Physics Seminar

March 25, 2019

>>: All right, can I have everyone's attention? I want to introduce our speaker for today. This is Margaret Cheung. She's a professor from the University of Houston, and a specialist in computational biophysics. She's done a lot with UH Clear Lake in the past with our physics program, and, you know, some of the things that we've done collaboratively. If anybody is interested in graduate school at UH, working on your Ph.D., she's a good person to talk to.

And also another thing is that she was the PI on a super computer that we use at UH called UHPC. And so if you'd like to know anything about super computing or about some of the resources that are available to us through the UH system, she's also a good person to talk to.

>>: Thank you, Professor Garrison. It's a great pleasure to be here. It's a wonderful, beautiful building. Just an advertisement about UHPC, it's free. You can basically just apply for an account and then you can have all the resources, one of the fastest computers on campus to run your simulations and your research project.

Today my talk will be to introduce or discuss the physics of memory and learning from the perspective of interacting biomolecules. So before I get into the nitty gritty details about my research, I like to provide a bigger picture. Why am I interested in this problem and why this problem is of interest to physicists or how could physicists ever try to tackle such a complex problem?

So what is the biology of memory and learning? And I quote from one of the Nobel laureate. In 2000 he received a Nobel Prize in medicine and physiology. He said in his autobiography that, "The new biology posits that consciousness is a biological process that will be eventually explained in terms of molecular signaling pathways used by interacting populations of nerve cells." So it's basically (inaudible) nailed down the computing unit that stores our memory and learning to be some small molecules inside a digital -- or inside the nerve cells.

Then the question is, how do neuron cells communicate? There's an interesting concept. Neurons, they are in touch without touching. So how do we know the physiology of these neuron cells? It was actually discovered by neurophysiologists, Santiago Ramon y Cajal, his favorite activity would be to stain different brain tissues, from either human or other mammals, and then slice it thinly and observe them under the microscope.

So at that time, it was believed that inside the brain, everything is just comprised of a bag of gooey stuff. But Cajal mentioned, well it's actually formed by very organized neurons. And these neurons talk with one another and they're in close proximity and there are very, very tiny gaps between these neurons and these gaps are called synapse.

So if we zoom in using a cartoon to represent what happened at these interface and we can see that we have a small synapse, called presynaptic material. The receiving cycle, postsynaptic neurons, and how the two communicate is through the release of chemicals or proteins or serotonins. So what happens is you can have a bag of these chemicals or biomolecules being synthesized and then delivered to the edge of the presynaptic terminals and release it upon stimulation.

So how do memory and learning occur is that once these type of activities happens again and then again, it will stimulate the growth and then the shape changes of this synapse. So there's a word in neuroscience, practice makes perfect. It is because it has shown by training the mouse multiple times to do a certain routine, that their synapse will become very different. And that marks the cellular level of learning that you need to have some kind of connectivity connecting to interacting neurons. And then the geometry and connectivity can be shaped by your learning experience. Okay.

So, the next question will be, how does a neuron decode extra cellular signals? So we know that in our brain is actually a bunch of neuron cells. They are in close proximity, there are chemicals diffusing across the synapse gaps. But how are they actually decoded?

So there are several things that have been known in the science community, that we use calcium as a form of signaling tools. So calciums are abundant in milk. That's why we drink milk. We know calciums are good for bones and muscles, but we also need them for neurons to communicate with one another.

And once they're able to communicate information, then what happens? Underneath that neurons lies proteins called calcium binding proteins. They will catch or try to recruit or bind these calcium molecules, and then encode into information, deciding whether they wanted to activate or deactivate a certain pathway.

For example, if a pathway for us to learn and a pathway to forget go through the same molecule. So how does this molecule know whether this information means remember or forget? So, in order to understand the connectivity, the physics of these pathways, we call protein mediated calcium signaling pathways. It sounds a lot like biology, but I'm going to break it down into a physics problem.

So, one thing that updated, a step closer to being a physics problem, is that we want to know how things change with time. How things change with space. So this is one of our mission is once we receive the signals -- so you have learned about signaling processings in your advanced labs. And we want to know how these patterns or how these frequencies or amplitude of these signals would translate into some sort of meaningful informations that tells the neuron cell whether you want it to forget or you want it to remember.

So let's get to a step closer. Zoom it in. Here is a stained neuron cell, it's about 20 micrometer. If you zoom it in, you see lots of protrusions growing on these dendrite and these are called dendritic spines. The dendritic spines are where -- it's where the connection with the other neurons through synapse. So if I blow this dendritic stem up

into a cartoon, as any typical pages shown in biological textbook, then it shows a bundle of calcium activating proteins.

Upon stimulation, calciums will be allowed to enter and calcium signal will be relayed, passing through a bunch of proteins through phosphorylation. And the outcome of this information relay is the reorganization of the cytoskeleton. They can either break down, so as a result the entire volume of a dendritic spine shrank, or they can grow triple the size and we can -- that denotes to memory learning. Okay?

So, I mentioned I want to break this down into a physics problem. Obviously learning and memory takes from seconds to days. And require repetition for multiple times, multiple hours in order to grow or consolidate a memory information. What I can do now today is I focus on what happened at the first second. So despite that learning and memory could last forever, you know that's why we still remember what happened a decade ago, but thinking about the protein's lifetime usually only lasts a few minute. How is this information retained for that long period of time?

So my first question is to understand it, what happened in the first second? The first second before the calcium was allowed in, there is a protein called calmodulin. And we call it in basal state. Everything is quiet. Calmodulin will be sequestered by a protein, neurogranin. So this protein will not randomly activate any meaningful downstream signals, proteins.

And upon stimulation such as binding of a lot of glutamate, serotonin, doing drugs, and you have sudden stimulation of these calcium cells coming in, it will activate this protein and then it will bind calcium and certain activate downstream kinases, we called calmodulin regulated proteins. And these cycles will lead to changes in synaptic efficacy, either grow to remember or shrink to forget. Okay.

So this is what we know, 2005. And after a decade, we know more about these proteins at the molecular level. We know that calmodulin is a protein with 148 amino acids. It's like a polymer. We know that using biochemical experiments, we can understand the binding affinity between calmodulin and its interacting partners, such as neurogranin. Or CaM kinase II, a calcium calmodulin dependent kinase. And we typically experimentally use a small peptide as a model to understand its kinetic binding mechanisms.

So, how do we actually try to understand these binding kinetics? Within the first second when calcium rush through, what happened? And we know calmodulin is structurally flexible, so there are not really hard as a solid. They can actually flex, just like a muscle. That's why we are made of soft materials. And then it can adopt distinct conformations when they needed to.

So let me tell you some story about calmodulin. It has two domains, connected by a link, shaped like a dumbbell. It binds to four calciums, two at each site. And what happens is it will -- change is conformational, so these lobe will wrap around the target and form a (inaudible) complex.

And interestingly for calmodulin, it can interact over 300 targets. So how the calmodulin know which target to interact, which pathway to activate? This is in close collaboration

with faculty at UTH at the Medical Center, Neal Waxham. So one of the neat things about doing research in Houston is we are at the center of the world's largest medical center. So there's a lot of interesting collaboration going on.

So the first questions we can ask, a simple -- we can ask ourself is that how does the sequence variation would affect binding? So we know there is a small peptide called CaM kinase two and there's another called CaM kinase one. They have the same number of amino acids. They have the same net charges. And the only differences are the position of the hydrophobic residues, meaning the pattern of these residues that hate water is different. They're marked by their numbers here.

So what he did is he measured the kinetic of association. So I'm explaining, the y axis is the observed rate and the x axis is the peptide target concentration. The slope here will tell us the rate, how fast it binds to calmodulin. So in experimental condition, you found that the association rate for CaM kinase one is twice as fast as CaM kinase two.

You would say, so what? That doesn't seem impressive for a physicist. We're dealing with billions and billions or years in the universe. But let me give you this perspective. If today your salary suddenly doubled, how do you feel? Pretty cool. That's basically how the cell feels because inside the cell is a very competitive environment. You have a small leg up is actually a large benefit.

Okay. So given the fact that they are the same length, they have the same net charge, according to the theory they should have the same association rate. So if there are significant difference between the two, of two, then it must happen after they form contacts.

So how do we actually investigate or test this hypothesis? That's why we need modeling, simulations and physics to understand how things work.

First of all we like to build models. Physicists, we build protein models, as well. One way to build it is to include every single atom. It's great. Such a model exists, but it has a big drawback. It takes too much resources to even move a local movement fluctuations and there's a limit in the number of amino acids you can simulate in that box.

So physicists way is to constrain that, meaning I make into low resolution protein model that allow longer integration times. So every amino acids I only use two beads. One bead traces the backbone and one bead captures the chemistry of that amino acids.

And once we have that representation, we need to develop the model to basically let them act like a protein or act like a calmodulin. So we have the interactions for calmodulin, interaction for target, and interaction between calmodulin and target. And it's important that in our model, we do not bias towards a known complex. So we wanted the model to really mimic how the calmodulin would act in solution.

So the interactions for calmodulin target, they break it down into several terms. The first term is to capture the connectivity, the fact that it did not fall apart. We do not have quantum mechanical nature in the simulations because it's coarse graining. If you want to

consider quantum chemical nature, you want the interaction has to be much faster. Over here we're talking about nanosecond.

So the binding between the two adjacent amino acids is captured by a harmonic spring. So they vibrate. And you have angular bonds that represent you fix the particular angle. And you have dihedrals, means you have the torsion or the backbones, and chirality capture the (inaudible) of the side chain.

Once we have the structures, then we also include the van der Waals interactions that captures the dispersion interaction or we say capture the chemistry of the amino acids. We capture the hydrogen bonding interactions. Hydrogen bonding is a unique feature of biological molecules to allow them to form directional interactions with each other and with the solvent.

We also include electrostatic interactions but because it's in solvents, not in vacuum. So we basically included the dielectric constant of 80 in water and also include the ionic strength to properly account for the screening effect. And for the interfacial interactions, they have van der Waals, hydrogen bonding, and electrostatic interactions.

Having all that, then we are ready to run the simulations. So let's take a look at what a typical simulation will look like. Here I'm going to show you in the middle is calmodulin. You have four calciums on the screen. And this is a little target. And you'll see that this target will first look around and then steered to a particular domain through electrostatic interactions. And then because of thermal fluctuations, they will further form, collapse, and bind.

And we do that for thousands of thousands of times. That's why we need super computing power to simulate these trajectories. And we collect them through a probabilistic distribution, and that is called the beta. Oops. Okay.

So how do I compute rate from these individual simulations? What I would do, this is basically an expression devised from Smoluchowski's equation. So this is a guy who actually figured out how things move, how fast the particles move in solution. They said that he kind of modeled the association fact of the two particle coming together into two zones. One zone, if I define have a collision zone, then the molecular details are important. But outside the zone if I can feel like the two doesn't really know each other, they're freely diffusing.

So as a result the rate is the multiplication of these two factor. One is the diffusional coefficient, rate constant outside of a sphere, sphere with a radius B. And the second part is about the molecular detail of how the two particle will collide. And beta represents the probability that they will successfully form in counter complex. Because each time they collide they may just bounce off. It may take several tries for them to really form a meaningful complex.

So basically this is just one equation. You're a physics crowd, I can show you some equations. So basically this is to show you collision zone. Q is actually much larger. I just basically make it into something smaller. And I simulate this process thousands of times and then collect the probability of chance of forming a contact.

And omega here is a quality control. You can see that if I cheat by having a very small sphere, of course I'm going to guarantee events, successful events all the time. So this give you a penalty.

And I also call it graduate student factor because in principle I like to set the boundary as far as possible, but the graduate students want to graduate in four or five years, so we can't really do that. So the students will figure out what is the optimal size of the omega. In our case, .2.

The question is, when this type of algorithm was developed years ago, it was tested on a hard sphere. Hard sphere is easy for them to identify what is a successful contact. But here, these two options are changing interfaces all the time. How do we define a successful event? What is an encounter complex? So we go back and ask the experimentalists. This is a way to do research is to communicate with experimentalists back and forth and get some ideas.

When they measure the rate, they actually essentially measure the intensity of the fluorescence dye tag on the position 75. So when there's an associated event, that intensity will change. So they can't really say they're forming counter complex or not.

Using the same idea, now I define a zone of interaction and if there are contacts between the two that formed within about two -- about 8 angstroms, then I consider a contact. And I call the other parameter Z 75. If I said today I count five contacts in the zone between the target and calmodulin, I can use that expression and compute my rate. You can see that is not much different.

But when I increase the threshold of Z 75, I say I have to form nine contacts between the two in order to call a successful event. You see that a factor of two appears. So obviously that forming for contacts and more tightly formed interface is critical to explain the differences in the experimental measurement.

Then the question is, okay, great, why was that? How does it happen? To do simulations, that's wonderful. That's the reason I got hooked when I do research as an undergrad. So when I was an undergrad I went to lab rotations, but not until I went to join a computational lab I got completely hooked. I love the fact that I know what I'm doing, most of the time. (laughter)

So when I think something interesting, I can load the trajectories and visualize it on my screen and what I see is this is an order parameter as a function of time. There are two major events. One is basically association, is attraction, and (inaudible) and electrostatic interactions. And both of them has that. But once they started to form a more elaborate contact, the two will have to wiggle and adjust to one another and form the final stage, another structural organization, and that part is (inaudible). This part is quite nontrivial.

Then the question is, why CaM kinase two is less successful than CaM kinase one? So we use a lot of the concepts from physics to analyze. We treat them as living matter. You have condensed matter, you have soft matter. Here we use living matter.

So Z represents nonspecific interactions between the target and the N domain. ZC is nonspecific interactions, any interactions. And Z in the x axis represents all contacts. If today the target making contact with N domain and C domain the same, then this line should collapse into that black diagonal line and this is called (inaudible) field model. Everything is the same.

But the fact that you have rise and fall, like, for example, for the C domain perform more than average and then the N domain is less than average. And eventually you have to make way for N domain coming in. It denotes a frustration. This frustration is due to topology. Meaning that some contacts formed early on has to open up to make ways for other contacts to form. And you can see that such topological frustration was more prominent in CaM kinase two.

So we see this variation great. That would delineate the factor of two, slower than CaM kinase one.

Then we can ask the question, so what's good we can analyze through statistics? So we count. We count the number of contacts formed between the target and calmodulin at the early stage and we count the contact between the target and calmodulin late stage, and then we subtract two. And that is the difference of the contact formation of each amino acids on the target.

As you can see that most of them increases, of course, as you move on from folding early stage to late stages, you should have more contacts form. But some contacts are negative, meaning that at late stage, it has to open up and make ways for others to form. But we see that for CaM kinase one, there's only one significant residues causing that frustration. For CaM kinase two, you have a lot. Particularly this notion at 297.

Then we realize that even though the two targets has the same net charge, but they're distribution is very different. But cam kinase one, all those are evenly spread out. Whereas cam kinase two, all of these amino acids are clustered at the N domain. So it forms a very strong Velcro. So it sticks to the calmodulin, it takes a lot more effort to shake it off.

So once we have that as a scientist, we haven't something new, the community might appreciate what we do and we come up with a new mechanism saying that, you know, how does a calmodulin recognize its target? It's through a conformational and mutually induced fit, meaning that because it has to interact, calmodulin has to interact with 300 different targets, it has to go through some secret handshake with its target in order to select which binding complex is reasonable.

So the question is now we know the path going to CaM kinase two. But what about path going to the other way? So obviously that we know that the signaling, the binding mechanism is influenced by calcium. So what is the role of calcium in all of this? And it is inspired by, motivated by early experimental work, about 15 years ago that my collaborator have done very beautiful work.

So how to read this diagram? What it has to do is he measured the amount of calcium released by calmodulin as in y axis and then he measured the time, of course, in x axis.

Note that on the left when it bind to neurogranin, everything happened in less than a second. Whereas if you take a look on the right, everything happened beyond a second. So let's go take a look.

Within a second in neurogranin, so basically you have a profile that traces the number of calciums being released over a period of time. But when he added neurogranin into a peptide, you see that rate suddenly increases, meaning that addition of neurogranin promotes the release of calcium. Target binding promotes calcium release.

However, on the one hand, if you take a look at it, so this profile is the same profile as this one but on a different time scale so they look a little different. When you add CaM kinase to a target peptide, you basically slow down the kinetics to release calcium. And if you add (inaudible) you find that the time to retain in a balanced state increases. That means target binding retain calcium binding. So obviously calmodulin can retain calcium -- can, you know, release calcium, modulated by different target.

Why was that? So usually as graduate students, most of the time during research involve reading literature because this is basically what we dig through the old work and see what are the pieces of information that is undiscovered. We basically flip over all the rocks and see what information has not (inaudible).

With neurogranin, there's not a lot of information. But we know if there's mice knock off the neurogranin -- so they can basically grow some type of mice without specific peptide. And then they can train this mouse through a maze, and train the mouse to go through the maze. And then after three days, they fill the maze with water so the mouse cannot see where the path will be. And then they go to record how successful the mouse or how long it take the mouse for them to reach to the exit.

So they can actually see, like, you know, what is -- the fact of this particular amino acids. It did actually -- some mices, if you have taken one -- if you have taken one peptide out, they will not learn. Some peptide removed, they can learn but cannot remember.

So for this particular knockout mice is if you take it out, it suddenly lost all the spatial understanding, cannot really recognize the markers on the maze that help them to go through the exit. Okay? And biochemically, neurogranin has a higher binding affinity for calmodulin without calcium called apoCaM, than holoCaM, meaning calmodulin with calcium. And this behavior is drastically different than most of the calmodulin binding targets.

And interestingly how the biochemists would do is now they know this peptide is important. What they do is they chop down different pieces and then they actually measure the biochemical property with these small segments. And they identify that without these particular segments in blue, they call acidic domain, if they only have the binding domain, then this peptide loses it ability to manipulate calcium binding. So somehow this is important, but we don't know.

Unfortunately there's no structure available for this complex that we have to rely on modeling and limited experimental approach to model, what does it look like? And how

do we do it? We still use the (inaudible) models, and that will allow us to model what a possible complex would look like.

So this is where the machine learning comes in. This seems like such a trendy name. But in physics we do that all the time. We just didn't call it machine learning. We have some non-sexy name called self-assembled neuro net clustering. Machine learning sounds better, doesn't it? So what it do is that using the nuclear magnetic resonance, this is an experimental technique to use it to decide the spatial distance between the isotopes in molecules. And then each little bar here represents a signal when neurogranin binds.

So we use this as a feature to train our database. You can see that we also have simulations if we presented in a complex system in terms of energy landscape. We have the y axis is the number of contacts and x axis is the distance between the two. And each data point here represents the distribution of data.

Then we -- among millions of complexes, we use the limited experiments as feature to fish out, to learn what the structure will look like and this is what we get. And what do we learn? After machine learning, you have to learn something, too. Sometimes not all the information are useful.

We found that calmodulin is still extended. It hasn't collapsed and neurogranin did not form a beautiful helix. It has a kink.

So once we have this structure, so we want to say why this structure is so peculiar that it affect calmodulin's ability to retain calcium? Then we turn in to do all-atom simulations. Now all-atom, we do quantum mechanical calculations. So these are the structures obtained from the protein databank and there's a structure from our simulations.

We basically solvate them, tune the right ph, set a correct ionic strength. And the system became very, very large. It contains hundreds of thousands of atoms, including water. And then we run it under room temperature. Then how to compute the binding energy of the calcium? Well we turn to physics. In physics there is a way to compute free energy through non-equilibrium work. It's called Jarzynski's equality. He's still alive and well. He's about 60 years old.

So what the algorithm is is that I have a state balanced, a calcium balanced released state. Basically I have this nano-tweezer, AKA some kind of forceps. I exert it on my bead, and then I pull it very, very fast thousands of times. Each pull costs about a month of calculations. That's why we need high performance computing, as well.

And we need to compute this process. There are random directions for thousands of times until data converge. Then we can get the free energy difference between the two and that translate into calcium affinity. So how to read the data?

So here we have delta G. G represent free energy. Calcium balanced state and calcium unbalanced state. The second delta -- you can see delta delta -- the second delta represents without a target and with a target. So if delta delta G is negative, meaning that calcium binding will retain calcium. If delta delta G is positive, meaning that target binding will basically release, promote the release of calcium.

So we do that and we do serious calculations and we compare with experiments. This is usually how we validate our approach. In the last rows are the experimental data. These are the free energy, delta delta G kCal per Mole per calcium binding site and for calmodulins but with calcium is minus 3.3 kCal per Mole and for neurogranin is 2.5.

Saying that, negative means CaM kinase binding promotes calcium, whereas neurogranin binding will decrease its affinity. And look at our calculations. For the holoCaM, the number is really close. It's not the same but let me tell you that if you're a computational scientist, you'll cry when you see this number. It's actually pretty good.

And for holoCaM and neurogranin, we didn't get it very close. You can see, wow, this is really a big number. It's exponential. But however it's still positive. So we think that it's probably due to the limitation of modeling and the sampling, but at least it's positive.

Doing simulations allow us to look at why? This is why I like molecular simulations. Now I take those structures, and I zoom it in. Why was that? So this is a canonical bound complex of calmodulin and target. The target is in translucent background. If you blow it up you see that this is a rod in helical form. And these two loops are the two calcium binding loop of calmodulin that wrap around the target. And you see that there's additional hydrogen binding between these two calcium biding sites, and this additional hydrogen binding further stabilizes the structure of the calcium binding loops, and that retains calcium. Does that make sense?

And then for the other one, because neurogranin does not form helical contact, so it does not have this benefit. On the one hand this acidic domain that I mentioned before, they actually stick outward and sit right between these calcium binding sites. You can see that, stick right in between them. So as a result, you can't form this nice beta scaffold between these two and that destabilized calcium binding loop. So as a result, they will promote calcium disassociation whenever neurogranin binds.

So we're able to understand this at the molecular level. So I like to -- how much time do I have?

>>: Let me check something. She says the slides aren't on Blackboard anymore. Oh, you know what? I think the sharing must have stopped.

>>: Do you have any students wanting to do summer research at the CTBP? If you want to, I'm actually receiving applications to recruit students to do research at the Center for Theoretical and Biological Physics.

So the next question is what is the biology of mind? "The new biology of mind is potentially more disturbing because it suggests that not only the body but also mind and the specific molecules that underlie our highest mental processes -- consciousness of self and of others, consciousness of the past and the future -- which evolved from our animal ancestors."

This is pretty scary. So let me tell you very scary information. So this is basically what we do is we try to find the similarity of calmodulin between human, cow, fruit fly, bacteria, rat, chicken, mouse, and frog. But it tells us that for all the vertebrates -- that

means us, pig, frog, fish, anything that has a vertebrae -- has the identical calmodulin. There's 148 of them. That means any mutation will lead to development. So this is a highly conserved molecules. You have the same calmodulin molecule that translate calciums in a fish, in a frog, in us. The only thing that's very different is soybeans, yeast, and bacteria.

So maybe we come up with an interesting idea. Perhaps physics is an evolutionary constraint. We heard about a lot of interactions could be constraint, but perhaps how things move is a constraint, as well.

So how do we actually test our hypothesis is that we basically study the function, the sequence, the structure of calmodulin using biophysics and evolutionary biology to do that. And luckily we are in the medical center. The world's best evolutionary biologist is at Baylor's. He developed a software called Evolutionary Trace that we can just load our sequence in and it can tell us how important each amino acid would be in an evolutionary scale.

And another one is that -- another great -- one of the best protein folding biophysicists at Rice University, he developed a frustratometer that tells us the physical properties of interactions for protein measured by a frustratometer. If the residue is happy to be in that position, it is marked green. If the residue is not happy because of the energetics come from frustration, then we've got highlight in red.

So once we have that, then we lay out those data. Right now people like to say big data. But data is big already. The only idea is it's not the data that's interesting, it's how to think about it, make it valuable is interesting. That's why you like to major in physics, right? How things work.

If we lay out this data, it's very colorful, has no meaning. But however if we want to project them into three dimensional structures, that would be something interesting. Let me first explain what that diagram would be. In the y axis is the evolutionary score. Anything less than five means super important in the evolutionary sense. Anything greater than five, still evolving.

X axis, positive, meaning minimally frustrated, meaning happy residues. They're happy where they are. And the negative represents unhappy residues, right? They will not be happy where they are because of energy frustrations. There are some conflicts and they don't want to be there.

So if we color these residues on three dimensional structures, look by human eyes. Human eyes are the best algorithms than anything. You immediately see clusters of stuff, didn't you? This is why when students want to do simulations, I say first look at it. Does it make sense or not? Right? Because it have to pass our human brain first.

So let's take a look. The first one has minimally frustrated, happy residues, highly conserved. Oh, they are really reasonable. They are conserved, they're forming folding scaffold. And we've known, proteins like to have a folding scaffold. That's why they fold in biological scale.

Another one we see, oh, they are highly frustrated. They're not happy where they are. They move a lot, but they are highly evolution conserved. So these are calcium binding sites. Of course these functional states, they move a lot.

>>: It stopped advancing again.

>>: So both of them are reasonable in a sense. Some places like to fluctuate a lot because they have to. It's a calcium binding site. Some of them have to remain, you know, stable -- okay. And what's interesting from this data is that we found some regions that are happy residues, they are not conserved. And these are the residues at the interface between target and calmodulin.

So if you think about these places are like cassettes. Like if you compile a software, you don't need all of the features in a software. Just compile some of them. So these are like cassettes. You can actually shuffle whenever the function is needed. So we call these modular amino acids.

The last one is pretty surprising, meaning they are not happy residues and they are still evolving. And we notice this is a particular residue called tyrosine 99 that is not located at interface between calmodulin and a protein. It's not located anywhere. It's actually located at the back of the calmodulin. And that particular place is where the calmodulin gets phosphorylated.

So if you think about it, wow, this is pretty cool. Because if we look at the evolutionary tree analysis -- let me tell you that. The nature has designed the best, most stable calmodulin in east. In east, this particular region, there is a highly oily residue. Very happy to be there. But when it started to evolve, become a soybean and this one became phenylalanine, meaning that it started to become less oily.

Interestingly when it evolved from phenylalanine to tyrosine 99, this is where the current, where the (inaudible) moved to fish. And this mutation happens at the sodium potassium ion pump. So this means that when a fish evolve in sea water, it need to start pumping out calciums and this particular amino acid's mutation would contribute to this more elaborative function.

And what's scary to tell you that it's still evolving. So there's still potential for us to evolve into another feature, a more powerful one. I don't know what that is.

So one thing that we learned is that we can also analyze the history of these mutations and we learn something. If we take a look at the -- (inaudible) amino acids from the ancient form, like east now, we found that the protein is less and less stable. The most stable ones will be in east, but it's not interesting.

But when you get into multicellular function, you involve the organization between different tissues, then this protein become more dynamic, more flexible, we call it oil out. So the fact that the protein has to move is an important feature for us to evolve.

Okay. So this is conclusion and outlook. Conformational and mutually induced fit as a mechanism for calmodulin to recognize target that lacks distinct structures. And

calmodulin's progressive mechanism target binding regulates its calcium binding affinity. We also understand the acidic region of neurogranin is key to lessen binding, need more calmodulin for calcium is bidirectional binding of calmodulin target is critical to the reciprocal relation to calcium affinity. And we think that a dynamic is an evolutionary driving force for promiscuous proteins like calmodulin to achieve their binding multi-specificity and diverse biological functions.

Okay. And acknowledgments, I'd like to thank my students, particularly former students, PhD student Pengzhi Zhang and Masters student Hoa Trinh, and former graduate students Qian Wang and post-docs Swarnendu Tripathi. And my collaborators at UTH, Neal Waxham, doing experiments (inaudible) bioinformatic and evolutionary studies on them. And the research funding from the National Science Foundation, National Institutes of Health, Department of Energy, and Center for Theoretical Biological Physics at Rice University. And I thank you for your attention. If you have any questions or are interested in doing summer research at CTBP, please let me know. Thank you.

(Applause.)

>>: Questions? (Inaudible).

>>: Oh, okay. Yeah.

>>: (Inaudible)?

>>: Good questions. I think that we use pretty low level. Mostly for C, Fortran to do the molecular simulations. But for visualization and analysis, we use Python, MATLAB for these purposes, modeling purposes. Yeah.

>>: What -- do you prebuild libraries also to facilitate the buildings of this model?

>>: Oh, yeah. We don't write from scratch anymore. Everything is very modular. So as long as you know how to write "hello world," you're very close to understanding how things work. Yeah. The barrier to do computational research is not that high.

Most of the part is to have confidence in yourself that you can self-learn and you can do it. Usually I learn programming by buying a book, teach yourself C in 21 days. Those are the kind of self-taught books. I think nowadays we have Coursera, Udemy, many free softwares or YouTube that can help you. And at UH we also have workshops that provide free tutorials for students to learn MATLAB or Python or things like that.

I think nowadays it's a must for any job requirement or research program or apply to any research program. You must have that, either experiments or theory. That's a pretty fair statement.

>>: This presentation was more (inaudible) have you ever researched what is it, the (inaudible) have you ever researched that?

>>: It's not in my specialty, but the answer is yes. A lot of (inaudible) engineering imaging using functional MRI. They can actually image the kinds of activities that are in your brain whenever some specific emotion is being provoked. So there are type of

research like that. If you're interested in doing this, Texan Medical Center has the world's best research in this aspect.

Any questions? Yeah?

>>: (Inaudible).

>>: Oh, of course. They actually have, you know, the birth of all the machine learning is actually started from the Seventies when the physicists wanted to understand how to model networks in brain. So there actually (inaudible) very active research using network theories or many different type of I would say modeling approach, how to retain memory, how memory is being associated. I'm using layman's language because I'm not an expert.

Layman's language, how does we learn? We learn through association with others. Right? So that's why some events will trigger a lot of emotions or trigger us to remember some kinds of things we've forgotten. Because in our brain, what the Hebb's theorem is saying neurons that stay together fire together. So a lot of these (inaudible) connectivity is being reused and rewired to relearn information for different activities.

So someone, you know, if you think about super memory people, how do they actually enhance their memories, trying to connect their thoughts and informations together so you can retrieve them better. So these are some of the, I think, interesting research areas. If I'm not doing this, I'd probably be doing that.

>>: (Inaudible) do you think that's true or (inaudible)?

>>: I don't know if this is something that's true or not. I think nobody can really experiment on the human brain. They only use mouse models and how does the results from mouse models can be applied to humans is another story. So I don't know. You can see that this is a highly delicate machine and then there are some glitches, I think. You can expect some glitches happens.

>>: So, you had experimental (inaudible) and there was the type one and the type two gene. They were actually able to remove those from mice and see what happened?

>>: Yeah, they can do that. There are lots of interesting things. So my collaborator -- it's actually pretty gross if you think about it. I said how do you actually get those cells is that you train the mice, and then you kill them and you slice them and observe their behavior. Or what you can do is that usually -- if you wanted to really, like, stimulate neuron cells, actually these nerve cell are harvested on mice when they're still in the belly of the pregnant mice. Yeah. So lots of very gross things going on. Yeah.

But (inaudible) some people like it. I don't know, that's why I think we major in physics. (laughter) I can tell you something interesting. When I was undergraduate student, I knew that I don't like wet lab. It's because when I joined a group and then the post-doc tell me, oh, meet me at 7:00 o'clock in the morning at a slaughterhouse. I was like, why am I going to a slaughterhouse? I went. It's because we wanted to harvest a bovine brain, cow brain from a slaughterhouse. So we need it fresh and put it into ice bucket

(laughter) and carried it and went to purify it to collect the proteins. And I said that's it. That's it. (laughter)

Sometimes you need that experience to know what you don't like. So that's why I think that it is important at your stage just to try out. Just put your foot in a door, whether it's something you like or not. And probably open doors, right? Sometimes you know okay you don't like this or like that and help you shape what you want to do in the future, professional school or grad school or even work. At the end, you can say that's enough for me. I don't like to do science at all. I want to be and Elon Musk and set up SpaceX. He got a degree in physics and then doing business, I think, very successful. So why not?

>>: (Inaudible)

>>: (Inaudible) or people with learning disabilities, like (inaudible) change the way research is done?

>>: Yeah, it is. This is the sole motivation for why these neuroscientists wanted to know, because they wanted to know how the mechanism work, how these mechanism actually regulated to what key proteins. And there are drugs actually targeting calmodulin to make them work more effectively.

And there's a reason why -- we heard about lead poisoning, right? And we know why is that? It's because if lead is being absorbed, it will displace calcium but it will never leave that place. So you have activated calmodulin all the time. See how exhausting it would be. So whenever you have lead poisoning, it's not reversible. It got permanent damage. So understanding these also have environmental impact. Yeah.

Yeah? Question?

>>: (Inaudible) does that mean that (inaudible)?

>>: Oh, yeah. My goodness. You're so up to the literature. This is great. So there are research studies that show that the so-called epigenetics is the decoration on the DNA, will affect how the DNA pack because once you pack it can express different proteins and that impacts the function or the phenyl type of the cells.

So it has been shown in various studies is how the epigenetics -- epi means outside environment -- how the fact that you were exposed to, how the history of your parents or how the habit of yourself impacted the way your gene unfold and eventually impacted how it was expressed. So not only neuron cells. I think that actually it can impact every single aspect of your -- of your expression.

Because you know that in our human body, even though they have different shape and different sizes, but all cells share the same DNA. Right? So my skin cells, my muscle cells, they share the same DNA. The reason they are different is because they are expressed differently. So how do they express differently? Because you can have the same sequence but they pack and fold differently.

So as a result, when they got expressed, the different part of the space is expressed, being compiled. So that's why, you know, it's very interesting, not everything can be explained in molecular level. You get more health conscious (laughter).

>>: So how many atoms are involved in your simulation (inaudible) computer simulations?

>>: This is coarse-grained models. Coarse-grained models, they kind of control with 1,000 beads. But I can go to 100K, 100,000. Yeah. It's a tiny piece compared to yours.

>>: Mine's low resolution.

>>: (Laughter).

(Applause.)

>>: If you haven't already, I have your papers here. Please make sure you get those. And for the graduate students working on your papers, you should be at least have (inaudible) your materials and methods by now and be moving forward on your paper. You might need to read it or (inaudible).

>>: I have a few business cards if you're interested in contacting me about the summer internship program.

(End of class)