

Did tRNA Synthetase Classes Arise on Opposite Strands of the Same Gene?

Structural homology of class II aminoacyl-tRNA synthetases to the HSP70 family and the existence of a gene whose sense and antisense strands code for a dehydrogenase and an HSP70 chaperonin justify reconsideration of a possible sense-antisense ancestry for the two synthetase classes.

The fidelity of protein synthesis resides almost entirely in the 20 aminoacyl-tRNA synthetases (aaRS), which acylate their cognate tRNAs with the appropriate amino acid. The origin of codon-dependent translation presents a challenging intellectual problem in biology, owing to the apparently irreducible complexity represented by their simultaneous appearance. Key to the conundrum is that contemporary aaRS divide into two classes each with ten enzymes (Eriani et al., 1990), whose respective architectures have quite unrelated homologies (Cusack et al., 1990). Moreover, catalytic domains in class I and II aaRS:tRNA complexes from corresponding subclasses have complementary shapes that recognize nonoverlapping surfaces on the tRNA acceptor stems (Ribas de Poublana and Schimmel, 2001). Pairwise binding between classes may therefore have protected tRNAs early in evolution, and the class division likely dates from the dawn of biology.

Class-defining amino acids in both aaRS classes occur in short motifs used by all class members for ATP binding (Carter, 1993). Class I aaRS use two signature sequences, PxxxxHIGH and KMSKS; class II aaRS mainly use the conserved motif 2 sequence, $\phi\phi\pm\phi\text{xxxFRxE/D}$, where ϕ is hydrophobic. Thus, the earliest aaRS function was probably ATP binding, arising from smaller peptides now embedded in contemporary aaRS, which grew larger to achieve amino acid and tRNA specificity as the amino acid repertoire expanded (Ribas de Poublana and Schimmel, 2001; Schimmel et al., 1993).

The remarkable symmetry of the aaRS class distinction, providing basic, acidic, and aromatic side chains within each class, may fall short of providing two functionally competent sets of ten amino acids. Class I lacks glycine and proline, which are closely associated with turns and frequently invariant in contemporary proteins. Class II amino acids, on the other hand, provide no large aliphatic side chains, which are essential for the hydrophobic cores that stabilize globular proteins. This asymmetry complements many other arguments (Rodin and Ohno, 1995) that the class division was intrinsic to the origin of translation and did not result from independent or serial origins.

Antisense translation of the complementary DNA associated with class-defining motifs generates nonrandom homology with class-defining motifs for the opposite class (Figure 1A). This relationship led Rodin and

Ohno (1995) to propose that the aaRS class distinction and equal subclass memberships originated from a primordial gene encoding ancestors of the two classes on opposite strands. Statistical significance of the sense-antisense homology between class-defining catalytic motifs, determined by the Jumble test, appeared reasonable. However, the null hypothesis for such fragmented alignments is hard to formulate and evaluate, so this ingenious solution to the dilemma posed by simultaneous origins of the two aaRS classes remained unconvincing and received little attention.

aaRS catalytic domains provide clues to their evolution. Notably, the class-defining catalytic motifs are positioned by distinct tertiary structural scaffolds. Class I signatures form loops at specific locations of a Rossmann dinucleotide binding fold (RF) built of parallel β structure and α helices, and also found in NAD⁺- and NADP⁺-dependent dehydrogenases. Class II motifs 1 and 2 are inserted into an antiparallel β structure, but no wider kinship has yet been recognized for this fold.

The Rodin-Ohno proposal would be substantially strengthened by demonstrating sense-antisense coding for significant portions of the two structurally distinct aaRS tertiary scaffolds. It is difficult to examine this key prediction using the extant aaRS genes. Evolutionary sequence divergence and codon redundancy have eliminated the most obvious of such relationships from all but the highly conserved signatures.

In this context, LéJohn (LéJohn et al., 1994a, 1994b) characterized what may be a rosetta stone gene in the freshwater mold *Achlya klebsiana* (Figure 1B). One strand encodes a glutamate dehydrogenase, GDH. The antisense strand encodes an HSP70-like chaperonin, whose expression is also stimulated by nitrogen imbalance. Dehydrogenases are based on the RF structure and hence have obvious homology to class I aaRS. Remarkably, we find similarly detailed homology between HSP70 and class II aaRS, justifying reconsideration of the Rodin-Ohno hypothesis.

The GDH was modeled on *H. sapiens* 3-hydroxyacyl-CoA dehydrogenase (HADH-II; 1E3W [Powell et al., 2000]; 20% identity) and D-glycerate dehydrogenase from *H. methylovorum* (1GDH [Goldberg et al., 1994]; 23% identity), the chaperonin on the ATPase domain of bovine HSP70 (2BUP [Flaherty et al., 1994]; 73% identity). ATP binding domains within aaRS classes are quite similar, so one example suffices for our purposes. We therefore selected from the smallest members of each class.

The shared RF topology is evident from the superposition of the first β - α - β crossover connection of HADH-II (rmsd = 1.54 Å for 24 C α atoms; Figure 1C) and 1GDH (rmsd = 2.1 Å for 26 C α atoms) onto that of class I tryptophanyl-tRNA synthetase (TrpRS) (Doublé et al., 1995). The second crossover is recognizably homologous despite assuming a different orientation in the two structures. Pairwise Needleman-Wunsch alignments of the two known dehydrogenase structures matched the superimposed segments to the same amino acids in the *Achlya* gene.

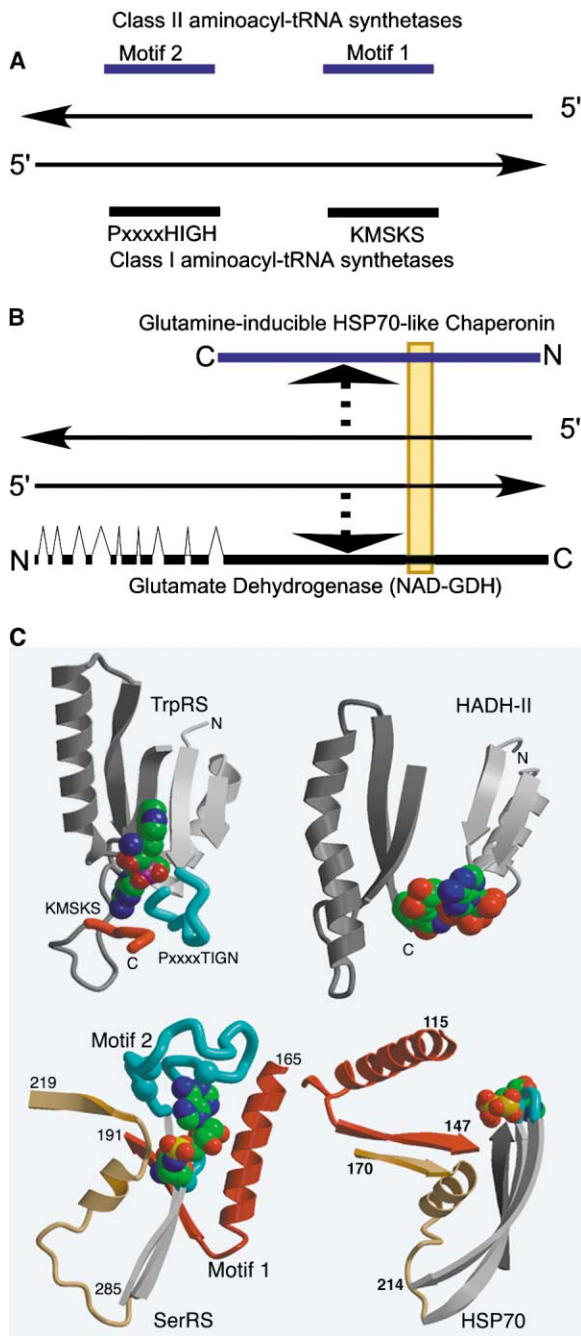


Figure 1. Sense-Antisense Relationships and the aaRS Class Distinction

(A) Antisense coding of class I (PxxxxHIGH; KMSKS) and class II (motifs 1 and 2) aaRS catalytic motifs (Rodin and Ohno, 1995).
(B) Contemporary proteins coded by in-frame, antisense sequences (LéJohn et al., 1994b). The beige box identifies sequences involved in structural superpositions with class I and class II aaRS.
(C) Nucleotide binding sites in models of the two contemporary sense-antisense proteins (right) and corresponding fragments of classes I and II aaRS (left). Superimposed fragments (CDSFIT [CCP4, 1991]) are light gray; aaRS ATP binding signatures are cyan (motif 2 and TIGN, the TrpRS variant of HIGH) and red (motif 1 and KMSKS). The class IIa TxE signature that orients the α -amino group is in the gold-colored turn connecting the β strand and α helix.

Quite surprisingly, residues 191–208 from the HSP70 ATP binding site superimpose almost exactly (rmsd = 1.1 Å) onto the motif 2-containing hairpin of class II SerRS (Cusack et al., 1990). Recognizable homology begins with the motif 1 α helix and β strand, (SerRS residues 163–191), whose orientation in HSP70 (residues 115–147) differs as if domain swapped (Bennett et al., 1994) and continues with a conserved TxE signature that orients the substrate's α -amino group in class IIa aaRS and an α helix that packs against the antiparallel β sheet in both proteins. The SerRS motif 2 fragment overlaps exactly with the HSP70 DLGGGT signature. The similarity of class II aaRS and HSP70/actin ATP binding domains represents an important, previously unrecognized homology analogous to that for the Rossmann fold superfamily.

Thus, 60–100 residue fragments homologous, respectively, to class I and class II aaRS catalytic domains are embedded in the sense and antisense proteins coded by the *Achlya* GDH gene. GDH uses NAD(H) in redox chemistry and HSP70 is an ATPase, while both synthetases release inorganic pyrophosphate. Not surprisingly, the chemically and functionally distinct nucleotides occupy similar, not identical, locations in each pair.

The Rodin-Ohno model would require that the two superimposed fragments align opposite one another. The superimposed fragments in Figure 1C align closely, within the same 5% of the \sim 650 amino acid sense-antisense coding region (Figure 2). The 35 amino acid offset of PxxxxHIGH from motif 2 in Figure 2 does not rule out an ancestral sense-antisense relationship between them. Carboxylate clusters apparently migrated comparable distances by mutation and selection as tropomyosin adapted to filamentous actin (McLachlan and Stewart, 1976).

If, as seems likely, we have selected appropriate homology models for the two *Achlya* proteins, then the structural homology in Figure 1C provides an existence proof that folding instructions for class I RF and class II antiparallel β sheet architectures can be coded by opposite strands of the same gene, as observed for the catalytic motifs. The superpositions establish the first independent support for the Rodin-Ohno hypothesis that the two aaRS classes evolved from sense-antisense ancestors coded by one gene.

The Rodin-Ohno proposal now rests on two qualitatively different, but complementary, examples of sense-antisense coding. They must be evaluated with respect to different null hypotheses, arising from (1) random generation of sequence homology in the catalytic motifs (Rodin and Ohno, 1995) and (2) random generation of tertiary structural homology with the two aaRS classes by convergent evolution, not common ancestry. The joint probability that catalytic motifs and tertiary structural scaffolds would both coincidentally be consistent with sense-antisense coding seems far smaller than that for either coincidence by itself.

Did the *Achlya* GDH gene result from convergent evolution? If so, it is not unique. Sense-antisense coding of HSP70 and GDH proteins occurs in widely dispersed phyla. BLAST searches based on translating the strands opposite those HSP70 genes with totally overlapping sense-antisense ORFs (SAS-ORFs) revealed homology to glutamate dehydrogenases from *H. mediterranei* and

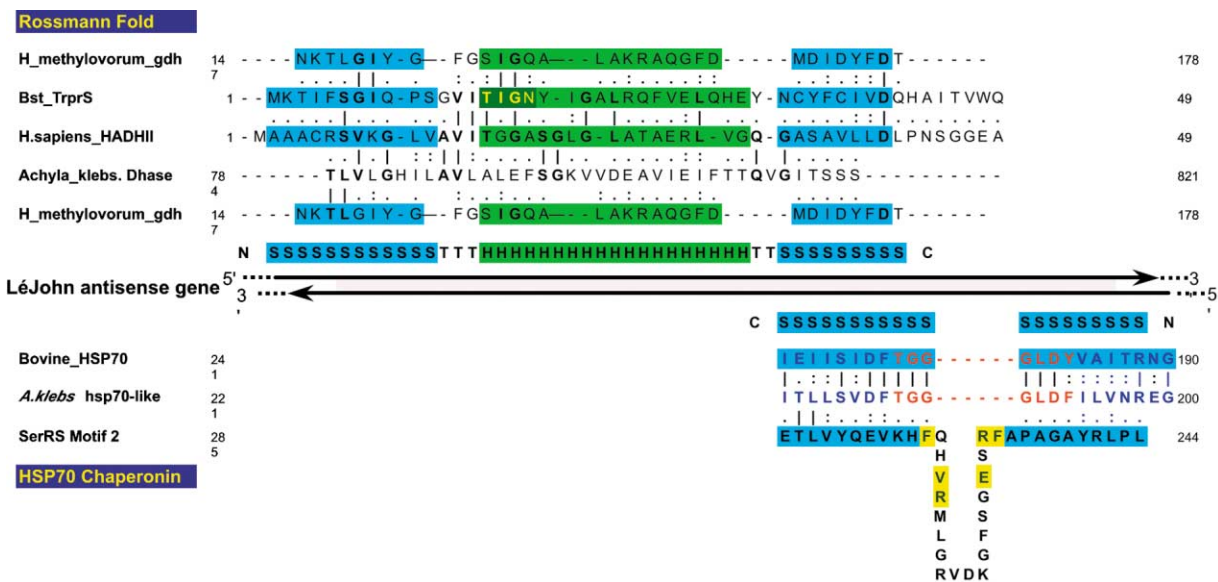


Figure 2. The *A. klebsiana* Sense-Antisense Gene Region Highlighted in Figure 1C
Secondary structures are blue for β strand and green for α helix. The DLGGGT HSP70 signature is highlighted by red letters. TIGN and motif 2 sequences are dark green and yellow, respectively. Alignments were performed using EMBOSS (Rice et al., 2000) and CLUSTALX (Thompson et al., 1997).

D. auraria, as well as *A. klebsiana*. Messenger RNA populations suggest rather widespread antisense coding (Lehner et al., 2002), so DNA sequences representing other large protein superfamilies may also possess residual and, as yet unrecognized, genetic complementarity.

LéJohn commented on the high fraction of GC-rich codons in the *Achlya* gene, and one of us has elaborated elsewhere (A. Addlagatta and W.L.D., unpublished data) the severe codon bias of both HSP70 and dehydrogenase gene families. At least 35 HSP70 sequences have SAS-ORFs with a very similar bias; 97% of the ~650 amino acids are encoded by only 32 codons with a 67% GC bias. Roughly 30% of the short-chain oxidoreductases (dehydrogenase homologs) have SAS ORFs in which 95% of the ~250 amino acids are encoded using only 32 codons with a 77% GC bias. Restriction to this codon set preserves antisense coding information that might be lost when synonymous codons with A or T in the wobble base are read from the antisense strand. It may therefore be a vestige of earlier sense-antisense coding.

The *Achlya* proteins suggest that base-pairing complementarity has important representations in protein folding and function, casting new light on successive levels of the central dogma of molecular biology. Folding instructions for the RF and HSP70 folds are evidently intrinsically related via antisense translation. Binding complementarity between sense and antisense proteins is also unexpected. Coimmunoprecipitation of the *A. klebsiana* GDH and chaperonin (Yang and LéJohn, 1994) implies quaternary surface complementarity, recalling that identified for pairs of class I and II aaRS (Ribas de Poublana and Schimmel, 2001) and earlier proposals of binding interactions between sense-antisense-related peptides (Tropsha et al., 1992).

How, under the Rodin-Ohno hypothesis, might we envision events linking a putative ancestral pair of aaRS to the full set of 20? Simultaneous appearance of a single progenitor for each aaRS class actually affords insight into a possible resolution of the classic chicken-and-egg dilemma afflicting any model for the evolution of codon-dependent translation. We consider three attributes that were probably key to a rudimentary ability to bind and orient ATP, an amino acid, and a tRNA acceptor stem: stability, globularity, and functionality.

Simple polypeptides and RNA may have recognized and interacted with one another, enabling chemical co-evolution from an earlier stage than in conventional RNA world scenarios (Carter and Kraut, 1974). Such behavior may have led to an era in which synthetase and tRNA precursors relaxed constraints on intrinsic stability by mutually stabilizing each other's folded conformations (Ribas de Poublana and Schimmel, 2001).

Could a single pair of class I and II aaRS have encoded globularity? The largest, and hence probably earliest, synthetase subclasses acylate tRNAs, respectively, for small polar (Ser, Thr, Gly, His, Pro; subclass IIa) and large aliphatic (Val, Leu, Ile, Met; subclass Ia) amino acids. Random binary coding using these two, complementary, amino acid types induces molten globule formation and hence considerable fold definition into a high percentage of patterned proteins (Kamtekar et al., 1993). Two synthetases, coding from that repertoire, might similarly have sufficed to produce recognizable protein folds. Subclass speciation via gene duplication then would have enriched the coding repertoire.

Could ancestral synthetases like those we envision have functioned sufficiently to initiate codon-dependent evolution from genes containing a binary coding repertoire? PP-50, a ~50 residue ATP synthase fragment including the Walker-A signature, binds ATP with $K_D \sim 10$

μM (Chuang et al., 1992), suggesting that limited folds like those in Figure 1C could have had emergent nucleotide binding functionality.

Antisense coding implies strict genetic linkage and hence must have had a considerable selective advantage. Early biotic systems likely exploited its high coding efficiency. Recombination and gene duplication may then have suppressed efficiency, relative to other selective pressures directing subsequent evolution. Strict linkage may then have become a liability. This suggests a curious inversion of conventional wisdom that discontinuous genes are a vestige of an initial RNA world. Introns, whose splice signals are not self-complementary, may actually have broken antisense coding and its constraints, freeing genes to evolve separately. Examples like the *Achlya* Gdh gene offer opportunities for their complementary binding interactions to link and regulate alternate parallel pathways, perhaps across cellular compartments. Such regulatory circuitry may account for their persistence in the proteome.

Acknowledgments

This work was supported at U.N.C. by GM48519 and at H.W.I. by NIDK26546. We thank H. Ke, H. Fried, A. Sancar, M. Edgell, A. Tropsha, L.S. Kaguni, P. Schimmel, and L. Ribas de Poublana for discussions, referees for their insights, and Anthony Adlagatta and Aram Kim for assistance with alignments.

Dedication

We dedicate this work to the memories of Susumu Ohno, an evolutionary pioneer, and Herb LéJohn, who with his research group made the pivotal discoveries of antisense coding, codon usage, and binding affinity of the two gene products with so many intriguing implications.

Charles W. Carter, Jr.¹ and William L. Duax²

¹Department of Biochemistry and Biophysics
CB 7260

University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599

²Hauptman-Woodward Medical Research Institute
73 High Street
Buffalo, New York 14203

References

- Bennett, M.J., Choe, S., and Eisenberg, D. (1994). *Proc. Nat. Acad. Sci. USA* 91, 3127–3131.
- Carter, C.W., Jr. (1993). *Annu. Rev. Biochem.* 62, 715–748.
- Carter, C.W., Jr., and Kraut, J. (1974). *Proc. Natl. Acad. Sci. USA* 71, 283–287.
- CCP4 (Collaborative Computational Project 4) (1991). The SRC(UK) Collaborative Computing Project No 4: A Suite of Programs for Protein Crystallography (Daresbury, U.K.: Daresbury Laboratory).
- Chuang, W.-J., Abeygunawardana, C., Pedersen, P.L., and Mildvan, A.S. (1992). *Biochemistry* 31, 7915–7921.
- Cusack, S., Berthet-Colominas, C., Härtlein, M., Nassar, N., and Leberman, R. (1990). *Nature* 347, 249–255.
- Doublé, S., Bricogne, G., Gilmore, C.J., and Carter, C.W., Jr. (1995). *Structure* 3, 17–31.
- Eriani, G., Delarue, M., Poch, O., Gangloff, J., and Moras, D. (1990). *Nature* 347, 203–206.
- Flaherty, K.M., Wilbanks, S.M., DeLuca-Flaherty, C., and McKay, D.B. (1994). *J. Biol. Chem.* 269, 12899–12907.
- Goldberg, J.D., Yoshida, T., and Brick, P. (1994). *J. Mol. Biol.* 236, 1123–1140.
- Kamtekar, S., Schiffer, J.M., Xiong, H., Babik, J.M., and Hecht, M.H. (1993). *Science* 262, 1680–1685.
- Lehner, B., Williams, G., Campbell, R.D., and Sanderson, C.M. (2002). *Trends Genet.* 18, 63–65.
- LéJohn, H.B., Cameron, L.E., Yang, B., MacBeath, G., Barker, D.S., and Willams, S.A. (1994a). *J. Biol. Chem.* 269, 4513–4522.
- LéJohn, H.B., Cameron, L.E., Yang, B., and Rennie, S.L. (1994b). *J. Biol. Chem.* 269, 4523–4531.
- McLachlan, A.D., and Stewart, M. (1976). *J. Mol. Biol.* 103, 271–298.
- Powell, A.J., Read, J.A., Banfield, M.J., Gunn-Moore, F., Yan, S.D., Lustbader, J., Stern, A.R., Stern, D.M., and Brady, R.L. (2000). *J. Mol. Biol.* 303, 311–327.
- Ribas de Poublana, L., and Schimmel, P. (2001). *Cell* 104, 191–193.
- Rice, P., Longden, I., and Bleasby, A. (2000). *Trends Genet.* 16, 276–277.
- Rodin, S.N., and Ohno, S. (1995). *Orig. Life Evol. Biosph.* 25, 565–589.
- Schimmel, P., Giegé, R., Moras, D., and Yokoyama, S. (1993). *Proc. Natl. Acad. Sci. USA* 90, 8763–8768.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). *Nucleic Acids Res.* 24, 4876–4882.
- Tropsha, A., Kizer, J.S., and Chaiken, I.M. (1992). *J. Mol. Recognit.* 5, 43–54.
- Yang, B., and LéJohn, H.B. (1994). *J. Biol. Chem.* 269, 4506–4512.