

A Conversation with Samie Jaffrey

INTERVIEWER: ANKE SPARMANN

Senior Editor, Nature Structural & Molecular Biology

Samie Jaffrey is the Greenberg–Starr Professor in the Department of Pharmacology at the Weill Cornell Medical College, Cornell University.

Anke Sparmann: Your lab combines tool development using both synthetic and chemical biology approaches and applies these approaches to questions of RNA biology. Can you give us an overview of your work?

Dr. Jaffrey: I'm going to talk about our work on RNA modifications and, in particular, methylated adenosine. Most of the mRNA [messenger RNA] in the cell is composed of A, C, G, and U, but a tiny bit—maybe one in every 300–400 adenosines—is methylated, and when it's methylated on the nitrogen of adenosine, it's called m⁶A. We normally don't think of mRNA as containing that many modifications; we think of tRNA [transfer RNA] and other RNAs as having modifications. This was discovered way back in 1974 when scientists were beginning to understand how mRNA capping and other phenomena occur. When they were doing metabolic labeling to study the m⁷G cap on mRNA, they ended up inadvertently discovering that there were actually methyl modifications inside mRNA, and with some clever biochemistry they figured out it was m⁶A. The people who did this were Dawn Kelley and Robert Perry at Fox Chase and Fritz Rottman and other researchers like Jim Darnell. Some of the pioneers of molecular biology were involved in discovering this modification.

m⁶A was known to be low abundance, but even though it was discovered way back then, almost nothing has been known since about what its function could be. It's very appealing to think that it might be a modification like phosphorylation is for proteins. The potential that m⁶A has some sort of regulatory function has been overlooked, and it's for many reasons. First, some people didn't even think that m⁶A was really in mRNA, because when you purify mRNA you actually have a lot of contaminants and they thought maybe m⁶A could have been from tRNA or ribosomal RNA. But people like Jim Darnell did experiments where he did exceptionally high-purity purification, at least to the ability that they had at the time, and they still were able to see m⁶A. So, there was belief—but still controversy—about whether it was really in mRNA.

The other thing is, how do you study m⁶A? m⁶A behaves exactly like adenosine when you do reverse transcription. If you take mRNA and you make cDNA

[complementary DNA], which is the reverse transcription step in molecular biology, you're going to lose any methyl marks that were there. A lot of people abandoned the field and quickly moved into splicing because that was discovered in the end of the '70s. A lot of this was lost and it wasn't even in textbooks, but there were a few researchers who stayed on it. Fritz Rottman at Case Western was really the pioneer who cloned the enzyme. It's called METTL3, or methyltransferase-like enzyme 3.

Then, as with so many other things in molecular biology the breakthroughs came from yeast and plants. Some scientists knocked out the enzyme in yeast and they found a very remarkable sporulation defect. In plants they found seeds would go to one stage of development and stop. That was Rupert Fray's paper in 2008. When we saw that paper we were just completely shocked; we never even heard of this modification.

I had a postdoc who started shortly thereafter; her name is Kate Meyer and she's now at Duke. I said, "We need to figure out what is this m⁶A. Is it really in mRNA? Where is it? Which mRNA? All mRNAs?" It could have been like the cap, where every RNA has it, or a tail or it could be selected. We had no idea. But that paper—and it related to seed development—told us that the effects were precise. The cells didn't just die. The yeast data, which had come out in 2002, really showed us that there was some connection between developmental and cell-fate decisions like sporulation and seed development; they're all developmental type processes. We wanted to figure out what was going on. That was when we applied for our first NIH grants on this topic.

There were m⁶A-binding antibodies that were being developed for other purposes to study DNA methylation in bacteria and things like this. So we got those antibodies and we pulled down mRNA fragments, and it turned out those antibodies bound very specific regions of mRNAs, which are the regions that have m⁶A in it. We were able to identify the transcripts in the transcriptome that have it. And it wasn't every transcript; it was very specific transcripts. These transcripts tend to have very unusual features. They tend to have huge exons, and we think those exons are the trigger for the methylation. They tend to occur in transcripts that encode regulators of cell fate

and development, just like had been predicted from the work in plants and yeast.

Even though we know now which transcripts have m⁶A based on the original mapping studies from 2012 and some newer methods that we've developed since then to really pinpoint the exact adenosine residue, the functions of m⁶A are still not completely clear. Jim Darnell's great work in the '70s showed the first—and even now the most well-established—function, which is that transcripts that have m⁶A are not as stable as transcripts that don't have it. So, m⁶A is a mark for instability. But besides this function of m⁶A controlling mRNA stability, ways to think about the purpose of m⁶A and how m⁶A transcripts are handled in this cell aren't quite clear.

Anke Sparmann: Is there any idea if the mark is dynamic? Is it removed and then put there?

Dr. Jaffrey: This is the source of a great deal of controversy and debate and a lot of people have different opinions on this. Let me just say that methylation patterns are very similar in almost every tissue. There may be subtle differences but the vast majority of modification seems to be the same, which would go against the idea of some sort of dynamics. The ability to be methylated seems to be hardwired based on gene structure, on the size of the exons. The exons are the same size everywhere because they come from the same DNA, so it doesn't seem that there are dramatic dynamics, but there may be subtle dynamics.

There is evidence for an eraser protein called ALKBH5 that does demethylate m⁶A. It's highly enriched in testes. Animals deficient in ALKBH5 are completely normal, so that does imply that most cell types don't need it, but testes and other germ tissues are abnormal. So, it may have specific roles, but it may not be general. One thing that is interesting is that ALKBH5 is up-regulated in certain cancers. It could be that these patterns of modification—which we call the epitranscriptome—could be shaped to some degree by demethylation, but most of the pattern of m⁶A seems to be constant and fixed. m⁶A therefore marks a set of transcripts that can be regulated in a specific way, but that regulation is going to occur on the same transcripts and in most tissues if those transcripts are expressed. We're trying to understand if there's a logic behind how these transcripts are being regulated.

The core thing that I'm focusing on today is what we discovered by looking at the m⁶A-binding proteins first discovered by Gideon Rechavi and his lab in Israel. They're called the YTH proteins; YTH is the name of the domain that binds the m⁶A. There are five genes in the genome that have YTH domains, but three of them are in the cytosol. The YTH proteins that are in the cytosol are almost identical: they're paralogs. They may have an identical function or a redundant function or a very similar function, but when you look at them, they basically are 450 amino acids or so with a YTH domain and just glutamines and glycines, and pretty much nothing else. Normally, proteins are filled with interesting domains that tell us about what their function is, but this is just filled with these fairly monotonous and uninformative sequences that researchers are starting to realize now are actually

biologically meaningful. These are sequences that people call "low-complexity domains" or "intrinsically disordered sequences"; this is what this protein is almost primarily composed of.

We started purifying this protein and looked at it under the microscope. At first it was a clear solution, but then we started to see little droplets form within the buffer. It turns out that the microscope lamp was warming up the buffer and causing a temperature-associated phase transition. These proteins go from a solution state where they're just floating around as, maybe, monomers. But then they come together as droplets within the liquid buffer due to the temperature being changed. And the temperature would get up to 30°C–34°C. Not too warm, but getting near physiological temperatures. We realized that all of these cytosolic m⁶A-binding proteins, which are nearly identical, do this. They undergo this phase separation, and it is regulated by salts and temperatures and other things.

That's interesting, but the more interesting thing was that when we diluted this protein so it did not form these little droplets and then added in methylated RNAs, these proteins quickly formed droplets induced by the RNA. The RNA triggered the phase transition. It turns out that m⁶A's function is to drive these proteins to undergo phase separations. And these proteins—now with RNAs bound to them—partitioned into these naturally occurring phase-separated structures that cells have, like P-bodies [processing bodies] or stress granules, and even other types of structures like the neuronal transport granules. What we find is these structures are highly enriched in m⁶A-modified RNAs, and m⁶A acts as a targeting signal because of this phase transition property because phase-transitioned structures can easily be partitioned into the larger, naturally occurring droplets. They are then driven into these structures. The function of m⁶A—and frankly the reason why some of these genes may have long exons is to encode m⁶A—is to ensure that the transcript then gets handled or triaged to different parts of the cell.

In some cells when they're triaged to P-bodies, these RNAs are probably degraded if P-bodies have a role in degradation. They may not be the actual sites of degradation, but they may be part of the process of the degradation. In other cells where there's stress, you form stress granules which are protein–RNA assemblies. Those are filled with methylated RNAs and the methylated RNAs are disproportionately trafficked to those structures. And in RNA granules in neurons as well, again we see that both the m⁶A-binding proteins and the m⁶A RNAs are enriched in those structures. What our work is telling us is that m⁶A is a signal that directs RNAs to specific structures, and it does the directing through these phase transition pathways. That's how specificity is conferred upon the directionality of these transcripts: By causing these RNAs to phase-separate they can get to their specific target sites like these P-bodies and stress granules, which are themselves a sort of liquid within a cell, liquids that can then fuse with each other and then accumulate and concentrate in specific areas.

Anke Sparmann: Do you think there's any specificity in that? Some go to P-bodies, others might go to stress

granules. Is that somehow encoded by the binding proteins or the RNAs themselves?

Dr. Jaffrey: This is going to be the big question: How can this process be regulated? The partitioning behavior of proteins when they first undergo phase separation and then when they partition is something that people don't fully understand. But these low-complexity domains aren't just glutamine/glycine. Some are arginine-rich, some are enriched in aromatic residues like tyrosine. There's actually a code associated with these low-complexity domains that we don't fully understand. These domains can also be phosphorylated and have other modifications that can further change their properties. I think when the YTH proteins bind and they have the low-complexity domain with a certain set of modifications they will partition maybe to P-bodies; but then when phosphorylated or modified in some way, then they maybe partition to other structures. Fundamentally, our ability to understand phase separation principles and the specificity with which these principles allow proteins to go to one com-

partment versus another will ultimately determine when m^6A has a destabilizing effect or when it has a translational repression effect or other types of effects that had been seen in this cell.

Anke Sparmann: So the specificity comes not from the mark itself?

Dr. Jaffrey: That's definitely something that I think the field is starting to believe: that the specificity is coming from the proteins that recognize the m^6A and how they are regulated rather than the m^6A marks themselves changing. That being said, there may be some transcripts that will have altered methylation but probably at the level of the proteins that bind it, which are sometimes called the "readers" of m^6A . Those readers may be regulated. Now that we had that framework of phase separation to understand how these proteins behave and what I think is their fundamental physical property, we can start to see how these modifications change their partitioning behavior. That might allow us to make predictions about what they would do to the methylated RNA that they're binding to.