Supplementary Materials for

Error-prone protein synthesis
in parasites with the smallest eukaryotic genome

Authors:

Sergey V. Melnikov[1], Keith D. Rivera[2], Denis Ostapenko[1], Arthur Makarenko[2], Neil D. Sanscrainte[3], James J. Becnel[3], Mark J. Solomon[1], Catherine Texier[4], Darryl J. Pappin[2], Dieter Söll[1,4#]

Author Affiliation:

[1] Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA
[2] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[3] United States Department of Agriculture–Agricultural Research Service, Center for Medical, Agricultural and Veterinary Entomology (CMAVE), Gainesville, FL 32608, USA
[4] Université Clermont Auvergne, CNRS, Laboratoire Microorganismes, Génome et Environement, F-63000, Clermont-Ferrand, France
[5] Department of Chemistry, Yale University, New Haven, CT 06520, USA

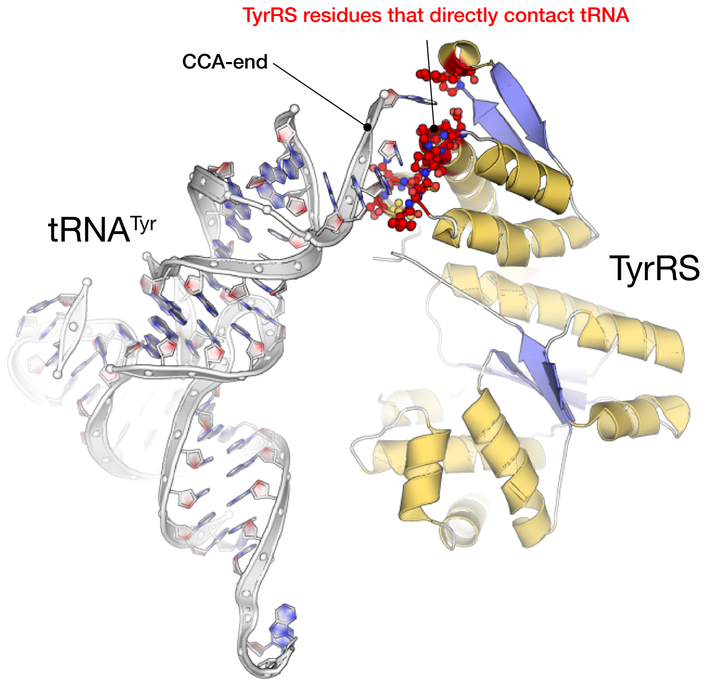
Corresponding Author:

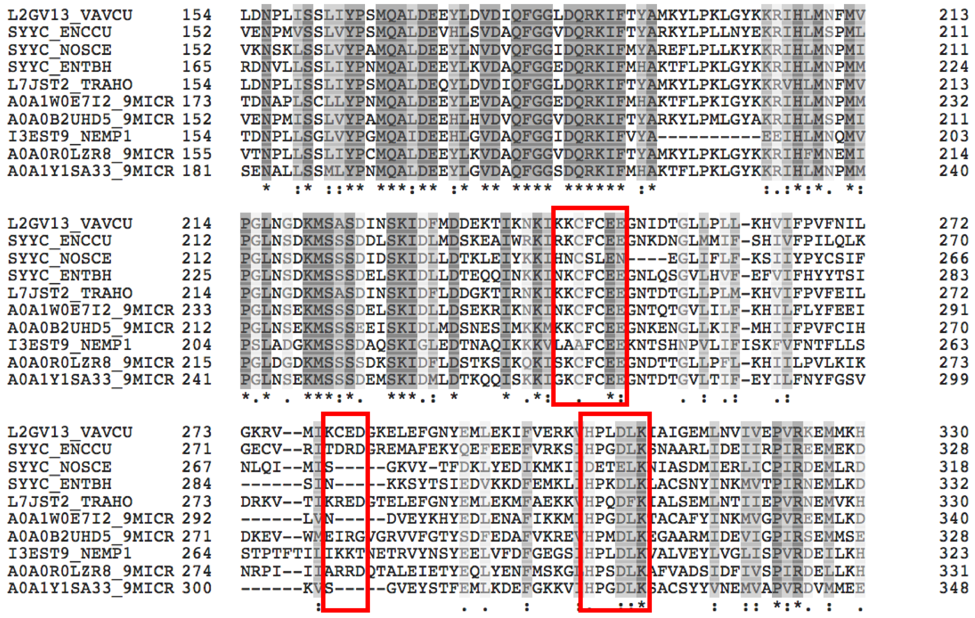
dieter.soll@yale.edu

www.pnas.org/cgi/doi/10.1073/pnas.1803208115

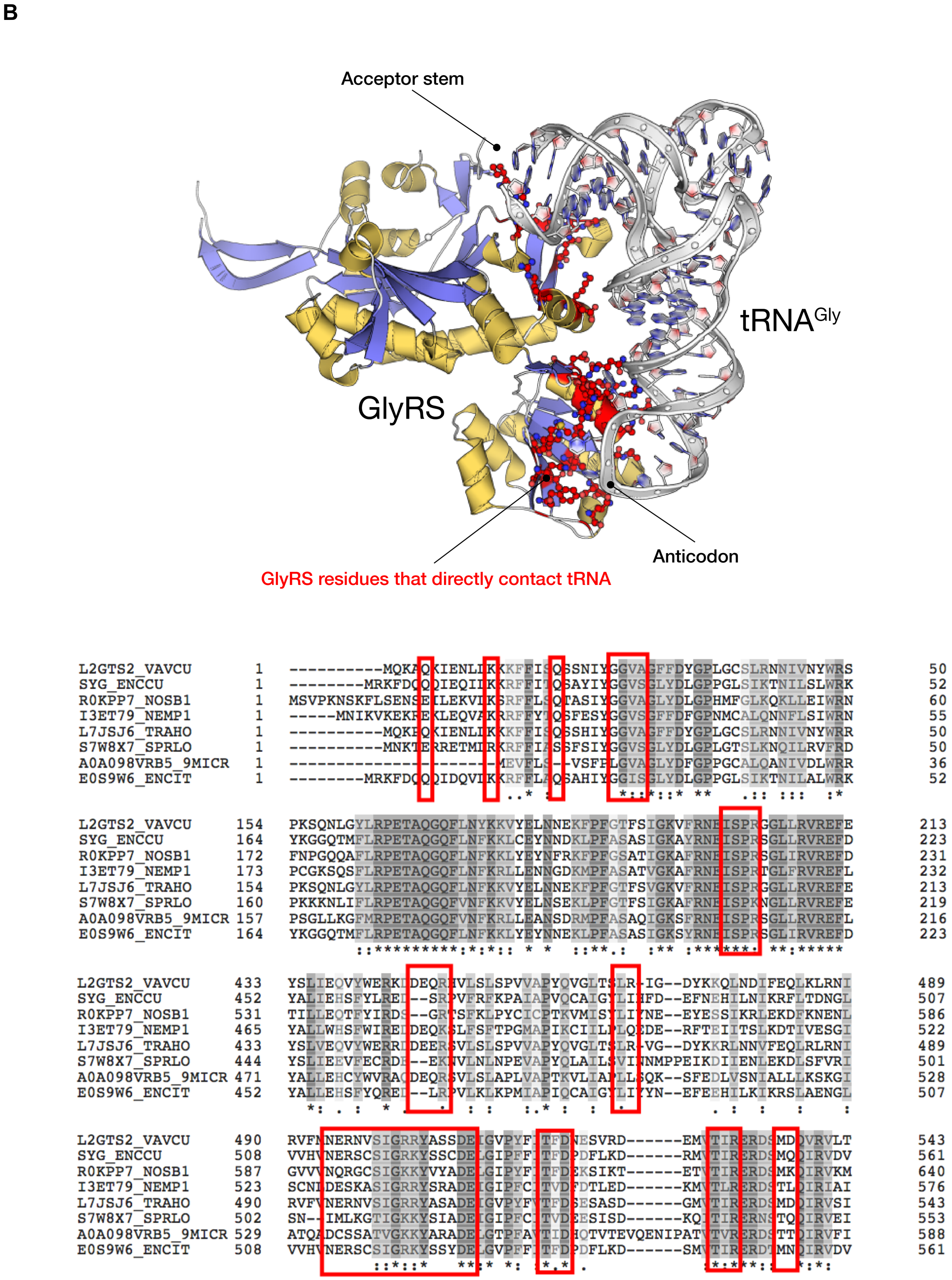
Figure S1 | Sequence variations in the catalytic and tRNA-binding domains of microsporidian aminoacyl-tRNA synthetases. (A-C) Crystal structures of eukaryotic aminoacyl-tRNA synthetases in complex with cognate tRNAs: TyrRS/tRNA^{Tyr} from yeast *S. cerevisiae* (pdb id 2dlc) (A), GlyRS/tRNA^{Gly} from human (pdb id 5e6m) (B), and TrpRS /tRNA^{Trp} from human (pdb id 2ake) (C). Below each structure, multiple sequence alignments highlight residues (in red) that are predicted to form direct contacts with cogante tRNAs. The figure illustrates that, despite close evolutionary relationships between microsporidian species, microsporidian synthetases carry potentially deleterious mutations within tRNA-binding and catalytic domains, suggesting altered tRNA affinity or specificity of tRNA recognition by microsporidian synthetases.

A



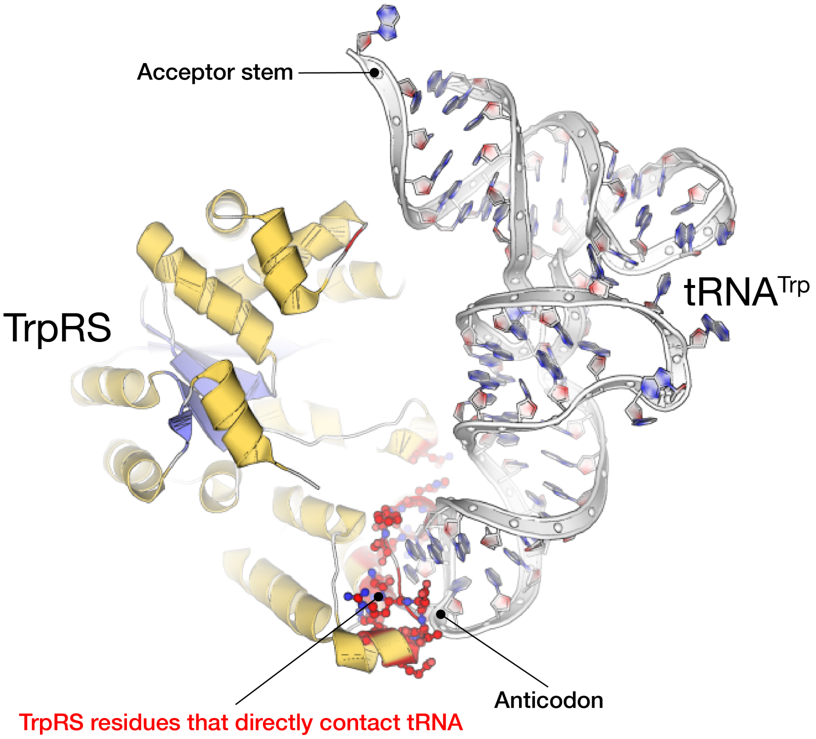


```
L2GV13_VAVCU     154   LDNPLISSLIYPSMQALDEEYLDVDIQFGGLDQRKIFTYAMKYLPKLGYKKRIHLMNFMV      213
SYYC_ENCCU       152   VENPMVSSLVYPSMQALDEVHLSVDAQFGGVDQRKIFTYARKYLPLLNYEKRIHLMSPML      211
SYYC_NOSCE       152   VKNSKLSSLVYPAMQALDEEYLNVDVQFGGIDQRKIFMYAREFLPLLKYKKRIHLMNPMI      211
SYYC_ENTBH       165   RDNVLLSSLIYPNMQALDEEYLKVDAQFGGEDQRKIFMHAKTFLPKLGYKKRIHLMNPMM      224
L7JST2_TRAHO     154   LDNPLISSLIYPSMQALDEQYLDVDIQFGGLDQRKIFTYAMKYLPKLGYKKRVHLMNFMV      213
A0A1W0E7I2_9MICR 173   TDNAPLSCLLYPNMQALDEEYLEVDAQFGGEDQRKIFMHAKTFLPKIGYKKRIHLMNPMM      232
A0A0B2UHD5_9MICR 152   VENPMISSLVYPAMQALDEEHLHVDVQFGGVDQRKIFTYARKYLPMLGYAKRIHLMSPMI      211
I3EST9_NEMP1     154   TDNPLLSGLVYPGMQAIDEEHLGVDAQFGGIDQRKIFVYA----------EEIHLMNQMV      203
A0A0R0LZR8_9MICR 155   VTNPLLSSLIYPCMQALDEEYLKVDAQFGGVDQRKIFTYAMKYLPKLGYKKRIHFMNEMI      214
A0A1Y1SA33_9MICR 181   SENALLSSMLYPNMQALDEEYLGVDAQFGGSDQRKIFMHAKTFLPKLGYKKRIHLMNPMM      240
                         *  :* ::** ***:** :* ** **** ****** :*          :.:*:*. *:

L2GV13_VAVCU     214   PGLNGDKMSASDINSKIDFMDDEKTIKNKIKKCFCEEGNIDTGLLPLL-KHVIFPVFNIL      272
SYYC_ENCCU       212   PGLNSDKMSSSDDLSKIDLMDSKEAIWRKIRKCFCEEGNKDNGLMMIF-SHIVFPILQLK      270
SYYC_NOSCE       212   PGLNSDKMSSSDIDSKIDLLDTKLEIYKKIHNCSLEN----EGLIFLF-KSIIYPYCSIF      266
SYYC_ENTBH       225   PGLNSDKMSSSDELSKIDLLDTEQQINKKINKCFCEEGNLQSGVLHVF-EFVIFHYYTSI      283
L7JST2_TRAHO     214   PGLNGDKMSASDINSKIDFLDDGKTIRNKIKKCFCEEGNTDTGLLPLM-KHVIFPVFEIL      272
A0A1W0E7I2_9MICR 233   PGLNSEKMSSSDELSKIDLLDSEKRIKNKINKCFCEEGNTQTGVLILF-KHILFLYFEEI      291
A0A0B2UHD5_9MICR 212   PGLNSEKMSSSEEISKIDLMDSNESIMKKMKKCFCEEGNKENGLLKIF-MHIIFPVFCIH      270
I3EST9_NEMP1     204   PSLADGKMSSSDAQSKIGLEDTNAQIKKKVLAAFCEEKNTSHNPVLIFISKFVFNTFLLS      263
A0A0R0LZR8_9MICR 215   PGLDGDKMSSSSDKSKIDFLDSTKSIKQKISKCFCEEGNDTTGLLPFL-KHIILPVLKIK      273
A0A1Y1SA33_9MICR 241   PGLNSEKMSSSDEMSKIDMLDTKQQISKKIGKCFCEEGNTDTGVLTIF-EYILFNYFGSV      299
                       *.* . ***:*.  ***.: *    * .*:  .  *:     . : .:   .:

L2GV13_VAVCU     273   GKRV--MIKCEDGKELEFGNYEMLEKIFVERKVHPLDLKIAIGEMLNVIVEPVRKEMMKH      330
SYYC_ENCCU       271   GECV--RITDRDGREMAFEKYQEFEEEFVRKSIHPGDLKSNAARLIDEIIRPIREEMEKD      328
SYYC_NOSCE       267   NLQI--MIS-----GKVY-TFDKLYEDIKMKIIDETELKNIASDMIERLICPIRDEMLRD      318
SYYC_ENTBH       284   ------SIN-----KKSYTSIEDVKKDFEMKLIHPKDLKLACSNYINKMVTPIRNEMLKD      332
L7JST2_TRAHO     273   DRKV--TIKREDGTELEFGNYEMLEKMFAEKKVHPQDFKIALSEMLNTIIEPVRNEMVKH      330
A0A1W0E7I2_9MICR 292   ------LVN-----DVEYKHYEDLENAFIKKMIHPGDLKTACAFYINKMVGPVREEMLKD      340
A0A0B2UHD5_9MICR 271   DKEV--WMEIRGVGRVVFGTYSDFEDAFVKREVHPMDLKEGAARMIDEVIGPIRSEMMSE      328
I3EST9_NEMP1     264   STPTFTIIIKKTNETRVYNSYEELVFDFGEGSIHPLDLKVALVEYLVGLISPVRDEILKH      323
A0A0R0LZR8_9MICR 274   NRPI--IIARRDQTALEIETYEQLYENFMSKGLHPSDLKAFVADSIDFIVSPIRDELLKH      331
A0A1Y1SA33_9MICR 300   ------KVS-----GVEYSTFEMLKDEFGKKVIHPGDLKSACSYYVNEMVAPVRDVMMEE      348
                              :             . .   :    :   ::*      :  :: *:*. :  .
```



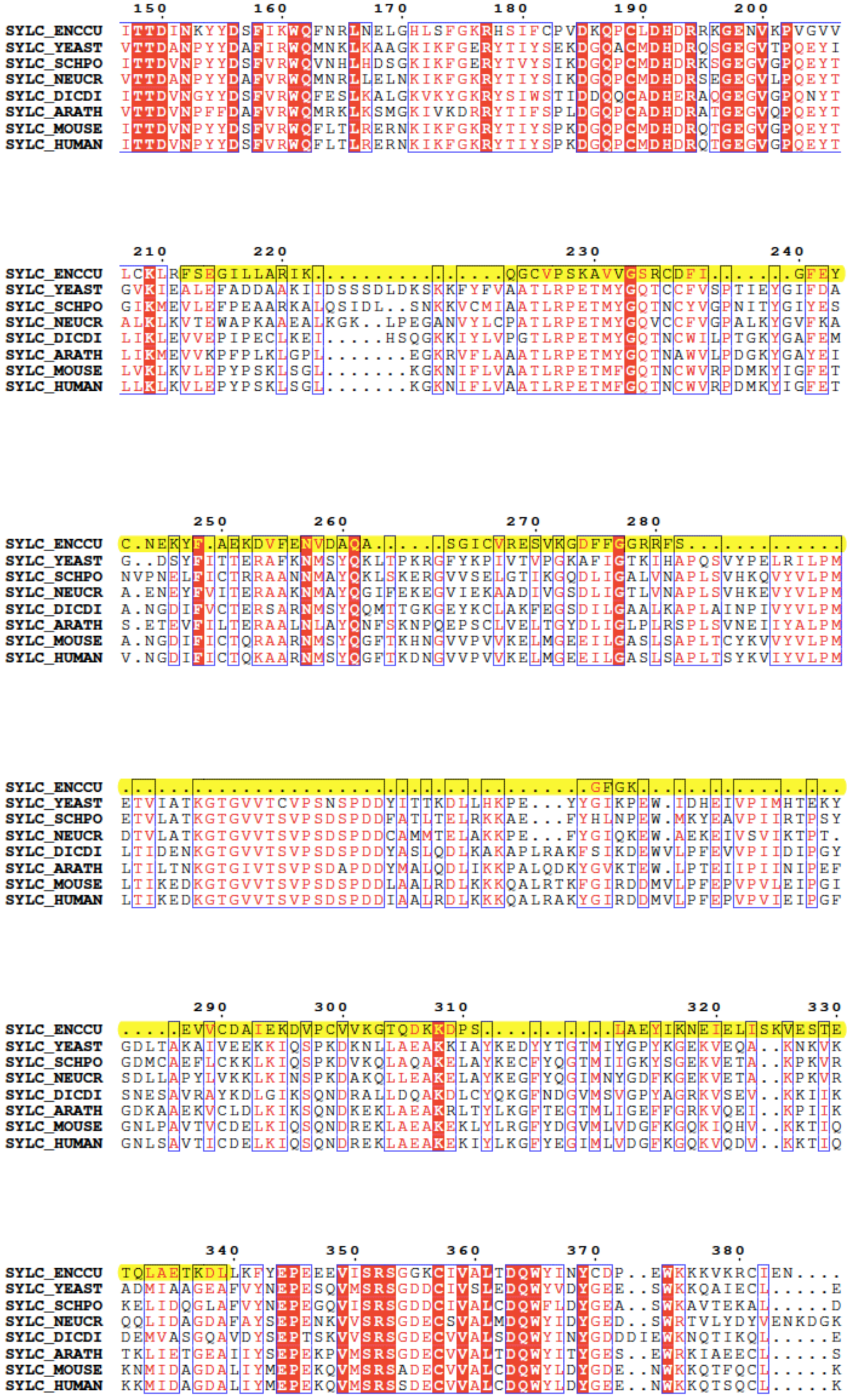
```
L2GTS2_VAVCU       1   ----------MQKAQKIENLIKKKFFISQSSNIYGGVAGFFDYGPLGCSLRNNIVNYWRS   50
SYG_ENCCU          1   --------MRKFDQQQIEQIIKKRFFITQSAYIYGGVSGLYDLGPPGLSIKTNILSLWRK   52
R0KPP7_NOSB1       1   MSVPKNSKFLSENSEILEKVIKSRFFLSQTASIYGGVAGLYDLGPHMFGLKQKLLEIWRN   60
I3ET79_NEMP1       1   -----MNIKVKEKREKLEQVAKRRFFYTQSFESYGGVSGFFDFGPNMCALQNNFLSIWRN   55
L7JSJ6_TRAHO       1   ----------MQKEQKIENLIKKKFFISQSSHIYGGVAGFFDYGPLGCSLRNNIVNYWRR   50
S7W8X7_SPRLO       1   ----------MNKTERRETMIRKRFFIASSFSIYGGVSGLYDLGPLGTSLKNQILRVFRD   50
A0A098VRB5_9MICR   1   ----------------------MEVFLS--VSFPLGVAGLYDFGPPGCALQANIVDLWRR   36
E0S9W6_ENCIT       1   --------MRKFDQQQIDQVIKKRFFLAQSAHIYGGISGLYDLGPPGLSIKTNILALWRK   52
                                              ..* :       *::*::* **   .:: :::  :*

L2GTS2_VAVCU     154   PKSQNLGYLRPETAQGQFLNYKKVYELNNEKFPFGTFSIGKVFRNEISPRGGLLRVREFE   213
SYG_ENCCU        164   YKGGQTMFLRPETAQGQFLNFKKLCEYNNDKLPFASASIGKAYRNEISPRSGLLRVREFD   223
R0KPP7_NOSB1     172   FNPGQQAFLRPETAQGQFLNFKKLYEYNFRKFPFGSATIGKAFRNEISPRSGLIRVREFD   231
I3ET79_NEMP1     173   PCGKSQSFLRPETAQGQFLNFKRLLENNGDKMPFASATVGKAFRNEISPRTGLFRVREFL   232
L7JSJ6_TRAHO     154   PKSQNLGYLRPETAQGQFLNFKKVYELNNEKFPFGTFSVGKVFRNEISPRGGLLRVREFE   213
S7W8X7_SPRLO     160   PKKKNLIFLRPETAQGQFVNFKKVYELNSEKLPFGTFSIGKAFRNEISPKNGLLRVREFE   219
A0A098VRB5_9MICR 157   PSGLLKGFMRPETAQGQFVNFKRLLEANSDRMPFASAQIGKSFRNEISPRSGLLRVREFL   216
E0S9W6_ENCIT     164   YKGGQTMFLRPETAQGQFLNFKKLYEYNNEKLPFASASIGKSYRNEISPRSGLIRVREFD   223
                              ::*********:*:*:: * *  ::**.:  :** :******. **:*****

L2GTS2_VAVCU     433   YSLIEQVYWERKIDEQRHVLSLSPVVAPYQVGLTSLR-IG--DYKKQLNDIFEQLKLRNI   489
SYG_ENCCU        452   YALIEHSFYLREI--SRPVFRFKPAIAPVQCAIGYLIHFD--EFNEHILNIKRFLTDNGL   507
R0KPP7_NOSB1     531   TILLEQTFYIRDS--GRTSFKLPYCICPTKVMISYLIYNE--EYESSIKRLEKDFKNENL   586
I3ET79_NEMP1     465   YALLWHSFWIREIDEQKSLFSFTPGMAPIKCIILPLQEDE--RFTEIITSLKDTIVESGI   522
L7JSJ6_TRAHO     433   YSLVEQVYWERRIDEERSVLSLSPVVAPYQVGLTSLR-VG--DYKKRLNNVFEQLRLRNI   489
S7W8X7_SPRLO     444   YSLIEEVFECRDE--EKNVLNLNPEVAPYQLAILSVINNMPPEIKDIIENLEKDLSFVRI   501
A0A098VRB5_9MICR 471   YALLEHCYWVRAQDEQRSVLSLAPLVAPTKVLIAPLLSQK--SFEDLVSNIALLLKSKGI   528
E0S9W6_ENCIT     452   YALLEHSFYQREI--LRPVLKLKPMIAPIQCAIGYLIYYN--EFEEHILKIKRSLAENGL   507
                         *: . :  * .   .  : :   :.* :  :  .         . :  :   :    :

L2GTS2_VAVCU     490   RVFMNERNVSIGRRYASSDEIGVPYFITFDNESVRD------EMVTIRERDSMDQVRVLT   543
SYG_ENCCU        508   VVHVNERSCSIGRKYSSCDELGIPFFITFDPDFLKD------RMVTIRERDSMQQIRVDV   561
R0KPP7_NOSB1     587   GVVVNQRGCSIGKKYVYADELGIPYFITFDEKSIKT------ETVTIRERDSMKQIRVKM   640
I3ET79_NEMP1     523   SCNLDESKASIGRRYSRADEIGIPFCITVDFDTLED------KMVTLRERDSTLQIRIAI   576
L7JSJ6_TRAHO     490   RVFVNERNVSIGRRYASADEIGVPYFVTFDSESASD------GMVTIRERDSMDQIRVSI   543
S7W8X7_SPRLO     502   SN--IMLKGTIGKKYSIADEIGIPFCITVDEESISD------KQITIRERNSTQQIRVEI   553
A0A098VRB5_9MICR 529   ATQADCSSATVGKKYARADELGTPFAVTIDHQTVTEVQENIPATVTVRERDSTTQIRVFI   588
E0S9W6_ENCIT     508   VVHVNERSCSIGRKYSSYDELGVPFFITFDPDFLKD------SMVTIRERDTMNQIRVDV   561
                                ::*::*   **:* *: :*.* .            :*:***::  *:*:
```

C



```
L2GSV8_VAVCU     282   SQFLPSLSNLGK-MSSS-RDKTIFLDDSDEDIKKKIRTYAHSGGRKTLQEHRALGGDLEN   339
SYW_ENCCU        256   VSLLPDLKGVNRKMSASDPNSSIYLDDAQDTIRKKIIAYAYSGGRKTLEEHREKGGDIDV   315
A0A0B2UIN1_9MICR 256   VSLLPDLKGVNKKMSASDQNSSISLDDTPENIRKKITTYAYSGGRKTMKEHKEMGGDISV   315
C4V9E6_NOSCE     261   VSLLPDLKGVNRKMSASDTTSSIFLTDTPAQIAKKIRKYAFSGGKETLEEHKRLGGDTAV   320
E0SAB3_ENCIT     256   VSLLPDLKGVNRKMSASDPNSSIYLDDTPETIRKKITVHAYSGGRETLEEHREKGGNIDV   315
L2GKT2_VITCO     256   LGLLPDLQGTNRKMSASDQKSAIFLNDGPNDVENKINKLAFSGGKETVEEHRRLGGDPSI   315
B7XK60_ENTBH     256   VDLLPSLKGIDKKMSASDATTSIYLSDTPQDIENKIKKHAFSGGQPTKELQEKLGANIHV   315
L7JY70_TRAHO     290   SQFLPSLMNLGK-MSSS-KDKTIFLDDSEESIKKKIRTYAHSGGRKTLQEHKMLGGDLDN   347
A0A177EF74_9MICR 251   TYFLPALQGTGVKMSASSSVSSIYLNDTPAEIKKKINKHAFSGGRDTLEEHRKHGGDPDV   310
S7XLV7_SPRLO     254   STFLPALQS-ESKMSASDEKTTIYLSDTRKKIKNKINKEAFSGGRDTLEEHRKYGGNPDI   312
                         :** * .    **:*   .:* * *    : :**   *.***: * : :.  *.:

L2GSV8_VAVCU     340   DVSYQYLRFFMEDEDKLQSVADRYKKGEMTSTEMKDLCIEELQRFIRTYKKMRDRVSDDD   399
SYW_ENCCU        316   DVPFEYLKYFLDDDQELEKYRSGYIKGEITSKEMKEKCVVVIQEFVSRYQESRKRVTDDD   375
A0A0B2UIN1_9MICR 316   DVPFEYLRYFLQEDDELEAYRSGYLSGNISSRDMKEKCISVIQKFVAEYQSSRQAVDDKE   375
C4V9E6_NOSCE     321   DISYQYLRYFYDDNIDLERIKNAYEKGEMSTGDIKNKCIDVLQKFVKDYQERRNSITEEL   380
E0SAB3_ENCIT     316   DVPFEYLRYFLDDDQELERYRLGYIKGEITSREMKEKCISIVQEFVLRYQESRKKITDED   375
L2GKT2_VITCO     316   DVAYQYLRYFLEDDEELLRLKEGYEKGAILSGEMKAYCIRVVQQFLEEYQRRRSEVTDEV   375
B7XK60_ENTBH     316   DVSYQYLKFFLEDDDELNRIGQLYSQGKIMTSDVKNICIQTIQKFIAAFQERRSKITEDI   375
L7JY70_TRAHO     348   DVSYQYLRFFMEDEDKLQSVADKYKKGEMSSMEMKDLCIEELQKFVRTYRKRRDRVSDED   407
A0A177EF74_9MICR 311   DIAYQYLRFFLEDDDQLKAYEEGYRNGSILSGEMKKACIETLQLFIKDFQARRAAITPEV   370
S7XLV7_SPRLO     313   DVSFQYLRYFLEDDKELEEIRTNYIKGIITSGEMKSKCIDVIAGFVEEYQMNRAKVTDED   372
                       *: ::**::* ::: .*      * .* : : ::*  *:  :  *:  ::  *  :  .
```

Figure S2 | Multiple sequence alignments reveal degeneration of the editing domain in LeuRS from microsporidian parasites. Position of the editing domain in the primary structure of microsporidian LeuRS (from *E. cuniculi*) is highlighted in yellow. To illustrate conservation of the editing domain across eukaryotic species, the sequences are shown for cytosolic LeuRS from *S. cerevisiae* (YEAST), *S. pombe* (SCHPO), *N. crassa* (NEUCR), *A. thaliana* (ARATH), *M. musculus* (MOUSE), *H. sapiens* (HUMAN).



```
               150       160       170       180       190       200
SYLC_ENCCU   ITTDINKYYDSFIKWQFNRLNELGHLSFGKRHSIFCPVDKQPCLDHDRRKGENVKPVGVV
SYLC_YEAST   VTTDANPYYDAFIRWQMNKLKAAGKIKFGERYTIYSEKDGQACMDHDRQSGEGVTPQEYI
SYLC_SCHPO   ITTDVNPYYDSFVRWQVNHLHDSGKIKFGERYTVYSIKDGQPCMDHDRKSGEGVGPQEYT
SYLC_NEUCR   VTTDANPYYDAFVRWQMNRLLELNKIKFGKRYTIYSIKDGQPCMDHDRSEGEGVLPQEYT
SYLC_DICDI   ITTDVNGYYDSFVRWQFESLKALGKVKYGKRYSIWSTIDDQQCADHERAQGEGVGPQNYT
SYLC_ARATH   VTTDVNPFFDAFVRWQMRKLKSMGKIVKDRRYTIFSPLDGQPCADHDRATGEGVQPQEYT
SYLC_MOUSE   ITTDVNPYYDSFVRWQFLTLRERNKIKFGKRYTIYSPKDGQPCMDHDRQTGEGVGPQEYT
SYLC_HUMAN   ITTDVNPYYDSFVRWQFLTLRERNKIKFGKRYTIYSPKDGQPCMDHDRQTGEGVGPQEYT

               210       220                       230              240
SYLC_ENCCU   LCKLRFSEGILLARIK................QGCVPSKAVVGSRCDFI.......GFEY
SYLC_YEAST   GVKIEALEFADDAAKIIDSSSDLDKSKKFYFVAATLRPETMYGQTCCFVSPTIEYGIFDA
SYLC_SCHPO   GIKMEVLEFPEAARKALQSIDL..SNKKVCMIAATLRPETMYGQTNCYVGPNITYGIYES
SYLC_NEUCR   ALKLKVTEWAPKAAEALKGK..LPEGANVYLCPATLRPETMYGQVCCFVGPALKYGVFKA
SYLC_DICDI   LIKLEVVEPIPECLKEI.....HSQGKKIYLVPGTLRPETMYGQTNCWILPTGKYGAFEM
SYLC_ARATH   LIKMEVVKPFPLKLGPL.......EGKRVFLAAATLRPETMYGQTNAWVLPDGKYGAYEI
SYLC_MOUSE   LVKLKVLEPYPSKLSGL.......KGKNIFLVAATLRPETMFGQTNCWVRPDMKYIGFET
SYLC_HUMAN   LLKLKVLEPYPSKLSGL.......KGKNIFLVAATLRPETMFGQTNCWVRPDMKYIGFET

                    250       260             270       280
SYLC_ENCCU   C.NEKYF.AEKDVFENVDAQA......SGICVRESVKGDFFGGRRFS.............
SYLC_YEAST   G..DSYFITTERAFKNMSYQKLTPKRGFYKPIVTVPGKAFIGTKIHAPQSVYPELRILPM
SYLC_SCHPO   NVPNELFICTRRAANNMAYQKLSKERGVVSELGTIKGQDLIGALVNAPLSVHKQVYVLPM
SYLC_NEUCR   A.ENEYFVITERAAKNMAYQGIFEKEGVIEKAADIVGSDLIGTLVNAPLSVHKEVYVLPM
SYLC_DICDI   A.NGDIFVCTERSARNMSYQQMTTGKGEYKCLAKFEGSDILGAALKAPLAINPIVYVLPM
SYLC_ARATH   S.ETEVFILTERAALNLAYQNFSKNPQEPSCLVELTGYDLIGLPLRSPLSVNEIIYALPM
SYLC_MOUSE   A.NGDIFICTQRAARNMSYQGFTKHNGVVPVVKELMGEEILGASLSAPLTCYKVVYVLPM
SYLC_HUMAN   V.NGDIFICTQKAARNMSYQGFTKDNGVVPVVKELMGEEILGASLSAPLTSYKVIYVLPM

SYLC_ENCCU   ......................................GFGK..................
SYLC_YEAST   ETVIATKGTGVVTCVPSNSPDDYITTKDLLHKPE...YYGIKPEW.IDHEIVPIMHTEKY
SYLC_SCHPO   ETVLATKGTGVVTSVPSDSPDDFATLTELRKKAE...FYHLNPEW.MKYEAVPIIRTPSY
SYLC_NEUCR   DTVLATKGTGVVTSVPSDSPDDCAMMTELAKKPE...FYGIQKEW.AEKEIVSVIKTPT.
SYLC_DICDI   LTIDENKGTGVVTSVPSDSPDDYASLQDLKAKAPLRAKFSIKDEWVLPFEVVPIIDIPGY
SYLC_ARATH   LTILTNKGTGIVTSVPSDAPDDYMALQDLIKKPALQDKYGVKTEW.LPTEIIPIINIPEF
SYLC_MOUSE   LTIKEDKGTGVVTSVPSDSPDDLAALRDLKKKQALRTKFGIRDDMVLPFEPVPVLEIPGI
SYLC_HUMAN   LTIKEDKGTGVVTSVPSDSPDDIAALRDLKKKQALRAKYGIRDDMVLPFEPVPVIEIPGF

                    290       300       310                  320       330
SYLC_ENCCU   .....EVVCDAIEKDVPCVVKGTQDKKDPS...........LAEYIKNEIELISKVESTE
SYLC_YEAST   GDLTAKAIVEEKKIQSPKDKNLLAEAKKIAYKEDYYTGTMIYGPYKGEKVEQA..KNKVK
SYLC_SCHPO   GDMCAEFLCKKLKIQSPKDVKQLAQAKELAYKECFYQGTMIIGKYSGEKVETA..KPKVR
SYLC_NEUCR   SDLLAPYLVKKLKINSPKDAKQLLEAKELAYKEGFYQGIMNYGDFKGEKVETA..KPKVR
SYLC_DICDI   SNESAVRAYKDLGIKSQNDRALLDQAKDLCYQKGFNDGVMSVGPYAGRKVSEV..KKIIK
SYLC_ARATH   GDKAAEKVCLDLKIKSQNDKEKLAEAKRLTYLKGFTEGTMLIGEFFGRKVQEI..KPIIK
SYLC_MOUSE   GNLPAVTVCDELKIQSQNDREKLAEAKEKLYLRGFYDGVMLVDGFKGQKIQHV..KKTIQ
SYLC_HUMAN   GNLSAVTICDELKIQSQNDREKLAEAKEKIYLKGFYEGIMLVDGFKGQKVQDV..KKTIQ

                     340       350       360       370         380
SYLC_ENCCU   TQLAETKDLLKFYEPEEEVISRSGGKCIVALTDQWYINYCDP..EWKKKVKRCIEN....
SYLC_YEAST   ADMIAAGEAFVYNEPESQVMSRSGDDCIVSLEDQWYVDYGEE..SWKKQAIECL.....E
SYLC_SCHPO   KELIDQGLAFVYNEPEGQVISRSGDDCIVALCDQWFLDYGEA..SWKAVTEKAL.....D
SYLC_NEUCR   QQLIDAGDAFAYSEPENKVVSRSGDECSVALMDQWYIDYGED..SWRTVLYDYVENKDGK
SYLC_DICDI   DEMVASGQAVDYSEPTSKVVSRSGDECVVALSDQWYINYGDDDIEWKNQTIKQL.....E
SYLC_ARATH   TKLIETGEAIIYSEPEKPVMSRSGDECVVALTDQWYITYGES..EWRKIAEECL.....S
SYLC_MOUSE   KNMIDAGDALIYMEPEKQVMSRSADECVVALCDQWYLDYGDE..NWKKQTFQCL.....K
SYLC_HUMAN   KKMIDAGDALIYMEPEKQVMSRSSDECVVALCDQWYLDYGEE..NWKKQTSQCL.....K
```

Figure S3 | The "rudimentary" editing domain of microsporidian LeuRS is poorly conserved even across microsporidian species. The multiple sequence alignment shows a fragment of LeuRS sequence from five microsporidian families: Encephalitozoonidea (*E. cuniculi*), Pleistophoridea (*T. hominis*), Nosematidea (*N. bombycus*), Ordosporidae (*O. colligata*) and Spraguidae (*S. lophii*). Position of the editing domain in the LeuRS primary structure is highlighted in yellow.

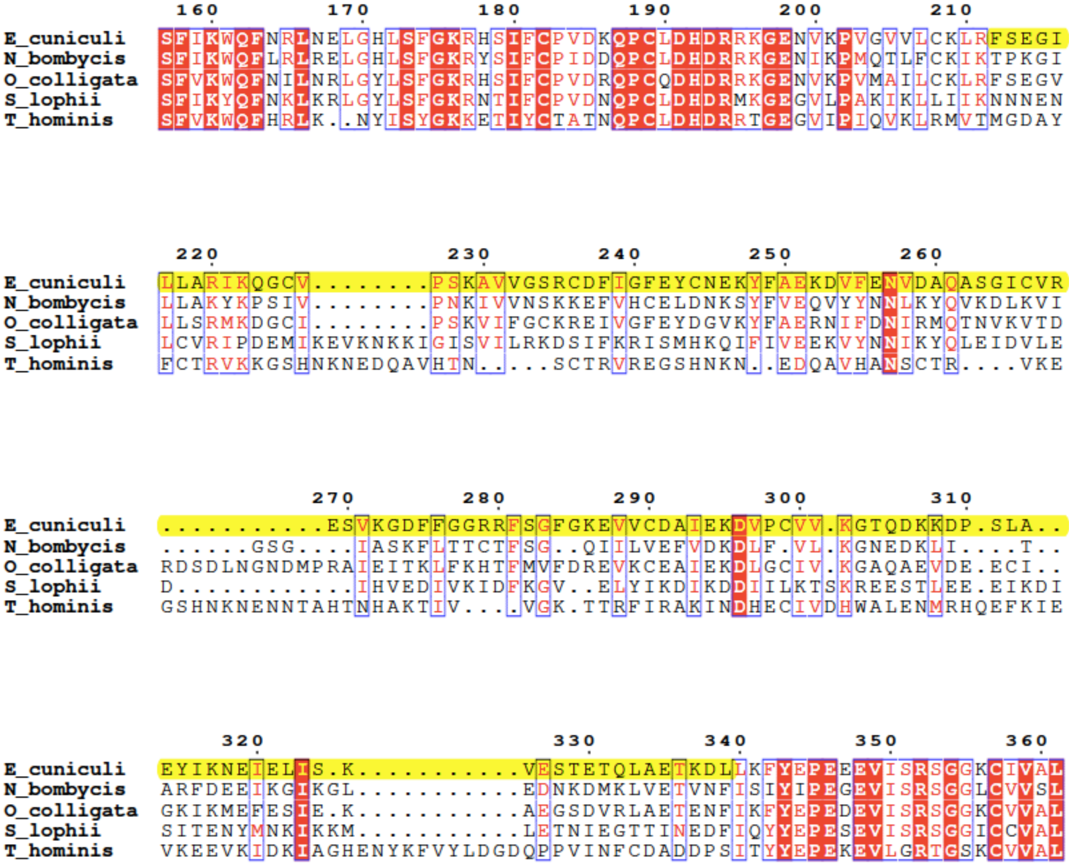


Figure S4 | Purification, active site titration of *E. cuniculi* LeuRS. (**a**) *E. cuniculi* LeuRS recombinantly expressed and purified from *E. coli* (1mg). (**b**) Active site titration to determine the active fraction of *E. cuniculi* LeuRS. (**c**) Side-by-side comparison of acid activation by *E. cuniculi* LeuRS (1 µM *E. cuniculi* LeuRS in 150 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM KF, 2.5 mM DTT, 0.05% BSA (w/v), 1 mM sodium pyrophosphate, 10 ATP mM, γ-[^{32}P]-ATP (1000 cpu/ul final) and 2 mM amino acid after 1 minute at 37°C).

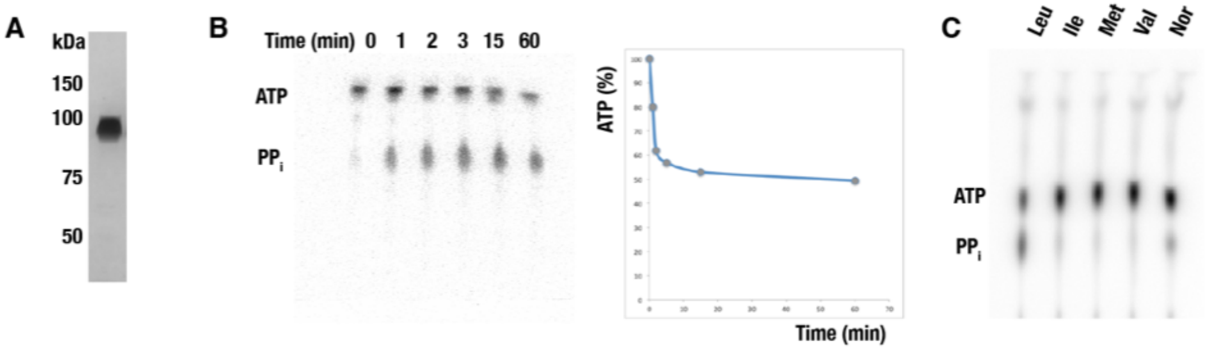
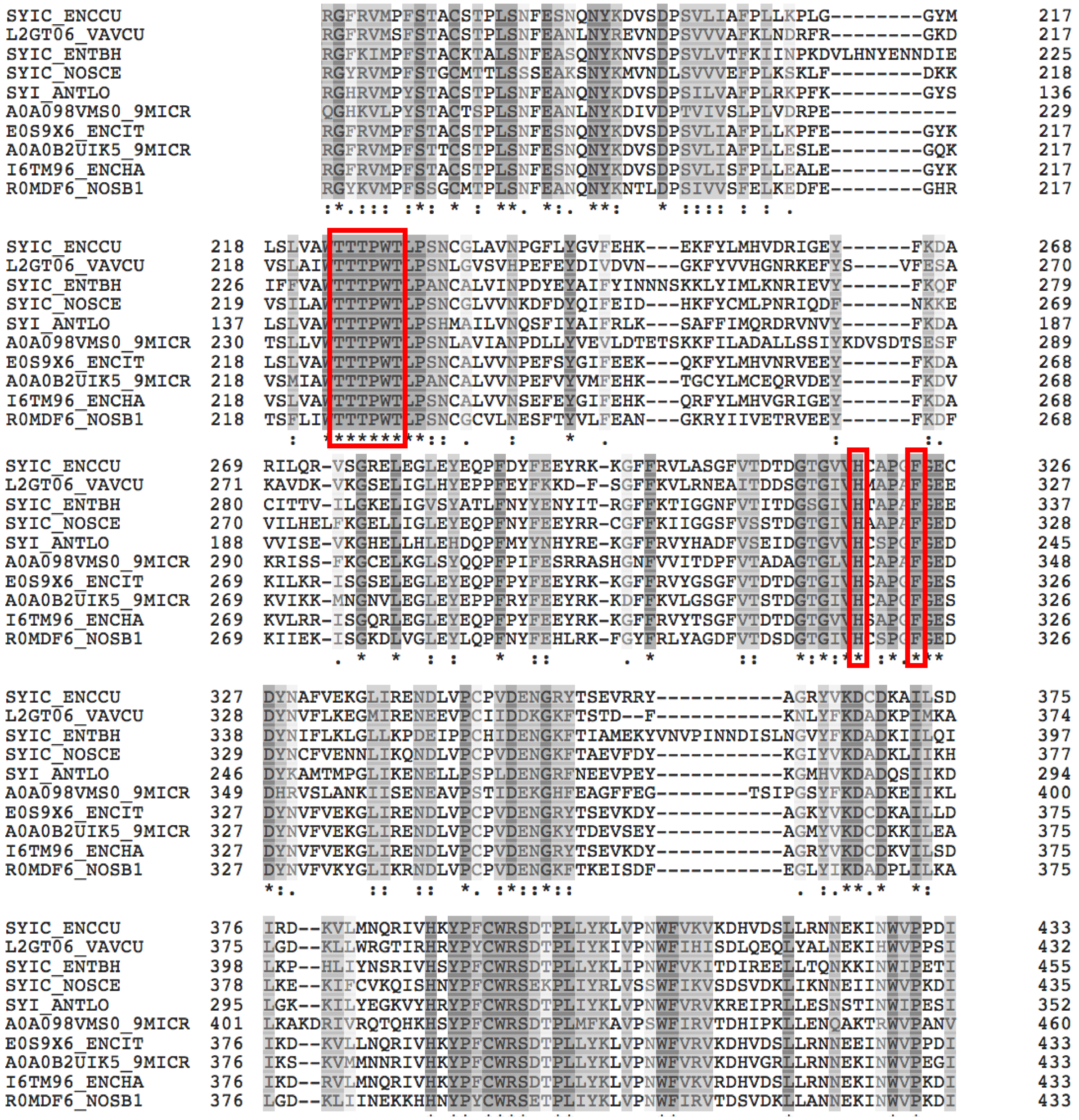


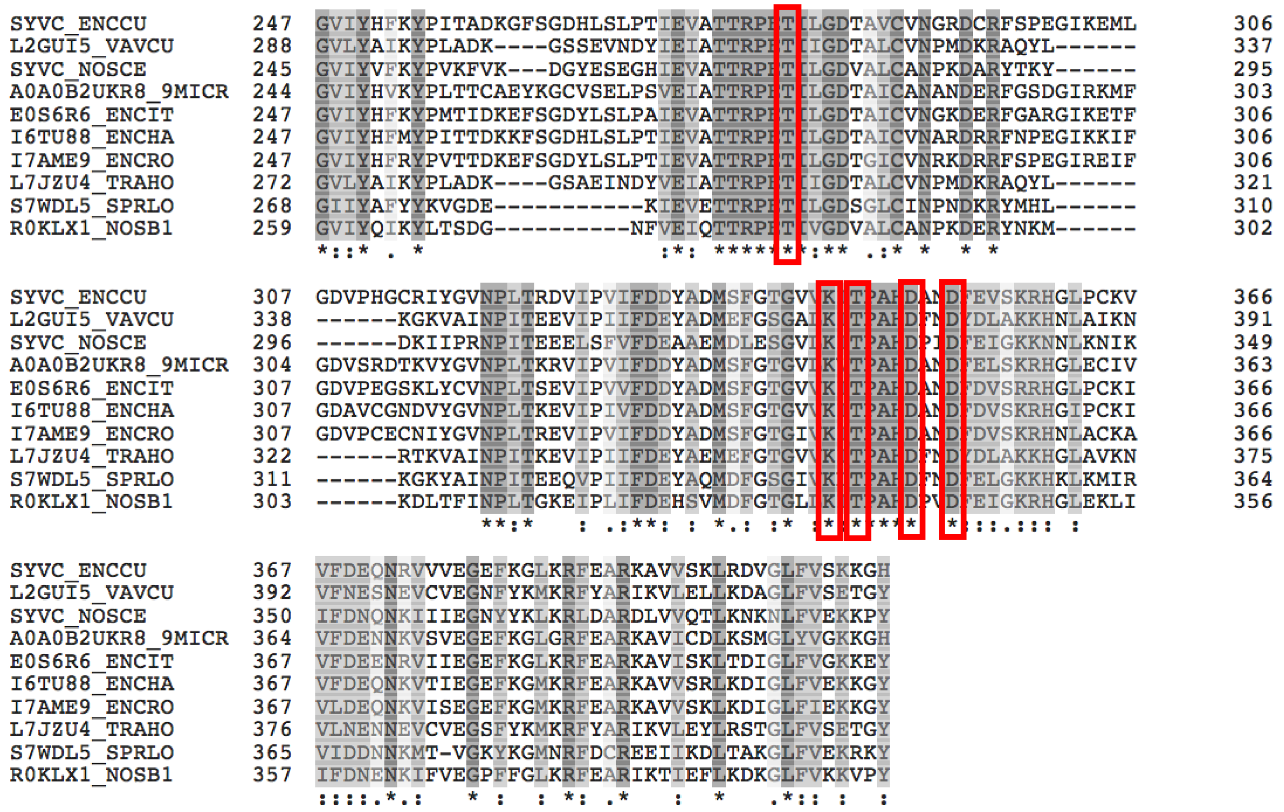
Figure S5 | High conservation of the editing domains in microsporidian IleRS and ValRS synthetases. Multiple sequence alignments illustrate that IleRS (A) and ValRS (B) have conserved editing domains. Active site residues that are involved in recognition of near-cognate amino acids are outlined in red.

A



```
SYIC_ENCCU                  RGFRVMPFSTACSTPLSNFESNQNYKDVSDPSVLIAFPLLKPLG--------GYM      217
L2GT06_VAVCU                RGFRVMSFSTACSTPLSNFEANLNYREVNDPSVVVAFKLNDRFR--------GKD      217
SYIC_ENTBH                  RGFKIMPFSTACKTALSNFEASQNYKNVSDPSVLVTFKIINPKDVLHNYENNDIE      225
SYIC_NOSCE                  RGYRVMPFSTGCMTTLSSSEAKSNYKMVNDLSVVVEFPLKSKLF--------DKK      218
SYI_ANTLO                   RGHRVMPYSTACSTPLSNFEANQNYKDVSDPSILVAFPLRKPFK--------GYS      136
A0A098VMS0_9MICR            QGHKVLPYSTACTSPLSNFEANLNYKDIVDPTVIVSLPLVDRPE-----------      229
E0S9X6_ENCIT                RGFRVMPFSTACSTPLSNFESNQNYKDVSDPSVLIAFPLLKPFE--------GYK      217
A0A0B2UIK5_9MICR            RGFRVMPFSTTCSTPLSNFESNQNYKDVSDPSVLIAFPLLESLE--------GQK      217
I6TM96_ENCHA                RGFRVMPFSTACSTPLSNFESNQNYKDVSDPSVLISFPLLEALE--------GYK      217
R0MDF6_NOSB1                RGYKVMPFSSGCMTPLSNFEANQNYKNTLDPSIVVSFELKEDFE--------GHR      217
                            :*.::: :*: * : **. *:. **:   * :::: : : .

SYIC_ENCCU        218  LSLVAWTTTPWTLPSNCGLAVNPGFLYGVFEHK---EKFYLMHVDRIGEY------FKDA      268
L2GT06_VAVCU      218  VSLAIWTTTPWTLPSNLGVSVHPEFEYDIVDVN---GKFYVVHGNRKEFYS----VFESA      270
SYIC_ENTBH        226  IFFVAWTTTPWTLPANCALVINPDYEYAIFYINNNSKKLYIMLKNRIEVY------FKQF      279
SYIC_NOSCE        219  VSILAWTTTPWTLPSNCGLVVNKDFDYQIFEID---HKFYCMLPNRIQDF------NKKE      269
SYI_ANTLO         137  LSLVAWTTTPWTLPSHMAILVNQSFIYAIFRLK---SAFFIMQRDRVNVY------FKDA      187
A0A098VMS0_9MICR  230  TSLLVWTTTPWTLPSNLAVIANPDLLYVEVLDTETSKKFILADALLSSIYKDVSDTSESF      289
E0S9X6_ENCIT      218  LSLVAWTTTPWTLPSNCALVVNPEFSYGIFEEK---QKFYLMHVNRVEEY------FKDA      268
A0A0B2UIK5_9MICR  218  VSMIAWTTTPWTLPANCALVVNPEFVYVMFEHK---TGCYLMCEQRVDEY------FKDV      268
I6TM96_ENCHA      218  VSLVAWTTTPWTLPSNCALVVNSEFEYGIFEHK---QRFYLMHVGRIGEY------FKDA      268
R0MDF6_NOSB1      218  TSFLIWTTTPWTLPSNCGCVLNESFTYVLFEAN---GKRYIIVETRVEEY------FKDF      268
                         :  *********:: .   :    *  .                   :       :.

SYIC_ENCCU        269  RILQR-VSGRELEGLEYEQPFDYFEEYRK-KGFFRVLASGFVTDTDGTGVVHCAPGFGEC      326
L2GT06_VAVCU      271  KAVDK-VKGSELIGLHYEPPFEYFKKD-F-SGFFKVLRNEAITDDSGTGIVHMAPAFGEE      327
SYIC_ENTBH        280  CITTV-ILGKELIGVSYATLFNYYENYIT-RGFFKTIGGNFVTITDGSGIVHTAPAFGEE      337
SYIC_NOSCE        270  VILHELFKGELLIGLEYEQPFNYFEEYRR-CGFFKIIGGSFVSSTDGTGIVHAAPAFGED      328
SYI_ANTLO         188  VVISE-VKGHELLHLEHDQPFMYYNHYRE-KGFFRVYHADFVSEIDGTGVVHCSPGFGED      245
A0A098VMS0_9MICR  290  KRISS-FKGCELKGLSYQQPFPIFESRRASHGNFVVITDPFVTADAGTGLVHCAPAFGED      348
E0S9X6_ENCIT      269  KILKR-ISGSELEGLEYEQPFPYFEEYRK-KGFFRVYGSGFVTDTDGTGIVHSAPGFGES      326
A0A0B2UIK5_9MICR  269  KVIKK-MNGNVLEGLEYEPPFRYFEEYRK-KDFFKVLGSGFVTSTDGTGIVHCAPGFGES      326
I6TM96_ENCHA      269  KVLRR-ISGQRLEGLEYEQPFPYFEEYRK-KGFFRVYTSGFVTDTDGTGVVHSAPGFGES      326
R0MDF6_NOSB1      269  KIIEK-ISGKDLVGLEYLQPFNYFEHLRK-FGYFRLYAGDFVTDSDGTGIVHCSPGFGED      326
                             . *  *  : :   *  ::      . *       ::   *:*:** :*.***

SYIC_ENCCU        327  DYNAFVEKGLIRENDLVPCPVDENGRYTSEVRRY-----------AGRYVKDCDKAILSD      375
L2GT06_VAVCU      328  DYNVFLKEGMIRENEEVPCIIDDKGKFTSTD--F-----------KNLYFKDADKPIMKA      374
SYIC_ENTBH        338  DYNIFLKLGLLKPDEIPPCHIDENGKFTIAMEKYVNVPINNDISLNGVYFKDADKIILQI      397
SYIC_NOSCE        329  DYNCFVENNLIKQNDLVPCPVDENGKFTAEVFDY-----------KGIYVKDADKLIIKH      377
SYI_ANTLO         246  DYKAMTMPGLIKENELLPSPLDENGRFNEEVPEY-----------KGMHVKDADQSIIKD      294
A0A098VMS0_9MICR  349  DHRVSLANKIISENEAVPSTIDEKGHFEAGFFEG--------TSIPGSYFKDADKEIIKL      400
E0S9X6_ENCIT      327  DYNVFVEKGLIRENDLVPCPVDENGRYTSEVKDY-----------AGKYVKDCDKAILLD      375
A0A0B2UIK5_9MICR  327  DYNVFVEKGLIRENDLVPCPVDENGKYTDEVSEY-----------AGMYVKDCDKKILEA      375
I6TM96_ENCHA      327  DYNVFVEKGLIRENDLVPCPVDENGRYTSEVKDY-----------AGRYVKDCDKVILSD      375
R0MDF6_NOSB1      327  DYNVFVKYGLIKRNDLVPCPVDENGKFTKEISDF-----------EGLYIKDADPLILKA      375
                       *:.      ::  ::  *. :*::*::                   . :.**.*  *:

SYIC_ENCCU        376  IRD--KVLMNQRIVHKYPFCWRSDTPLLYKLVPNWFVKVKDHVDSLLRNNEKINWVPPDI      433
L2GT06_VAVCU      375  LGD--KLLWRGTIRHRYPYCWRSDTPLIYKIVPNWFIHISDLQEQLYALNEKIHWVPPSI      432
SYIC_ENTBH        398  LKP--HLIYNSRIVHSYPFCWRSDTPLLYKLIPNWFVKITDIREELLTQNKKINWIPETI      455
SYIC_NOSCE        378  LKE--KIFCVKQISHNYPFCWRSEKPLIYRLVSSWFIKVSDSVDKLIKNNEIINWVPKDI      435
SYI_ANTLO         295  LGK--KILYEGKVYHRYPFCWRSDTPLIYKLVPNWFVRVKREIPRLLESNSTINWIPESI      352
A0A098VMS0_9MICR  401  LKAKDRIVRQTQHKHSYPFCWRSDTPLMFKAVPSWFIRVTDHIPKLLENQAKTRWVPANV      460
E0S9X6_ENCIT      376  IKD--KVLLNQRIVHKYPFCWRSDTPLLYKLVPNWFVRVKDHVDSLLRNNEEINWVPPDI      433
A0A0B2UIK5_9MICR  376  IKS--KVMMNNRIVHKYPFCWRSDTPLLYKLVPNWFVRVKDHVGRLLRNNEKINWVPEGI      433
I6TM96_ENCHA      376  IKD--RVLMNQRIVHKYPFCWRSDTPLLYKLVPNWFVKVRDHVDSLLRNNEKINWVPKDI      433
R0MDF6_NOSB1      376  LGD--KLIINEKKHHNYPYCWRSETPLIYKLVPNWFIRVTDSVDKLLANNEKINWVPKDI      433
```

B



```
SYVC_ENCCU        247   GVIYHFKYPITADKGFSGDHLSLPTIEVATTRPETILGDTAVCVNGRDCRFSPEGIKEML   306
L2GUI5_VAVCU      288   GVLYAIKYPLADK----GSSEVNDYIEIATTRPETIIGDTALCVNPMDKRAQYL------   337
SYVC_NOSCE        245   GVIYVFKYPVKFVK---DGYESEGHIEVATTRPETILGDVALCANPKDARYTKY------   295
A0A0B2UKR8_9MICR  244   GVIYHVKYPLTTCAEYKGCVSELPSVEIATTRPETILGDTAICANANDERFGSDGIRKMF   303
E0S6R6_ENCIT      247   GVIYHFKYPMTIDKEFSGDYLSLPAIEVATTRPETILGDTAICVNGKDERFGARGIKETF   306
I6TU88_ENCHA      247   GVIYHFMYPITTDKKFSGDHLSLPTIEVATTRPETILGDTAICVNARDRRFNPEGIKKIF   306
I7AME9_ENCRO      247   GVIYHFRYPVTTDKEFSGDYLSLPTIEVATTRPETILGDTGICVNRKDRRFSPEGIREIF   306
L7JZU4_TRAHO      272   GVLYAIKYPLADK----GSAEINDYVEIATTRPETIIGDTALCVNPMDKRAQYL------   321
S7WDL5_SPRLO      268   GIIYAFYYKVGDE-----------KIEVETTRPETILGDSGLCINPNDKRYMHL------   310
R0KLX1_NOSB1      259   GVIYQIKYLTSDG----------NFVEIQTTRPETIVGDVALCANPKDERYNKM------   302
                        *::* . *                 :*: *******:** .:* *  * *

SYVC_ENCCU        307   GDVPHGCRIYGVNPLTRDVIPVIFDDYADMSFGTGVVK TPAHDANDFEVSKRHGLPCKV   366
L2GUI5_VAVCU      338   ------KGKVAINPITEEVIPIIFDEYADMEFGSGAIK TPAHDFNDYDLAKKHNLAIKN   391
SYVC_NOSCE        296   ------DKIIPRNPITEEELSFVFDEAAEMDLESGVIK TPAHDPIDFEIGKKNNLKNIK   349
A0A0B2UKR8_9MICR  304   GDVSRDTKVYGVNPLTKRVIPVIFDDYADMSFGTGVIK TPAHDANDFELSKRHGLECIV   363
E0S6R6_ENCIT      307   GDVPEGSKLYCVNPLTSEVIPVVFDDYADMSFGTGVVK TPAHDANDFDVSRRHGLPCKI   366
I6TU88_ENCHA      307   GDAVCGNDVYGVNPLTKEVIPIVFDDYADMSFGTGVVK TPAHDANDFDVSKRHGIPCKI   366
I7AME9_ENCRO      307   GDVPCECNIYGVNPLTREVIPVIFDDYADMSFGTGIVK TPAHDANDFDVSKRHNLACKA   366
L7JZU4_TRAHO      322   ------RTKVAINPITKEVIPIIFDEYAEMEFGTGVVK TPAHDFNDYDLAKKHGLAVKN   375
S7WDL5_SPRLO      311   ------KGKYAINPITEEQVPIIFDEYAQMDFGSGIVK TPAHDFNDFELGKKHKLKMIR   364
R0KLX1_NOSB1      303   ------KDLTFINPLTGKEIPLIFDEHSVMDFGTGLIK TPAHDPVDFEIGKRHGLEKLI   356
                                    **:*   : .:**: : *.: :* :* *****  *:::.::: :

SYVC_ENCCU        367   VFDEQNRVVVEGEFKGLKRFEARKAVVSKLRDVGLFVSKKGH
L2GUI5_VAVCU      392   VFNESNEVCVEGNFYKMKRFYARIKVLELLKDAGLFVSETGY
SYVC_NOSCE        350   IFDNQNKIIIEGNYYKLKRLDARDLVVQTLKNKNLFVEKKPY
A0A0B2UKR8_9MICR  364   VFDENNKVSVEGEFKGLGRFEARKAVICDLKSMGLYVGKKGH
E0S6R6_ENCIT      367   VFDEENRVIIEGEFKGLKRFEARKAVISKLTDIGLFVGKKEY
I6TU88_ENCHA      367   VFDEQNKVTIEGEFKGMKRFEARKAVVSRLKDIGLFVEKKGY
I7AME9_ENCRO      367   VLDEQNKVISEGEFKGMKRFEARKAVVSKLKDIGLFIEKKGY
L7JZU4_TRAHO      376   VLNENNEVCVEGSFYKMKRFYARIKVLEYLRSTGLFVSETGY
S7WDL5_SPRLO      365   VIDDNNKMT-VGKYKGMNRFDCREEIIKDLTAKGLFVEKRKY
R0KLX1_NOSB1      357   IFDNENKIFVEGPFFGLKRFEARIKTIEFLKDKGLFVKKVPY
                        ::::.*.:   * :  : *: .*   :  *   .*:: :  :
```

Organism	Strain	Genome size (Mb)	Total gene count
Anncaliia algerae	PRA109	17.53	5330
Anncaliia algerae	PRA339	12.16	3661
Anncaliia algerae	Undeen	13.82	
Edhazardia aedis	USNM 41457	51.35	4262
Encephalitozoon cuniculi	EC1	2.24	1898
Encephalitozoon cuniculi	EC2	2.24	1902
Encephalitozoon cuniculi	EC3	2.24	1898
Encephalitozoon cuniculi	GB-M1	2.5	2126
Encephalitozoon hellem	ATCC 50504	2.25	2006
Encephalitozoon hellem	Swiss	2.18	1864
Encephalitozoon intestinalis	ATCC 50506	2.22	2011
Encephalitozoon romaleae	SJ-2008	2.19	1883
Enterocytozoon bieneusi	H348	3.86	3806
Hamiltosporidium tvaerminnensis	OER-3-3	13.27	
Mitosporidium daphniae	UGP3	5.64	3428
Nematocida ausubeli	ERTm2	4.7	2831
Nematocida ausubeli	ERTm6	4.27	2544
Nematocida parisii	ERTm1	4.07	2724
Nematocida parisii	ERTm3	4.15	2788
Nosema bombycis	CQ1	15.69	4643
Nosema ceranae	BRL01	7.86	2678
Ordospora colligata	OC4	2.29	1879
Pseudoloma neurophilia	MK1	5.25	3676
Spraguea lophii	42_110	4.98	2596
Trachipleistophora hominis	Unknown strain	8.5	3253
Vavraia culicissubsp. floridensis		6.12	2875
Vittaforma corneae	ATCC 50505	3.21	2340

Table S1 | Microsporidian species with known genome sequences. The table summarizes microsporidian species with known genome sequences, according to Microsporidia Data Bank (http://microsporidiadb.org/).

Enzyme	***E. coli***	***V. culicis***	***S. cerevisiae***	***H. sapiens***
AlaRS	876	870	982	967
ArgRS	577	613	607	661
AsnRS	465	451	492	552
AspRS	590	457	557	501
CysRS	461	485	767	748
GlnRS	550	697	809	775
GluRS	471	656	708	1714*
GlyRS	689+303	565	687	739
HisRS	424	486	526	509
IleRS	939	1021	1073	1266
LeuRS	860	832	1090	1176
LysRS	505	508	591	597
MetRS	677	532	751	900
PheRS	795+327	501+557	595+503	508+589
ProRS	572	497	688	1714*
SerRS	430	445	462	505
ThrRS	642	640	734	724
TrpRS	334	411	432	475
TyrRS	424	337	394	528
ValRS	951	917	1058	1265

Table S2 | Size of aminoacyl-tRNA synthetases in microsporidia compared to other species. In *H. sapiens*, GluRS and ProRS are fused in one protein (marked with asterisk).

Table S3 | Steady-state parameters for amino acid activation by LeuRS from yeast *S. cerevisiae* and microsporidium *E. cuniculi*. The values represent the best fit value of three independent experiments in which a synthetase variant was freshly purified before the measurements. N/D stands for Not Detected due to insufficient activity.

	S. cerevisiae		***E. cuniculi***	
Substrate	Km (mM)	Vmax	Km (mM)	Vmax (s^{-1})
Leu	0.05 ± 0.01	96 ± 8	0.11 ± 0.02	9.8 ± 1.2
Nva	2,6 ± 0.3	89 ± 9	2,3 ± 0.2	8.7 ± 1.1
Val	57 ± 7	22 ± 3	78 ± 12	2.9 ± 0.4
Ile	26 ± 3.4	3.7 ± 0.3	36 ± 4.8	0.35 ± 0.05
Met	47 ± 9	5.8 ± 0.6	66 ± 6	0.23 ± 0.03
Asp	N/D	N/D	N/D	N/D