Supporting Information

Rocha et al. 10.1073/pnas.1102835108

SI Methods

Analysis of CUG-Encoded Residue Distribution in Fungal Proteins. For multiple sequence alignments, the *Candida albicans* proteins containing at least one CUG-encoded residue (corresponding to 3,998 ORFs) were selected using ANACONDA (1). Putative/identified *S. cerevisiae* orthologues (expectation value $>10^{-5}$) were obtained from the Candida Genome Database (http://www.candidagenome.org/). Finally, orthologous proteins from yeast species (*Candida glabrata*, *Kluyveromyces lactis*, *Ashbya gossypii*, *Debaryomyces hansenii*, *Yarrowia lipolytica*, and *Schizosaccharomyces pombe)* sufficiently distant to allow an accurate identification of similarity regions within the selected proteins were chosen using the Génolevures web resource (http://www.genolevures.org/). Multiple sequence alignments were preformed with ClustalW (2), using the BLOSUM62 similarity score matrix, and displayed with ESPript (3). The final sample comprised a total of 680 CUG-containing proteins for which the positions of the CUG-encoded residues were scored according to their location in the multiple sequence alignment (within columns of strictly identical residues, of residues with conserved chemical properties, or of nonconserved residues).

Available experimental three-dimensional models for the selected *C. albicans* proteins were identified in the Protein Data Bank (http://www.rcsb.org/pdb/home/home.do) and whenever possible theoretical models were generated by comparative homology modeling using SWISS-MODEL (4) (http://swiss-model.expasy.org). All models with less than 30% target-to-template sequence identity were rejected. Local model quality was assessed with ANOLEA and GROMOS, and only models for which the CUG-encoded residue was located in a reliable region and had an energetically favorable conformation (5) were considered. This defined a subset of 72 *C. albicans* ORFs for which orthologous sequences for CTG-clade organisms (*Candida lusitaniae*, *Candida tropicalis*, *Candida parapsilosis*, *Candida guilliermondii*, *Debaryomyces hansenii*, and *Lodderomyces elongisporus*), WGD-clade (*Saccharomyces mikatae*, *Saccharomyces bayanus*, *Candida glabrata*, *Saccharomyces castellii*) and other fungi (*Kluyveromyces lactis*, *Ashbya gossypii*, and *Kluyveromyces waltii*) were obtained from the Broad Institute Fungal Orthogroups Repository (http://www.broadinstitute.org/regev/orthogroups/). The distribution of CUG-encoded residues in CTG-clade fungi and *Saccharomyces cerevisiae* was determined from the alignment of this subset of sequences, as described above. Three-dimensional models were obtained for all CTG-clade and *S. cerevisiae* orthologs using the procedure and quality criteria described above for *C. albicans* ORFs. This new subset includes 1,392 CUG-encoded residues of which 702 could be structurally analyzed.

For all models, solvent-accessibility and secondary structure location of each CUG-encoded residue was calculated with WhatIf (6) or DSSP (7), classifying the residues as buried, intermediate or exposed, according to the criteria defined in ESPript (3).

Analysis of the Expression Levels of Genes Related to Protein Quality Control Mechanisms upon Increase of CUG-Ambiguity. Proteasome activity was assayed as previously reported (8). Briefly, 2 mL of assay buffer (10 mM Tris-HCl, 20 mM KCl, 5 mM $MgCl_2$, pH 8.0), 150 μL of protein extract and 10 μL of proteasome substrate succinyl-leucine-leucine-valine-tyrosine-MCA (Sigma) were mixed. The variation in fluorescence intensity (emission at 435 nm, excitation at 365 nm) over 1 h incubation at 37 °C was measured with a luminescence spectrometer (Perkin Elmer) and used to calculate the normalized proteasome activity (fluorescence per μg of total protein).

Heat shock protein mRNA levels were determined by quantitative PCR (qPCR). Total RNA was prepared using the hot acidic phenol method (9) with some modifications to completely remove DNA contamination. First-strand cDNA synthesis was carried out in 20 μL reactions using the SuperScript II RT kit (Invitrogen) and quantifed using an Applied Biosystems 7500 Real-Time PCR system with SYBR Green I dye (Applied Biosystems, Power SYBR Green® PCR Master Mix), according to the manufacturer's instructions. The real-time qPCR reactions were performed using default conditions except for the number of cycles (45–50 cycles). The presence of nonspecific products was verified by dissociation curve analysis. Every target gene was quantified using the mean value of nine independent replicates; outliers were rejected according to the critical values of Dixon's "Q" parameters at 95% confidence level (10). The mean values were normalized to the amplification efficiency and to the expression levels of Actin-1 (endogenous reference gene). Samples displaying statistically significant differences according to the Fourier test of variances were further analyzed with Student's *t* test or with the Welch test for unequal variances.

Structure Determination and Refinement. The *C. albicans* SerRS isoforms SerRS_Ser197 and SerRS_Leu197 were purified as previously described (11). Crystals of the apo-forms of both SerRS variants and of SerRS_Ser197 complexes with SerSA and ATP were grown using ammonium sulfate as a precipitant, and X-ray data were collected using synchrotron radiation at beamline ID14-1 of the European Synchrotron Radiation Facility (ESRF) (Grenoble), processed and the initial model was obtained as previously described (11). Refinement was carried out with a single step of rigid-body refinement, followed by several cycles of energy-gradient minimization, simulated annealing, and restrained individual B factor refinement (as implemented in CNS) (12), alternating with rounds of manual model building with COOT (13). An additional step of translation/libration/screw (TLS) refinement was included in the final refinement stage with PHENIX (14), using four TLS groups (residues 1–101, 102–162, 163–297 and 298–452) defined using the TLS motion determination (TLSMD) server (15). The final model comprises residues 1–452; two segments (residues 284–288 and 388–393) are not visible in the electron density map and were therefore omitted. The structure of SerRS_Leu197, SerRS–SerSA, and SerRS–ATP were solved using SerRS_Ser197 as a search model, and refined similarly. In the SerRS–SerSA model region 284–288 (motif 2 loop) is visible in the electron density, but residues 69–71 (at the tip of the helical coil) and 396–388 are not and were omitted. In the SerRS_Leu197 structure residues 68–74 (at the tip of the helical coil), 284–290, and 388–393 were not visible in the electron density. A significant residual electron density was observed in the active site pocket of SerRS_Leu197 variant, suggesting the presence of a discretely disordered adenine moiety, although no nucleotide was added to the crystallization buffer. However, no adenine moiety was included in the final model due to positional instability during refinement. Stereochemical quality of the refined models was calculated with MOLPROBITY (16) as implemented in PHENIX (14). Refinement statistics are summarized in Table S2. For pairwise or multiple structure alignment, the programs LSQMAN (http://X-ray.bmc.uu.se/usf/lsqman_man.html) or MUSTANG (17) were used. Figures of protein structures were prepared with PyMOL (http://www.pymol.org). The final 3D coordinates and

structure factors have been deposited in the Protein Data Bank (PDB) with ID codes 3QNE (SerRS_Ser197), 3QO5 (SerRS_Leu197), 3QO7 (SerRS_Ser197 in complex with ATP), and 3QO8 (SerRS_Ser197 in complex with seryl-adenylate analog). *C. albicans* SerRS closest structural homologues are the SerRS from *Pyrococcus horikoshii* [PDB ID 2DQ0 (18), rmsd 2.2 Å for 414 aligned Cα residues, 40% sequence identity] and *Thermus thermophilus* [PDB ID 1SRY (19), rmsd 2.1 Å for 408 aligned Cα residues, 28% sequence identity].

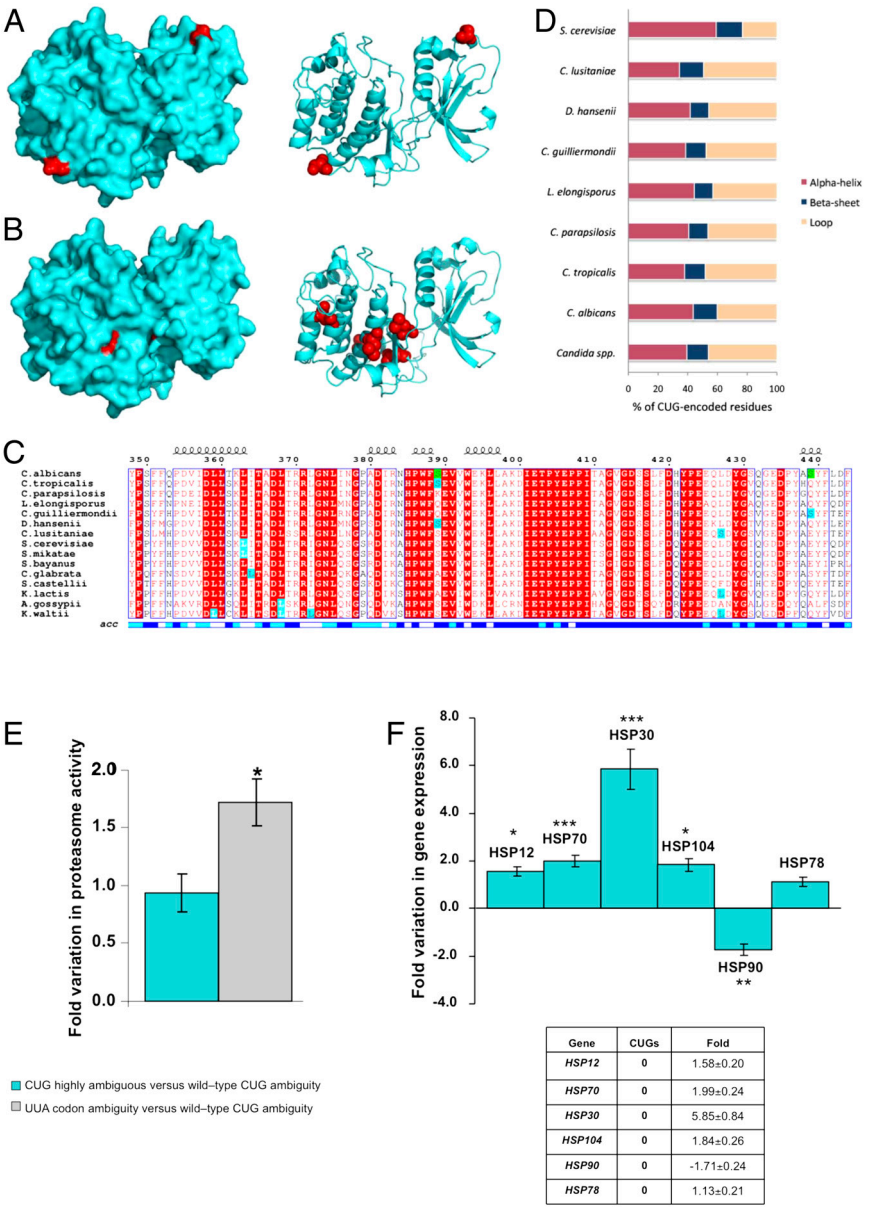
In Vitro SerRS Activity. The seryl-adenylate formation was followed by monitoring pyrophosphate (PPi) release at 360 nm at 25 °C as described by Lloyd et al. (20) with the EnzChek PP_i assay kit (Molecular Probes) according to the manufacturer's instructions. The reaction mixture contained 2 mM ATP, 1 mM L-serine, 50 mM Tris·HCl (pH 7.5), 10 mM $MgCl_2$, 40 mM KCl and 0.1 mM sodium azide. Enzyme concentrations ranged from 0.1–0.8 μM (SerRS monomer). To analyze the effect of tRNA on the specific activity of SerRS 4 mg/mL of yeast unfractioned tRNA (Roche) was used. For Km determination the concentration of serine was varied between 0.1 and 20 × Km. In all cases the reaction mixture was incubated at 25 °C for 5′ prior to starting the reaction by L-serine addition. All data was corrected against a reference background measured without addition of the substrate. The initial velocities obtained for each substrate concentration were fitted to a Michaelis–Menten equation and kinetic parameters were calculated using nonlinear regression. Active site titration was performed according to the method of Fersht et al. (21) using the buffer described above and showed that the fraction of active enzyme (70%) was similar for both SerRS isoforms.

In Vivo SerRS Activity. The activity of *C. albicans* SerRS in vivo was studied by expressing the orthogonal synthetase/$tRNA^{Ser}_{CAG}$ pair in *Escherichia coli* and measuring the mistranslation phenotype. For this purpose, the *C. albicans* CAI4 strain $tRNA^{Ser}_{CAG}$ gene was amplified by PCR from genomic DNA using primers 5′-GTACTCGGGAACACCAAACAAAAGATGC-3′ and 5′-TCTACCTTAGGTGATTGACTTTATTACATGC-3′ and the tRNA-encoding fragment was cloned into the Eco88I and Eco81I restriction sites of the pACYC184 vector (New England Biolabs).

The compatible pACYC-tRNA and pT7-SerRS plasmids were cotransformed into *E. coli* BL21(DE3) (Stratagene). To avoid the negative pressure caused by CUG-ambiguity *C. albicans* SerRS CUG codon (that codes for residue 197) was mutated to UCG, to create the SerRS_Ser197 isoform, or to CUG, to express the SerRS_Leu197 isoform. Overnight cultures of fresh transformants (grown at 30 °C in LB medium supplemented with 1% glucose and antibiotics) were used to inoculate LB supplemented with 100 $\mu g\,mL^{-1}$ ampicilin, 25 $\mu g\,mL^{-1}$ chloramphenicol and 1 mM IPTG. IPTG was added to induce the expression of *C. albicans* SerRS, whereas the expression of *C. albicans* $tRNA^{Ser}_{CAG}$ was constitutive. Bacterial cells grew at 37 °C and cell growth (A_{600}) was monitored as a function of time to determine the specific growth rate (average of 3–5 independent experiments). The amount of expressed SerRS was quantified by western blot analysis. For this, *E. coli* BL21 (DE3) lysates were prepared and protein extract was quantified by measuring A_{280}.

Equivalent amounts of total protein (20 μg) were separated on a 12.5% SDS-PAGE, transferred to a nitrocellulose membrane (Hybond ECL, Amersham), and analyzed by Western blot using an antiHis monoclonal antibody (Genscript), and the IRDye800 conjugated secondary antibody antimouse IgG (Li-Cor Biosciences). Signals from four-independent western blots were quantified (intensity 10.0) with an Odyssey Infrared Imaging System (Li-Cor Biosciences). The total β-galactosidase activity present in *E. coli* BL21(DE3) cells cotransformed with the orthogonal synthetase/tRNA pair was determined (average of 12–17 independent experiments) as previously described (22).

1. Moura G, et al. (2005) Comparative context analysis of codon pairs on an ORFeome scale. *Genome Biol* 6:R28.
2. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties, and weight matrix choice. *Nucleic Acids Res* 22:4673–4680.
3. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res* 31:3320–3323.
4. Arnold K, Bordoli L, Kopp J, Schwede T (2006) The SWISS-MODEL workspace: A web-based environment for protein structure homology modelling. *Bioinformatics* 22:195–201.
5. Bordoli L, et al. (2009) Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 4:1–13.
6. Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J Mol Graph* 8:52–56.
7. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
8. Demasi M, Davies KJ (2003) Proteasome inhibitors induce intracellular protein aggregation and cell death by an oxygen-dependent mechanism. *FEBS Lett* 542:89–94.
9. Kohrer K, Domdey H (1991) Preparation of high molecular weight RNA. *Methods Enzymol* 194:398–405.
10. Rorabacher DB (1991) Statistical treatment for rejection of deviant values: Critical values of Dixon Q parameter and related subrange ratios at the 95 percent confidence level. *Anal Biochem* 63:139–146.
11. Rocha R, Pereira PJB, Santos MA, Macedo-Ribeiro S (2010) Purification, crystallization, and preliminary X-ray diffraction analysis of the seryl-tRNA synthetase from *Candida albicans*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 67:153–156.
12. Brunger AT, et al. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54:905–921.
13. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66:486–501.
14. Zwart PH, et al. (2008) Automated structure solution with the PHENIX suite. *Methods Mol Biol* 426:419–435.
15. Painter J, Merritt EA (2006) Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr D Biol Crystallogr* 62:439–450.
16. Chen VB, et al. (2010) MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* 66:12–21.
17. Konagurthu AS, et al. (2010) MUSTANG-MR structural sieving server: Applications in protein structural analysis and crystallography. *PLoS ONE* 5:e10048.
18. Itoh Y, et al. (2008) Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon *Pyrococcus horikoshii*. *RNA Biol* 5:169–177.
19. Fujinaga M, Berthet-Colominas C, Yaremchuk AD, Tukalo MA, Cusack S (1993) Refined crystal structure of the seryl-tRNA synthetase from *Thermus thermophilus* at 2.5 Å resolution. *J Mol Biol* 234:222–233.
20. Lloyd AJ, Thomann HU, Ibba M, & Soll D (1995) A broadly applicable continuous spectrophotometric assay for measuring aminoacyl-tRNA synthetase activity. *Nucleic Acids Res* 23:2886–2892.
21. Fersht AR, et al. (1975) Active site titration and aminoacyl adenylate binding stoichiometry of aminoacyl-tRNA synthetases. *Biochemistry* 14:1–4.
22. Santos MA, Perreau VM, Tuite MF (1996) Transfer RNA structural change is a key element in the reassignment of the CUG codon in *Candida albicans*. *EMBO J* 15:5060–5068.



Gene	CUGs	Fold
HSP12	0	1.58±0.20
HSP70	0	1.99±0.24
HSP30	0	5.85±0.84
HSP104	0	1.84±0.26
HSP90	0	-1.71±0.24
HSP78	0	1.13±0.21

Fig. S1. CUG-encoded residue location in nonconserved positions is optimized to prevent misfolding induced by Ser-to-Leu exchange. (*A*, *B*) Spatial distribution of CUG-encoded residues in the catalytic subunit of cAMP-dependent protein kinase (isoform 2, TPK2) from *C. albicans* and *S. cerevisiae*. Surface (*Left*) and cartoon (*Right*) representations of *C. albicans* TPK2 (*A*) and *S. cerevisiae* (*B*) homology models, with the CUG-encoded residues (serine for *C. albicans* and leucine for *S. cerevisiae* TPK2) highlighted in red. *S. cerevisiae* TPK1 (1FOT (1), 74% identity with both TPK2 orthologs) was used as template. The CUG-encoded leucines in *S. cerevisiae* TPK2 are predominantly buried in the protein core, whereas CUG-encoded residues are exposed at the surface in the *C. albicans* protein. The divergent CUG localization suggests a relocation of the CUG-encoded residues from the protein core to the surface to avoid protein structure destabilization and misfolding induced by serine/leucine ambiguous CUG decoding in *C. albicans*. (*C*) Representative fragment of the multiple sequence alignment of fungal TPK2 highlighting the CUG-encoded residues in *C. albicans* (green) and in other members of the CTG- and WGD-clades (cyan). Strictly identical residues are shown in white on a red background and conserved residues are shown in red type. (*D*) Distribution of CUG-encoded residues by secondary structure elements in *S. cerevisiae* and CTG-clade organisms shows that in the latter most of them are found in loops or within α-helical regions. (*E*) Activation of the ubiquitin-proteasome pathway. Highly ambiguous cells were engineered by expressing a mutant *S. cerevisiae* $tRNA^{Leu}_{CAG}$ that efficiently decodes CUG codons as leucine (2), increasing leucine misincorporation to 28% (3). A second *C. albicans* cell-line was created with induced-genetic code ambiguity in standard leucine-UUA via heterologous expression of a mutant *S. cerevisiae* $tRNA^{Ser}_{UAA}$, which recognizes leucine-UUA codons and translates them as serine. Hydrolysis of the fluorogenic substrate succinyl-Leu-Leu-Val-Tyr-MCA was used to assess proteasome function in these two *C. albicans* lines (4). *C. albicans* UUA ambiguous cells showed higher proteasome activity comparatively to wild-type and highly ambiguous CUG decoding cells, which have identical activities. Hence, conversely to other genetic code ambiguities, CUG codon ambiguity in *C. albicans* does not seem to cause massive protein structural changes/misfolding. Fold variation represents the variation in fluorescence intensity (average) for each strain comparative to *C. albicans* wild-type and is expressed as fold ± SD of five independent experiments. $^{*}P < 0.05$. (*F*) Quantification of the heat shock (HSP) mRNA expression levels by real-time qPCR in wild-type and highly ambiguous *C. albicans* cells. Fold induction or repression shown for each gene represents the variation in gene expression (average) for *C. albicans* pUA15 relative to the control situation (pUA12 cells) and is expressed as fold ± SD. of 7–9 independent experiments. $^{*}P < 0.01$; $^{**}P < 0.001$; $^{***}P < 0.0001$. None of the HSPs measured contain CUG codons in their coding sequences.

1 Mashhoon N, Carmel G, Pflugrath JW, Kuret J (2001) Structure of the unliganded cAMP-dependent protein kinase catalytic subunit from *Saccharomyces cerevisiae*. *Arch Biochem Biophys* 387:11–19.

2 Miranda I, et al. (2007) A genetic code alteration is a phenotype diversity generator in the human pathogen *Candida albicans*. *PLoS ONE* 2:e996.

3 Gomes AC, et al. (2007) A genetic code alteration generates a proteome of high diversity in the human pathogen *Candida albicans*. *Genome Biol* 8:R206.

4 Demasi M, Davies KJ (2003) Proteasome inhibitors induce intracellular protein aggregation and cell death by an oxygen-dependent mechanism. *FEBS Lett* 542:89–94.

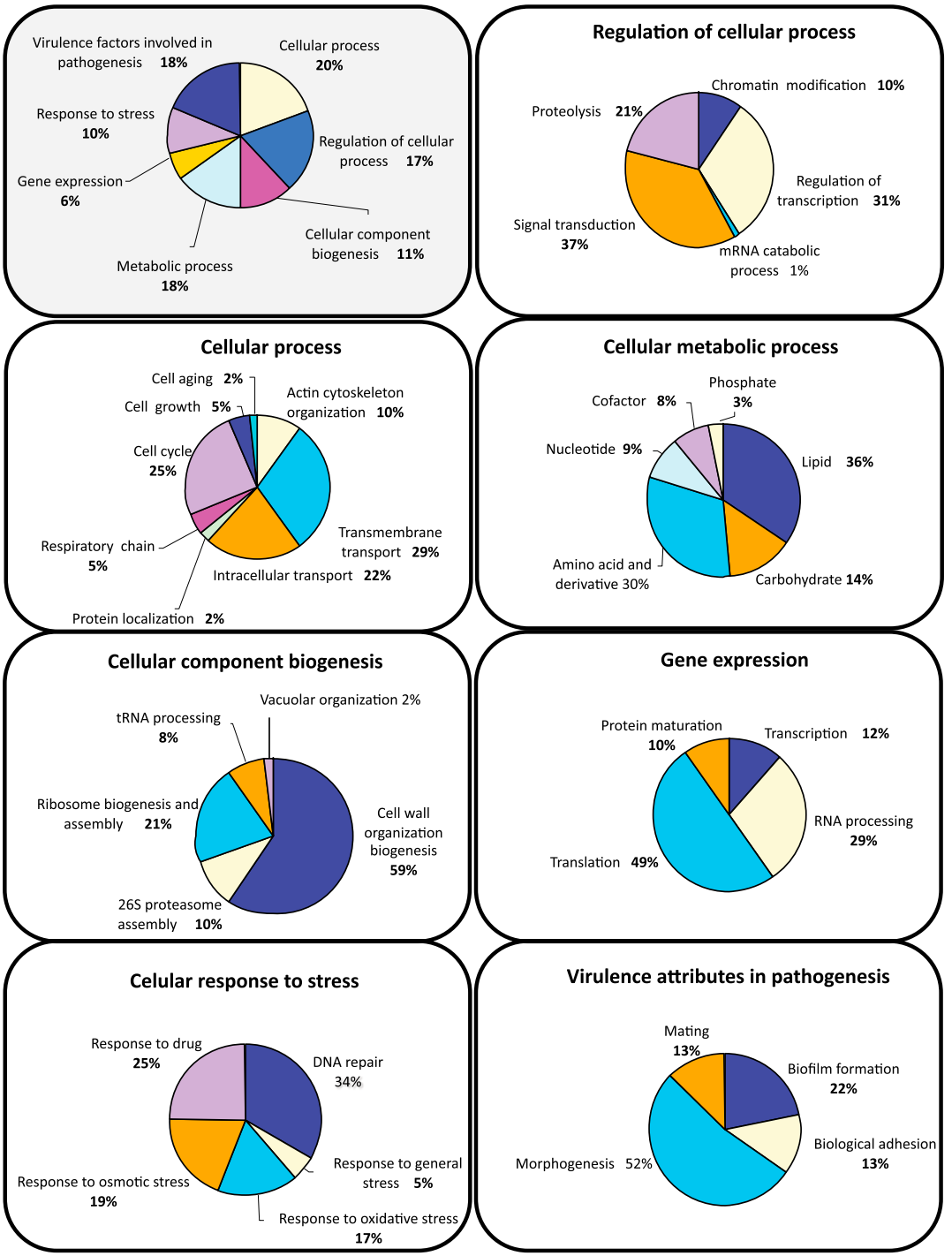


Fig. S2. CUG-containing ORFs are widely distributed by multiple functional categories. Despite the wide distribution of *C. albicans* proteins containing CUG-encoded residues over all cellular processes, a representative group is involved in membrane and cell-wall structure and composition, and in other virulence attributes such as filamentous growth.

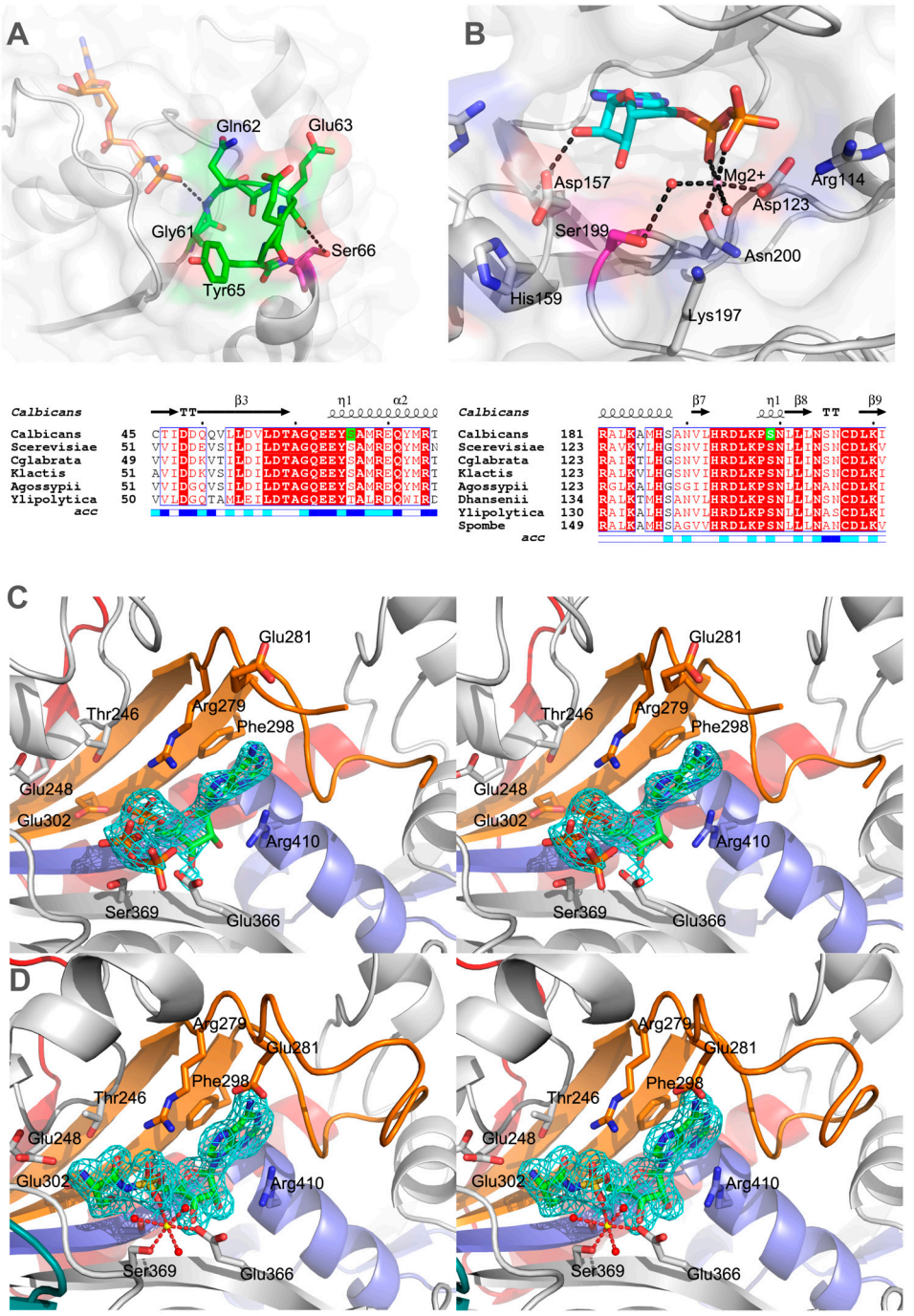


Fig. S3. Functional impact of serine/leucine incorporation in *C. albicans* proteins (*A*, *B*) and plasticity of the SerRS active site in the first steps of the aminoacylation reaction (*C*, *D*). (*A*) Cartoon representation of *C. albicans* Ras1 GTPase homology model (based on the human Ras structure 3K8Y (); 65% amino acid sequence identity) in the vicinity of the CUG-encoded Ser66 (carbon atoms in magenta). The active site bound nonhydrolyzable GTP analogue (GNP) is shown in orange (carbon atoms). Ser66 is within the conserved switch II region (2) (residues 62–65, carbon atoms in green). In human Ras, the side-chain of the structurally equivalent serine participates in an extended hydrogen bond network, correctly orienting Glu62 (Glu63 in this model) to stabilize the catalytic Gln61 (Gln62 in this model) into the active site poised for GTP hydrolysis (1). Mutations in this region, locking Ras in the active GTP-bound conformation, have been identified in human tumours (3) and shown to generate transient activated phenotypes in *S. cerevisiae* (4). The incorporation of leucine at position 66 will disturb this hydrogen bond network, likely interfering with GTP hydrolysis. (*B*) Cartoon representation of *C. albicans* Cek1 protein kinase homology model (based on the yeast FUS3 structure 2B9F (5); 58% amino acid sequence identity). The invariant CUG-encoded serine (carbon atoms in magenta) is within the ATP-binding pocket (carbon atoms of bound ADP molecule in blue), and its main chain accepts a hydrogen bond from the ribose hydroxyl group, whereas the side-chain directly positions a water molecule (red sphere) coordinating the active site magnesium ion (pale pink sphere). Introduction of a leucine at position 199, will compromise the CEK1 active site. Residues represented as sticks are color-coded (nitrogen in blue, oxygen in red phosphate in orange, carbon in gray unless stated otherwise), hydrogen bonds are represented by dashed black lines. Representative fragments of Ras1 and CEK1 multiple sequence alignments highlighting the CUG-encoded serines (boxed in green) are shown below each structure. Identical residues are shown in white letters against a red background and conserved residues are represented by red letters. Secondary structure elements from the *C. albicans* Ras1 and CEK1 homology models, are displayed above the alignment. The calculated solvent-accessibility for each residue is indicated by a colored bar below the aligned sequences (blue: accessible; cyan: intermediate; white: buried). (*C*, *D*) Stereoview of the SerRS–Ser197 active site in the SerRS–ATP and SerRS–SerSA complexes. Close-up view of SerRS active site with the conserved structural motifs coloured as in Fig. 2*A*, with ATP (*C*) and SerSA (*D*) molecules as well as interacting active site residues shown as sticks (color code:

red, oxygen; blue, nitrogen; yellow, sulfur; orange, phosphorous; green, carbon in active site ligands). The $2F_o - F_c$ maps for ATP and SerSA, contoured at 1σ, are shown as green meshes. The magnesium ion in the SerRS–SerSA complex (*D*) is shown as a yellow sphere and links to the coordinating atoms represented as red dashed lines. Notice the large inward movement of Glu281, and concomitant ordering of motif 2 loop (A-conformation) upon SerSA binding. SerSA binding induces a closed conformation (or A-conformation) of the active site motif 2 loop (ordering of residues 283–290), a region also involved in tRNA acceptor stem binding (6). In the SerRS–ATP complex (panel C), nucleotide binding does not induce the switch to the A-conformation (motif 2 loop is partially disordered in this complex). The side chains of Glu281 (180° rotation), Glu230, Arg279, and Arg294 in the active site show the largest differences in orientation between the two complexes (panels *C* and *D*). All SerSA and ATP interacting residues are highly conserved in all cytoplasmic SerRSs (Fig. S4), underscoring the similarity of the interactions and ligand binding modes between this fungal enzyme and its archaeal and bacterial homologues (6–8).

1 Buhrman G, Holzapfel G, Fetics S, Mattos C (2010) Allosteric modulation of Ras positions Q61 for a direct role in catalysis. *Proc Natl Acad Sci USA* 107:4931–4936.
2 Krengel U, et al. (1990) Three-dimensional structures of H-ras p21 mutants: Molecular basis for their inability to function as signal switch molecules. *Cell* 62:539–548.
3 Buhrman G, Wink G, Mattos C (2007) Transformation efficiency of RasQ61 mutants linked to structural features of the switch regions in the presence of Raf. *Structure* 15:1618–1629.
4 Marshall MS, Gibbs JB, Scolnick EM, Sigal IS (1987) Regulatory function of the *Saccharomyces cerevisiae* RAS C-terminus. *Mol Cell Biol* 7:2309–2315.
5 Remenyi A, Good MC, Bhattacharyya RP, Lim WA (2005) The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network. *Mol Cell* 20:951–962.
6 Cusack S, Yaremchuk A, Tukalo M (1996) The crystal structure of the ternary complex of *T. thermophilus* seryl–tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J* 15:2834–2842.
7 Belrhali H, et al. (1995) The structural basis for seryl-adenylate and Ap4A synthesis by seryl–tRNA synthetase. *Structure* 3:341–352.
8 Cusack S, Hartlein M, Leberman R (1991) Sequence, structural, and evolutionary relationships between class 2 aminoacyl–tRNA synthetases. *Nucleic Acids Res* 19:3489–3498.

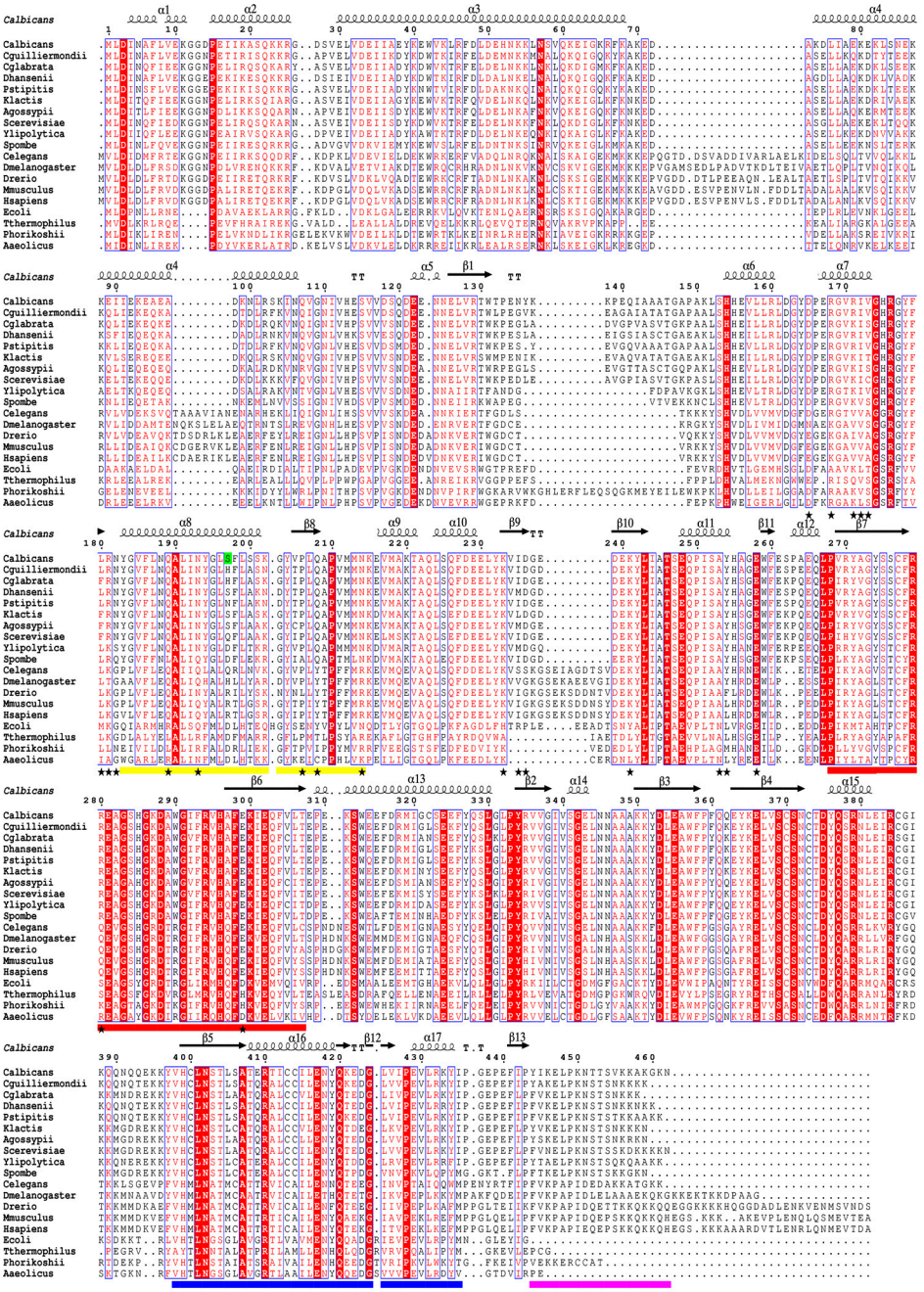


Fig. S4. Cytoplasmic seryl-tRNA synthetases from three kingdoms of life. Amino acid sequence alignment of canonical SerRS from archaea (*Aquifex aeolicus*), bacteria (*Escherichia coli*, *Thermus thermophilus*, and *Pyrococcus horikoshi*), higher eukaryotes (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Dario rerio*, *Mus musculus*, and *Homo sapiens*) and lower eukaryotes (*Candida albicans*, *Candida guilliermondii*, *Candida glabrata*, *Debaryomyces hansenii*, *Pichia stipitis*, *Kluyveromyces lactis*, *Ashbya gossypii*, *Sacharomyces cerevisiae*, *Yarrowia lipolytica*, and *Schizosaccharomyces pombe*). Strictly identical residues are shown in white letters against a red background and conserved residues are represented in red type. The secondary structure elements of *C. albicans* SerRS are shown above the alignment; numbers refer to the *C. albicans* SerRS sequence. The conserved class II synthetase motifs 1, 2, and 3, and the C-terminal tail are underlined in yellow, red, blue, and pink, respectively. The residues belonging to the dimer interface are highlighted by asterisks. The CUG-encoded residue (serine, in the alignment) is boxed in green.

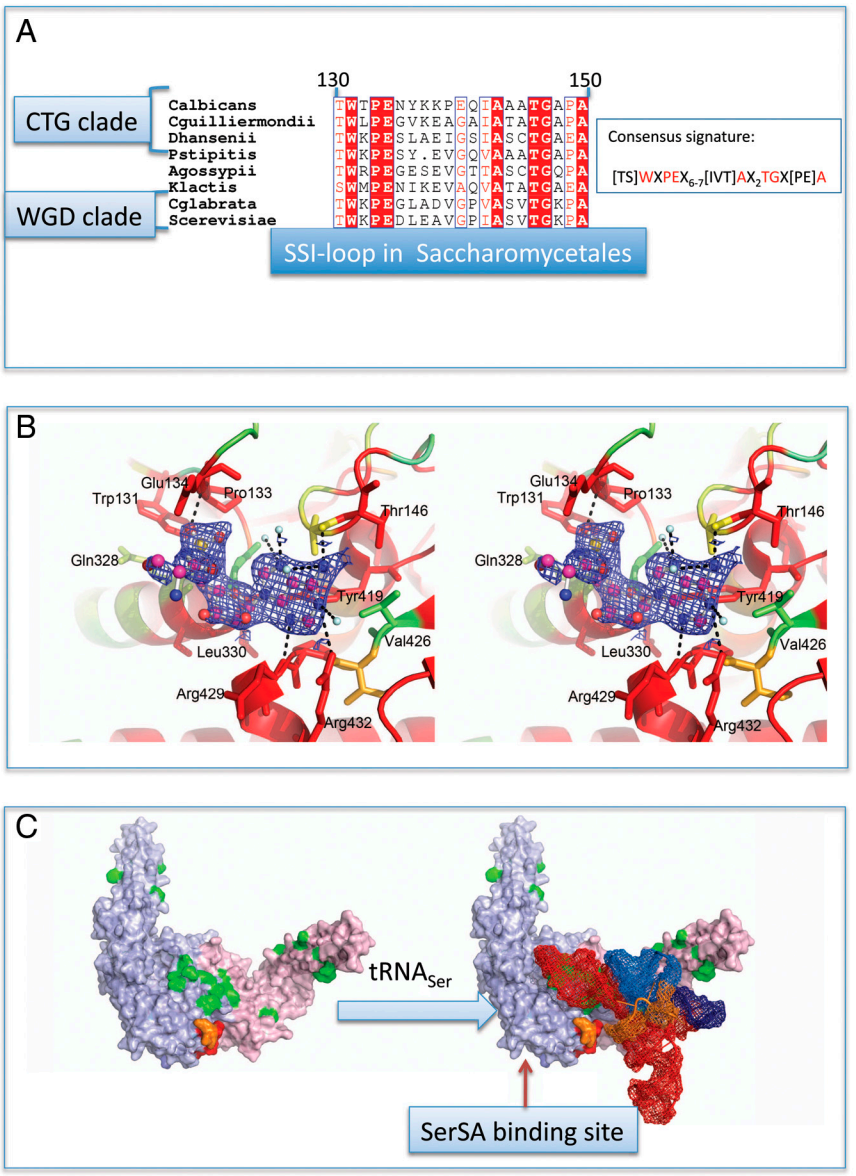


Fig. S5. The *Saccharomycetales*-specific insertion loop and a model for tRNA binding to *C. albicans* SerRS. (*A*) Sequence alignment of the SSI loop connecting strand β1 and helix α6, highly conserved in the CTG-clade and *Saccharomycetales* in general, with the exception of the SerRS of the early diverging lineage *Yarrowia lipolytica* (see Fig. S4). Strictly identical residues are shown in white letters against a red background and conserved residues are represented in red type. Numbers refer to the *C. albicans* SerRS amino acid sequence. (*B*) Stereoview of the surface hydrophobic pocket, highlighting the conservation of the residues interacting with the SerSA molecule (carbon in magenta, oxygen in red, nitrogen in blue, sulfur in orange). Strictly conserved residues in *Saccharomycetales* are represented as red sticks, water molecules as light blue spheres and hydrogen bonds as black dashes. The electron density map, contoured at 1σ, is represented as a dark blue mesh. (*C*) A docking model of SerRS–tRNA^{Ser} complex was generated by superposing the catalytic core domain of *C. albicans* SerRS with *E. coli* ThrRS in complex with its cognate tRNA [1QF6 (1)]. As some regions are missing in the crystal structure of *T. thermophilus* tRNA^{Ser}, including the acceptor arm and anticodon loop (2, 3), a theoretical model of yeast tRNA^{Ser} (5TRA) was further overlaid onto the tRNA^{Thr} in the initial docking model of the SerRS–tRNA^{Thr} complex. The tRNA binds across the two monomers and this rigid docking model emphasizes the putative role of the proximal region of the C-terminal tail (residues 443–449 of the proximal region, coloured in red and residues 450–452 coloured in orange) as a structural element for tRNA interaction. The dimeric *C. albicans* SerRS surface is coloured in violet (monomer A) and pink (monomer B), and residues corresponding to those involved in *T. thermophilus* SerRS-tRNA^{Ser} interaction (2–4) are highlighted in green. The tRNA^{Ser} is coloured red, except for the D-loop (orange), T-loop (blue), and variable loop (purple). The pockets at the protein surface binding SerSA in the SerRS–SerSA structure are shown.

1 Sankaranarayanan R, et al. (1999) The structure of threonyl–tRNA synthetase–tRNA(Thr) complex enlightens its repressor activity and reveals an essential zinc ion in the active site. *Cell* 97:371–381.

2 Cusack S, Yaremchuk A, Tukalo M (1996) The crystal structure of the ternary complex of *T. thermophilus* seryl–tRNA synthetase with tRNA(Ser) and a seryl-adenylate analogue reveals a conformational switch in the active site. *EMBO J* 15:2834–2842.

3 Biou V, Yaremchuk A, Tukalo M, Cusack S (1994) The 2.9 Å crystal structure of *T. thermophilus* seryl–tRNA synthetase complexed with tRNA(Ser). *Science* 263:1404–1410.

4 Itoh Y, et al. (2008) Crystallographic and mutational studies of seryl-tRNA synthetase from the archaeon *Pyrococcus horikoshii*. *RNA Biol* 5:169–177.

Rocha et al. www.pnas.org/cgi/doi/10.1073/pnas.1102835108

Table S1. Accessible surface area* of *C. albicans* CUG-encoded residues

Name	CUG position† Residue	CUG position† Conservation	Identity (PDB code)	Accessibility (Å²) Total	Accessibility (Å²) Backbone	Accessibility (Å²) Side-chain	H-bonds	Secondary structure
ADE4	149		51	20.2	5.54	14.65	0	L
ADE12	322		55	26.03	11.3	14.73	1	HELIX
ADE13	98		65	20.47	9.92	10.55	0	L
AGM1	113		98 (2DKD)	6.32	0.31	6.01	0	L
ARC35	46		37	3.92	0	3.92	0	STRAND
ARG4	361		52	17.4	1.57	15.83	0	L
ATM1	472		46	8.16	0.17	7.99	0	STRAND
ATM1	534		46	20.3	5.13	15.16	1	L
BNI1	1455		51	12.05	1.4	10.66	0	L
CAR2	296		54	19.25	6.36	12.89	0	HELIX
CCP1	314		55	22.88	4.6	18.28	0	HELIX
CDC5	321		47	25.84	8.02	17.82	0	HELIX
CDC68	374		39	22.42	3.94	18.47	0	HELIX
CEK1	199	Serine	58	13.07	2.74	10.32	0	HELIX
CHT4	341		36	23.45	9.96	13.49	0	HELIX
CLB2	301		36	20.09	3.19	16.9	0	HELIX
CLB2	383		36	0	0	0	1	HELIX
CMD1	144		71	21.75	5.75	16	0	HELIX
CPY1	336		73	7.57	1.55	6.01	1	HELIX
CRL1	193		34	11.95	3.57	8.37	0	HELIX
CSE4	170		76	10.9	1.83	9.06	1	HELIX
CYB2	199		45	2.98	1.06	1.92	1	HELIX
CYB2	201		45	1.33	0.11	1.22	0	L
DFR1c	2		99 (1AOE)	32.62	22.83	9.79	0	L
DOA4	663	Serine	34	0	0	0	0	STRAND
ERO1	76		37	6.32	2	4.32	0	L
ERO1	192		37	20.73	6.07	14.66	0	HELIX
FBP1	250	Serine	45	0	0	0	1	L
FOL1	550		40	22.46	6.36	16.09	0	HELIX
GSH2	109		44	5.63	3.2	2.42	1	L
GSH2	392		44	17.58	1.9	15.68	0	HELIX
HAM1	122		32	8.52	0.11	8.41	1	STRAND
HCA4	402		32	31.04	11.23	19.82	0	END
HEM3	163		39	15.71	3.89	11.81	0	L
HOG1	59		49	19.77	9.45	10.32	1	L
HOS1	307		31	26.87	10.61	16.26	0	HELIX
HOS1	377		31	2.64	2.43	0.21	1	HELIX
HST7	264		36	3.08	0	3.08	1	STRAND
ILS1	390		38	18.38	4.57	13.81	0	HELIX
ILV1	278		46	10.59	5.41	5.18	0	L
IPK2	86		35	16.11	2.88	13.23	0	L
KAR3	405		36	19.51	6.76	12.75	0	STRAND
KTR2	425		45	25.34	7.26	18.08	0	L
LAP3	127		39	18.14	5.07	13.07	0	L
LAP3	415		39	13.5	1.39	12.1	0	HELIX
MDL1	620		42	17.93	7.65	10.27	1	HELIX
MKK2	282		40	24.95	10.65	14.3	0	HELIX
MNS1	90		50	21.55	1.69	19.86	0	HELIX
MNS1	216		50	4.2	1.05	3.15	0	L
MSM1	74		31	13.43	0.91	12.51	1	HELIX
MTR2	23		100 (1Q40)	38.19	3.41	34.78	0	L
MYO5	308		41	0.17	0	0.17	1	HELIX
NAM7	777		59	11.56	2.6	8.96	1	HELIX
OSM2	201		30	1.44	1.44	0	1	L
PBS2	219		42	3.65	0	3.65	1	STRAND
PEP8	80		45	10.89	1.92	8.97	0	L
PKC1	793		48	31.61	9.91	21.7	0	L
PPE1	215		34	20.23	5.23	15	0	L
PPR1	67		48	30.56	9.99	20.57	0	L
PRE8	172		63	0.35	0	0.35	2	HELIX
PRP5	362		36	23.85	7.5	16.36	1	L
RAD6	58		81	11	1.83	9.17	0	STRAND
RAS1	66		65	18.85	4.19	14.65	1	HELIX
RFC4	103		68	27.94	7.71	20.23	0	L
RFC4	273		68	3.53	1.01	2.52	2	HELIX
RTS1	389	Serine	63	3.2	2.11	1.09	0	L
SEC23	63		60	0	0	0	0	L
SES1	197		100	13.23	1.82	11.42		HELIX
SGA1	226		36	7.96	0	7.96	1	HELIX
SPE1	286		40	6.53	0.64	5.89	0	HELIX
TYS1	323		73	11.75	3.44	8.31	1	HELIX
UAP1	216		98 (2YQC)	12.56	0.52	12.04	1	L
URA3	29		71	18.17	8.83	9.34	2	HELIX
VPS4	127		77	13.43	5.45	7.98	0	STRAND
VPS24	117		33	13.52	5.48	8.04	2	L
STE4	357		36	14.24	8.4	5.84	0	STRAND
TBP1	227		91	15.5	5.17	10.33	1	HELIX
TCP1	459		37	16.28	4.67	11.61	2	HELIX
THI6	371		30	4.39	0	4.39	2	HELIX
TLG2	298		32	15.13	2.85	12.28	0	HELIX
TOM1	3186		45	19.85	4.2	15.65	0	L
TOP2	476		62	17.14	3.89	13.25	2	L
TOP2	703		62	0.32	0	0.32	2	L
TOP2	955		62	22.6	4.55	18.05	1	L

Rocha et al. www.pnas.org/cgi/doi/10.1073/pnas.1102835108

	CUG position†		Identity	Accessibility (Å²)				Secondary		CUG position†		Identity	Accessibility (Å²)				Secondary
Name	Residue	Conservation	(PDB code)	Total	Backbone	Side-chain	H-bonds	structure	Name	Residue	Conservation	(PDB code)	Total	Backbone	Side-chain	H-bonds	structure
KIC1	221		38	17.66	6.22	6.22	0	HELIX	TPK2	389		74	23.96	8.57	15.39	0	HELIX
KIP2	309	**Serine**	39	0.98	0.77	0.21	2	HELIX	TPK2	439		74	21.62	6.02	15.59	0	L
KIP2	369		39	8.88	1.92	6.96	0	STRAND	XOG1‡	102		99 (1EQC)	22.02	0	22.02	0	HELIX
KNS1	448		32	0.46	0	0.46	1	STRAND	YTA6	632	**Serine**	46	0.21	0	0.21	1	STRAND
KNS1	473		32	21.68	5.78	15.9	1	L									

*Calculated with WhatIf (1).

†The CUG-encoded residue position was scored as nonconserved (black), conserved (gray), or strictly conserved (bold text) in the multiple sequence alignments.

‡Structural analysis of CUG-residue position also described in ref. 2.

1 Vriend G (1990) WHAT IF: A molecular modeling and drug design program. *J Mol Graph* 8:52–56.

2 Cutfield JF, Sullivan PA, Cutfield SM (2000) Minor structural consequences of alternative CUG codon usage (Ser for Leu) in *Candida albicans* exoglucanase. *Protein Eng* 13:735–738.

Table S2. Refinement statistics

	SerRS–Ser197	SerRS–ATP	SerRS–SerSA	SerRS–Leu197
Refinement statistics				
Space group	$P6_122$	$P6_122$	$P6_122$	$P6_122$
Unit-cell parameters (Å)	$a = b = 90.1$, $c = 276.7$	$a = b = 91.2$, $c = 275.8$	$a = b = 91.0$, $c = 275.8$	$a = b = 92.5$, $c = 273.6$
Resolution (Å)	78.1–2.00	79.1–2.55	78.8–2.00	80.1–2.30
R_{factor}/R_{free}*† (%)	16.5/20.7	18.3/21.1	18.1/20.6	19.0/22.0
No. of reflections (working/test set)	43,422/2,197	22,755/1,105	45*m*132/2,277	31,060/1,528
Number of atoms (Average *B*-factor, Å^2)				
Total	3,903 (31.1)	3,669 (51.7)	3,913 (35.3)	3,670 (48.7)
Protein all atoms	3,577 (30.3)	3,580 (51.8)	3,598 (35.0)	3,518 (48.8)
Protein main chain	1,769 (26.0)	1,767 (49.5)	1,772 (31.2)	1,738 (45.3)
Protein side-chain	1,808 (34.5)	1,813 (54.1)	1,826 (38.6)	1,780 (52.1)
Water	326 (39.4)	58 (44.7)	255 (40.2)	152 (46.4)
SerSA—active site	—	—	29 (18.3)	—
ATP—active site	—	31 (50.1)	—	—
SerSA—Surface	—	—	29 (45.4)	—
Mg^{2+}—active site	—	—	1 (29.1)	—
Mg^{2+}	—	—	1 (48.6)	—
Structure quality				
Rmsd bonded lengths (Å)	0.01	0.01	0.01	0.01
Rmsd bonded angles (°)	1.2	1.4	1.0	1.3
Rmsd bonded Bs (Å^2)	5.3	4.8	5.1	5.8
Ramachandran core/allowed/gener allowed/ outliers(%)	92.1/7.1/0.8/0.0	92.9/6.3/0.8/0.0	92.6/6.9/0.5/0.0	92.2/7.0/0.5/0.3
Luzzati coordinate error (Å)	0.23	0.26	0.28	0.31
Solvent accessible surface analysis‡				
Solvent accessible area (Å^2):TotalDimer interface	216,072,287	216,092,273	212,202,283	213,692,265

*$R_{factor} = \sum ||F_{obs}| - |F_{calc}|| / \sum |F_{obs}|$, where $|F_{obs}|$ and $|F_{calc}|$ are observed and calculated structure factor amplitudes, respectively.
†R_{free} is the cross-validation computed for a randomly selected subset of 5% of the total number of reflections, which were not used during refinement.
‡Values obtained using the PISA server (Protein Interfaces, Surfaces and Assemblies) at the European Bioinformatics Institute (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) (1).

1 Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372:774–797.

Other Supporting Information Files

Dataset S1
Dataset S2