

1 **Title:**

2 **Classification of the treble clef zinc finger: noteworthy lessons for structure and function**
3 **evolution**

4

5 **Authors:**

6 Gurmeet Kaur, Srikrishna Subramanian¹

7

8 **Contact Information**

9 CSIR-Institute of Microbial Technology (IMTECH), Sector 39-A, Chandigarh, 160036, India. Phone:
10 +911726665483; Fax: +91-1722695215; Email: krishna@imtech.res.in

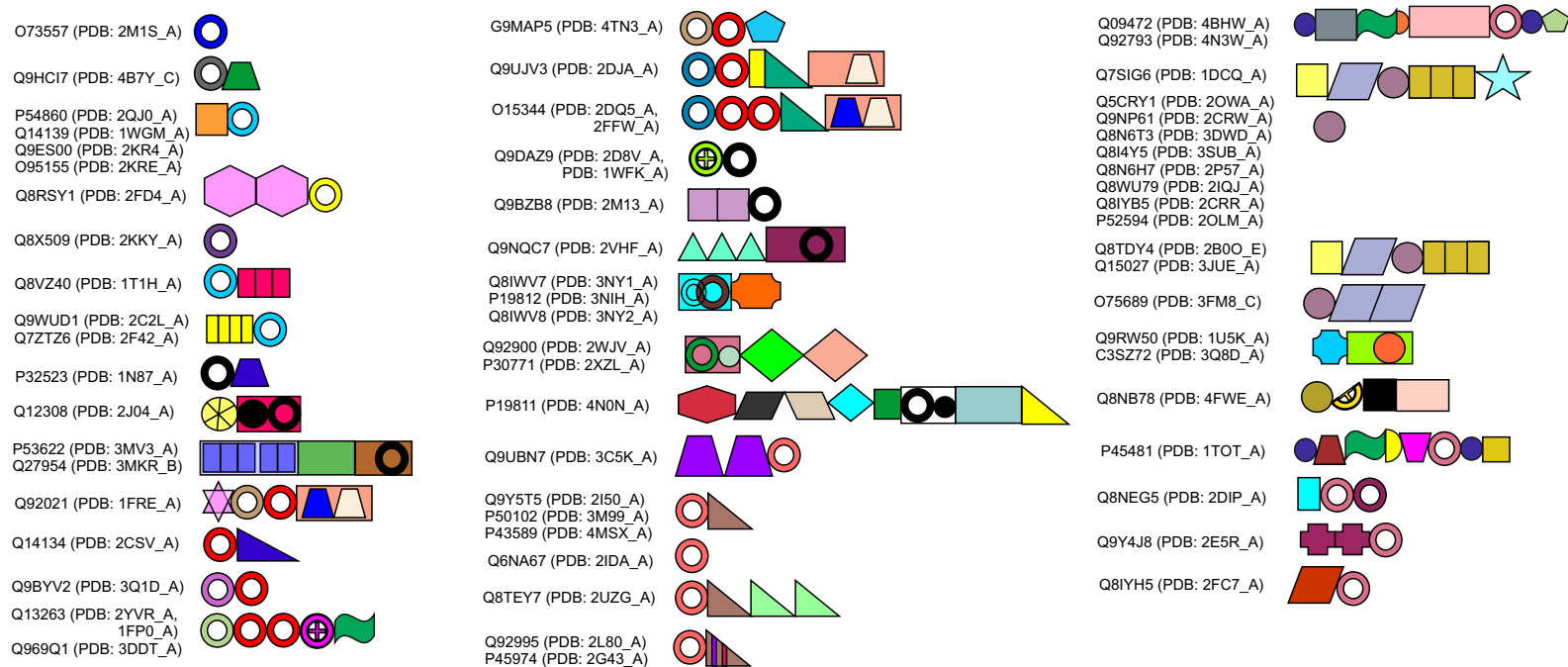
11 ¹ Corresponding author

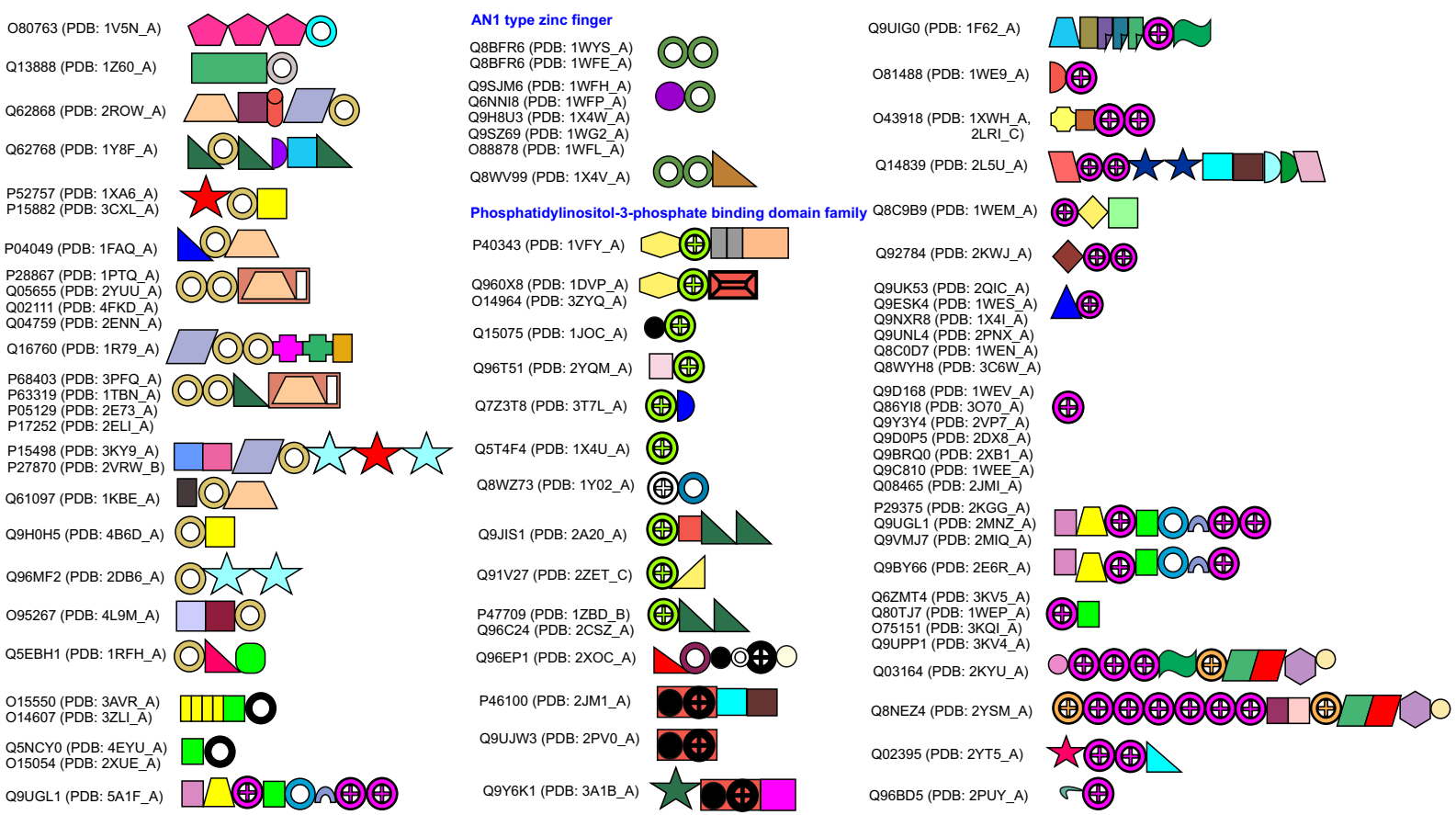
12

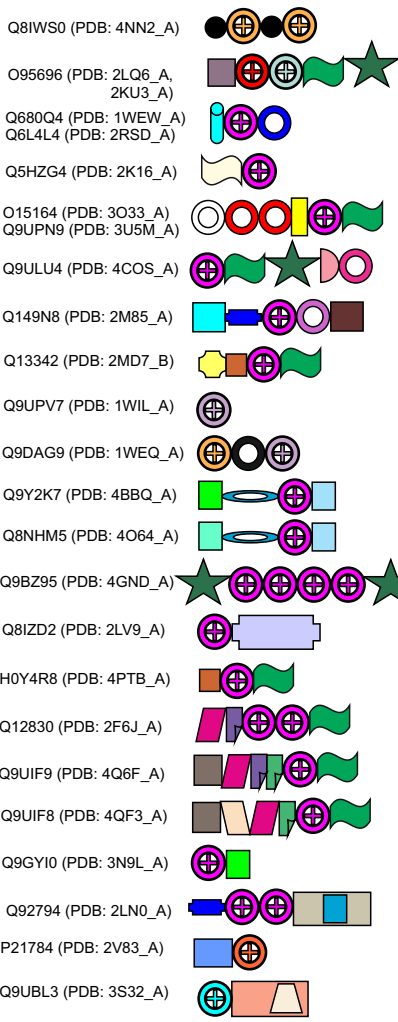
13

14 **Supplementary File 3**

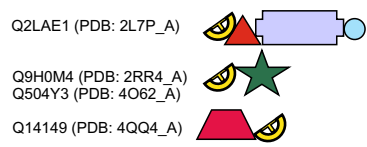
15 **Co-occurring domains with the treble clef zinc fingers.** Domain co-occurrence for proteins with
16 structurally characterised TC ZFs is shown. The UniProtId, selected PDBid(s) and the domains
17 present in the full-length sequence of the protein are depicted. Different shapes have represented the
18 different domains.



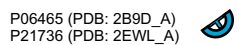




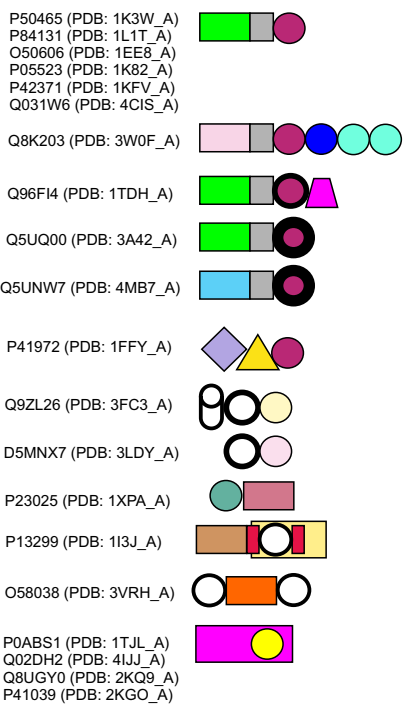
CW domain family



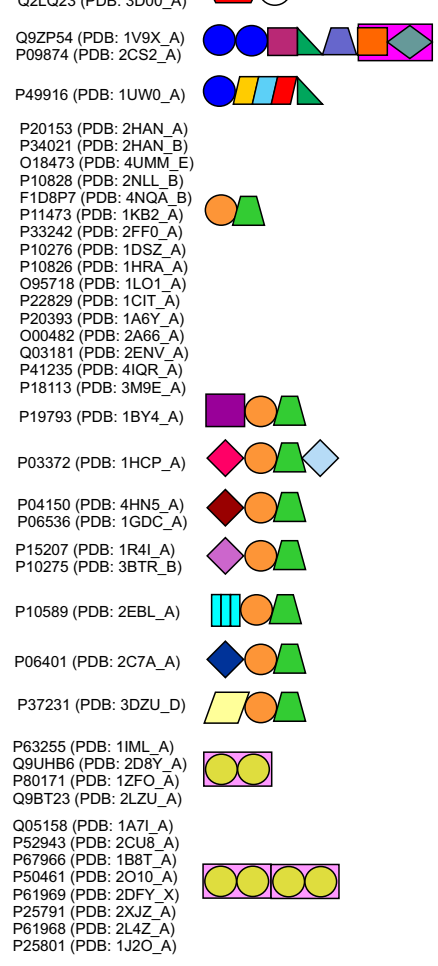
Zinc binding region of E7 oncoprotein



Nuclear receptor-like finger family



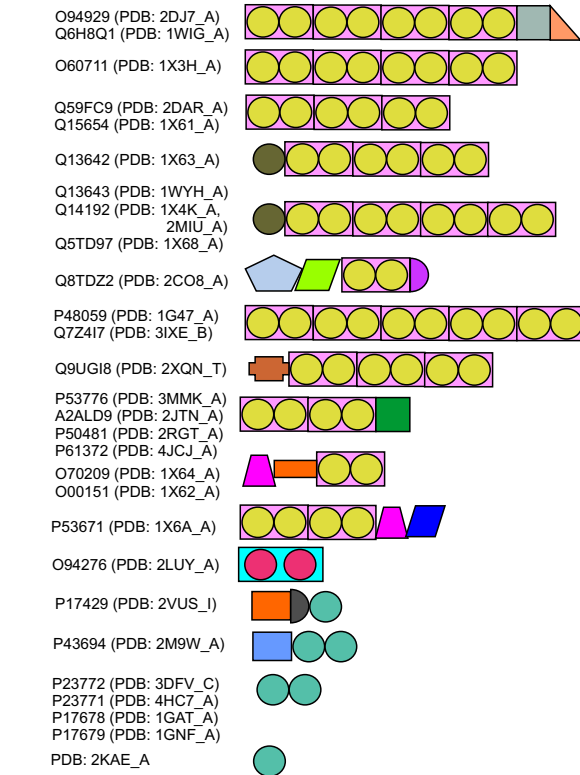
Q9HJ63 (PDB: 2GVI_A)



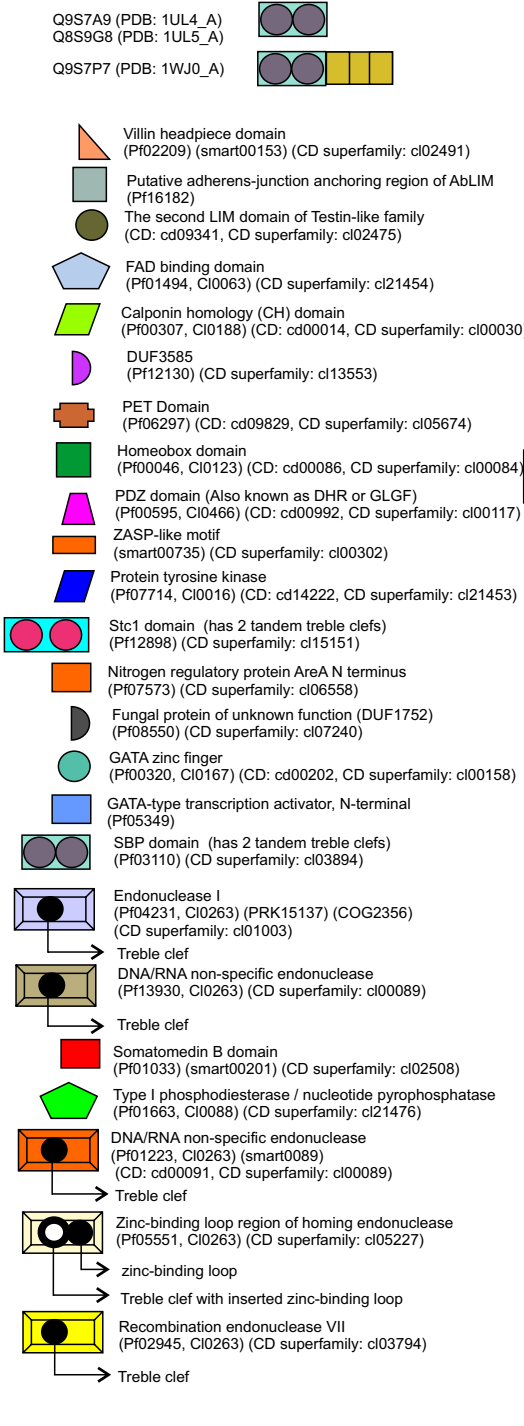
- Enhancer of polycomb-like (PF10513) (CD superfamily: cl111500)
- PHD-finger (PF13831, CI0390) (CD superfamily: cl17040)
- PHD-finger (PF13832, CI0390) (CD superfamily: cl17040)
- Bromodomain associated (PF07524, CI0012) (smart00576) (CD superfamily: cl17378)
- Domain of unknown function (DUF3544) (PF12064) (CD superfamily: cl13494)
- Histone H1 (PF00538) (smart00526) (CD: cd00073, CD superfamily: cl00073)
- FYVE/PHD zinc finger (annotated by SCOP; does not find sequence similarity hits in CD or Pfam)
- CXXC zinc finger domain (PF02008) (CD superfamily: cl03401)
- F-box-like (PF12937, CI0271) (CD superfamily: cl02535)
- associated with SET domains (smart00570) (CD superfamily: cl02673)
- DDT domain (PF02791) (smart00571) (CD superfamily: cl02674)
- Methyl-CpG binding domain (PF01429, CI0081) (CD: cd01397, CD superfamily: cl00110)
- Outer membrane protein (OmpH-like) (PF03938) (CD superfamily: cl21485)
- MOZ/SAS family (PF01853, CI0257) (CD superfamily: cl19766)
- N-Acyltransferase superfamily (CD: cd04301, CD superfamily: cl17182)
- Recombination activating protein 2 (PF03089, CI0186) (CD superfamily: cl20243)
- RAG2 PHD domain (PF13341, CI0390) (CD superfamily: cl16246)
- PHD-finger (only 2nd zinc binding site structure solved, but sequence for complete PHD finger is present)
- PostSET (smart00508) (CD superfamily: cl02636)
- Histidine kinase-, DNA gyrase B-, and HSP90-like ATPase (PF13589, CI0025) (CD: cd00075, CD superfamily: cl00075)
- E7 protein, Early protein (PF00527) (CD superfamily: cl02891)
- Formamidopyrimidine-DNA glycosylase N-terminal domain (PF01149) (CD: cd08965/cd08967/cd08966, CD superfamily cl03119)
- N-terminal domain of metazoan Nei-like glycosylase 3 (NEIL3) (CD: cd08969, CD superfamily: cl03119)
- N-terminal domain of Fpg (formamidopyrimidine-DNA glycosylase, MutM)/_Nei (endonuclease VIII) base-excision repair DNA glycosylases (CD: cd8773, CD superfamily: cl03119)
- Formamidopyrimidine-DNA glycosylase H2TH domain (PF06831, CI0303) (CD superfamily: cl06065)
- Zinc finger found in FPG and IleRS (PF06827, CI0167) (CD superfamily: cl06062)

- Degraded zinc finger-type found in FPG and IleRS (annotated by SCOP; does not find sequence similarity hits in CD or Pfam)
- zinc finger-type found in FPG and IleRS (annotated by SCOP; does not find sequence similarity hits in CD or Pfam)
- Zn-finger in Ran binding protein and others (PF00641, CI0167) (CD superfamily: cl02656)
- GRF zinc finger (PF06839, CI0167) (CD superfamily: cl06070)
- Endonuclease VIII-like 1, DNA bind (PF09292) (CD superfamily: cl07797)
- tRNA synthetases class I (I, L, M and V) (PF00133, CI0039) (CD: cd00818, CD superfamily: cl00015)
- Anticodon-binding domain of tRNA (PF08264) (CD: cd07960, CD superfamily: cl12020)
- Beta-barrel domain in structure (Not annotated anywhere, doesn't find hits to any CD)
- Treble clef (Not annotated anywhere, doesn't find hits to any CD, present in structure)
- Recombination endonuclease VII (PF02945, CI0263) (CD superfamily: cl03794)
- Catalytic Treble clef (Not annotated anywhere, doesn't find hits to any CD, present in structure)
- XPA protein N-terminal (PF01286) (Cd superfamily: cl03148)
- XPA protein C-terminus (PF05181) (CD superfamily: cl04970)
- G1Y-YIG catalytic domain (PF01541, CI0418) (CD: cd10437, CD superfamily: cl15257)
- G1YX(10-11)YIG family of class I homing endonucleases C-terminus (CD: cd00283, CD superfamily: cl19093)
- NUMOD3 motif (2 copies) (PF07460)
- Treble clef (Not annotated anywhere, doesn't find hits to any CD)
- PP-loop family (PF01171, CI0039) (CD: cd01993, CD superfamily: cl00292)
- RNA polymerase-binding transcription factor (PRK10778, PRK11019) (COG1734) (TIGR02420) (CD superfamily: cl00755)
- Prokaryotic dksA/trar C4-type zinc finger (PF01258, CI0167)
- FmdE, Molybdenum formylmethanofuran dehydrogenase operon (PF02663) (CD superfamily: cl12269)
- Treble clef (annotated by SCOP)
- Poly(ADP-ribose) polymerase and DNA-Ligase Zn-finger region (PF00645) (CD superfamily: cl02913)
- PADR1 (NUC008) domain (PF08063) (CD superfamily: cl06892)

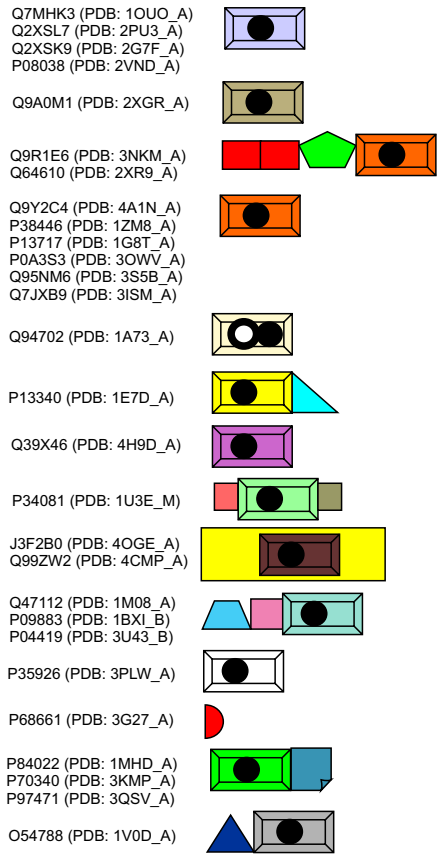
- BRCA1 C Terminus (BRCT) domain (PF00533, CI0459) (smart00292) (CD: cd00027, CD superfamily: cl00038)
- WGR domain (PF05406) (CD: cd08001, CD superfamily: cl01581)
- Poly(ADP-ribose) polymerase (parp)-like (CD: cd01437, CD superfamily: cl00283)
- Poly(ADP-ribose) polymerase catalytic domain (PF00644, CI0084)
- Poly(ADP-ribose) polymerase, regulatory domain (PF02877)
- DNA ligase N terminus (PF04675) (CD superfamily: cl20272)
- DNA ligase N terminus (PF04675) (CD superfamily: cl20272)
- ATP dependent DNA ligase C terminal region (PF04679) (CD: cd07967, CD superfamily: cl08424)
- Zinc finger, C4 type (two domains) (CD: cd06956/cd07161/cd06961/cd07160/cd06955/cd07167/cd06964/cd07170/cd07171/cd07172/cd06969/cd07166/cd07167/cd06965/cd06960/cd07173/cd06958/cd07172/cd06965, CD superfamily: cl02596)
- Ligand-binding domain of nuclear hormone receptor (PF0104) (CD: cd06943/cd06938/cd06935/cd06954/cd06933/cd07070/cd06937/cd06946/cd06949/cd07076/cd07348/cd06940/cd07069/cd06932/cd06931/cd07073/cd06948/cd07074/cd06932, CD superfamily: cl11397)
- Nuclear/hormone receptor activator site AF-1 (PF181825) (CD superfamily: cl13292)
- Oestrogen receptor (PF02159) (CD superfamily: cl03471)
- Oestrogen-type nuclear receptor final C-terminal (PF12743) (CD superfamily: cl15062)
- Glucocorticoid receptor (PF02155) (CD superfamily: cl03469)
- Androgen receptor (PF02166)
- Collagen triple helix repeat (20 copies) (PF01391, CD superfamily: cl19732)
- Progesterone receptor (PF02161)
- PPAR gamma N-terminal region (PF12577) (CD superfamily: cl13952)
- LIM domain (has 2 tandem treble clefs) (PF00412) (CD: cd09478/cd09480/cd09840/cd09476/cd09479/cd09403/cd09481/cd09482/cd09327/cd09328/cd09329/cd09330/cd09344/cd09424/cd09429/cd09423/cd09427/cd09346/cd09434/cd09422/cd09426/cd09431/cd09433/cd09358/cd09406/cd09408/cd09410/cd09339/cd09331/cd09332/cd09333/cd09334/cd09335/cd09453/cd09457/cd09459/cd09386/cd09387/cd09384/cd09385/cd09413/cd09416/cd09419/cd09468/cd09376/cd09359/cd09450/cd09343/cd09428/cd09347/cd09467/cd09366/cd09374/cd09350/cd09354/cd09357/cd09463/cd09465/cd09448/cd09486/, CD superfamily: cl02475)



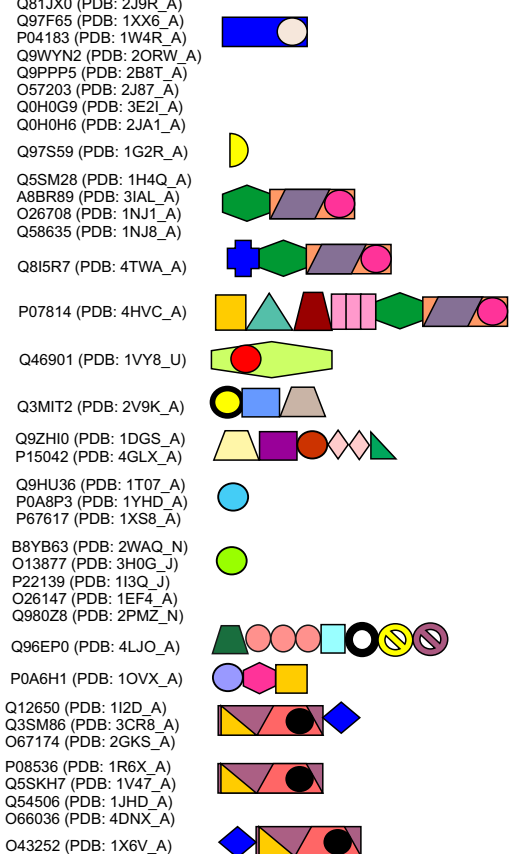
SBT domain family



His-Me family



Other small families



- Villin headpiece domain (PF02209) (smart00153) (CD superfamily: cl02491)
- Putative adherens-junction anchoring region of AblIM (Pf16182)
- The second LIM domain of Testin-like family (CD: cd09341, CD superfamily: cl02475)
- FAD binding domain (PF01494, Cl0063) (CD superfamily: cl21454)
- Calponin homology (CH) domain (PF00307, Cl0188) (CD: cd00014, CD superfamily: cl00030)
- DUF3585 (Pf12130) (CD superfamily: cl13553)
- PET Domain (PF06297) (CD: cd09829, CD superfamily: cl05674)
- Homeobox domain (PF00046, Cl0123) (CD: cd00086, CD superfamily: cl00084)
- PDZ domain (Also known as DHR or GLGF) (PF00595, Cl0466) (CD: cd00992, CD superfamily: cl00117)
- ZASP-like motif (smart00735) (CD superfamily: cl00302)
- Protein tyrosine kinase (PF07714, Cl0016) (CD: cd14222, CD superfamily: cl21453)
- Stc1 domain (has 2 tandem treble clefs) (Pf12898) (CD superfamily: cl15151)
- Nitrogen regulatory protein AreA N terminus (PF07573) (CD superfamily: cl06558)
- Fungal protein of unknown function (DUF1752) (PF08550) (CD superfamily: cl07240)
- GATA zinc finger (PF00320, Cl0167) (CD: cd00202, CD superfamily: cl00158)
- GATA-type transcription activator, N-terminal (PF05349)
- SBP domain (has 2 tandem treble clefs) (Pf03110) (CD superfamily: cl03894)
- Endonuclease I (PF04231, Cl0263) (PRK15137) (COG2356) (CD superfamily: cl01003)
- Treble clef
- DNA/RNA non-specific endonuclease (PF13930, Cl0263) (CD superfamily: cl00089)
- Treble clef
- Somatomedin B domain (PF01033) (smart00201) (CD superfamily: cl02508)
- Type I phosphodiesterase / nucleotide pyrophosphatase (PF01663, Cl0088) (CD superfamily: cl21476)
- DNA/RNA non-specific endonuclease (PF01223, Cl0263) (smart0089) (CD: cd00091, CD superfamily: cl00089)
- Treble clef
- Zinc-binding loop region of homing endonuclease (PF05551, Cl0263) (CD superfamily: cl05227)
- zinc-binding loop
- Treble clef with inserted zinc-binding loop
- Recombination endonuclease VII (PF02945, Cl0263) (CD superfamily: cl03794)
- Treble clef

- T4 recombination endonuclease VII, dimerisation (PF09124, Cl0306) (CD superfamily: cl07663)
- HNH endonuclease (PF01844, Cl0263) (CD: cd00085, CD superfamily: cl00083)
- Treble clef
- NUMOD4 motif (PF07463) (CD superfamily: cl06480)
- HNH endonuclease (PF13392, Cl0263) (CD superfamily: cl16264)
- Treble clef
- Intron encoded nuclease repeat motif (smart00497) (CD superfamily: cl11508)
- CRISPR/Cas system-associated protein Cas9 (CD: cd09643, CD superfamily: cl00083)
- HNH endonuclease (PF13395, Cl0263)
- Treble clef
- Colicin-like bacteriocin tRNase domain (PF03515, Cl0446) (CD: superfamily: cl04127)
- Coiled-coil receptor-binding R-domain of colicin E2 (PF11570) (CD superfamily: cl13083)
- DNase/tRNase domain of colicin-like bacteriocin (PF12639, Cl0263) (CD superfamily: cl15861)
- Treble clef
- HNH endonuclease (finds no sequence similarity to any other known protein domains; not annotated by Pfam, UniProt, SCOP)
- Treble clef
- DUF1364 (is a treble clef) (PF07102) (CD superfamily: cl06229)
- Mh1 domain (PF03165, Cl0263) (CD: cd10491/cd10490/cd10492, CD superfamily: cl00055)
- Treble clef
- Mh2 domain (PF03166, Cl0357) (CD: cd10985/cd10497/cd10498, CD superfamily: cl00056)
- CIDE-N domain (PF02017, Cl0072) (CD: cd06535, CD superfamily: cl02541)
- DNA fragmentation factor 40 kDa (PF09230) (CD superfamily: cl07748)
- Treble clef
- Thymidine kinase (PF00265, Cl0263) (PRK04296) (PTZ00293) (COG1435) (CD superfamily: cl21455)
- Treble clef
- DUF448 (is a treble clef in structure) (PF04296) (CD: cd00279, CD superfamily: cl00189)
- tRNA synthetase class II core domain (G, H, P, S and T) (PF0587, Cl0040) (CD: cd00778, CD superfamily: cl00268)
- ProRS Prolyl-anticodon binding domain (CD: cd00862, CD superfamily: cl00266)
- Prolyl-tRNA synthetase, C-terminal (PF09180, PF09181)
- Anticodon binding domain (PF03129, Cl0458)

- Aminoacyl-tRNA editing domain (PF04073) (CD: cd04335, CD superfamily: cl00022)
- Glutathione S-transferase, C-terminal domain (PF14497, Cl0497) (CD: cd10309, Cl02776)
- tRNA synthetases class I (E and Q), catalytic domain (PF00749, Cl0039) (CD: cd00807, CD superfamily: cl00015)
- tRNA synthetases class I (E and Q), anti-codon binding domain (PF03950) (CD superfamily: cl04341)
- WHEP-TRS domain (PF00458) (CD: cd00936, CD superfamily: cl00349)
- CRISPR-associated protein Cse1 (CRISPR_cse1) (PF09481) (CD: cd09669, CD superfamily: cl08044)
- Circularly permuted treble clef
- Circularly permuted treble clef (not annotated, does not find sequence similarity to any other domains in Pfam and CDD)
- THUMP domain (annotated by authors)
- putative pseudouridylylase (PRK14554)
- NAD-dependent DNA ligase adenylation domain (PF01653, Cl0078) (CD: cd00114, CD superfamily: cl03295)
- NAD-dependent DNA ligase OB-fold domain (PF03120, Cl0021) (CD superfamily: cl03901)
- NAD-dependent DNA ligase C4 zinc finger domain (PF03119) (CD superfamily: cl03900)
- Helix-hairpin-helix domain (PF14520/PF12826, Cl0198) (CD superfamily: cl14815)
- Bacterial Fe(2+) trafficking (PF04362) (PRK05408) (COG2924) (CD superfamily: cl01104)
- RNA polymerases N / 8 kDa subunit (PF01194) (PRK04016) (PLN00032) (CD superfamily: cl00712)
- PUB domain (PF09409) (CD: cd10464, CD superfamily: cl15262)
- RanBP2-type zinc finger (annotated by UniProt)
- UBA domain found in E3 ubiquitin-protein ligase RING finger protein 31 and similar proteins (CD: cd14325, CD superfamily: cl21463)
- ClpX C4-type zinc finger (PF06889) (CD superfamily: cl05964)
- AAA domain (Cdc48 subfamily) (PF07724, Cl0023) (CD: cd00009, CD superfamily: cl21455)
- C-terminal, D2-small domain, of ClpB protein (PF10431) (smart01086) (CD superfamily: cl11083)
- ATP-sulfurylase (CD: cd00517, CD superfamily: cl00015)
- ATP-sulfurylase (with a treble clef-like zinc finger) (PF01747, Cl0039)
- PUA-like domain (PF14306, Cl0178)
- Adenylylsulphate kinase (PF01583, Cl0023) (CD: cd02027, CD superfamily: cl17190)

