Supplementary Material

SF1 and SF2 helicases: family matters

Margaret E. Fairman-Williams, Ulf-Peter Guenther
& Eckhard Jankowsky

Current Opinion in Structural Biology, 2010

Supplementary Figure S1

Supplementary Figure S1

Alignment of the helicase core of SF1 and SF2 proteins. SF1 and SF2 proteins from *H. sapiens* (Black), *S. cerevisiae* (violet), *E. coli* (green), and representative members of Flavi- Poty- and Pox-viruses (red) were identified and retrieved from Uniprot and NCBI databases. The helicase core sequences were aligned using ClustalW [1] and Seaview [2] alignment and visualization programs. The resulting alignments were further refined by hand using structural information. For clarity, large insertions are removed from the alignment. These sections are marked by cyan highlights of the number of deleted residues. Proteins are listed in alphabetical order by organism. Helicase motif designations are indicated at the top of the alignment. Characteristic residues in each motif are highlighted in yellow.

The numbering of the motifs is based largely on the original designation by Gorbalenya and Koonin [3]. Exceptions are as follows: motifs V, Va, and Vb were designated as a single motif V, and motif IIIa, present in SF1 proteins, but absent in most SF2 proteins. Motif IIIa in the SF1 was historically termed motif IV. However, structural comparisons of SF2 and SF1 proteins revealed that the position of motif IV in SF1 differed from the position of motif IV in SF2 proteins [4]. The SF1 motif corresponding to motif IV in SF2 was then termed motif IVa. For consistency, we chose here to assign the SF2 designation to both groups, i.e., the SF1 motif IVa is labeled here motif IV as well.

Several motifs that are primarily involved in nucleic acid binding show little sequence conservation (e.g. motifs Ib, Vb). In protein families with little structural information, appropriate designation of these motifs was ambigous. For this reason, no or only minimal highlighting of these motifs is shown. Similarly, motif Q is difficult to assign with limited structural information, and the corresponding sequence assignment should be viewed with care.

We only included certain viral SF2 proteins, even though many viruses encode SF1 and SF2 helicases [5,6]. However, sequence and structural information for most of these proteins is very limited. In contrast, the flaviviral NS3 proteins and the poxviral NPH-II helicase from vaccinia virus are among the biochemically and structurally best characterized SF2 helicases. In addition, flaviviridae NS3 and poxiviridae NPH-II, together with the CI helicases from potyviridae, bear interesting resemblance in sequence to the DEAH/RHA family. We therefore included in the alignment two viral proteins from each of these virus types (poxviruses, potyviruses, and three

flavivirus families). Comparing these well characterized proteins to other SF2 helicases may provide valuable insight into the function of other SF2 proteins.

Note: Because of the large size of the alignment layout, it is best viewed and printed at 400 % of its original size.

SF	1									2														
Family	**UvrD/Rep**	**Pif1-like**		**Upf1-like**						**Swi/Snf**														
S. cerevisiae	Srs2p Hmi1p		Pif1p Rrm3p	Sen1p		Upf1p	Hcs1p	Dna2p	Mtt1p	Mot1p	Rad54p	Rdh54p	Rad26p	Isw1p	Isw2p	Chd1p	Sth1p	Yfk8p	Ino80p	Yab9p	Swr1p		Rad16p Rad5p Yl427p	Yl427p
Human	FBH LBA1	HELB	PIF1	SETX	IBP160 ZNFX1	UPF1	IGHMBP2	DNA2	MOV10 MOV10L1 HELZ PR285 ptMRDFL1	BTAF1	Rad54 ATRX	RA54B	ERCC6	SMAC1	SMAC5	CHD1 CHD2 CHD3 CHD4 CHD5 CHD6 CHD7 CHD8 CHD9	SMCA2 SMCA4	HELLS	INOC1	SMRCD	EP400 SRCAP	SMAL1 ZRAB3	HLTF TTF2	SHRPH

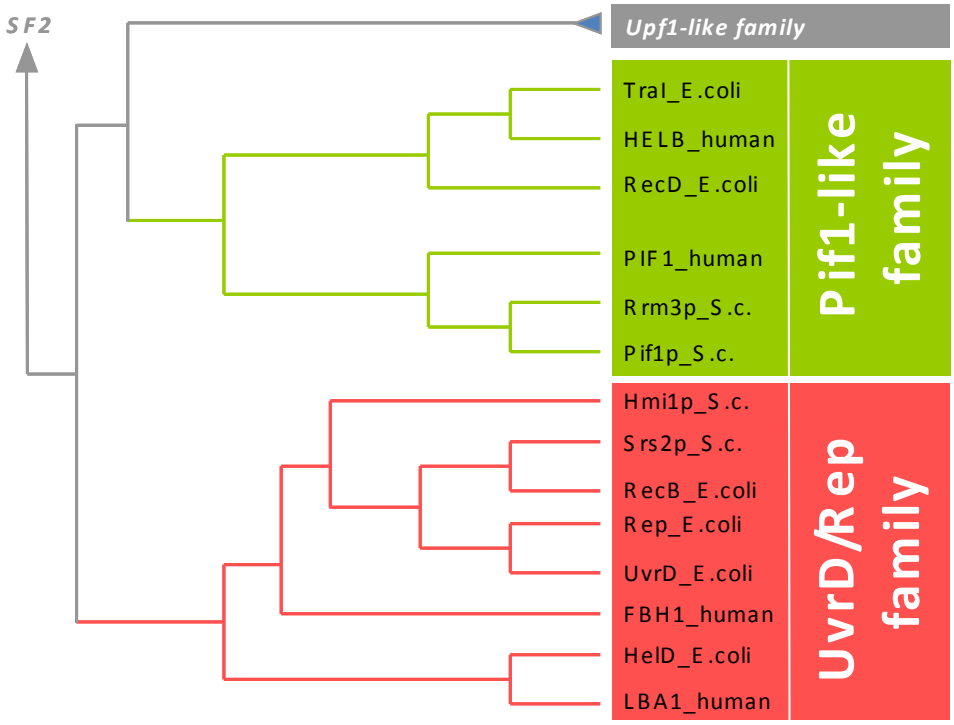
SF	2																							
Family	**Rad3/Xpd**		**Ski2-like**							**DEAH/RHA**									**RIG-I-like**			**RecQ-like**		
S. cerevisiae	Chl1p	Rad3p	Ski2p	Mtr4p	Hfm1p		Slh1p	Brr2p		Prp22p	Prp16p	Prp43	Dhr2p	Dhr1p	YL419p		Prp2p		Mph1p		Sqs1p	Hrq1p		
Human	DDX11 DDX12 RTEL1 FANCJ	ERCC2	SKIV2	SK2L2	HFM1	HL308	HELC1	U5200	DDX60	DHX8 DHX16	DHX38	DHX15	DHX33 DHX35	DHX37 DHX40	DHX9 DHX36 DHX57 DHX29 DHX30	DQX1	DHX34	RIG-I IFIH1 LPG2	FANCM	Dicer	RecQ1 BLM WRN RecQ4	RecQ5	

SF	2																							
Family	**DEAD-box**																							
S. cerevisiae	Prp5p	Dbp9p	Dbp6p		Tif1p/Tif2p	Fal1p	Dhh1p	Dbp5p	Sub2p	Has1p	Mss116p	Dbp4p	Sbp4p	Dbp7p	Rrp3p	Dbp8p	Drs1p	Dbp10p	Dbp3p	Ded1p Dbp1p	Dbp2p	Prp28p	Mak5p	Rok1p	Mrh4p
Human	DDX46 DDX42	DDX56	DDX51	DDX21 DDX50	DDX2A	RecQ5	DDX6	DDX19 DDX20 DDX25	DDX39 UAP56	DDX18		DDX10	DDX55	DDX31	DDX47	DDX49	DDX27	DDX54	DDX17	DDX3Y DDX3X DDX4	DDX53 DDX43 DDX5 DDX1 DDX59	DDX23 DDX41	DDX24	DDX52	DDX28

Supplementary table TS1. List of eukaryotic proteins used in the SF1/2 alignment. Proteins from *S. cerevisae* and *H. sapiens* are listed by family, with homologous/orthologous proteins shown in the same column in each subtable. SUV3 homologs from both organisms could not be assigned to a family.

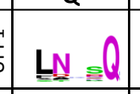
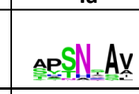
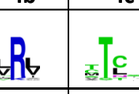
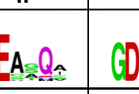
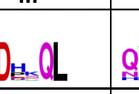
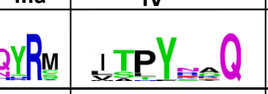
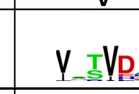
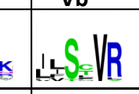
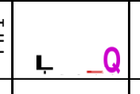
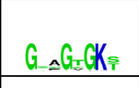
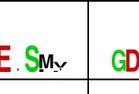
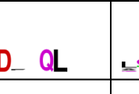
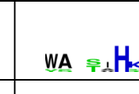
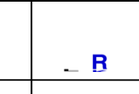
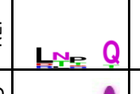
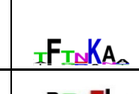
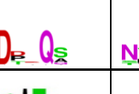
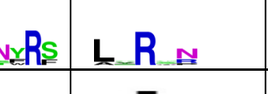
SF	1			2							
Family	**UvrD/Rep**	**Pif1-like**	**Upf1-like**	**Swi/Snf**	**Rad3/Xpd**	**Ski2-like**	**DEAH/RHA**	**RIG-I-like**	**RecQ-like**	**DEAD-box**	**RecG-like**
E. Coli	RecB Rep UvrD HelD	Tral RecD	Z1202 UL1	RapA	DinG YoaA	Q6XGE6	HrpA HrpB		RecG PriA MFD	DbpA RhlB Dead RhlE SrmB	RecG PriA MFD

Supplementary table TS2. List of SF1/2 proteins from *E.coli* used in the alignment. The highly similar t ype III restrictions enzymes BPP1 and BP15I could not be assigned to any family.



Supplementary Figure S2

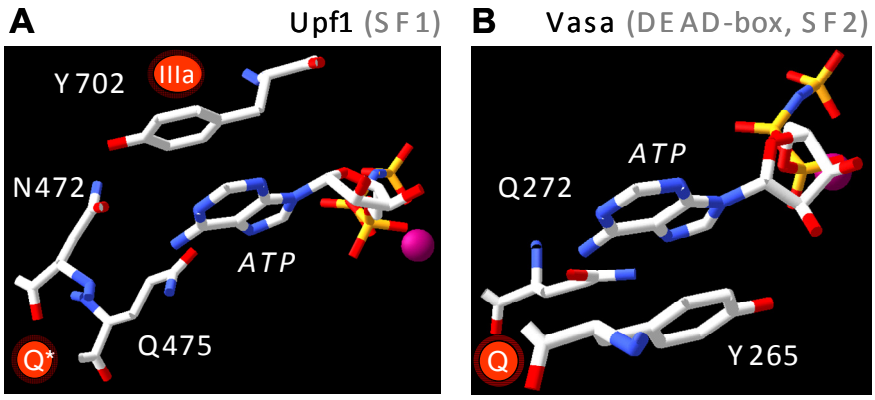
Cladogram for UvrD/Rep and Pif1-like families. The cladogram shows an example of the phylogenetic analyses for two SF1 protein families, based on the sequence alignment of the helicase core (**Suppl. Fig.S1**). The figure shows the result of an analysis using distance- and character-based program options (BioNJ, protpars) of Seaview4 [7,8]. Consistent grouping of proteins into family clusters, such as those seen here, were obtained with various phylogenetic methods. However, topologies of phylogenetic trees varied on occasion, dependent on the algorithm used. This variation is presumably due to the large number of unavoidable sequence gaps in the alignment (**Suppl. Fig.S1**). In several instances, branching was not supported by high bootstrap values. Consequently, the evolutionary relationship between individual family members and the phylogenetic relationship between the families (**Fig.1**) should thus be viewed with care. Nonetheless, the cladograms revealed several phylogenetic relations that are in excellent with structural data (**Figs.3,4**), for example the close relation of the NS3/NPH-II group to the DEAH/RHA proteins (**Figs.3,4**).

SF	F	Q	I	Ia	Ib	Ic	II	III	IIIa	IV	IVa	V	Va	Vb	VI
1	UPF1														
	PIF1														
	UVRD/ REP														
2	DEAD														
	DEAH/ RHA														
	RIG-I														
	SKI2														
	SWI2/ SNF2														
	RECQ														
	RAD3														
	T1R														
	RECG														
	Viral NS3 /NPH-II														

Supplementary Figure S3

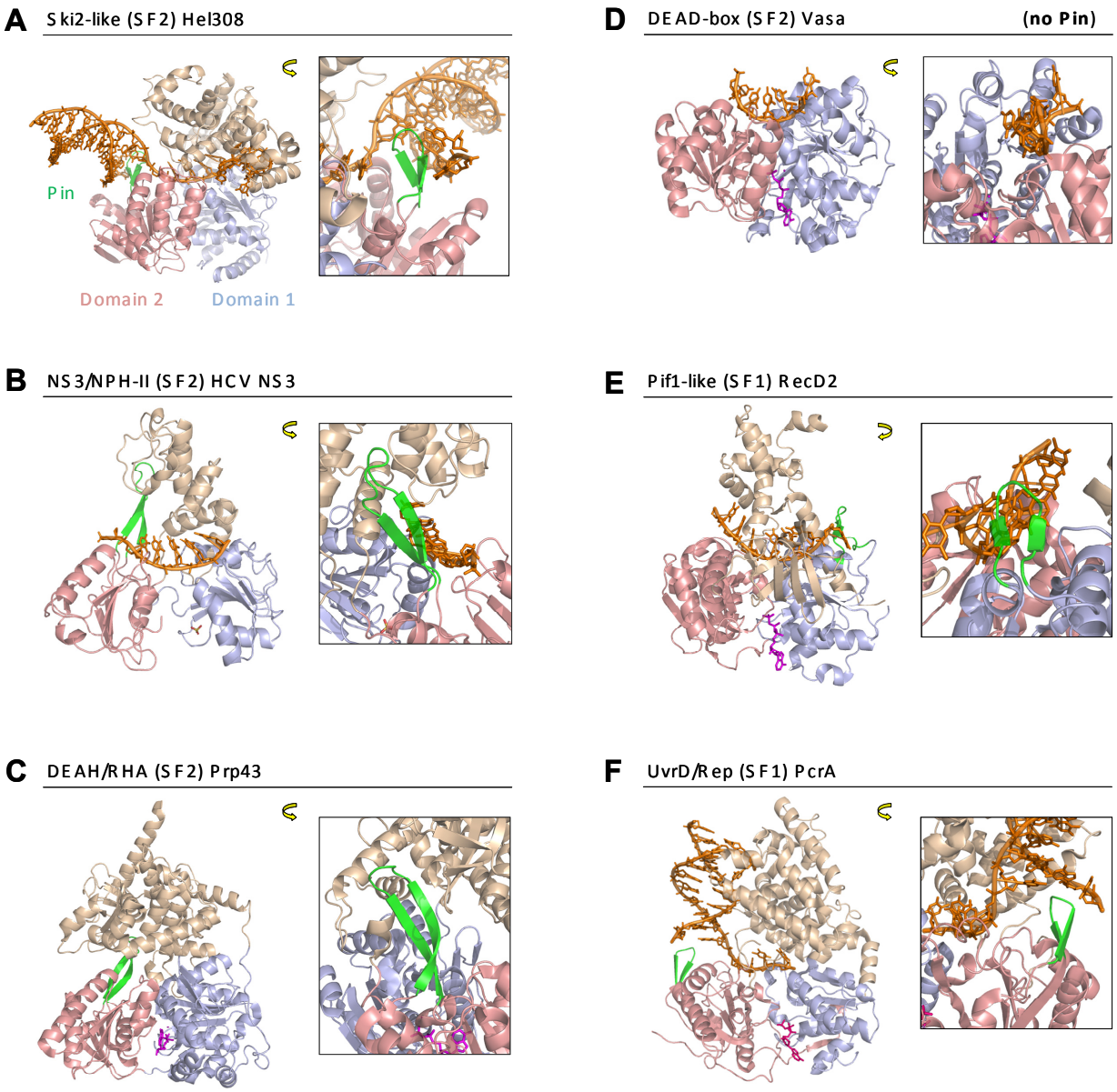
Supplementary Figure S3

Characteristic sequence motifs for the individual SF1 and SF2 protein families. Sequence conservation on the family/group level of the characteristic helicase motifs is shown as sequence logo [9]. The height of the amino acid symbols reflects the sequence conservation and the color their chemical properties: polar - green, basic - blue, acidic – red, and hydrophobic -black. Sequence positions that are only conserved within a given family/group are not displayed.



Supplementary Figure S4

The putative adenine stacking platform in motif IIIa of SF1 proteins. Sidechains of conserved residues that provide specificity for the adenosine base are shown from similar angles in structures of the SF1 helicase UPF1 [10] (**A**) and SF2 DEAD-box protein *Vasa* [11] (**B**). Atoms are colored as follows: C – white, O – red and N – blue. Residues of the Q-motif that contact the adenosine in a specific manner are positioned similarly in both structures (Q475, N472 in UPF1; Q272 in Vasa). A conserved aromatic residue stacking on the adenine base is located in motifs IIIa in UPF1, but in the Q-motif in Vasa (Y702 in UPF1, Y265 in Vasa). Notably, this residue approaches the adenine base from opposite direction in SF1 (top) and SF2 (bottom). In addition, motif IIIa contains a highly conserved arginine.



Supplementary Figure S5

Characteristic "pins" in structures of SF1 and SF2 helicases. Representative structures of proteins from different SF2 and SF1 families, containing "pin" domains, a prominent β-hairpin. The helicase domain 1A is blue, 2A pink, accessory/ inserted domains are light-beige. The "pin" is green in all structures. ATP and nucleic acid, are magenta or orange, respectively. In the left panels, all structures are oriented as shown in **Fig.4**. The right panels show the structures rotated by roughly 90^0 (indicated by a yellow arrow), zoomed into the region containing the "pin". Panel

D shows the DEAD-box protein Vasa as example for a family lacking a "pin". The following structures are shown: (**A**) Hel308 (*Archaeoglobus fulgidus*) [12], (**B**) HCV NS3 (*Hepatitis C virus*) [13], (**C**) Prp43p (*S. cerevisiae*) [14], (**D**) Vasa (*D. melanogaster*) [15], (**E**) RecD2 (*D. radiodurans*) [15], (**F**) PcrA (*B. stearothermophilus*) [16]. We note that Rad3/XPD proteins, instead of a β-hairpin, have a 2-helix insertion in a similar location between motifs Va and VI [17]. It is unclear whether this motif serves a function similar to the "pin".

REFERENCES

1. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, A. W, R. L, et al.: **Clustal W and Clustal X version 2.0**. *Bioinformatics* 2007, **23**:2947-2948.

2. Gouy M, Guindon S, Gascuel O: **SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building.** *Mol. Biol. Evol.* 2010, **27**:221-224.

3. Gorbalenya AE, Koonin EV: **Helicases: amino acid comparisons and structure-function relationships.** *Curr Opin Struct Biol* 1993, **3**:419-429.

4. Korolev S, Yao N, Lohman TM, Weber PC, Waksman G: **Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases.** *Protein Sci.* 1998, **7**:605-610.

5. Kadare G, Haenni AL: **Virus-encoded RNA helicases.** 1997, **J. Virol**:2583-2590.

6. Frick DN, Lam AM: **Understanding helicases as a means of virus control.** *Curr. Pharm. Des.* 2006, **12**:1315-1338.

7. Gascuel O: **BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data.** *Mol. Biol. Evol.* 1997, **14**:685-695.

8. Felsenstein J: **PHYLIP (Phylogeny Inference Package) version 3.52.** *distributed by the author* 1993, **Department of Genome Sciences, University of Washington, Seattle.**

9. Crooks GE, Hon G, Chandonia JM, Brenner SE: **WebLogo: a sequence logo generator.** *Genome Res.* 2004, **14**:1188-1190.

10. Clerici M, Mourão A, Gutsche I, Gehring NH, Hentze MW, Kulozik A, Kadlec J, Sattler M, Cusack S: **Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2.** *EMBO J.* 2009, **28**:2293-2306.

11. Sengoku T, Nureki O, Nakamura A, Kobayashi S, Yokoyama S: **Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa.** *Cell* 2006, **125**:287-300.

12. Büttner K, Nehring S, Hopfner KP: **Structural basis for DNA duplex separation by a superfamily-2 helicase.** *Nat. Struct. Biol.* 2007, **14**:647-352.

13. Yao N, Reichert P, Taremi SS, Prosise WW, Weber PC: **Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase.** *Structure* 1999, **15**:1353-1363.

14. He Y, Andersen GR, Nielsen KH: **Structural basis for the function of DEAH helicases**. *EMBO Rep.* 2010:in press.

15. Singleton MR, Dillingham MS, Gaudier M, Kowalczykowski SC, Wigley DB: **Crystal structure of RecBCD enzyme reveals a machine for processing DNA breaks.** *Nature* 2004, **432**:187-193.

16. Velankar SS, Soultanas P, Dillingham MS, Subramanya HS, Wigley DB: **Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism.** *Cell* 1999, **97**:75-84.

17. Fan L, Fuss JO, Cheng QJ, Arvai AS, Hammel M, Roberts VA, Cooper PK, Tainer JA: **XPD helicase structures and activities: insights into the cancer and aging phenotypes from XPD mutations.** *Cell* 2008, **133**:789-800.