domain of transcription cofactor PC4 reveals dimeric ssDNA binding site. *Nat Struct Biol*, **4**, 900-903.
9. Steigemann, B., Schulz, A. and Werten, S. (2013) Bacteriophage T5 encodes a homolog of the eukaryotic transcription coactivator PC4 implicated in recombination-dependent DNA replication. *J Mol Biol*, **425**, 4125-4133.
10. Rossi, P., Barbieri, C.M., Aramini, J.M., Bini, E., Lee, H.W., Janjua, H., Xiao, R., Acton, T.B. and Montelione, G.T. (2013) Structures of apo- and ssDNA-bound YdbC from Lactococcus lactis uncover the function of protein domain family DUF2128 and expand the single-stranded DNA-binding domain proteome. *Nucleic Acids Res*, **41**, 2756-2768.
11. Graebsch, A., Roche, S., Kostrewa, D., Soding, J. and Niessing, D. (2010) Of bits and bugs--on the use of bioinformatics and a bacterial crystal structure to solve a eukaryotic repeat-protein structure. *PLoS One*, **5**, e13402.
12. Li, S., Lu, G., Fang, X., Ramelot, T.A., Kennedy, M.A., Zhou, X., Gong, P., Zhang, X., Liu, M., Zhu, J. *et al.* (2020) Structural insight into the length-dependent binding of ssDNA by SP_0782 from Streptococcus pneumoniae, reveals a divergence in the DNA-binding interface of PC4-like proteins. *Nucleic Acids Res*, **48**, 432-444.
13. Chao, J.A., Patskovsky, Y., Almo, S.C. and Singer, R.H. (2008) Structural basis for the coevolution of a viral RNA-protein complex. *Nat Struct Mol Biol*, **15**, 103-105.

14. Huang, J., Zhao, Y., Liu, H., Huang, D., Cheng, X., Zhao, W., Taylor, I.A., Liu, J. and Peng, Y.L. (2015) Substitution of tryptophan 89 with tyrosine switches the DNA binding mode of PC4. *Sci Rep*, **5**, 8789.
15. Zhao, Y., Zhang, Y., Huang, J., Wang, S., Yi, L., Zhang, X., Xu, M., Fang, X. and Liu, J. (2019) The effect of phosphate ion on the ssDNA binding mode of MoSub1, a Sub1/PC4 homolog from rice blast fungus. *Proteins*, **87**, 257-264.
16. Weber, J., Bao, H., Hartlmuller, C., Wang, Z., Windhager, A., Janowski, R., Madl, T., Jin, P. and Niessing, D. (2016) Structural basis of nucleic-acid recognition and double-strand unwinding by the essential neuronal protein Pur-alpha. *Elife*, **5**.
17. Cappadocia, L., Marechal, A., Parent, J.S., Lepage, E., Sygusch, J. and Brisson, N. (2010) Crystal structures of DNA-Whirly complexes and their role in Arabidopsis organelle genome repair. *Plant Cell*, **22**, 1849-1867.
18. Werten, S. and Moras, D. (2006) A global transcription cofactor bound to juxtaposed strands of unwound DNA. *Nat Struct Mol Biol*, **13**, 181-182.