The Danger of Annotation by Analogy: Most “*thiI*” Genes Play No Role in Thiamine Biosynthesis[▽]

Robert A. Bender*

Department of Molecular, Cellular, and Developmental Biology, The University of Michigan, Ann Arbor, Michigan 48109

In the book of Genesis, God gave Adam the power to name all the creatures that he encountered. From the moment that he named the first apple “delicious,” he ran into trouble. It may have looked delicious, but until he actually did the experiment, he had no idea of its real properties. More recently, bioinformatics programs have given us the power to name all the genes whose sequences we encounter. A paper by Martinez-Gomez et al. in this issue (2) shows that inappropriate nomenclature can still get us into trouble. Until we do the experiment, we may have no idea of the real properties of a gene product. Specifically, they show that only the rhodanese domain of the ThiI protein is required for a key thiolation reaction in the synthesis of thiamine. The other two domains (THUMP and AANH) are dispensable. And yet, nearly three-quarters of all genes annotated as *thiI* by the OMNIOME.pep resource encode only the first two domains and lack the rhodanese domain entirely. Worse yet, Pfam, another of the more common resources for information about protein domains, calls the AANH domain a “ThiI domain” and fails to identify a rhodanese domain in either the *Escherichia coli* or *Salmonella enterica* ThiI protein. Thus, proteins with an AANH domain but no rhodanese domain are likely to be annotated as ThiI and those with a rhodanese domain but no AANH domain are likely to be annotated as “unknown function,” exactly the reverse of reality.

To understand where this confusion begins, we need to recall that the *thiI* gene is allelic to the *nuvA* gene of *E. coli*. It has been shown that ThiI/NuvA is responsible for thiolating the uridine residue in tRNA. Thiolation of tRNA requires all three domains of ThiI, especially the second domain, which is involved in activating the target (by adenylation) and transferring the sulfur from the rhodanese domain to the target. The thiolation reaction in thiamine biosynthesis also requires activation by adenylation and transfer of a disulfide bond. Thus, it was tempting to predict (incorrectly) that the chemistry of the thiolation in thiamine synthesis would be completely analogous.

The bioinformatic method of “prediction by analogy” offers great utility and convenience but tenuous logic for naming genes and proteins. Thus, if the *thiI* (*nuvA*) gene from *E. coli* contains a domain for binding, a domain for adenylation, and a domain for transferring sulfur, then the biochemical function of ThiI (NuvA) protein in *S. enterica* or any organism with a similar gene can be predicted (by analogy): it will bind, adenylate, and transfer sulfur. Moreover (and even more dan-

gerously), the physiological role of such similar proteins can also be predicted (by analogy): they will be required for thiamine synthesis. Each of these predictions involves a probabilistic argument, and the compounding of multiple probabilities degrades the confidence level of such “predictions by analogy.”

Several facts about ThiI are well established: *thiI* mutants are auxotrophic for thiamine, specifically for the thiazole component of this essential vitamin (7). *thiI* is allelic to a previously characterized *E. coli* gene, *nuvA* (3). ThiI/NuvA is the enzyme responsible for the thiolation of a uridine residue in tRNA. Three domains of ThiI are essential for the thiolation of tRNA: a THUMP domain that binds tRNA, an AANH domain that activates the uridine residue by adenylation (9), and a rhodanese domain that transfers sulfur to the activated uridine residue (4). To accomplish this transfer of sulfur, a sulfur transferase (IscS) moves the sulfhydryl group from free cysteine to a specific cysteine residue in the rhodanese domain (Cys456), generating a persulfide at this site (1). The ThiI-persulfide then attacks the adenylated uridine residue, releasing the adenylyl group and forming a disulfide bond between Cys456 and the uridine residue. Then a cysteine within the AANH domain (Cys344) attacks the disulfide, leaving a thiouridine on the tRNA and a disulfide between Cys456 and Cys344 (4). Finally, the disulfide between Cys456 and Cys344 is rereduced and the cycle is ready to repeat. Thus, the AANH domain plays two roles, adenylating the uridine residue and breaking the disulfide bond between Cys456 of the rhodanese domain and the tRNA.

In thiamine synthesis, the ThiFS protein complex is the ultimate sulfur donor to thiazole synthase. The thiolation of ThiFS requires activation of a C-terminal glycine residue by adenylation (6) and an IscS-dependent formation of a persulfide on ThiI (5, 6). Thus, it seemed logical to assume that the thiolation chemistry in thiazole synthesis could be predicted by analogy to thiolation of tRNA. The AANH domain of ThiI would adenylate (activate) the ThiFS protein. Thiolation of Cys456 of ThiI by IscS would generate the required persulfide. The persulfide would attack and release the adenylyl group, forming a disulfide bond between the Cys456 of ThiI and the terminal glycine of ThiS. Then transfer of the disulfide bond from ThiS to the Cys344 of the AANH domain of ThiI would leave a thiolated ThiFS and a ThiI with a disulfide that could be rereduced to restart the cycle.

Unfortunately, this prediction by analogy turns out to be incorrect. It had already been shown that ThiS is adenylated by ThiF, not by the AANH domain of ThiI (8). So the only role for the AANH domain appeared to be the ability of Cys344 to accept the disulfide bond from Cys456. However, Martinez-Gomez et al. have now shown that neither the THUMP nor the AANH domain of ThiI is essential. The rhodanese domain

* Mailing address: Department of MCD Biology, The University of Michigan, Ann Arbor, MI 48109-1048. Phone: (734) 975-1776. Fax: (734) 647-0884. E-mail: rbender@umich.edu.

[▽] Published ahead of print on 8 July 2011.

alone can carry out the thiolation reaction *in vivo*. The disulfide between ThiI and ThiS is resolved when a cysteine from ThiS attacks and forms a disulfide link between ThiI and ThiS (8). The Cys344 of ThiI plays no role at all. Thus, the AANH domain is not involved in either the adenylation reaction or the disulfide transfer.

This shines a bright light on a problem in the annotation of *thiI* genes in general. Most of the genes annotated as *thiI* contain two of the three domains of the *E. coli* enzyme (THUMP and AANH), but they lack the rhodanese domain, the only domain actually involved in thiamine biosynthesis. So, the work of Martinez-Gomez et al. shows a need to revise the annotation of most of the genes currently annotated as *thiI*. Equally important, their work also suggests some novel chemistry at work in thiazole synthesis. When a genetic approach failed to find a suppressor that could replace ThiI in thiazole synthesis, they looked for physiological conditions that might bypass the ThiI activity and found that the addition of exogenous cysteine to the growth medium allowed *thiI* mutants to grow. This suppression of the *thiI* defect was seen even in double mutants with *iscS*, *gshA*, and *sufS* but did not bypass the need for ThiFS. Thus, there must exist a novel way to mobilize sulfur from cysteine into thiazole (and perhaps elsewhere). This demonstrates that even in organisms whose physiology is as well studied as the enteric bacteria, there are still mysteries to be found.

The work of the Downs lab has long argued that there is more to physiology than the core pathways on our metabolic charts. The “leaky reactions” and “lack of specificity” seen in many enzymes are not errors to be overlooked; they are the stuff of evolution and of survival. Their demonstration that there is a novel, cysteine-dependent sulfur transferase pathway

available in *S. enterica* opens up a whole path for investigation. The work of Martinez-Gomez et al. cautions us all to be careful of the article that we use in describing an enzyme. It is always safer to say “a thiolase” than “the thiolase.” The facts of life are seldom as simple as they seem. Just because it looks like a duck and quacks like a duck, we should not assume anything without experimental verification. After all, ducks and dinosaurs are more related than we might have expected.

REFERENCES

1. Kambampati, R., and C. T. Lauhon. 2000. Evidence for the transfer of sulfane sulfur from IscS to ThiI during the *in vitro* biosynthesis of 4-thiouridine in *Escherichia coli* tRNA. *J. Biol. Chem.* **275**:10727–10730.
2. Martinez-Gomez, N. C., L. D. Palmer, E. Vivas, P. L. Roach, and D. M. Downs. 2011. The rhodanese domain of ThiI is both necessary and sufficient for synthesis of the thiazole moiety of thiamine in *Salmonella enterica*. *J. Bacteriol.* **193**:4582–4587.
3. Mueller, E. G., C. J. Buck, P. M. Palenchar, L. E. Barnhart, and J. L. Paulson. 1998. Identification of a gene involved in the generation of 4-thiouridine in tRNA. *Nucleic Acids Res.* **26**:2606–2610.
4. Mueller, E. G., P. M. Palenchar, and C. J. Buck. 2001. The role of the cysteine residues of ThiI in the generation of 4-thiouridine in tRNA. *J. Biol. Chem.* **276**:33588–33595.
5. Palenchar, P. M., C. J. Buck, H. Cheng, T. J. Larson, and E. G. Mueller. 2000. Evidence that ThiI, an enzyme shared between thiamin and 4-thiouridine biosynthesis, may be a sulfurtransferase that proceeds through a persulfide intermediate. *J. Biol. Chem.* **275**:8283–8286.
6. Taylor, S. V., et al. 1998. Thiamin biosynthesis in *Escherichia coli*. Identification of ThiS thiocarboxylate as the immediate sulfur donor in the thiazole formation. *J. Biol. Chem.* **273**:16555–16560.
7. Webb, E., K. Claas, and D. M. Downs. 1997. Characterization of thiI, a new gene involved in thiazole biosynthesis in *Salmonella typhimurium*. *J. Bacteriol.* **179**:4399–4402.
8. Xi, J., Y. Ge, C. Kinsland, F. W. McLafferty, and T. P. Begley. 2001. Biosynthesis of the thiazole moiety of thiamin in *Escherichia coli*: identification of an acyldisulfide-linked protein-protein conjugate that is functionally analogous to the ubiquitin/E1 complex. *Proc. Natl. Acad. Sci. U. S. A.* **98**:8513–8518.
9. You, D., T. Xu, F. Yao, X. Zhou, and Z. Deng. 2008. Direct evidence that ThiI is an ATP pyrophosphatase for the adenylation of uridine in 4-thiouridine biosynthesis. *ChemBiochem* **9**:1879–1882.

The views expressed in this Commentary do not necessarily reflect the views of the journal or of ASM.