Landes Bioscience
www.landesbioscience.com

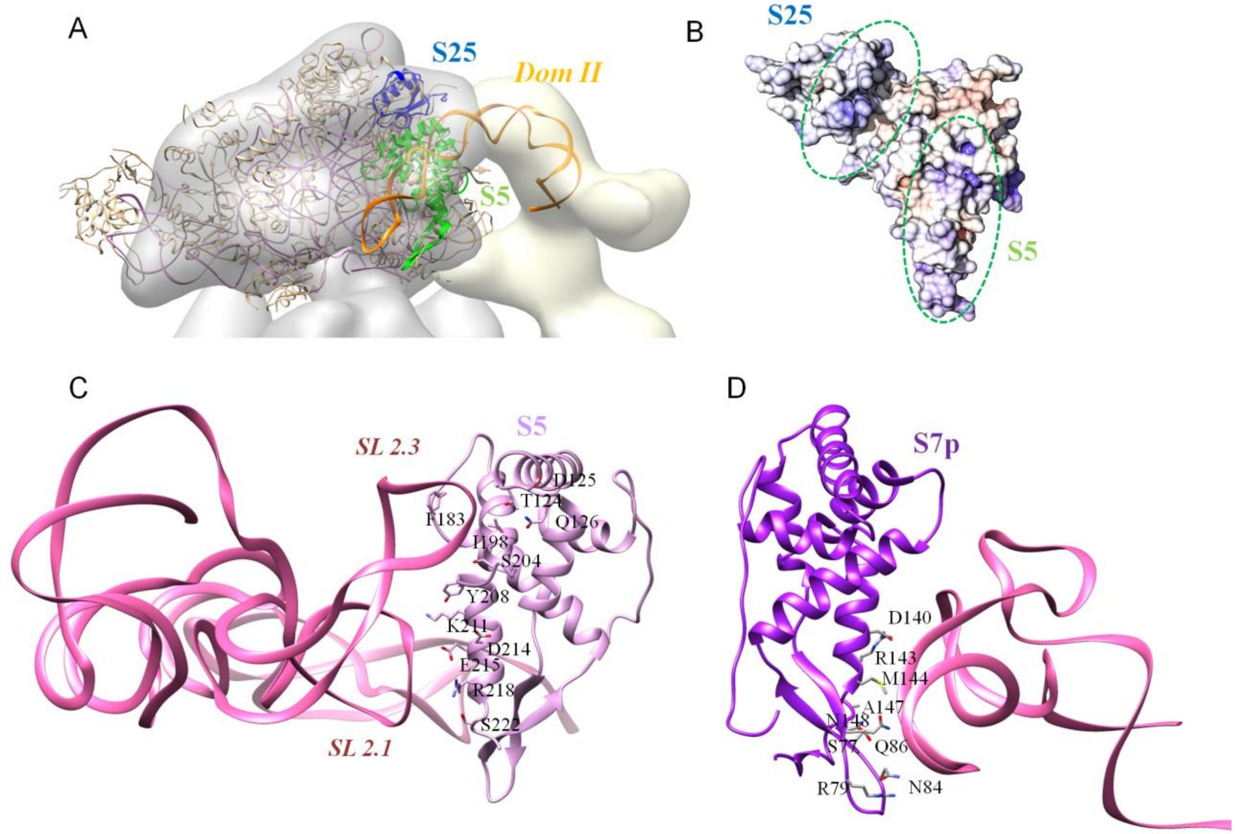
Supplemental Material to:

Agnel Joseph, Prasanna Bhat, Saumitra Das,
Narayanaswamy Srinivasan

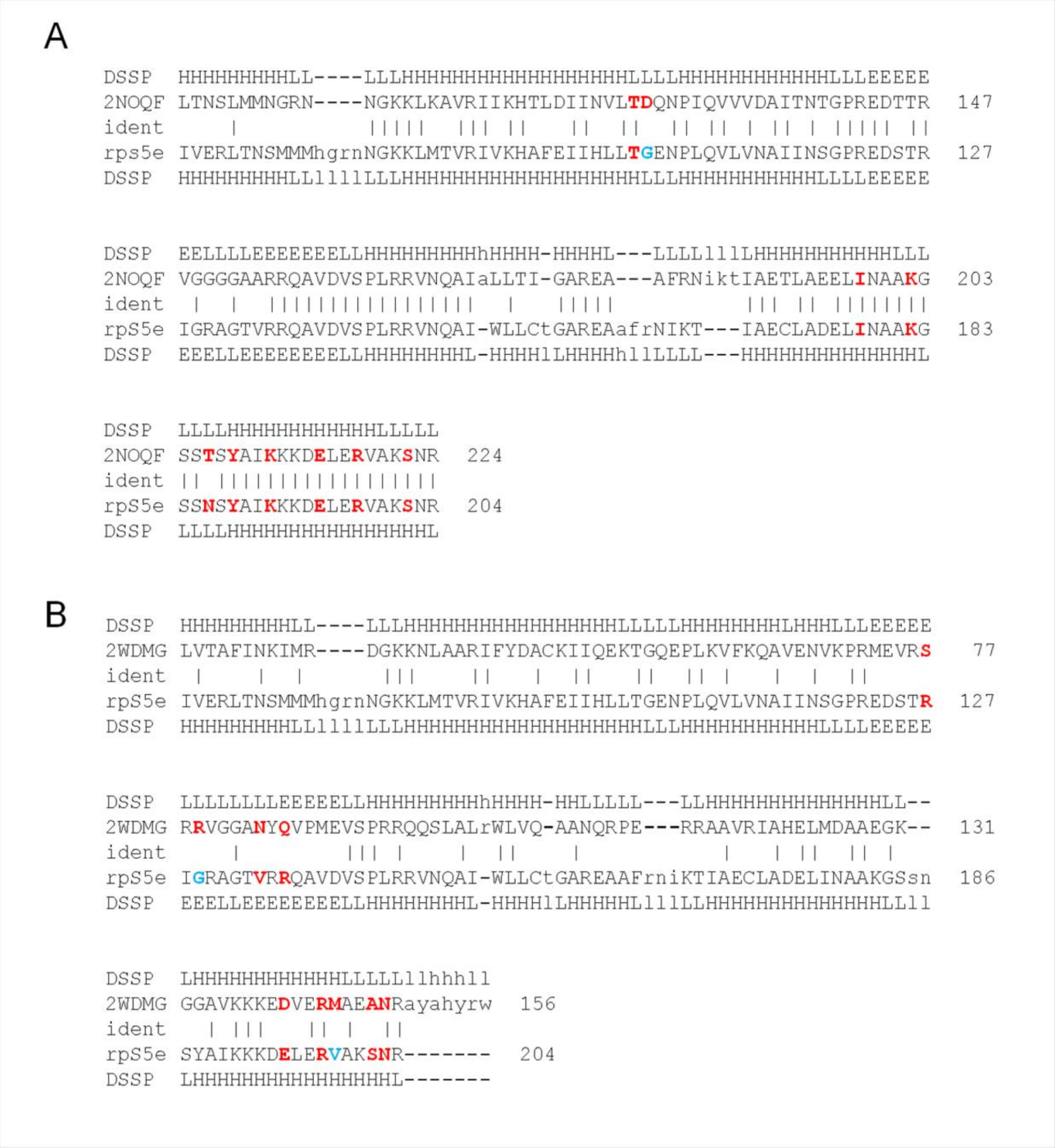
Re-analysis of cryoEM data on HCV IRES bound to 40S subunit of human ribosome integrated with recent structural information suggests new contact regions between ribosomal proteins and HCV RNA

2014; 11(7)
http://dx.doi.org/10.4161/rna.29545

www.landesbioscience.com/journals/rnabiology/article/29545/

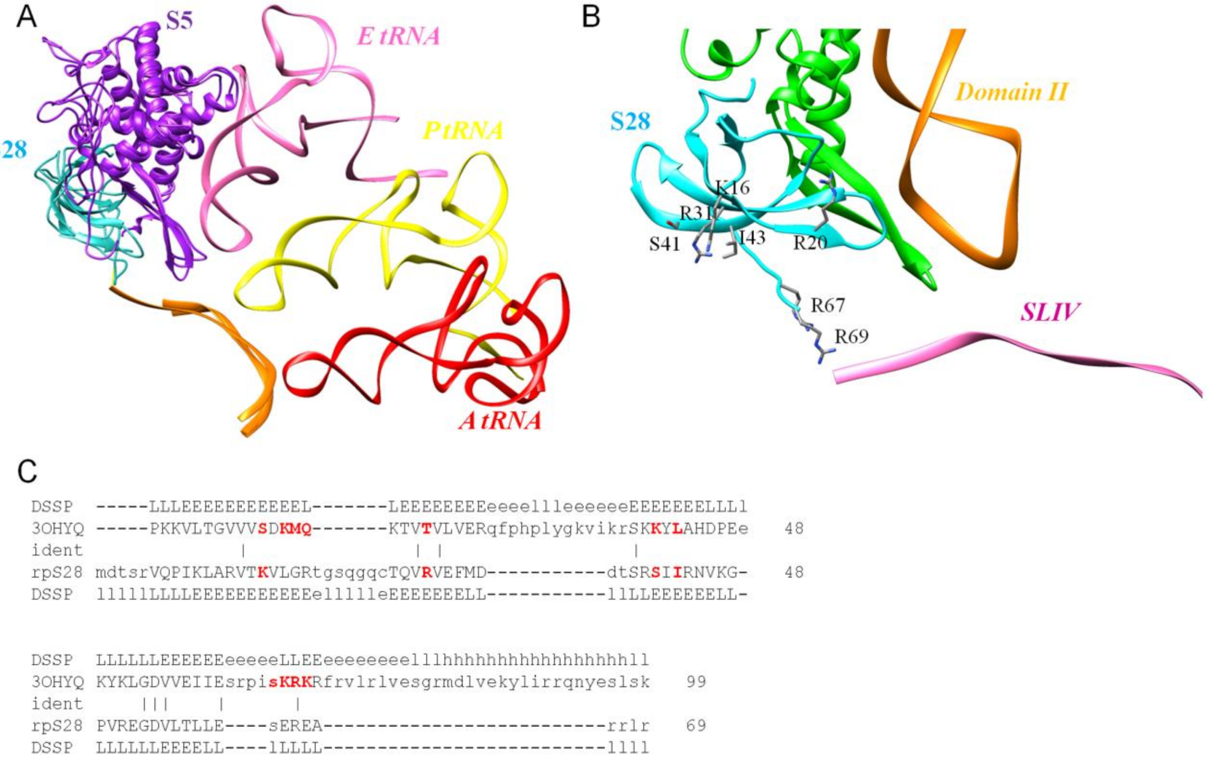


Supplementary Figure 1. *Interaction of HCV IRES domain II with ribosomal S5 and S25*. (A) The locations of S5 and S25 based on the fit of ribosomal head (light brown) [1], are shown. The fitted structure of domain II [2] relative to these proteins, is given. (B) Electrostatic potential at the interface with domain II is shown. The potentials are derived using Delphi [3] and surface is displayed in Chimera [4]. (C) Residue interactions at the interface between ribosomal protein S5 (purple) and CrPV IRES RNA (pink) based on a cryoEM derived model [5] (PDB ID: 2NOQ). (D) Interface residues of prokaryotic ribosomal protein S7 (S5p) (purple) [6] that bind E-site tRNA (pink).



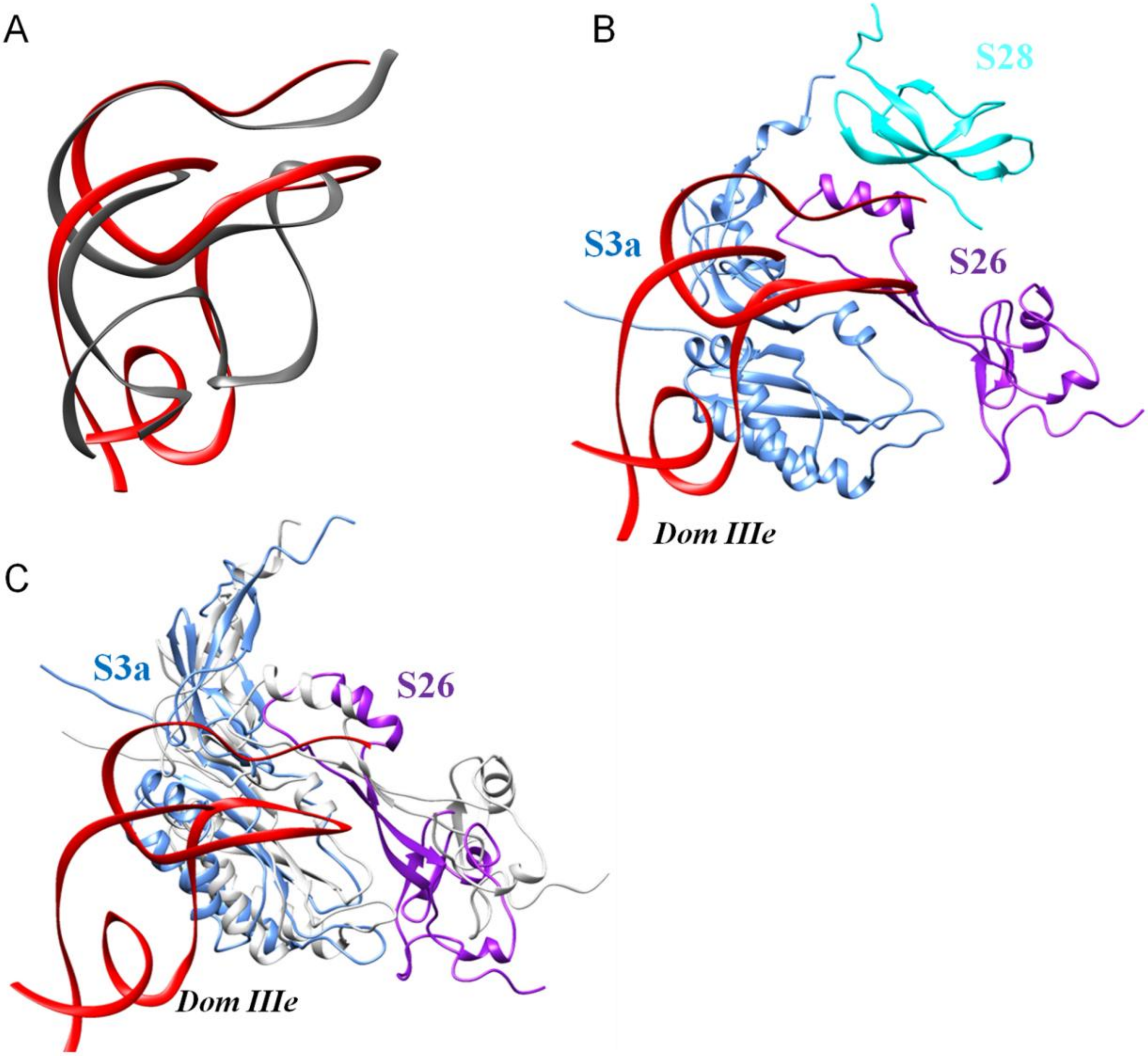
```
A

DSSP  HHHHHHHHHLL----LLLHHHHHHHHHHHHHHHHHHHLLLLHHHHHHHHHHHLLLEEEEE
2NOQF LTNSLMMNGRN----NGKKLKAVRIIKHTLDIINVLTDQNPIQVVVDAITNTGPREDTTR   147
ident     |          |||||  ||| ||   ||  ||  || || | || | ||||| ||
rps5e IVERLTNSMMMhgrnNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDSTR   127
DSSP  HHHHHHHHHLLllllLLLHHHHHHHHHHHHHHHHHHHHLLLHHHHHHHHHHLLLLEEEEE

DSSP  EELLLLEEEEEEEELLHHHHHHHHhHHHH-HHHHL---LLLLlllLHHHHHHHHHHHLLL
2NOQF VGGGGAARRQAVDVSPLRRVNQAIaLLTI-GAREA---AFRNiktIAETLAEELINAAKG   203
ident  |  |  |||||||||||||||||  |   |||||          ||| || ||||||||
rpS5e IGRAGTVRRQAVDVSPLRRVNQAI-WLLCtGAREAafrNIKT---IAECLADELINAAKG   183
DSSP  EEELLEEEEEEEEELLHHHHHHHL-HHHHlLHHHHhllLLLL---HHHHHHHHHHHHHHL

DSSP  LLLLHHHHHHHHHHHHLLLLL
2NOQF SSTSYAIKKKDELERVAKSNR   224
ident || ||||||||||||||||||
rpS5e SSNSYAIKKKDELERVAKSNR   204
DSSP  LLLLHHHHHHHHHHHHHHHHL

B

DSSP  HHHHHHHHHLL----LLLHHHHHHHHHHHHHHHHLLLLLHHHHHHHHHLHHHLLLEEEEE
2WDMG LVTAFINKIMR----DGKKNLAARIFYDACKIIQEKTGQEPLKVFKQAVENVKPRMEVRS    77
ident  |    |  |      |||    ||   |  ||   ||  || |   |  |  ||
rpS5e IVERLTNSMMMhgrnNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDSTR   127
DSSP  HHHHHHHHHLLllllLLLHHHHHHHHHHHHHHHHHHHHLLLHHHHHHHHHHLLLLEEEEE

DSSP  LLLLLLLEEEEEELLHHHHHHHHHhHHHH-HHLLLLL---LLHHHHHHHHHHHHHHLL--
2WDMG RRVGGANYQVPMEVSPRRQQSLALrWLVQ-AANQRPE---RRAAVRIAHELMDAAEGK--   131
ident     |        ||| |    |  ||    |           |   | ||  || |
rpS5e IGRAGTVRRQAVDVSPLRRVNQAI-WLLCtGAREAAFrniKTIAECLADELINAAKGSsn   186
DSSP  EEELLEEEEEEEEELLHHHHHHHL-HHHHlLHHHHHLlllLLHHHHHHHHHHHHHHLLll

DSSP  LHHHHHHHHHHHHLLLLLllhhhll
2WDMG GGAVKKKEDVERMAEANRayahyrw   156
ident   | |||   || |  ||
rpS5e SYAIKKKDELERVAKSNR-------   204
DSSP  LHHHHHHHHHHHHHHHHL-------
```

Supplementary Figure 2. (A) Structure based sequence alignment of human 40S S5 and yeast S5 in the CrPV bound form [5] (PDB ID : 2NOQ). Residues of 2NOQ, chain F (yeast S5) at the interface with CrPV IRES are highlighted in red and corresponding residues in human 40S S5 that were used as interaction restraints for docking, are also highlighted in red. Residues not added as restraints are in blue (B) Structure based sequence alignment of human 40S S5 and prokaryotic S7 (S5p) in the crystal structure of *Thermus thermophilus* ribosome [6] (PDB ID : 2WDM). Residues at the interface with E-site tRNA are highlighted in red and corresponding residues of human 40S S5 are also highlighted in red and blue, depending on whether or not they are used interaction restraints. Substitutions to Gly and residues not observed at the interface of model obtained by map density fit, are not used as restraints.

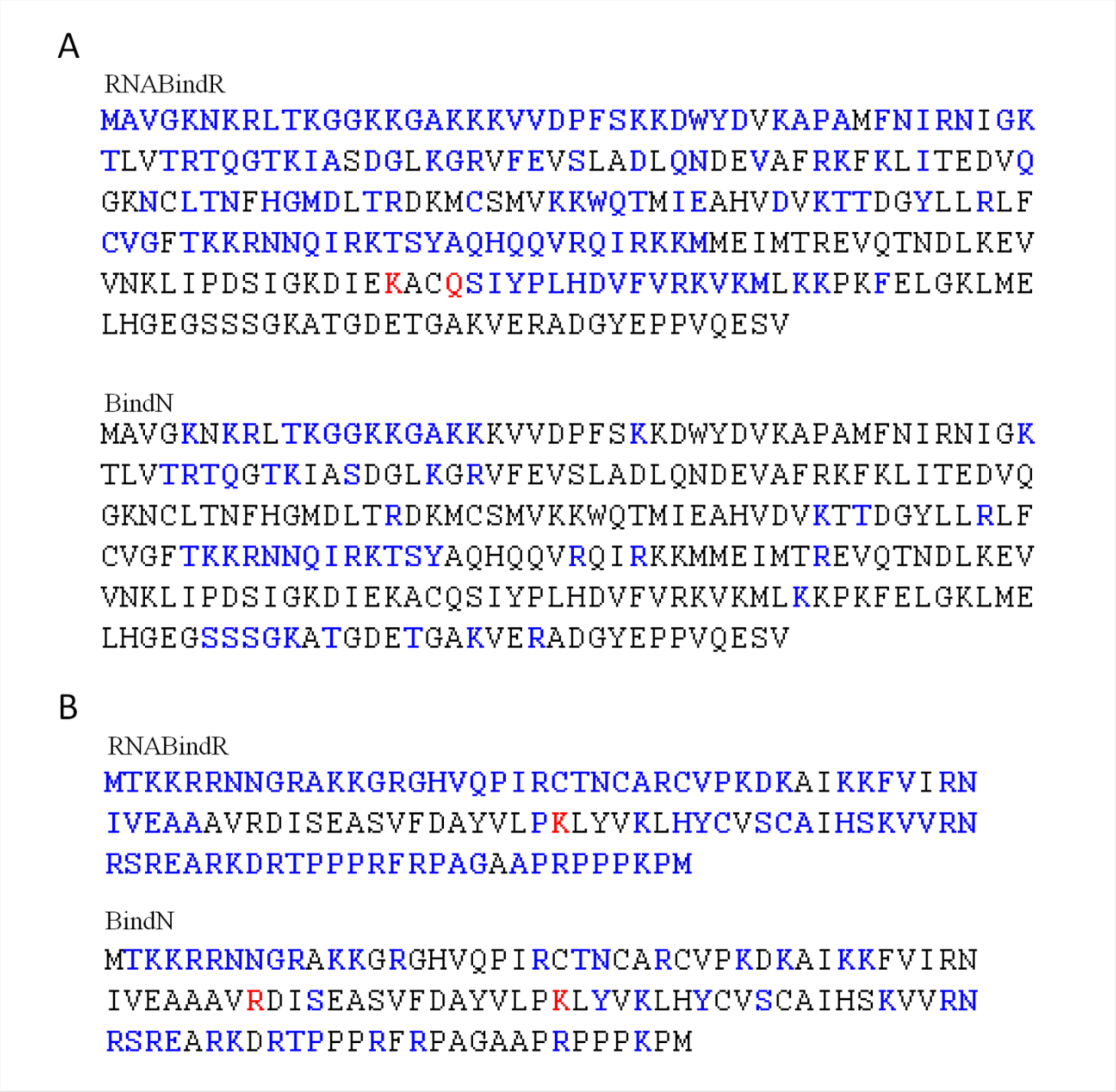


```
DSSP  -----LLLEEEEEEEEEEEL-------LEEEEEEEEeeeellleeeeeeEEEEEEELLLl
3OHYQ -----PKKVLTGVVVSDKMQ-------KTVTVLVERqfphplygkvikrSKKYLAHDPEe   48
ident              |               | |                 |
rpS28 mdtsrVQPIKLARVTKVLGRtgsqgqcTQVRVEFMD-----------dtSRSIIRNVKG-   48
DSSP  lllllLLLLEEEEEEEEEEEellllleEEEEEEELL-----------llLLEEEEEELL-

DSSP  LLLLLLEEEEEEeeeeeLLEEeeeeeeeeelllhhhhhhhhhhhhhhhhll
3OHYQ KYKLGDVVEIIEsrpisKRKRfrvlrlvesgrmdlvekylirrqnyeslsk   99
ident     |||    |      |
rpS28 PVREGDVLTLLE----sEREA--------------------------rrlr   69
DSSP  LLLLLLEEEELL----lLLLL--------------------------llll
```

Supplementary Figure 3. (A) Locations of A (red), P (yellow) and E (pink) site tRNAs relative to 40S ribosomal protein S5 (purple) and S28 (cyan). The figure shows the structural alignment of coordinates from Thermus thermophilus ribosome [6] (PDB ID : 2WDM), Triticum aestivum ribosome [7] (PDB ID: 3IZ7) and rabbit 40S [8] (PDB ID : 4KZZ). (B) Probable interface residues of ribosomal protein S28 involved in contact with domain IV of HCV IRES. The residues were identified by comparison with mRNA bound S28 structure [8] and 30S ribosomal protein S12 [9] – rRNA interactions (Figure 4). (C) Structure based sequence alignment of 30S ribosomal protein S12 (PDB ID : 3OHY) [9] and human ribosomal S28. Interface residues observed at S12 – rRNA interface are highlighted in red and equivalent residues likely to occur at the human 40S S28 – HCV IRES domain IV interface, are highlighted. Structurally equivalent residues are in uppercase.



Supplementary Figure 4. *Model of IRES RNA involving domains IIIef and pseudoknot.* (A) Structural alignment of the crystal structure of region of IRES RNA involving the pseudoknot and domains IIIef [10] (dark grey) with the model generated for IIIef+pseudoknot (red). (B) The relative orientations of S26 (purple), S28 (cyan) and S3a (blue) based on fitted co-ordinates of ribosomal head and body. The fitted model of IIIef+pseudoknot (red) is also shown. (C) Comparison of docked pose with the conformation prior to docking (grey).



A

RNABindR

MAVGKNKRLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMFNIRNIGK
TLVTRTQGTKIASDGLKGRVFEVSLADLQNDEVAFRKFKLITEDVQ
GKNCLTNFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLRLF
CVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEV
VNKLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLME
LHGEGSSSGKATGDETGAKVERADGYEPPVQESV

BindN

MAVGKNKRLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMFNIRNIGK
TLVTRTQGTKIASDGLKGRVFEVSLADLQNDEVAFRKFKLITEDVQ
GKNCLTNFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLRLF
CVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEV
VNKLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLME
LHGEGSSSGKATGDETGAKVERADGYEPPVQESV

B

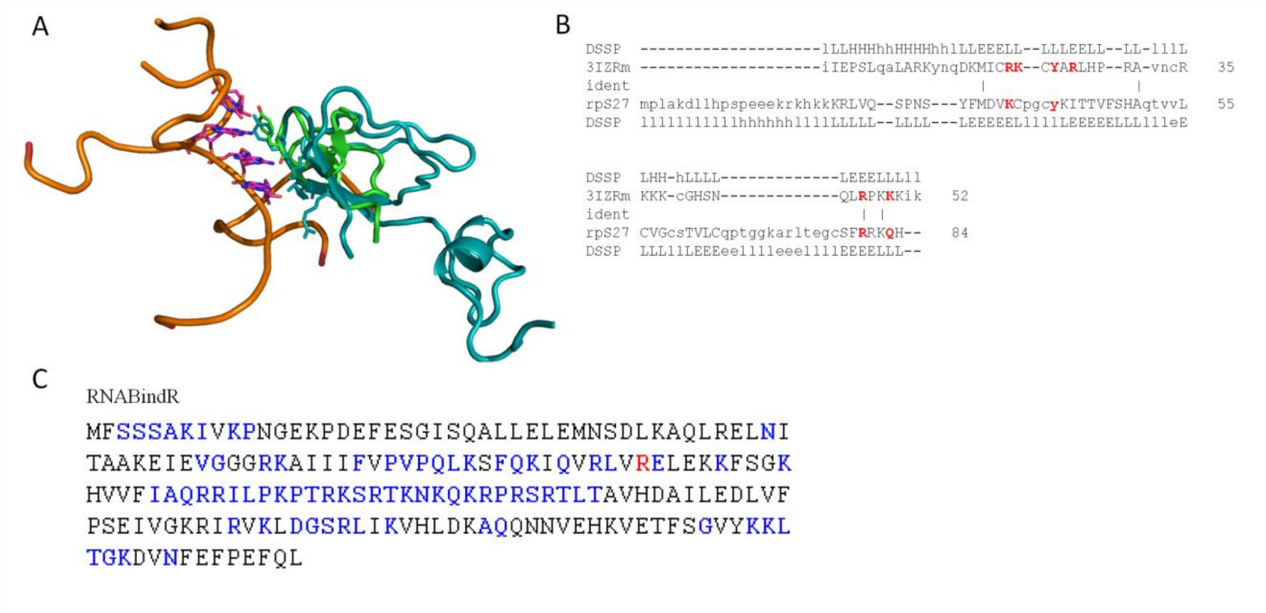
RNABindR

MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRN
IVEAAAVRDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRN
RSREARKDRTPPPRFRPAGAAPRPPPKPM

BindN

MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRN
IVEAAAVRDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRN
RSREARKDRTPPPRFRPAGAAPRPPPKPM

Supplementary Figure 5. *Interactions involving S26 and S3a.* Prediction of RNA binding residues for (A) S3a and (B) S26 using RNABindR [11] and BindN [12]. Positive predictions are in blue and residues occurring at interface with IIIef+pseudoknot model, are highlighted in red.

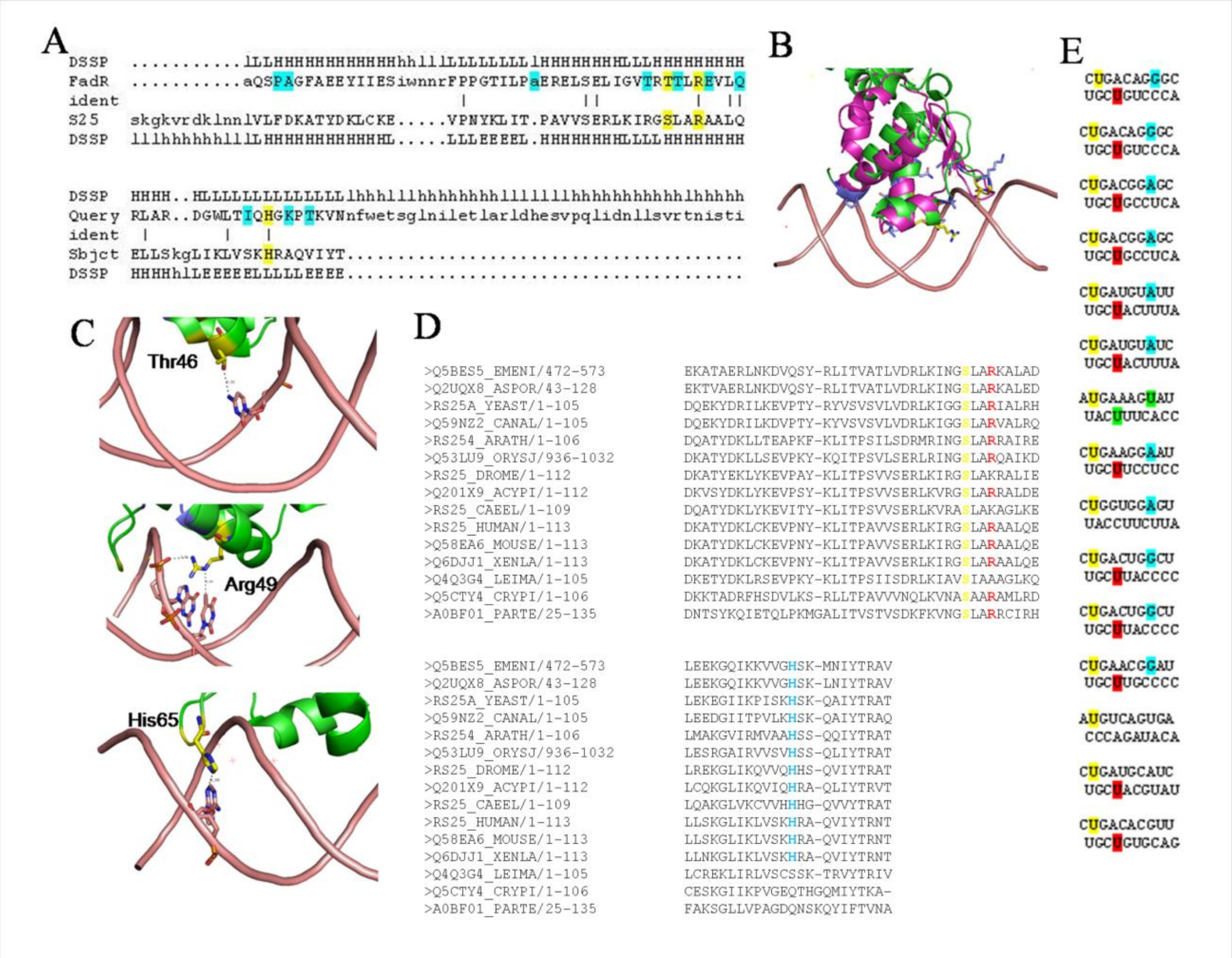


```
DSSP  --------------------lLLHHHhhHHHHhhlLLEEELL--LLLEELL--LL-lllL
3IZRm --------------------iIEPSLqaLARKynqDKMICRK--CYARLHP--RA-vncR    35
ident                                      |                |
rpS27 mplakdllhpspeeekrkhkkKRLVQ--SPNS---YFMDVKCpgcyKITTVFSHAqtvvL    55
DSSP  llllllllllllhhhhhhllllLLLLL--LLLL---LEEEEELllllLEEEEELLLllleE

DSSP  LHH-hLLLL-------------LEEELLLll
3IZRm KKK-cGHSN-------------QLRPKKKik   52
ident                         | |
rpS27 CVGcsTVLCqptggkarltegcSFRRKQH--   84
DSSP  LLLllLEEEeelllleeellllEEEELLL--
```

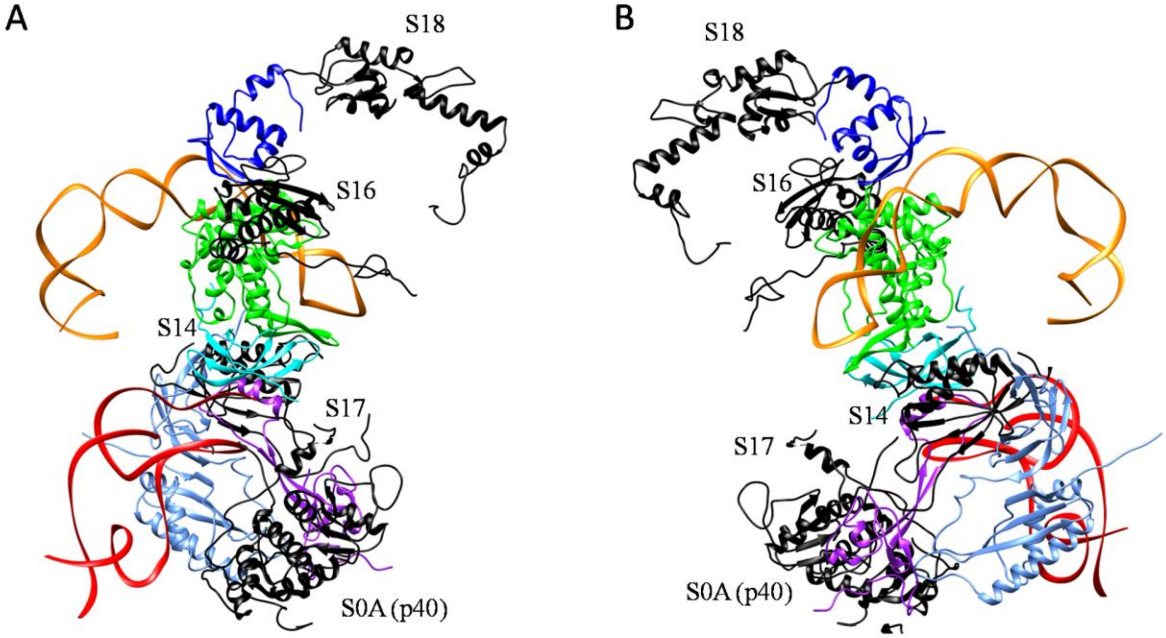
RNABindR

```
MFSSSSAKIVKPNGEKPDEFESGISQALLELEMNSDLKAQLRELNI
TAAKEIEVGGGRKAIIIFVPVPQLKSFQKIQVRLVRELEKKFSGK
HVVFIAQRRILPKPTRKSRTKNKQKRPRSRTLTAVHDAILEDLVF
PSEIVGKRIRVKLDGSRLIKVHLDKAQQNNVEHKVETFSGVYKKL
TGKDVNFEFPEFQL
```

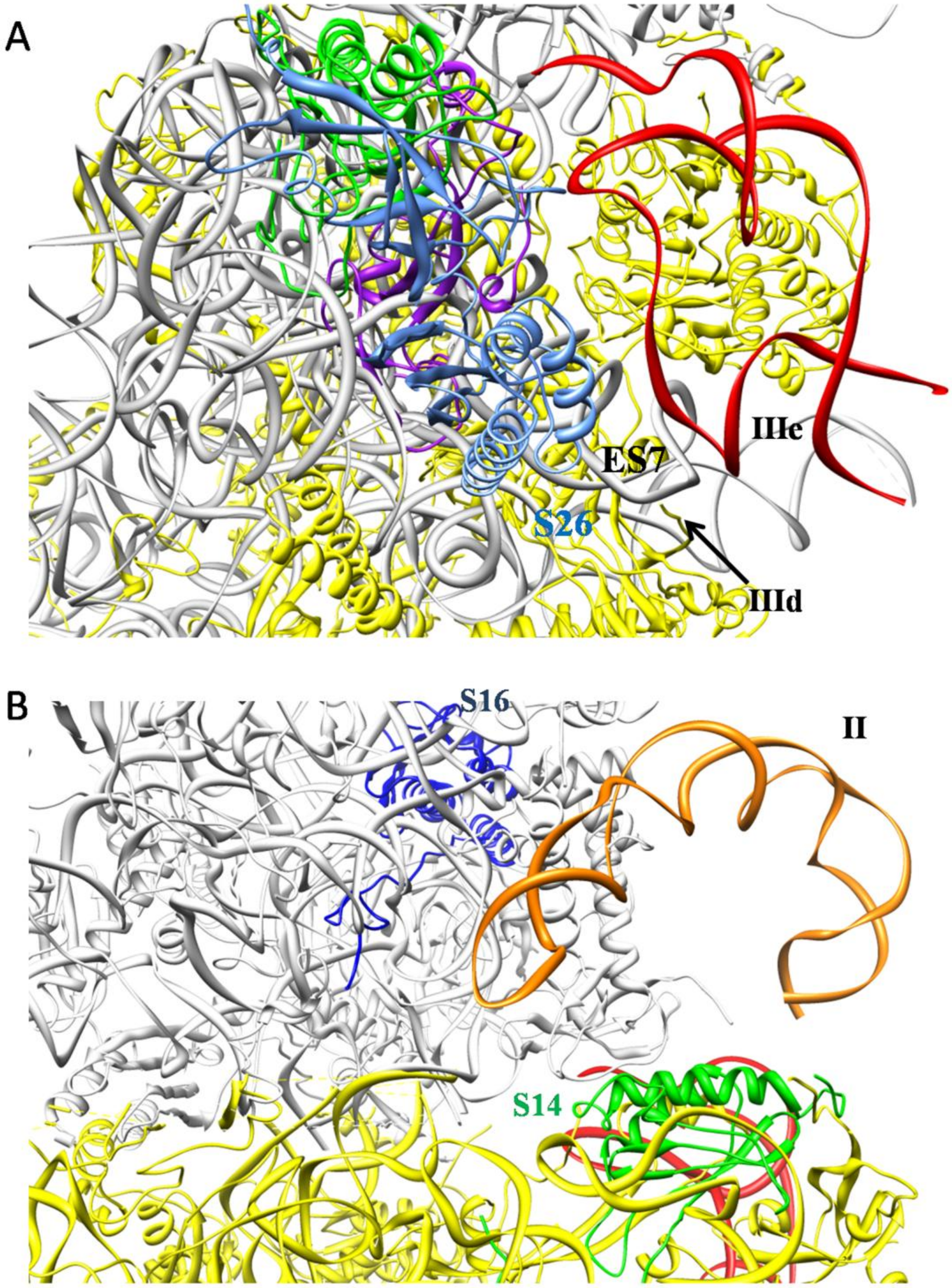
Supplementary Figure 6. *Contacts with S27*. (A) Structural alignment of human S27 (blue) and 60S ribosomal protein l40 [7] (green). The probable RNA contact is represented in the context of l40 (B) Structure based sequence alignment of human S27 with ribosomal l40 (PDB ID: 3IZR [7]). RNA binding residues of the l40 observed at the interface of interest (Interface of contact of S27 and IRES), are highlighted. Structurally equivalent residues are in uppercase. (C) 40S ribosomal protein S7 residues predicted to interact with RNA by RNABindR [11]. Positive predictions are in blue and residue highlighted in red is in the vicinity of the RNA bulge between jIIIabc and IIId.



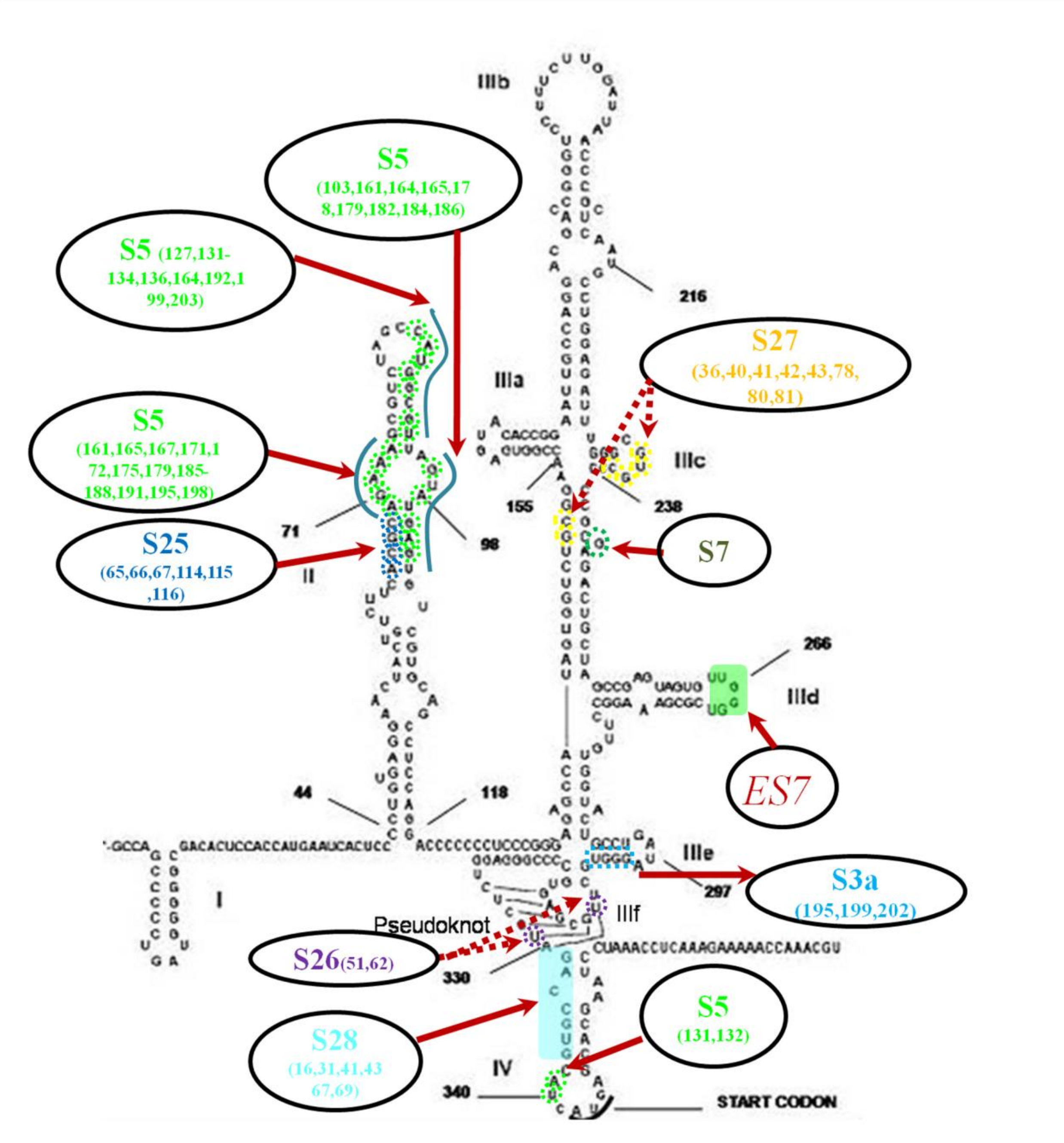
Supplementary Figure 7. (A) Structure based sequence alignment obtained by the comparison of structures of S25 and FadR transcription factor [13]. Structurally equivalent residues are in uppercase. (B) Structural superposition corresponding to the alignment (A). Structure of human ribosomal S25 is shown in pink, while the structure of DNA bound FadR is in green. (C) Three important base specific contacts found in FadR- DNA complex. These residues are also found in S25, at structurally equivalent positions in the previous alignment (A). (D) & (E) Multiple alignment of S25 sequences & 18S rRNA segments (1534-1544 & 1589-1599) from different species. The residues that take part in base specific contacts, are highlighted. Probable RNA bases to which these residues make contact, are also shown. These bases are chosen referring to the interactions observed in the FadR-DNA complex [13]. The amino acid residue of S25 and the rRNA base that it interacts with are highlighted in the same color. Non-preferred bases are highlighted in green.



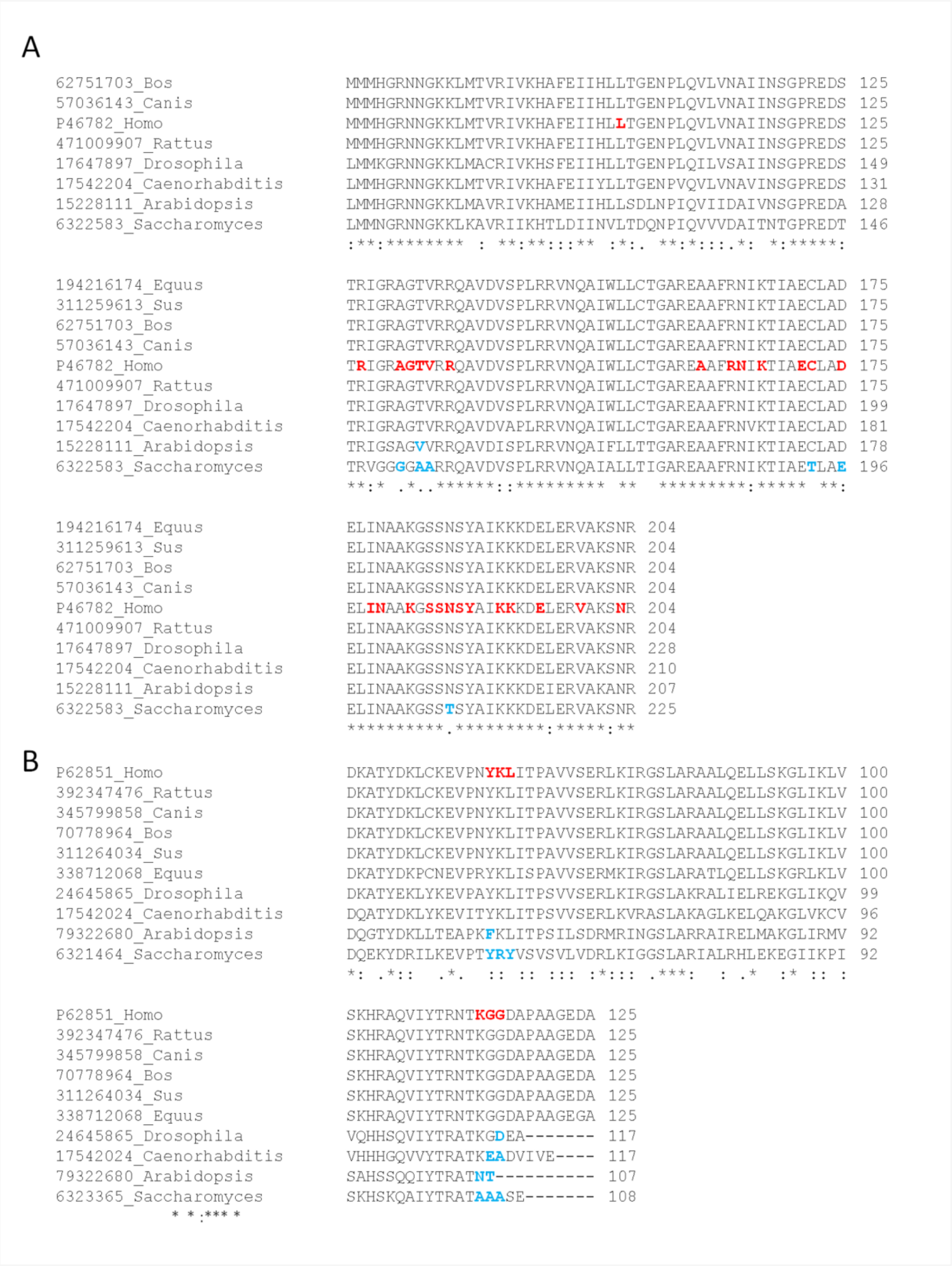
Supplementary Figure 8. *Proteins in the vicinity of HCV IRES*. Different 40S ribosomal proteins occurring in the vicinity of HCV IRES but not likely to participate in direct contacts, are indicated in black. (A) and (B) shows different views related by 180° rotation.



Supplementary Figure 9. *Proteins in the vicinity of domain II and IIIef+pseudoknot*. (A) Relative locations of proteins S3a (blue), S14 (green), expansion segment 7 (ES7), IRES domain IIId and IIIe are shown. (B) Relative locations of HVC IRES domain II, S16 (blue) and S14 (green).



Supplementary Figure 10. *Interactions involving 40S ribosomal proteins and HCV IRES RNA*. Residue-base interactions involving HCV IRES domains and human 40S ribosomal proteins S5, S25, S28, S26, S27 and S3a are shown. Residue numbers are given below the protein names and interacting bases are highlighted.



```
A

62751703_Bos                 MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS 125
57036143_Canis               MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS 125
P46782_Homo                  MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS 125
471009907_Rattus             MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS 125
17647897_Drosophila          LMMKGRNNGKKLMACRIVKHSFEIIHLLTGENPLQILVSAIINSGPREDS 149
17542204_Caenorhabditis      LMMHGRNNGKKLMTVRIVKHAFEIIYLLTGENPVQVLVNAVINSGPREDS 131
15228111_Arabidopsis         LMMHGRNNGKKLMAVRIVKHAMEIIHLLSDLNPIQVIIDAIVNSGPREDA 128
6322583_Saccharomyces        LMMNGRNNGKKLKAVRIIKHTLDIINVLTDQNPIQVVVDAITNTGPREDT 146
                             :**:******** : **:**:::** :*:. **:*:::.*: *:*****:

194216174_Equus              TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 175
311259613_Sus                TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 175
62751703_Bos                 TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 175
57036143_Canis               TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 175
P46782_Homo                  TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 175
471009907_Rattus             TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 175
17647897_Drosophila          TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD 199
17542204_Caenorhabditis      TRIGRAGTVRRQAVDVAPLRRVNQAIWLLCTGAREAAFRNVKTIAECLAD 181
15228111_Arabidopsis         TRIGSAGVVRRQAVDISPLRRVNQAIFLLTTGAREAAFRNIKTIAECLAD 178
6322583_Saccharomyces        TRVGGGGAARRQAVDVSPLRRVNQAIALLTIGAREAAFRNIKTIAETLAE 196
                             **:* .*..******::********* **  *********:***** **:

194216174_Equus              ELINAAKGSSNSYAIKKKDELERVAKSNR 204
311259613_Sus                ELINAAKGSSNSYAIKKKDELERVAKSNR 204
62751703_Bos                 ELINAAKGSSNSYAIKKKDELERVAKSNR 204
57036143_Canis               ELINAAKGSSNSYAIKKKDELERVAKSNR 204
P46782_Homo                  ELINAAKGSSNSYAIKKKDELERVAKSNR 204
471009907_Rattus             ELINAAKGSSNSYAIKKKDELERVAKSNR 204
17647897_Drosophila          ELINAAKGSSNSYAIKKKDELERVAKSNR 228
17542204_Caenorhabditis      ELINAAKGSSNSYAIKKKDELERVAKSNR 210
15228111_Arabidopsis         ELINAAKGSSNSYAIKKKDEIERVAKANR 207
6322583_Saccharomyces        ELINAAKGSSTSYAIKKKDELERVAKSNR 225
                             **********.*********:*****:**
B
P62851_Homo                  DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV 100
392347476_Rattus             DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV 100
345799858_Canis              DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV 100
70778964_Bos                 DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV 100
311264034_Sus                DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV 100
338712068_Equus              DKATYDKPCNEVPRYKLISPAVVSERMKIRGSLARATLQELLSKGRLKLV 100
24645865_Drosophila          DKATYEKLYKEVPAYKLITPSVVSERLKIRGSLAKRALIELREKGLIKQV 99
17542024_Caenorhabditis      DQATYDKLYKEVITYKLITPSVVSERLKVRASLAKAGLKELQAKGLVKCV 96
79322680_Arabidopsis         DQGTYDKLLTEAPKFKLITPSILSDRMRINGSLARRAIRELMAKGLIRMV 92
6321464_Saccharomyces        DQEKYDRILKEVPTYRYVSVSVLVDRLKIGGSLARIALRHLEKEGIIKPI 92
                             *: .*::  .*.  :: :: ::: :*::: .***:  : .*  :* :: :

P62851_Homo                  SKHRAQVIYTRNTKGGDAPAAGEDA 125
392347476_Rattus             SKHRAQVIYTRNTKGGDAPAAGEDA 125
345799858_Canis              SKHRAQVIYTRNTKGGDAPAAGEDA 125
70778964_Bos                 SKHRAQVIYTRNTKGGDAPAAGEDA 125
311264034_Sus                SKHRAQVIYTRNTKGGDAPAAGEDA 125
338712068_Equus              SKHRAQVIYTRNTKGGDAPAAGEGA 125
24645865_Drosophila          VQHHSQVIYTRATKGDEA------- 117
17542024_Caenorhabditis      VHHHGQVVYTRATKEADVIVE---- 117
79322680_Arabidopsis         SAHSSQQIYTRATNT---------- 107
6323365_Saccharomyces        SKHSKQAIYTRATAAASE------- 108
            * *:*** *
```

Supplementary Figure 11. *Conservation of interface residues of S5 ad S25*. (A) and (B) highlights (red) interface residues of S5 and S25 that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

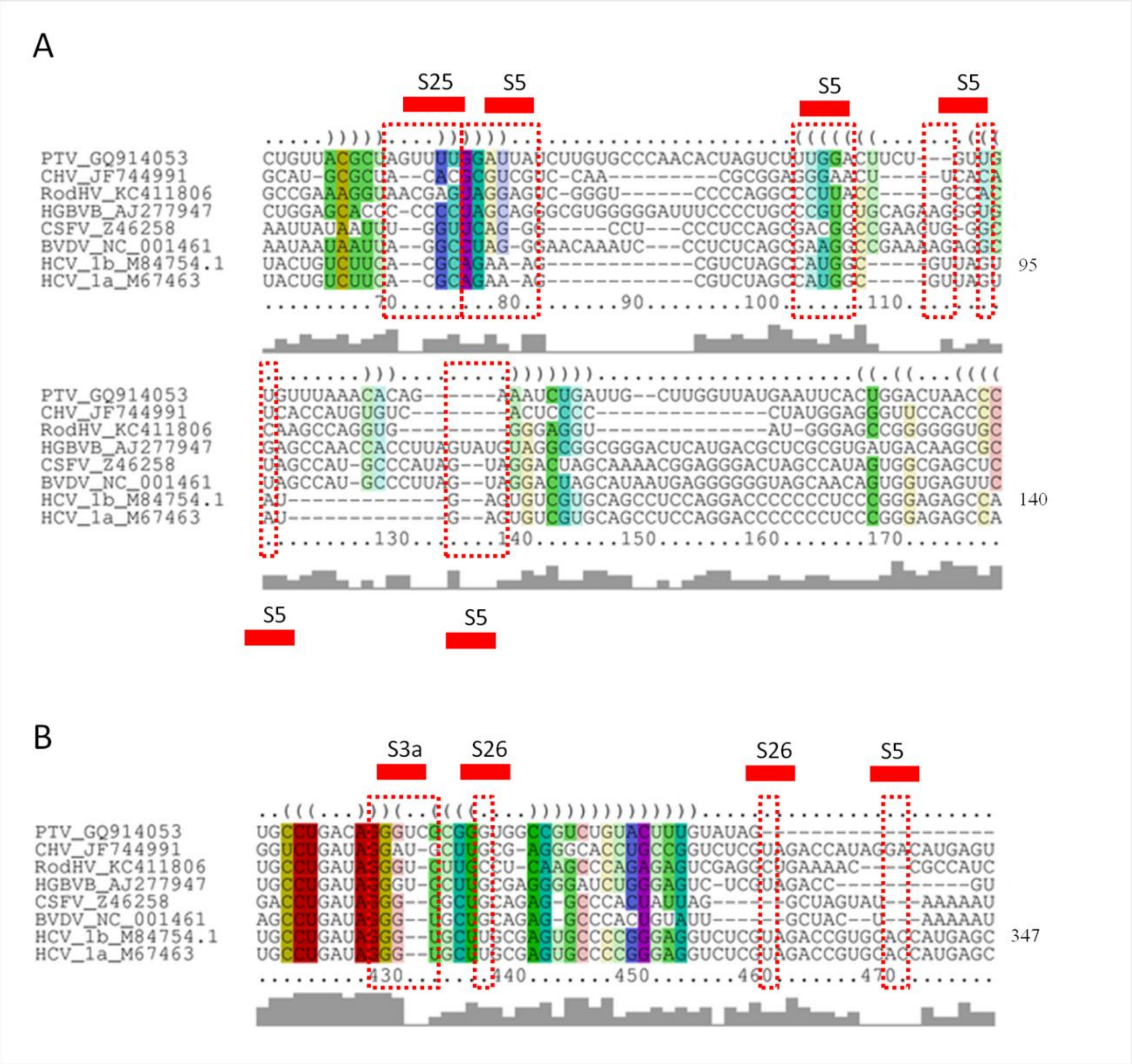
```
15228895_Arabidopsis          RDVQEASVYEGYTLPKLYAKTQYCVSCAIHSHVVRVRSRTNRRVRTPP--  98
17508699_Caenorhabditis       RDIGDASAYTQYALPKLYHKLHYCIACAIHSKVVRNRSREARRDRNPP--  98
P62854_Homo                   RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--  98
6981488_Rattus                RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--  98
57092393_Canis                RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--  98
62751960_Bos                  RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--  98
255522917_Equus               RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--  98
147899011_Sus                 RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--  98
17647893_Drosophila           RDITEASIWDSYVLPKLYAKLHYCVSCAIHSKVVRNRSREARRIRTPP--  98
6320978_Saccharomyces         RDLSEASVYPEYALPKTYNKLHYCVSCAIHARIVRVRSREDRKNRAPPQR
100
                              **: :** :  *.*** * * :**::****:::** ***  *: * **
```

```
P61247_Homo_sapiens           KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS 236
545210000_Equus               KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS 347
296470536_Bos                 KLIPDSIGKDIEKACQSIYPFHDVFVRKVKMLKKPKFELGKLMELHGEGS 236
8394221_Rattus                KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGG 236
545557812_Canis               KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS 236
212549625_Sus                 KLIPDSIGKDIEKACQPIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS 236
17864162_Drosophila           KLALDSIAKDIEKSCQRIYPLHDVYIRKVKVLKKPRFDVSKLLELHGDGG 239
17553700_Caenorhabditis       KLIPDSIGKDIEKTCSKLYPLQEVYIRKVKIIKRPKVDLGRLHDLHGDSI 234
15236171_Arabidopsis          KFIPEAIGREIEKATQGIYPLQNVFIRKVKILKAPKFDLGKLMEVHGDYT 236
6323474_Saccharomyces         KLIPEVINKEIENATKDIFPLQNIHVRKVKLLKQPKFDVGALMALHGEGS 236
                              *:  : * ::**:: . ::*::::.:****::* *:.::. *  :**:
```

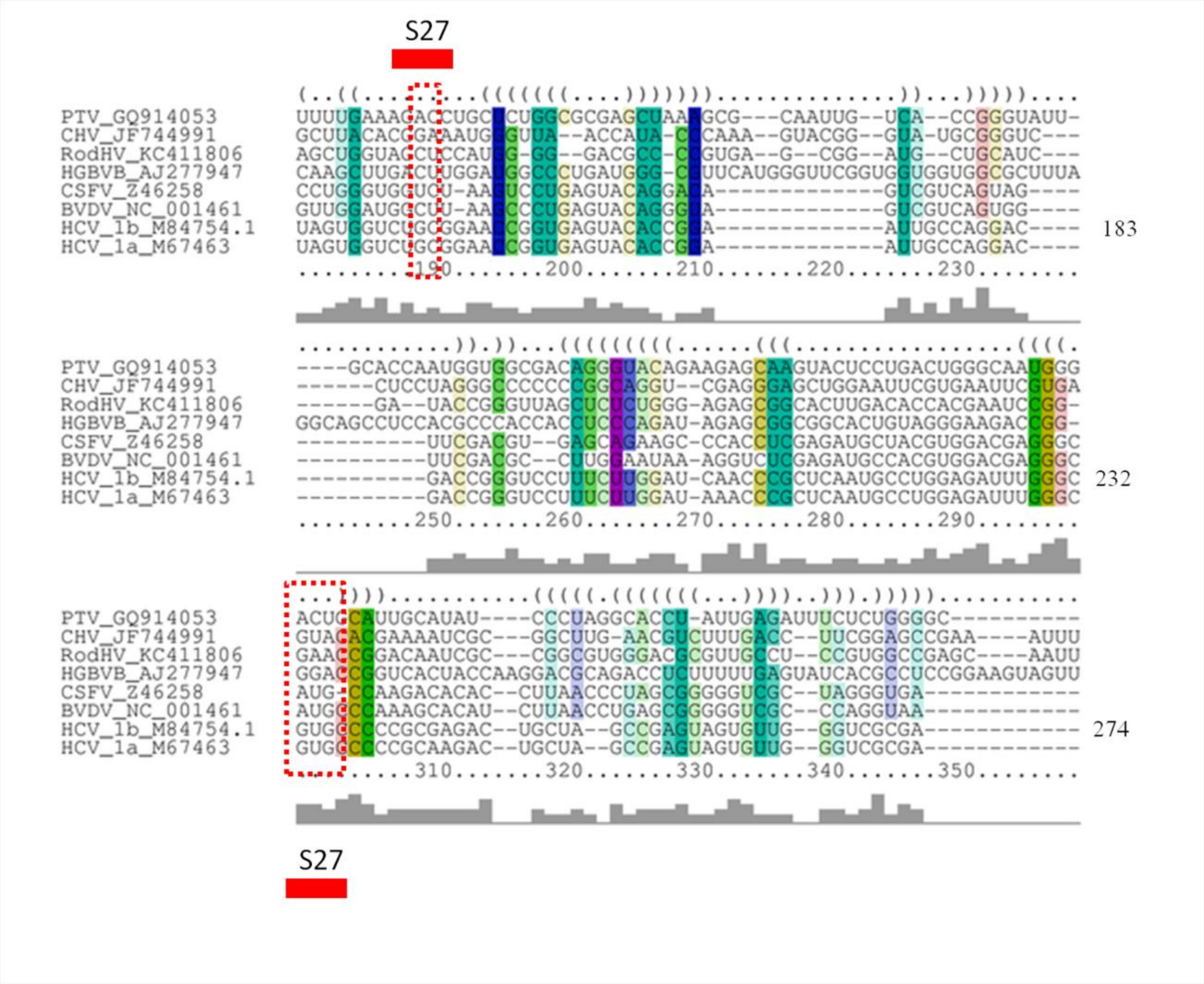
Supplementary Figure 12. *Conservation of interface residues of S26 ad S3a*. (A) and (B) highlights (red) interface residues of S26 and S3a that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

```
545519486_Canis             MPLA--KDLLHPSAEEEKRKHKKQRLVQSPNSYFMDVKCPGCYKITTVFS 48
94966948_Bos                MPLA--RDLLHPSLDEEKKKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS 48
P42677_Homo                 MPLA--KDLLHPSPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS 48
16758382_Rattus             MPLA--KDLLHPSPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS 48
545216788_Equus             MPLA--KDLLHPSPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS 48
335302490_Sus               MPLA--KDLLHSSPEEEKRKHKK-YLVQSPNSYFMDVKCPGCYKITTVFS 47
24649976_Drosophila         MPLA--KDLLHPLPAEEKRKHKLKRLVQHPNSYFMDVKCPGCYRITTVFS 48
17563244_Caenorhabditis     MPLA--VDLLHPEPQREIRCHKLKRLVQHPNSYFMDVKCSGCFKISTVFS 48
15238845_Arabidopsis        MVLQNDIDLLHPPPELEKRKHKLKRLVQSPNSFFMDVKCQGCFNITTVFS 50
6321809_Saccharomyces       MVLV--QDLLHPTAASEARKHKLKTLVQGPRSYFLDVKCPGCLNITTVFS 48
                            * *    ****.    * : **   *** *.*:*:**** ** .*:****

545519486_Canis             HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRQKQH 84
94966948_Bos                HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH 84
P42677_Homo                 HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH 84
16758382_Rattus             HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH 84
545216788_Equus             HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH 84
335302490_Sus               HAQTVVLYVGCSAVLCQPTGGKARLTEGCSFRRKQR 83
24649976_Drosophila         HAQGVVVCAGCATILCQPTGGRAKLTEGCSFRRKPQ 84
17563244_Caenorhabditis     HATTVVVCVGCNTVLCQPTRGKAKLTEGCSFRKKQ- 83
15238845_Arabidopsis        HSQTVVVCGNCQTVLCQPTGGKARLQEGCSFRKK-- 84
6321809_Saccharomyces       HAQTAVTCESCSTVLCTPTGGKAKLSEGTSFRRK-- 82
                            *:  .*   .* ::** ** *:*:* ** ***:*
```

Supplementary Figure 13. *Conservation of interface residues of S27*. Interface residues of S27 (red) that contact IRES RNA are highlighted. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.



Supplementary Figure 14. Conservation of IRES RNA bases. Bases that occur at the interface with ribosomal proteins S5 & S25 are highlighted in (A) while those interacting with S3a & S26 are in (B). The figure shows multiple alignment of IRES RNA sequences from Porcine teschovirus (PTV), Canine herpesvirus (CHV), Rodent herpesvirus (RodHV), Hepatitis GBVB (HGBVB), Classical swine fever virus (CSFV), Bovine viral diarrhea virus (BVDV) and HCV IRES subtypes. Alignment was carried out using LocaRNA [14] and the coloring reflects compatibility for base-pairing. A consensus secondary structure is also provided on top and extent of conservation is indicated below the alignment. The GI accessions are given for each sequence



Supplementary Figure 15. Conservation of IRES RNA bases. Bases that occur at the interface with ribosomal proteins S27 are highlighted. The figure shows multiple alignment of IRES RNA sequences from Porcine teschovirus (PTV), Canine herpesvirus (CHV), Rodent herpesvirus (RodHV), Hepatitis GBVB (HGBVB), Classical swine fever virus (CSFV), Bovine viral diarrhea virus (BVDV) and HCV IRES subtypes. Alignment was carried out using LocaRNA [14] and the coloring reflects compatibility for base-pairing. A consensus secondary structure is also provided on top and extent of conservation is indicated below the alignment. The GI accessions are given for each sequence

Supplementary Table 1. *Structural relatives with nucleic acid binding properties.* For the ribosomal proteins S28, S7 and S27, the *rmsd* and Z-scores of the folds identified as structural relatives, are listed. DALI [15] was used for doing the structure comparison and search.

Ribosomal Protein	**Related Fold (PDB code)**	***rmsd***	***Z score***
S28	3OHY	3.4 Å	3.5
S27	3IZR	2.0 Å	2.6

Supplementary Table 2. *Dope Score [16] of models selected for human ribosomal 40S proteins.*

Ribosomal Protein	**Dope Score**
S5	-17590.8
S25	-7006.4
S26	-5486.6
S28	-4346.3
S3a	-18404.8
S27	-5264.0
S7	-17603.2

Supplementary data 1: Structure based sequence alignments

1. Structure based sequence alignment of human 40S S5 and yeast S5 in the CrPV bound form 5 (PDB ID : 2NOQ). Residues of 2NOQ, chain F (yeast S5) at the interface with CrPV IRES are highlighted in red and corresponding residues in human 40S S5 that were used as interaction restraints for docking, are also highlighted in red. Residues not added as restraints are in blue.

```
>2NOQF
-------------------------------------------rYANKRFRKAqcpiieR
LTNSLMMNGRN----NGKKLKAVRIIKHTLDIINVLTDQNPIQVVVDAITNTGPREDTTR
VGGGGAARRQAVDVSPLRRVNQAIaLLTI-GAREA---AFRNiktIAETLAEELINAAKG
SSTSYAIKKKDELERVAKSNR

>rps5e
tpdiklfgkwstddvqindislqdyiavkekyakylphsagryaAKRFRKAQC------P
IVERLTNSMMMhgrnNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDSTR
IGRAGTVRRQAVDVSPLRRVNQAI-WLLCtGAREAafrNIKT---IAECLADELINAAKG
SSNSYAIKKKDELERVAKSNR
```

2. Structure based sequence alignment of human 40S S5 and prokaryotic S7 (S5p) in the crystal structure of *Thermus thermophilus* ribosome (PDB ID : 2WDM). Residues at the interface with E-site tRNA are highlighted in red and corresponding residues of human 40S S5 are also highlighted in red and blue, depending on whether or not they are used interaction restraints. Substitutions to Gly and residues not observed at the interface of model obtained by map density fit, are not used as restraints.

```
>2WDMG
----------------------------------------arrrrAEVRQLQPdlvygdV
LVTAFINKIMR----DGKKNLAARIFYDACKIIQEKTGQEPLKVFKQAVENVKPRMEVRS
RRVGGANYQVPMEVSPRRQQSLALrWLVQ-AANQRPE---RRAAVRIAHELMDAAEGK--
GGAVKKKEDVERMAEANRayahyrw

>rpS5e
tpdiklfgkwstddvqindislqdyiavkekyakylphsagryaaKRFRKAQC------P
IVERLTNSMMMhgrnNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDSTR
IGRAGTVRRQAVDVSPLRRVNQAI-WLLCtGAREAAFrniKTIAECLADELINAAKGSsn
SYAIKKKDELERVAKSNR-------
```

3. Structure based sequence alignment of 30S ribosomal protein S12 (PDB ID : 3OHY) and human ribosomal S28. Interface residues observed at S12 – rRNA interface are highlighted in red and equivalent residues likely to occur at the human 40S S28 – HCV IRES domain IV interface, are highlighted. Structurally equivalent residues are in uppercase.

```
>3OHYQ
-----PKKVLTGVVVSDKMQ-------KTVTVLVERqfphplygkvikrSKKYLAHDPEe
KYKLGDVVEIIEsrpisKRKRfrvlrlvesgrmdlvekylirrqnyeslsk

>rpS28
```

```
mdtsrVQPIKLARVTKVLGRtgsqgqcTQVRVEFMD-----------dtSRSIIRNVKG-
PVREGDVLTLLE----sEREA--------------------------rrlr
```

4. Structure based sequence alignment of human S27 with ribosomal l40 (PDB ID: 3IZR). RNA binding residues of the l40 observed at the interface of interest (Interface of contact of S27 and IRES), are highlighted. Structurally equivalent residues are in uppercase.

```
>3IZRm
--------------------iIEPSLqaLARKynqDKMICRK--CYARLHP--RA-vncR
KKK-cGHSN-------------QLRPKKKik
>rpS27
mplakdllhpspeeekrkhkkKRLVQ--SPNS---YFMDVKCpgcyKITTVFSHAqtvvL   55
CVGcsTVLCqptggkarltegcSFRRKQH--
```

5. Structure based sequence alignment obtained by the comparison of structures of S25 and FadR transcription factor. Structurally equivalent residues are in uppercase.

```
>1H9Ta
-----------aQSPAGFAEEYIIESiwnnrFPPGTILPaERELSELIGVTRTTLREVLQ
RLAR--DGWLTIQHGKPTKVNnfwetsglniletlarldhesvpqlidvllsvrtnisti

>rpS27
skgkvrdklnnlVLFDKATYDKLCKE-----VPNYKLIT-PAVVSERLKIRGSLARAALQ
ELLSKgLIKLVSKHRAQVIYT---------------------------------------
```

6. Multiple alignment of S25 sequences highlighting conserved residues involved base specific contacts with 18S rRNA interaction.

```
>Q5BES5_EMENI/472-573
EKATAERLNKDVQSY-RLITVATLVDRLKINGSLARKALADLEEKGQIKKVVGHSK-MNIYTRAV
>Q2UQX8_ASPOR/43-128
EKTVAERLNKDVQSY-RLITVATLVDRLKINGSLARKALEDLEEKGQIKKVVGHSK-LNIYTRAV
>RS25A_YEAST/1-105
DQEKYDRILKEVPTY-RYVSVSVLVDRLKIGGSLARIALRHLEKEGIIKPISKHSK-QAIYTRAT
>Q59NZ2_CANAL/1-105
DQEKYDRILKDVPTY-KYVSVSVLVDRLKIGGSLARVALRQLEEDGIITPVLKHSK-QAIYTRAQ
>RS254_ARATH/1-106
DQATYDKLLTEAPKF-KLITPSILSDRMRINGSLARRAIRELMAKGVIRMVAAHSS-QQIYTRAT
>Q53LU9_ORYSJ/936-1032
DKATYDKLLSEVPKY-KQITPSVLSERLRINGSLARQAIKDLESRGAIRVVSVHSS-QLIYTRAT
>RS25_DROME/1-112
DKATYEKLYKEVPAY-KLITPSVVSERLKIRGSLAKRALIELREKGLIKQVVQHHS-QVIYTRAT
>Q201X9_ACYPI/1-112
DKVSYDKLYKEVPSY-KLITPSVVSERLKVRGSLARRALDELCQKGLIKQVIQHRA-QLIYTRVT
>RS25_CAEEL/1-109
DQATYDKLYKEVITY-KLITPSVVSERLKVRASLAKAGLKELQAKGLVKCVVHHHG-QVVYTRAT
>RS25_HUMAN/1-113
DKATYDKLCKEVPNY-KLITPAVVSERLKIRGSLARAALQELLSKGLIKLVSKHRA-QVIYTRNT
```

```
>Q58EA6_MOUSE/1-113
DKATYDKLCKEVPNY-KLITPAVVSERLKIRGSLARAALQELLSKGLIKLVSKHRA-QVIYTRNT
>Q6DJJ1_XENLA/1-113
DKATYDKLCKEVPNY-KLITPAVVSERLKIRGSLARAALQELLNKGLIKLVSKHRA-QVIYTRNT
>Q4Q3G4_LEIMA/1-105
DKETYDKLRSEVPKY-KLITPSIISDRLKIAVSIAAAGLKQLCREKLIRLVSCSSK-TRVYTRIV
>Q5CTY4_CRYPI/1-106
DKKTADRFHSDVLKS-RLLTPAVVVNQLKVNASAARAMLRDCESKGIIKPVGEQTHGQMIYTKA-
>A0BF01_PARTE/25-135
DNTSYKQIETQLPKMGALITVSTVSDKFKVNGSLARRCIRHFAKSGLLVPAGDQNSKQYIFTVNA
```

7. The alignment highlights (red) interface residues of S5 that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

```
 >194216174_Equus
------------------------MTEWETAAPAVAETPDIKLFGKWSTD
DVQINDISLQDYIAVK-EKYAKYLPHSAGRYAAKRFRKAQCPIVERLTNS
MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>311259613_Sus
------------------------MTEWETAAPAVAETPDIKLFGKWSTD
DVQINDISLQDYIAVK-EKYAKYLPHSAGRYAAKRFRKAQCPIVERLTNS
MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>62751703_Bos
------------------------MTEWETAAPAVAETPDIKLFGKWSTD
DVQINDISLQDYIAVK-EKYAKYLPHSAGRYAAKRFRKAQCPIVERLTNS
MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>57036143_Canis
------------------------MTEWETAAPAVAETPDIKLFGKWSTD
DVQINDISLQDYIAVK-EKYAKYLPHSAGRYAAKRFRKAQCPIVERLTNS
MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>P46782_Homo
------------------------MTEWETAAPAVAETPDIKLFGKWSTD
DVQINDISLQDYIAVK-EKYAKYLPHSAGRYAAKRFRKAQCPIVERLTNS
MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>471009907_Rattus
------------------------MTEWETATPAVAETPDIKLFGKWSTD
DVQINDISLQDYIAVK-EKYAKYLPHSAGRYAAKRFRKAQCPIVERLTNS
MMMHGRNNGKKLMTVRIVKHAFEIIHLLTGENPLQVLVNAIINSGPREDS
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>17647897_Drosophila
MAEVAENVVETFEEPAAPMEAEVAETILETNVVSTTELPEIKLFGRWSCD
DVTVNDISLQDYISVK-EKFARYLPHSAGRYAAKRFRKAQCPIVERLTCS
LMMKGRNNGKKLMACRIVKHSFEIIHLLTGENPLQILVSAIINSGPREDS
```

```
TRIGRAGTVRRQAVDVSPLRRVNQAIWLLCTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>17542204_Caenorhabditis
------------------MADNWGSENVVADAAPATEAPEVALFGKWSLQ
SVNVSDISLVDYIPVK-EKSAKYLPHSAGRFQVRRFRKAACPIVERLANS
LMMHGRNNGKKLMTVRIVKHAFEIIYLLTGENPVQVLVNAVINSGPREDS
TRIGRAGTVRRQAVDVAPLRRVNQAIWLLCTGAREAAFRNVKTIAECLAD
ELINAAKGSSNSYAIKKKDELERVAKSNR
>15228111_Arabidopsis
----------------------MAASAEIDAEIQQQLTNEVKLFNRWSFD
DVSVTDISLVDYIGVQPSKHATFVPHTAGRYSVKRFRKAQCPIVERLTNS
LMMHGRNNGKKLMAVRIVKHAMEIIHLLSDLNPIQVIIDAIVNSGPREDA
TRIGSAGVVRRQAVDISPLRRVNQAIFLLTTGAREAAFRNIKTIAECLAD
ELINAAKGSSNSYAIKKKDEIERVAKANR
>6322583_Saccharomyces        MSDTEAPVEVQEDFEVVEEFTPVVLATPIPEEVQQAQT-
EIKLFNKWSFE
EVEVKDASLVDYVQVR---QPIFVAHTAGRYANKRFRKAQCPIIERLTNS
LMMNGRNNGKKLKAVRIIKHTLDIINVLTDQNPIQVVVDAITNTGPREDT
TRVGGGGAARRQAVDVSPLRRVNQAIALLTIGAREAAFRNIKTIAETLAE
ELINAAKGSSTSYAIKKKDELERVAKSNR
```

8. The alignment highlights (red) interface residues of S26 that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

```
>15228895_Arabidopsis
MTFKRRNGGRNKHNRGHVKPIRCSNCGKCCPKDKAIKRFIVRNIVEQAAI
RDVQEASVYEGYTLPKLYAKTQYCVSCAIHSHVVRVRSRTNRRVRTPP—
PRFARRKEDTPKPAQPGQAPRPAGGAPAAPRA
>17508699_Caenorhabditis
MTFKRRNHGRNKKNRGHVAFIRCTNCGRCCPKDKAIKKFVVRNIVEAAAV
RDIGDASAYTQYALPKLYHKLHYCIACAIHSKVVRNRSREARRDRNPP--
PRFGQR-AAAARPGAPG--PRP----------
>P62854_Homo
MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRNIVEAAAV
RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--
PRFRPAGAAPRPPPKPM---------------
>6981488_Rattus
MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRNIVEAAAV
RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--
PRFRPAGAAPRPPPKPM---------------
>57092393_Canis
MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRNIVEAAAV
RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--
PRFRPAGAAPRPPPKPM---------------
>62751960_Bos
MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRNIVEAAAV
RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--
PRFRPAGAAPRPPPKPM---------------
>255522917_Equus
MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRNIVEAAAV
RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--
PRFRPAGAAPRPPPKPM---------------
>147899011_Sus
MTKKRRNNGRAKKGRGHVQPIRCTNCARCVPKDKAIKKFVIRNIVEAAAV
```

```
RDISEASVFDAYVLPKLYVKLHYCVSCAIHSKVVRNRSREARKDRTPP--
PRFRPAGAAPRPPPKPM---------------
>17647893_Drosophila
MTKKRRNGGRNKHNRGHVKPVRCTNCARCVPKDKAIKKFVIRNIVEAAAV
RDITEASIWDSYVLPKLYAKLHYCVSCAIHSKVVRNRSREARRIRTPP—
LRSFPKDMARNNQNRK----------------
>6320978_Saccharomyces
MPKKRASNGRNKKGRGHVKPVRCVNCSKSIPKDKAIKRMAIRNIVEAAAV
RDLSEASVYPEYALPKTYNKLHYCVSCAIHARIVRVRSREDRKNRAPPQR
PRFNRDNKVSPAAAAKKAL-------------
```

9. The alignment highlights (red) interface residues of S25 that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

```
>P62851_Homo
MPPKDDKKKKDAGKSAKKDKDPVNKSGGKAKKKKWSKGKVRDKLNNLVLF
DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV
SKHRAQVIYTRNTKGGDAPAAGEDA
>392347476_Rattus
MLPKDDKKKKDAGKSAKKDKDPVNKSGGKAKKKKWSKGKVRDKLNNLILF
DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV
SKHRAQVIYTRNTKGGDAPAAGEDA
>345799858_Canis
MPPKDDKKKKDAGKSAKKDKDPVNKSGGKAKKKKWSKGKVRDKLNNLVLF
DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV
SKHRAQVIYTRNTKGGDAPAAGEDA
>70778964_Bos
MPPKDDKKKKDAGKSAKKDKDPVNKSGGKAKKKKWSKGKVRDKLNNLVLF
DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV
SKHRAQVIYTRNTKGGDAPAAGEDA
>311264034_Sus
MPPKDDKKKKDAGKSAKKDKDPVNKSGGKAKKKKWSKGKVRDKLNNLVLF
DKATYDKLCKEVPNYKLITPAVVSERLKIRGSLARAALQELLSKGLIKLV
SKHRAQVIYTRNTKGGDAPAAGEDA
>338712068_Equus
MPPKDNKKKKDAGKSAKKDKDPVNKSGGKATKKKWSKGKVQDKLNNLVLF
DKATYDKPCNEVPRYKLISPAVVSERMKIRGSLARATLQELLSKGRLKLV
SKHRAQVIYTRNTKGGDAPAAGEGA
>24645865_Drosophila
MPPKKDAKS-SAKQPQKTQKKKEGSGGGKAKKKKWSKGKVRDKLNNQVLF
DKATYEKLYKEVPAYKLITPSVVSERLKIRGSLAKRALIELREKGLIKQV
VQHHSQVIYTRATKGDEA-------
>17542024_Caenorhabditis
MPPKKDPKG-GKAPP---SKKKEGSGGGKAKKKKWSKGKVRDKLNNMVLF
DQATYDKLYKEVITYKLITPSVVSERLKVRASLAKAGLKELQAKGLVKCV
VHHHGQVVYTRATKEADVIVE----
>79322680_Arabidopsis
MAPKKDKVPPPSSKPAK-------SGGGKQ-KKKWSKGKQKEKVNNMVLF
DQGTYDKLLTEAPKFKLITPSILSDRMRINGSLARRAIRELMAKGLIRMV
SAHSSQQIYTRATNT----------
>6321464_Saccharomyces
MPPKQQLSKAAKAAAA--------LAGGKKSKKKWSKKSMKDRAQHAVIL
DQEKYDRILKEVPTYRYVSVSVLVDRLKIGGSLARIALRHLEKEGIIKPI
SKHSKQAIYTRATAAASE-------
```

10. The alignment highlights (red) interface residues of S27 that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

```
>545519486_Canis
MPLA--KDLLHPSAEEEKRKHKKQRLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRQKQH
>94966948_Bos
MPLA--RDLLHPSLDEEKKKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH
>P42677_Homo
MPLA--KDLLHPSPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH
>16758382_Rattus
MPLA--KDLLHPSPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH
>545216788_Equus
MPLA--KDLLHPSPEEEKRKHKKKRLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLCVGCSTVLCQPTGGKARLTEGCSFRRKQH
>335302490_Sus
MPLA--KDLLHSSPEEEKRKHKK-YLVQSPNSYFMDVKCPGCYKITTVFS
HAQTVVLYVGCSAVLCQPTGGKARLTEGCSFRRKQR
>24649976_Drosophila
MPLA--KDLLHPLPAEEKRKHKLKRLVQHPNSYFMDVKCPGCYRITTVFS
HAQGVVVCAGCATILCQPTGGRAKLTEGCSFRRKPQ
>17563244_Caenorhabditis
MPLA--VDLLHPEPQREIRCHKLKRLVQHPNSYFMDVKCSGCFKISTVFS
HATTVVVCVGCNTVLCQPTRGKAKLTEGCSFRKKQ-
>15238845_Arabidopsis
MVLQNDIDLLHPPPELEKRKHKLKRLVQSPNSFFMDVKCQGCFNITTVFS
HSQTVVVCGNCQTVLCQPTGGKARLQEGCSFRKK--
>6321809_Saccharomyces
MVLV--QDLLHPTAASEARKHKLKTLVQGPRSYFLDVKCPGCLNITTVFS
HAQTAVTCESCSTVLCTPTGGKAKLSEGTSFRRK--
```

11. The alignment highlights (red) interface residues of S3a that contact IRES RNA. Residues that are not conserved in another species, are shown in blue. The GI accessions are given for each sequence.

```
>P61247_Homo_sapiens
-----------MAVGKNKRLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMF
NIRNIGKTLVTRTQGTKIASDGLKGRVFEVSLADLQND---EVAFRKFKL
ITEDVQGKNCLTNFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLR
LFCVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEVVN
KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS
S-SGKATGDETGAKVERADGYEPPVQESV-----
>545210000_Equus
HFRPFGSLSSVMAVGKNKRLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMF
NIRNIGKTLVTRTQGTKIASDGLKGRVFEVSLADLQND---EVAFRKFKL
ITEDVQGKNCLTNFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLR
LFCVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEVVN
KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS
S-SGKATGDETGAKVERADGYEPPVQESV-----
>296470536_Bos
-----------MAVGKNKCLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMF
NIRNIGKTSVTRTQGTKITSDGLKGRVFEVSLADLQND---EVAFRKFKL
ITEDVQGKNCLTNFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLR
LFCVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEVVN
KLIPDSIGKDIEKACQSIYPFHDVFVRKVKMLKKPKFELGKLMELHGEGS
S-SGKATGDETGAKVERADGYEPPVQESV-----
>8394221_Rattus
-----------MAVGKNKRLTKGGKKGAKKKVVDPFSKKDWYDVKAPAMF
NIRNIGKTLVTRTQGTKIASDGLKGRVFEVSLADLQND---EVAFRKFKL
ITEDVQGKNCLTNFHGMDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLR
LFCVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEVVN
KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGG
S-SGKTTGDETGAKVERADGYEPPVQESV-----
>545557812_Canis
-----------MVVGKNKRLTKGGRKGAKKKVVDPFSKKDWYDVKAPAMF
NIRNIGKTLVTRTQGTKIASDGLKGRVFEVSLADLQND---EVAFRKFKL
ITEDVQGKNCLTNFHDVDLTRDKMCSMVKKWQTMIEAHVDVKTTDGYLLH
LFCVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEVVN
KLIPDSIGKDIEKACQSIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS
S-SGKAMGDETGAKVERADGYKPPVQESV-----
>212549625_Sus
-----------MAVGKNKRLTKGGKKGAKKKVVDPFSKRDWYDVKAPAMF
NIRNIGKTLVTRTQGTKIASDGLKGRVFEVSPADLQND---EVAFRKSKL
ITEDVQGKNCLTNFHGMDLTRDKMCSMVKKWQTVIEAHVDVKTTDGYLLR
LFCVGFTKKRNNQIRKTSYAQHQQVRQIRKKMMEIMTREVQTNDLKEVVN
KLIPDSIGKDIEKACQPIYPLHDVFVRKVKMLKKPKFELGKLMELHGEGS
S-SGKATGDETGAKVERADGYEPPVQESV-----
>17864162_Drosophila
-----------MAVGKNKGLSKGGKKGGKKKVVDPFSRKDWYDVKAPNMF
QTRQIGKTLVNRTQGQRIASDYLKGRVFEVSLADLQKDIDPERSFRKFRL
IAEDVQDRNVLCNFHGMDLTTDKYRSMVKKWQTLIEAIVEAKTVDGYLLR
VFCIGFTAKDQQSQRKTCYAQQSQVRKIRARMTDIITNEVSGADLKQLVN
KLALDSIAKDIEKSCQRIYPLHDVYIRKVKVLKKPRFDVSKLLELHGDGG
GKSVEAVVSSEGAVIDRPEGYEPPVQEAV-----
>17553700_Caenorhabditis
-----------MAVGKNN—NKMGKKGGKKKAVDPFSRKEWYDIKAPNMF
NTRQVGKTLINRTQGTKIASEGLKGRVFEVSLGDLNNS---EADFRKFKL
IAEDVQGKNVLTNFHAMSMTHDKLCSIVKKWHTLIEANTAVKTTDGYTLR
VFVIAFTKKSVNQVKKTSYTKTSKIRKIRSEMIGCIEKEVTGCDLKEVVS
KLIPDSIGKDIEKTCSKLYPLQEVYIRKVKIIKRPKVDLGRLHDLHGDSI
T------VGADGEKVDRPDDYEPPVQQEV-----
>15236171_Arabidopsis
```

```
-----------MAVGKNKRISK-GRKGGKKKAVDPFSKKDWYDVKAPGSF
TNRNVGKTLVSRTQGTKIASEGLKHRVFEVSLADLQND—EDNAYRKIRL
RAEDVQGRNVLTQFWGMDFTTDKLRSLVKKWQTLIEAHVDVKTTDGYTLR
MFCIAFTKRRANQVKRTCYAQSSQIRQIRRKMSEIMVKEASSCDLKELVA
KFIPEAIGREIEKATQGIYPLQNVFIRKVKILKAPKFDLGKLMEVHGDYT
A-------EDVGVKVDRP-ADETMVEEPTEIIGA
>6323474_Saccharomyces
-----------MAVGKNKRLSK-GKKGQKKRVVDPFTRKEWFDIKAPSTF
ENRNVGKTLVNKSTGLKSASDALKGRVVEVCLADLQGS—EDHSFRKIKL
RVDEVQGKNLLTNFHGMDFTTDKLRSMVRKWQTLIEANVTVKTSDDYVLR
IFAIAFTRKQANQVKRHSYAQSSHIRAIRKVISEILTKEVQGSTLAQLTS
KLIPEVINKEIENATKDIFPLQNIHVRKVKLLKQPKFDVGALMALHGEGS
G-------EEKGK---KVTGFKDEVLETV-----
```

References

1. Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. Structures of the human and Drosophila 80S ribosome. Nature 2013; 497:80-5.
2. Lukavsky PJ, Kim I, Otto GA, Puglisi JD. Structure of HCV IRES domain II determined by NMR. Nat Struct Biol 2003; 10:1033-8.
3. Li L, Li C, Sarkar S, Zhang J, Witham S, Zhang Z, Wang L, Smith N, Petukh M, Alexov E. DelPhi: a comprehensive suite for DelPhi software and associated resources. BMC Biophys 2012; 5:9.
4. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem 2004; 25:1605-12.
5. Schuler M, Connell SR, Lescoute A, Giesebrecht J, Dabrowski M, Schroeer B, Mielke T, Penczek PA, Westhof E, Spahn CM. Structure of the ribosome-bound cricket paralysis virus IRES RNA. Nat Struct Mol Biol 2006; 13:1092-6.
6. Voorhees RM, Weixlbaumer A, Loakes D, Kelley AC, Ramakrishnan V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. Nat Struct Mol Biol 2009; 16:528-33.
7. Armache JP, Jarasch A, Anger AM, Villa E, Becker T, Bhushan S, Jossinet F, Habeck M, Dindar G, Franckenberg S, et al. Localization of eukaryote-specific ribosomal proteins in a 5.5-A cryo-EM map of the 80S eukaryotic ribosome. Proc Natl Acad Sci U S A 2010; 107:19754-9.
8. Lomakin IB, Steitz TA. The initiation of mammalian protein synthesis and mRNA scanning mechanism. Nature 2013; 500:307-11.
9. Bulkley D, Innis CA, Blaha G, Steitz TA. Revisiting the structures of several antibiotics bound to the bacterial ribosome. Proc Natl Acad Sci U S A 2010; 107:17158-63.
10. Berry KE, Waghray S, Mortimer SA, Bai Y, Doudna JA. Crystal structure of the HCV IRES central domain reveals strategy for start-codon positioning. Structure 2011; 19:1456-66.
11. Terribilini M, Sander JD, Lee JH, Zaback P, Jernigan RL, Honavar V, Dobbs D. RNABindR: a server for analyzing and predicting RNA-binding sites in proteins. Nucleic Acids Res 2007; 35:W578-84.
12. Wang L, Huang C, Yang MQ, Yang JY. BindN+ for accurate prediction of DNA and RNA-binding residues from protein sequence features. BMC Syst Biol 2010; 4 Suppl 1:S3.
13. van Aalten DM, DiRusso CC, Knudsen J. The structural basis of acyl coenzyme A-dependent regulation of the transcription factor FadR. Embo J 2001; 20:2041-50.

14. Smith C, Heyne S, Richter AS, Will S, Backofen R. Freiburg RNA Tools: a web server integrating INTARNA, EXPARNA and LOCARNA. Nucleic Acids Res 2010; 38:W373-7.
15. Holm L, Kaariainen S, Rosenstrom P, Schenkel A. Searching protein structure databases with DaliLite v.3. Bioinformatics 2008; 24:2780-1.
16. Shen MY, Sali A. Statistical potential for assessment and prediction of protein structures. Protein Sci 2006; 15:2507-24.