

Secondary structure adventures with Carl Woese

Harry F Noller

Center for Molecular Biology of RNA and Department of Molecular, Cell and Developmental Biology; University of California at Santa Cruz; Santa Cruz, CA USA

Keywords: Carl Woese, sequence analysis, ribosome, translation, humor

Not long after my arrival at UCSC as an assistant professor, I came across Carl Woese's paper "Molecular Mechanics of Translation: A Reciprocating Ratchet Mechanism."¹ In the days before the crystal structure of tRNA was known, Fuller and Hodgson² had proposed two alternative conformations for its anticodon loop; one was stacked on the 3' side (as later found in the crystal structure) and the other on the 5' side. In an ingenious and elegant model, Woese proposed that the conformation of the loop flips between Fuller and Hodgson's 5'- and 3'-stacked forms during protein synthesis, changing the local direction of the mRNA such that the identities of the tRNA binding sites alternated between binding aminoacyl-tRNA and peptidyl-tRNA. The model predicted that there are no A and P sites, only two binding sites whose identities changed following translation of each codon, and that there would be no translocation of tRNAs in the usual sense—only binding and release. I met Carl in person the following year when he presented a seminar on his ratchet model in Santa Cruz. He was chatting in my colleague Ralph Hinegardner's office in what Carl termed a "Little Jack Horner appointment" (the visitor sits and listens to his host describing "What a good boy am I"). He was of compact stature, and bore a striking resemblance to Oskar Werner in Truffaut's film "Jules and Jim." He projected the impression of a New-Age guru—a shiny black amulet suspended over the front of his black turtleneck sweater and a crown of prematurely white hair. Ralph asked me to explain to Carl what we were doing with ribosomes. I quickly summarized our early experiments that were pointing to a functional role for 16S rRNA. Carl regarded me silently, with a penetrating stare. He then turned to Ralph and said, in an ominous low voice, "I'm going to have some more tanks made as soon as I get back." Carl's beautiful model was, unfortunately, wrong—it was simpler and more elegant than the complex mechanism that Nature actually uses. Unyielding, Carl railed against the A-site-P-site model at every opportunity,^{3,4} and although we ended up enjoying a long, intense, and fruitful collaboration, and became close, life-long friends, I finally gave up trying to describe to him our biochemical and crystallographic results on the A, P, and E sites.

Four years after our first meeting, while on sabbatical in Berlin in 1975, worrying over my dreams of cloning and sequencing the

16S and 23S rRNA genes, my mind began to leap-frog ahead to the next step—if we had their sequences, how would we work out their secondary structures? At first, I had naïvely imagined that getting the primary structure, combined with Nacho Tinoco's nearest-neighbor free-energy rules,⁵ would quickly yield the secondary structure. To this end, two undergrads, Julie Howe and Nigel Crawford, wrote Fortran programs that ran on the IBM 360–40 to produce 2-d matrices of all possible intramolecular base-pairing combinations that gave runs of four consecutive base pairs or more (Fig. 1). It was a step up from a simple dot-matrix plot, in that each dot was replaced by a number: 3 for G-C, 2 for A-U, and 1 for G-U pairs. You could quickly scan the matrix to identify the strongest helices. The problem was that there were a lot of them. During a visit from the mathematician Mike Waterman, I told him that Julie's program estimated that there would be on the order of 10000 possible helices of four or more base pairs in the ~1500-nucleotide long 16S rRNA; how many potential 16S rRNA secondary structures would the combinatorics predict if there are a total of 100 actual helices? Using Stirling's approximation, Mike quickly came up with the number 10^{15} ; to put this in perspective, 10^{80} was the current estimate of the total number of fundamental particles in the known universe. Clearly, we were not going to Xerox all of the possible structures. And the project was not going to yield simply to the Tinoco rules.

Although deducing the cloverleaf structure for tRNA had been relatively easy, there was already an ongoing controversy over the secondary structure of 5S rRNA, which had only 120 nucleotides. Several papers had been published from reputable labs using everything from chemical probing to NMR spectroscopy; but apart from the 5'–3' terminal helix, the proposed structures were all different. Flipping through the journals in the library of the Max Planck Institute, I came across Carl's paper on 5S rRNA⁶ and there it was! His approach originated as a spinoff from a study aimed primarily at determining the phylogenetic relationships between organisms by differences between their 5S rRNA sequences (although Carl soon realized that he would need the much larger 16S rRNA to do the job). By comparing the sequences of only six different bacterial 5S rRNAs, George Fox and Carl had found a common secondary structure compatible with all six sequences!

It was clear to me that this was the way to go, so as Jürgen Brosius was deducing the DNA sequence of an *E. coli* 16S rRNA gene, I contacted Carl and proposed that we collaborate on determining the secondary structure of 16S rRNA. He was enthusiastic, and immediately committed himself to the project. By then, Carl had obtained catalogs of the sequences of T1 RNase

Correspondence to: Harry F Noller; Email: noller@biology.ucsc.edu
Submitted: 01/17/2014; Accepted: 01/23/2014; Published Online: 03/05/2014
<http://dx.doi.org/10.4161/rna.27970>

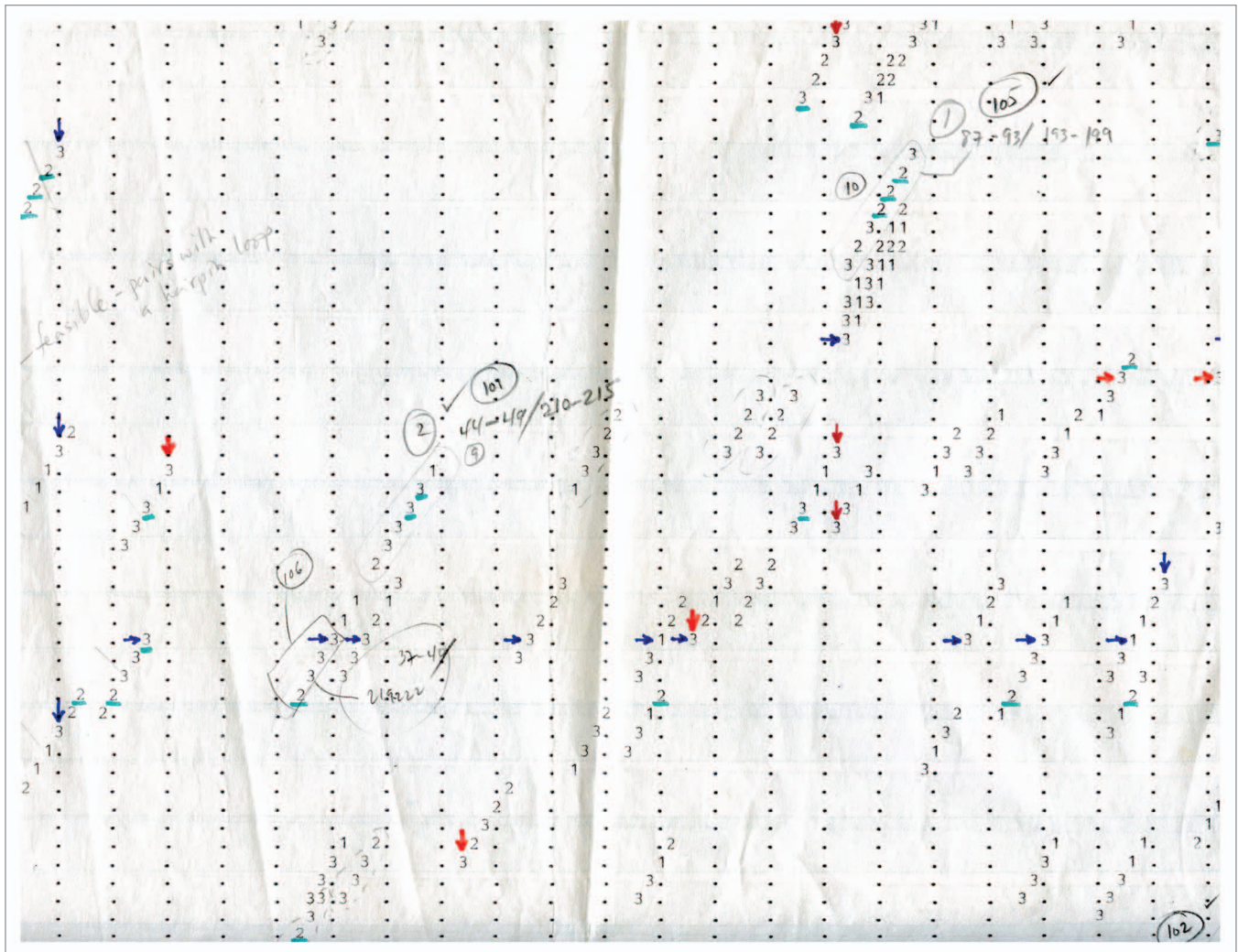


Figure 1. A section of a printout from Nigel Crawford's base-pairing matrix, showing the numerous competing pairing possibilities for a local region of 16S rRNA.

oligonucleotides for the 16S rRNAs from about 100 different bacteria, which I imagined would be sufficient to do the job. Unfortunately, because the T1 oligos are created by cleavage at G residues, and most real RNA helices are G-rich, one strand or the other of any given helix was usually cut up to give oligos that were too short to assign to unique positions in the complete sequence. So Carl's catalogs gave us around eight or so "proven" helices. We defined "proven" as helices that had two or more compensating base changes between two organisms, and importantly, no non-compensated changes, which we would call "disproofs." I asked Mike Waterman what he thought of our seat-of-the-pants rules, and he came to similar conclusions using rigorous statistical methods. We were on our way.

During this time, I visited Carl in Urbana for the first time. He met me at the airport and took me to his home, where he introduced me to his gracious and lovely wife Gay, and poured us each a scotch on the rocks, his cocktail of choice. He then showed me the binder that he kept on the history of his "reciprocating ratchet" model, including a letter from Francis Crick. He

continued to believe in his model and make disparaging remarks about the existence of A and P sites for the rest of his life, long after they appeared in the crystal structures of ribosome complexes. We talked late into the evening sitting in his cramped study, listening to jazz. He had been an amateur jazz pianist (Fig. 2), and we had a few chances to play some tunes together. On one occasion, he even surprised me by hiring a rhythm section (piano, bass, and drums) and a rented saxophone for me to play one evening in his living room. He especially revered Art Tatum, Miles Davis (Carl named his cat "Miles"), and Ella Fitzgerald (especially her rendition of "Miss Otis Regrets"). When he was feeling content, this part of Carl's personality would sometimes emerge in the form of a quiet scat chorus or two to himself.

As we ran out of possibilities using Carl's catalogs, we faced up to the inevitable conclusion that the next step was to obtain complete sequences for additional 16S rRNAs. The optimum sequences would be ones that diverged by about 20% or so from that of *E. coli*, so that you could align them accurately on the conserved parts and there would be not too many insertions and



Figure 2. Carl Woese plays the piano at the author's home in Bonny Doon, ca. 1990.

deletions. Carl had a good read on this from his oligo sequences. We decided on *Bacillus brevis* (a Gram-positive bacterium) and *Halobacterium volcanii* (an archaeon). Following our experience with *coli*, we first attempted to clone a *B. brevis* 16S rRNA gene. We were pretty good at cloning by this time, so we were frustrated and puzzled that we were unable to clone its 16S DNA. After several maddening months, we discovered that the genomic DNA from the *B. brevis* strain we were using contained modifications that prevented cutting by restriction enzymes. In a burst of youthful hubris, I thought, “Hey, let’s sequence the *B. brevis* 16S RNA directly!” Debra Peattie had developed a gel method for direct sequencing of RNA in Wally Gilbert’s lab at Harvard, and we thought we would be able to purify large fragments of 16S rRNA on gels. Plus, we had Carl’s catalog of *B. brevis* oligos. Well, we almost got the whole sequence by direct sequencing—about 95% or so, and a very useful amount for secondary structure analysis. The result of this heroic effort was rejected by *Nucleic Acids Research*, who clearly didn’t appreciate its implications (or the heroics)—it was finally published in the *Journal of Biological*

Chemistry.⁷ On the way, Alexei Kopylov, on a 1-yr postdoctoral fellowship from the Soviet Union, created a spectacular 2-d RNA gel method that resolved hundreds of fragments. And JoAnn Kop did another kind of 2-d diagonal gel in which one dimension was native and the other denaturing, designed to reveal base-paired pairs of sequences; her gel revealed the very elusive pairing of the penultimate stem (the now-famous helix 44), which had so many non-canonical pairs that it had stumped us until then. The *H. volcanii* sequence went more smoothly—Ramesh Gupta came from Carl’s lab to help with its cloning (straight-forward) and sequencing. Growing a 20-L culture of the halophile *H. volcanii* in the lab to get its genomic DNA produced a memorable aroma, not unlike that of Salt Lake City on a hot summer day.

As the *B. brevis* and *H. volcanii* sequences emerged, we established many new helices and fit them into the secondary structure. As we progressed, the problem stayed curiously at the same level of difficulty. This was because, as we fitted new helices, the unstructured parts got smaller. At the same time, the remaining helices were harder, because they were more conserved,

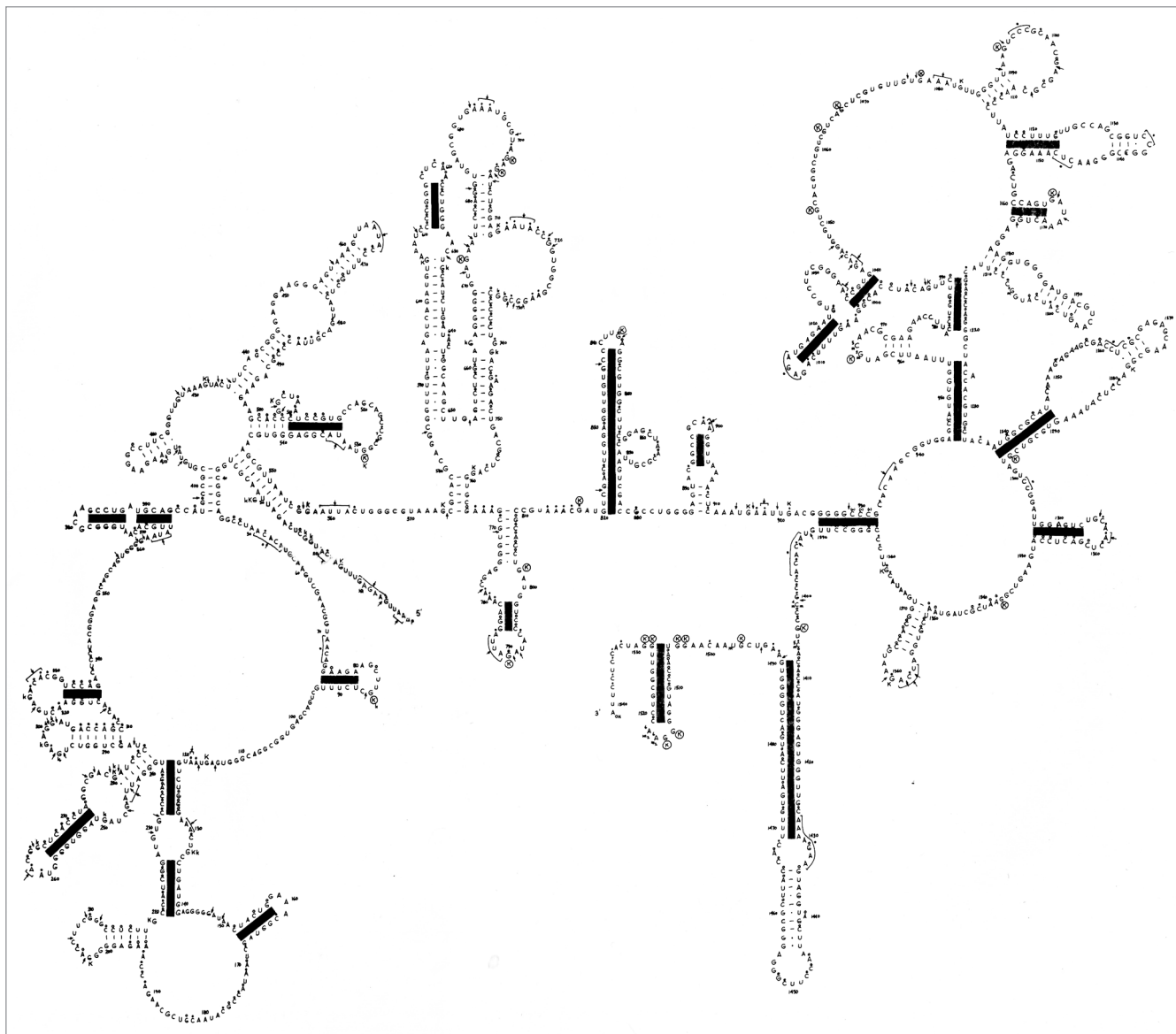


Figure 3. An early version of the 16S rRNA secondary structure (hand-drawn by the author in September, 1979). Black bars indicate the helices that were considered proven by two or more pairs of compensating base changes at that time. Experimental validations of the model, including bases reactive toward kethoxal (K) or susceptible to cleavage by single-strand-specific RNases (arrows), are notated.

so it was harder to find variation to establish “proof.” Eventually, people in other labs began to sequence mitochondrial rRNA gene sequences, and these had sufficient sequence divergence to provide evidence for the final helices. When we were done, I went back and reexamined our Tinoco-based matrices. One surprise was that the most thermodynamically stable helix in *E. coli* 16S rRNA was clearly disproven by the phylogenetic analysis—a further reason for using sequence comparison rather than just ΔG values. As the sequences emerged, we fit in Carl’s T1 RNase oligo sequences, and they agreed with our assignments to the extent that they could weigh in on them. We also had extensive chemical modification data from kethoxal and other single-strand-specific reagents and of course data from partial digestion with RNases from our labs and elsewhere that provided experimental support

for the phylogenetically derived secondary structure. We knew we had it^{8,9} (Fig. 3).

About the time we finished the secondary structure of 16S rRNA, we invited Carl to Santa Cruz for a small celebration with students and postdocs from the lab. Someone had brought a batch of brownies, of the Toklas variety, and set them out on the kitchen table. Later that evening, one of the students approached me with a sense of urgency. “I just saw Carl eat four or five brownies,” she whispered. I checked with Carl to see if he realized what he had done, and he merely nodded as if to indicate that this was nothing I needed to be concerned about. Some time later, a loud “*Ho! Ho! Ho!*” erupted from the corner where Carl had been sitting quietly. “*Ho! Ho! Ho!*” It was impossible to talk to him, as he kept exploding with laughter. I convinced myself

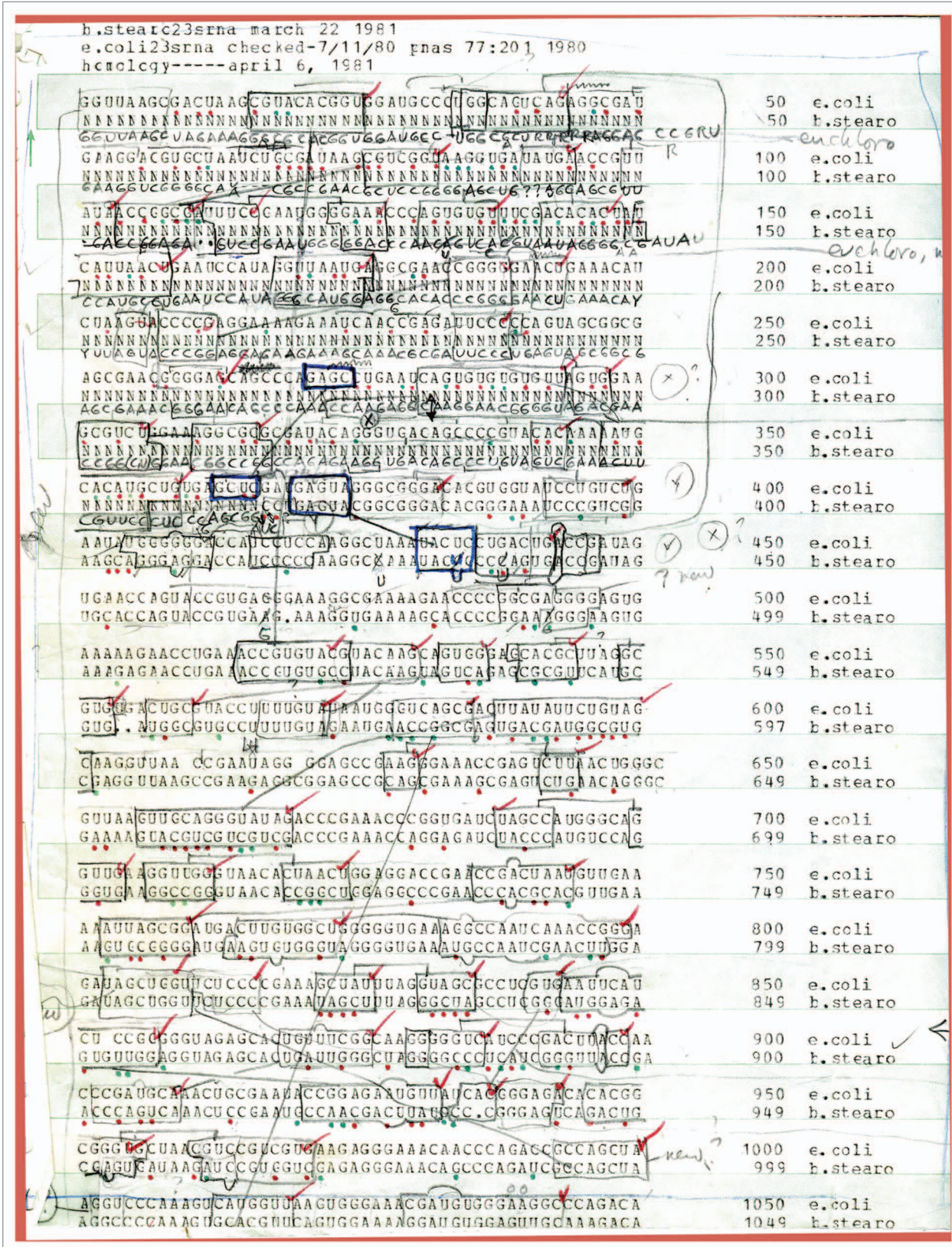


Figure 4. Using "red-dot-green-dot" to work out the secondary structure of 23S rRNA from an alignment of the *E. coli* and *B. stearotherophilus* sequences. Sequences were aligned by Robin Gutell using Ted Goldstein's STREAM alignment editor. A red dot was then drawn by hand at each transversion and a green dot under each transition. The two complementary strands of each helix (boxed sequences connected by pencil lines) were identified by their mirror-symmetric patterns of red and green dots.

that he was basically OK, in spite of the tears flowing down his cheeks. As the guests went home and we went to bed, Carl's laughter continued undiminished. In the morning, I found him in a somber mood, pacing thoughtfully back and forth across the living room. "Good morning, Carl," I said. He nodded a silent acknowledgment and continued to pace. I began to make coffee, but sensed that he was about to say something revelatory. With furrowed brow, he at last spoke. "Last night," came his measured words, "I discovered humor."

Our collaboration flourished from the mid-1970s through the mid-1980s, working over a distance of more than 1800 miles during a time before E-mail or even Fax. We talked on the phone almost daily and sent sequences, data, and secondary structure diagrams (some of which were rubber-stamped with the statement "May a gang of nomadic barbers gang-lather your sister") back and forth by regular mail (4 d from Santa Cruz to Urbana, or the reverse). During this time, of course, Carl also discovered the Archaea. He perennially considered himself underappreciated, even after his election to the National Academy, a MacArthur Award, the National Medal of Science, and many other prizes and medals. When the discovery of the Archaea was published, he called a press conference. The *New York Times* front-page picture showed him sitting in his office with his feet up, passionately waving his arms. The following day, when he treated himself to an ice-cream cone at the local Baskin-Robbins, the pretty young girl at the counter gave him a smile of recognition. Fixing her with his best Profound Genius stare, he asked, "Do you know who I am?" "Sure!" she replied, "You're Bobby Woese's Dad!" In spite of the clear evidence from sequence comparisons, not everyone accepted that the Archaea were indeed a third domain of life. For many years, Carl continued to take a merciless pounding from critics and competitors, which troubled him greatly. It took more than a decade for the dissent to die down, and for the Archaea to finally take their rightful place in all the textbooks.

Besides the secondary structures, interesting biological insights emerged almost as side-products of the project. Carl had recognized a couple of years earlier that the kethoxal-reactive sequences that our lab had identified in 16S rRNA contained many universally conserved sequences. The most conserved, invariant, sequences, of course turned out to be the most interesting ones. At the time, this came as a heterodox observation, because conventional wisdom held that the most important rRNA sequences would be those that formed binding sites for the ribosomal proteins. Instead, the kethoxal reactivity of these conserved sequences meant that they were exposed on the surface of the ribosome, available for interaction with its functional ligands. Carl was one of the few who embraced the idea that ribosome function was based on rRNA. We now know that among the conserved sequences were the decoding center of the 30S subunit (the 530 loop and the 1490 region of 16S rRNA); and in 23S rRNA, the peptidyl transferase catalytic site—sequences of around 15–20 consecutive nucleotides with hardly a single change between the *E. coli* and *H. sapiens* sequences, perhaps the most conserved nucleotide sequences in all of biology. The larger sizes of the eukaryotic rRNAs

sequences were mainly due to a handful of large insertions. In the mitochondria (especially the tiny animal versions of the rRNAs), you could see what would turn out to be the skeleton of the three-dimensional structure, barely holding together the universally conserved functional sites. And of course, Carl recognized that the Archaeal sequences stood apart as a qualitatively distinct and coherent group, leading to his discovery of the third kingdom of life.

When it came to the secondary structure of 23S rRNA, we went straight to getting additional sequences. By the time we cloned and sequenced the *B. stearothermophilus* 23S rRNA, more 23S sequences were beginning to emerge from other labs. One day, during a phone conversation with Carl, he told me that he was looking at an alignment of two sequences and was making a green dot wherever there was a transition and a red dot for every transversion. He said, "I'm not sure what I'm doing here, but I think this might somehow be useful." The following day, I tried doing it, and realized that the secondary structure nearly popped out of the page (Fig. 4). Wherever there was a true helix, its two complementary strands had mirror-symmetric patterns of red and green dots. Using the "red-dot-green-dot" method, we quickly deduced the secondary structure without the help of computer programs, other than in hand-aligning the sequences.¹⁰ In our frenzy of "binge structuring," visual recognition of complementary sequences became so reflexive that it began to have its side effects. One afternoon, sitting at the kitchen table with Carl, I realized that I had stopped listening to what he was saying. My attention had become riveted on a case of Augsburger beer lying on the floor, searching in vain for a sequence complementary to the 5'-AUG in the label. "It's such a short sequence, why can't I see it?" my head seemed to be asking itself. I sensed that it was time to write the paper.

Two decades later, when the crystal structures of the ribosome were solved, the secondary structures of the rRNAs were confirmed in three dimensions. In fact, the secondary structure of 16S rRNA, together with a wealth of biochemical data, made it possible for us to fit the 16S rRNA secondary structure into a 5.5Å electron density map in the absence of any high-resolution X-ray data, with only a couple of errors. Most significantly, the beauty of the phylogenetic approach is that you get the "biological" secondary structure. Physical or biochemical approaches depend on having the physiological structure in your sample, and preserving it throughout your experiment. The physiological structure may depend on many parameters, including the nature and concentration of ions (many biophysical experiments were done in such non-physiological buffers as 1M NaCl and/or in the absence of Mg²⁺ or polyamines) and of course the influence of ribosomal proteins (and indeed ribosome assembly) on rRNA structure. Comparative sequence analysis gives you the bottom line: organisms that violated this pairing died, and have not been heard from since. Carl Woese was one of the first to realize the fundamental importance of RNA to ribosome function, and to life. He also made me realize that thinking about mechanism from an evolutionary point of view is essential to understanding biology. Carl was a profoundly creative and fiercely uncompromising scientist and thinker, who

stood apart from the rest of his contemporaries. To be able to work with him and share the excitement of discovery was a rare and wonderful time.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Woese C. Molecular mechanics of translation: a reciprocating ratchet mechanism. *Nature* 1970; 226:817-20; PMID:5444622; <http://dx.doi.org/10.1038/226817a0>
2. Fuller W, Hodgson A. Conformation of the anticodon loop in tRNA. *Nature* 1967; 215:817-21; PMID:6049728; <http://dx.doi.org/10.1038/215817a0>
3. Woese CR. Just So Stories and Rube Goldberg Machines: Speculations on the Origin of the Protein Synthetic Machinery. In *Ribosomes: Structure, Function and Genetics* (eds. Chambliss, G. et al.) University Park Press, Baltimore, p. 357 (1980).
4. Woese CR. Translation: in retrospect and prospect. *RNA* 2001; 7:1055-67; PMID:11497425; <http://dx.doi.org/10.1017/S1355838201010615>
5. Tinoco I Jr, Borer PN, Dengler B, Levin MD, Uhlenbeck OC, Crothers DM, Bralla J. Improved estimation of secondary structure in ribonucleic acids. *Nat New Biol* 1973; 246:40-1; PMID:4519026; <http://dx.doi.org/10.1038/newbio246040a0>
6. Fox GE, Woese CR. 5S RNA secondary structure. *Nature* 1975; 256:505-7; PMID:808733; <http://dx.doi.org/10.1038/256505a0>
7. Kop J, Kopylov AM, Magrum L, Siegel R, Gupta R, Woese CR, Noller HF. Probing the structure of 16S ribosomal RNA from *Bacillus brevis*. *J Biol Chem* 1984; 259:15287-93; PMID:6439718
8. Woese CR, Magrum LJ, Gupta R, Siegel RB, Stahl DA, Kop J, Crawford N, Brosius J, Gutell R, Hogan JJ, et al. Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 1980; 8:2275-93; PMID:6159576; <http://dx.doi.org/10.1093/nar/8.10.2275>
9. Noller HF, Woese CR. Secondary structure of 16S ribosomal RNA. *Science* 1981; 212:403-11; PMID:6163215; <http://dx.doi.org/10.1126/science.6163215>
10. Noller HF, Kop J, Wheaton V, Brosius J, Gutell RR, Kopylov AM, Dohme F, Herr W, Stahl DA, Gupta R, et al. Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* 1981; 9:6167-89; PMID:7031608; <http://dx.doi.org/10.1093/nar/9.22.6167>

Acknowledgments

I am grateful to the many exceptional colleagues and friends who participated in and contributed to this story, including Jürgen Brosius, Nigel Crawford, Tom Dull, Ted Goldstein, Ramesh Gupta, Robin Gutell, Winship Herr, Jim Hogan, Julie Howe, Poindexter Kennedy, JoAnn Kop, Alexei Kopylov, Lindy Palmer, Don Sleeter, Mike Waterman, and Ginny Wheaton.