Comprehensive Molecular Diagnostics and Advanced Gene Expression Analysis

Prof. Arindam Ghosh

Dr. B.C. Roy Multi-Speciality Medical Research Centre

Indian Institute of Technology Kharagpur

Week 02

Lecture 08 : Realtime PCR (Part 2)

Namaskar. Welcome back everybody. We are today discussing about real time PCR which is a continuation of the last lecture as we are covering the whole real time PCR under tools of molecular diagnostic and gene expression analysis. And today we will be discussing the next part the remaining part that has been covered in the previous lecture under the following headings. We will be discussing what are the components of a real time PCR machine alright it is denoted as qPCR mind it not RT PCR which is reverse transcriptase PCR we discussed in the last day. So, henceforth as I told you qPCR means real time PCR.

We will be discussing the amplification plot, what are the stages of the amplification amplification plot, what do we mean by the c t value in real time PCR. We will be looking to the quantitation methods absolute relative quantitation, we will be looking to the melt curve analysis and finally, the applications of real time PCR. So, there is a lot to discuss strap yourselves. So, first recap from a previous class where we discussed how the fluorescence is quantitated all real time PCR systems uses a fluorophore or a fluorescent dye which is proportional to the amount of the PCR product.

So, what are the components that are present in quantitating fluorescence you all know a lamp which is sent through a I mean an excitation wavelength is sent through a filter which hits a fluorescent molecule in a sample which is again collected through a secondary filter which is amplified using photo multiplier and photo detector and detected by a CCD or a fluoro detector camera alright. So, all of this is incorporated in the real time PCR machine. So, how does a real time PCR machine differs from a traditional thermal cycler normal traditional PCR machine well it is the thermal cycler. So, all these are components of a thermal cycler along with an inbuilt fluorescent emission detection system and an inbuilt computer that can translate the fluorescence data into meaningful results. So, all three are combined into one hence a real time PCR machine is much more costly compared to a traditional PCR machine.

We can roughly say it is 5 times costlier than the traditional PCR machine. So, how does the real time PCR work ok I mean the detection. So, let us play a I mean thought experiment sort of thing. So, finally, we will try to comprehend how a real time PCR works in a step by step manner without going into graphs ok. We will be breaking down a simplistic real world scenario.

So, that you will be able to construct the graphs and follow along. So, imagine we are looking at a PCR tube the reaction and the reaction is in its 25th cycle. You all know these are small PCR tubes that are placed in a block of thermal cycler. So, at this time in real time in the 25th cycle the let us say let us presume this PCR tube contains a master mix or a mixture or a soup of the various components of a PCR that we need right. And currently let us hypothesis state that there are 1 million copies of the amplicon that is present in 25th cycle of the PCR tube all right.

So, from our basic concept that we have discussed till now what was it in the last cycle 24th cycle almost identical all other components will remain as such except the number of amplicons or the copies will be half right because it doubles in each cycle it is supposed to double in each cycle. So, in cycle 24th it was 500,000 copies right. Now what about cycle 3 very similar everything will appear same, but the copies would have been 250,000 before that 125,000. So, you get the point right. So, with this I hope this concept is clear with one cycle the amplicon replicates itself 2 to the power right exponentially it replicates.

So, if we are to plot or draw curve right what curve is this the amount of DNA and the cycle number ok. So, our plot stands here this is 1 million the graph the axis is 1 million the y axis and the number of cycle is 25. So, let me ask you what will happen when the cycle is in 26th cycle all right. So, definitely it will go up in which direction likely it will be 2 million amplicons simple math we multiply it we double it and in next cycle 27 to be 4 million amplicon. So, if we keep on multiplying right.

So, what will happen in 200 cycle you can easily do the math and it will appear with this number 1 and 19 zeros right there will be this much copy of amplicon. Well there is a weight of DNA right every molecule has a molecular weight and if we calculate the weight of this shear amount of amplicon will amount to 10 to the power 35 tons of DNA whose weight is equal to 10 billion times the size of the earth can you imagine yes the raw molecular weight will be this this is the beauty of numbers, but quite obviously, practically this amount of DNA will not be impossible to accommodate in a PCR tube right. You are absolutely right because this thing does not happen we cannot get unlimited amplification as we move up the cycles why because of something known as limiting factors. Realistically the raw materials will be depleted by the time we have so much amplification. So, the exponential growth does not go forever that is the point of

this thought experiment right.

So, in reality yes there is a zone where the reaction efficiency is as we desire it doubles itself in every cycle, but soon it reaches a point where it will not amplify doubly ok. So, with this concept I hope now you are able to understand the real time PCR amplification plot alright rather we need to notice that here the y axis was the number of copy of DNA. In reality we do not measure the number of copy we measure the fluorescence that part has been established in last class. We measure the fluorescence which is corresponding to the amount of PCR product or the number of amplicons. So, what happens when it starts to amplify initially there is a little change very less change in fluorescence signal, but after a certain while there is a huge increase in the fluorescent signal ok this is the nature of the curve.

So, you see here we have defined an unit which is R n which is R n. R n is actually the relative the fluorescence of the reporter dye normalized by the passive reference dye we discussed in the last slide of last lecture that there are three types of dye reporter, quencher and a reference dye reference dye helps us to negate the background fluorescence. So, background fluorescence is always there. So, imagine R n plus as the total fluorescence and R n minus as the background fluorescence given by the reference dye. So, when we subtract that we will get the delta R n which is this thing which is actually plotted.

So, delta R n negates the background signal and we get the amount of fluorescent which is actually emitted by the reporter dye. What is the use of it? Delta R n plotted against the cycle numbers to produce amplification curve right amplification curve and to estimate the C t value. What is C t value? will be you will be knowing it very soon. So, from the amplification plot nature we can easily classify it is actually classified into roughly four stages linear ground phase, early exponential phase, late exponential or log phase and then the plateau phase. So, let us study them one by one linear ground phase is the first phase where the PCR is starting the fluorescence signal has not risen above the background is almost same this defines the baseline of the amplification this this part all right.

What is early exponential phase? Here the number starts to kick in and the background fluorescence the reporter the fluorescence of the reporter crosses the I mean exceeds the background fluorescence all right. So, at some point of time it will hugely exceed at a detectable level ok. This is where so, this is an arbitrary threshold all right the cycle at which this phenomena occurs is marked ok. So, this is a cycle where the amplification cuts an arbitrarily threshold mark. So, we call it cycle threshold or C t.

I will again define and redefine it to understand real time PCR quantitation you must

know the C t value it is also referred to as C q in some cases ok both are the same. So, next phase linear exponential phase or log phase this phase here we can see the PCR amplification is the maximum where each and each amplicon is actually doubling it is size doubling it is number in every cycle and after some time when the limiting factors come into play all the primers polymerase the dNTPs anything everything will be depleted then it reaches a plateau phase where it can no longer double with each cycle and the signal does not proportionately increase right. So, these 4 phases are essential right. So, how can we I mean measure DNA measure quantity of DNA with this information? Now, again let us go back to our previous example. So, we are a team of 2.

So, I let us let me do an experiment with you all right. So, imagine moment that we have got 2 PCR tubes in our hand all right and you have 4 times the amount of DNA compared to what I have all right. So, you have one at any stage at any given stage you have I have got 4 times more amount of DNA less amount of DNA than what you have got. So, imagine at cycle number 25 you have already achieved 1 million copies. However, it will take me some time to get there right.

So, what will happen? I will reach at your level when I am in cycle 28 right. You see check this number see this is very easy. So, we every cycle multiplies itself twice. So, when I have got 4 times less DNA compared to you the amount of product that you have achieved in cycle number 23 it will take 25 cycles for me to go there fine. Why? Because to start with I have only 250000 copy I need at least 2 more cycles to reach a million you already have got a million.

So, there is a difference right. So, if we plot this it will appear like this. Again imagine another scenario. So, you started with 8 times less DNA than I did this time I have more and you have less all right. Similarly, it will take 3 cycles for you to catch me right the number being 1 million as soon as the cycle hits 1 million the amplicon hits 1 million we get a exponential curve right.

So, in this case what happens you reach there in cycle number 28 fine I have already reached the cycle number 25. Whereas, in earlier case when you had a more DNA in previous sample you already reached the 1 million in cycle number 23 right. So, every time we are targeting one value of an amplicon arbitrarily arbitrary value that is 1 million and the set of experimental tube which started with more amount of DNA more copies of DNA will definitely reach their first right. This concept is used to calculate the CT value by now you have understood that the 1 million threshold is actually crossed by the one set of tube which has got more DNA to start with. So, in this case if my threshold I have set to 1 million the first case will be crossing it in 23 cycle the second situation will be crossing it in 25 cycle and the third one in which you had 8 times less DNA will also get there in 28 cycles all right.

So, this is called the CT value and the quantity is calculated as 2 to the power the CT value of the DNA right very very very important. So, make sure you understand this CT value concept clearly it is the cycle in which the total amount of PCR product crosses a detection thresholds fluorescence signal fine. So, as you have understood in order to relate or mathematically relate the various set of tubes we study the CT value in the log phase why we study the CT value in the log phase we set the CT value in the log phase the amount of fluorescence because there the reaction efficiency is maximum it is in the log phase where the each and every cycle is showing double replication right. So, a sample whose CT is 3 cycles earlier has actually 2 to the power 3 that is 8 times more template this concept should absolutely be clear. Now one thing I should mention just as a trivia that if a CT value is 40 or more we presume that there is no amplification.

So, even after 40 cycles if the PCR product has not taken off or the fluorescence has not taken off we presume that there is no amplification. So, if something is present we have if we have got a target sequence target DNA it will usually start to amplify much earlier. Mind it we will discuss detection of infectious disease much I mean in detail in later part of our discussion, but since we are in the post pandemic era and almost all of you have heard or seen or interpreted or you have discussed regarding COVID 19 positive report where the results are given in CT value right. And the concept was the same the lower the CT value it means more and more viral load alright and above cut off CT value say 2 to the power 35. So, cut off value more than 35 we would have said no viral load COVID 19 negative or it is not problematic whereas, a very high viral load say 23 24 it means so much more virus to start with mind.

So, as we discussed in the exponential phase the reactions are occurring very efficiently it is doubling every 2 cycle doubling twice in every cycle. So, the CT values are very much reproducible we can actually compare the result of 2 tubes whereas, in case of this plateau phase not so much ok. So, we should target this area. So, with this we can now claim that if we say 1 million as 1 unit. So, a CT value of 23 or CT value of 1 if it is 21 unit.

So, CT value of 23. So, 2 cycles less will be 4 times and 3 cycles more is 1 by 8th time right I am repeating myself just to make sure to enforce the concept to you alright . So, with this concept let us now understand how we quantify real time PCR. There are 2 methods one is absolute quantitation and another is the relative quantitation. So, absolute quantitation requires construction of an absolute standard curve for each target gene or gene of interest mind it there are 2 whenever we say GOI in this molecular biology section it refers gene of interest. So, each target gene of interest needs to have one standard curve.

This standard curve is based on what it is based on serial dilution of the sample with known copy number. So, we will take known number of samples in PCR tube and we will amplify them we will note the cycles in which they cross the threshold amplification that is the CT value is achieved and we will plot. So, CT of each standard is plotted against logarithm of known concentration and then a standard curve is drawn. So, it looks like this ok. So, the y axis is CT value and the x axis log of copy number it gives us many information alright it gives us information about slope correlation coefficient and many things.

So, what are those terms let us understand correlation coefficient basically quantifies the extent of accuracy of the standard curve. You do not need to know how it is calculated you do not need to know the map, but you need to know that ideally it should be 1 it is denoted as R square. However, practically the highest achievable value is 0.999 it is very close to 1, but we cannot achieve 1. Regarding the efficiency if we consider the slope of the standard curve if the standard curve is 100 percent efficient the slope will be of course, negative slope it is going in this direction.

So, it is minus 3.32 where the efficiency the reaction efficiency is considered to be 100 percent. It means ideally it is replicating in every step mind it is an ideal situation in all biological experiment an ideal situation can seldom be achieved. So, for what do you mean by this it means exact 2 might not be produced in the next cycle it may be 1.

9999 like that right. So, this is the thing. So, however, experimental variables like length of the primer secondary structure G C G C content of the primer target of the amplicon everything affects the reaction efficiency, but in ideal scenario where the efficiency is 100 percent the slope will be minus 3.32. Now as you already have seen there is a direct relation to the amount of the C T value and the DNA which is 2 to the power the cycle number 2 to the power the C T value will give us the exact copy of DNA. Now reference standards with known copy numbers are run alongside the sample suppose we need to detect any target gene from any unknown sample ok. So, what will we do we will run reference known samples.

So, that same DNA in multiple known concentrations will be run alongside the unknown sample all right and from that we will have the C T value of the known standards as well as the unknown sample with the known standards we will point the plot and suppose we get a C T value we get a of here 20 or say we can get a C T value anywhere here between 20 and 25. So, unknown sample will have a C T value. So, we will plot that C T value in the standard curve and we will draw a perpendicular to the x axis and then we can from the log scale we can easily get the copy number right. This is the principle of absolute quantitation in real time PCR. Now what do we mean by relative quantitation? See in absolute quantitation you the number we get is absolute we

get the actual number of the copy of DNA that is in the sample and mind it for that we need reference copies of known DNA or known concentration of the same target of interest that will be run parallelly with the sample, but this might be cumbersome.

So, to solve this we have got a relatively easy way of having an idea of the quantitation that is known as real time relative quantitation. This uses a gene that is known as housekeeping gene. So, what is a housekeeping gene? A gene that is constantly expressed and it is present in abundance everywhere and expression of these genes are unchanged with any experimental situation or any physiological alteration. For example, 18S rRNA, GAPDH that is glyceral dehydre phosphate dehydrogenase, beta actin all those these are few examples of housekeeping gene ok often abbreviated as HKG. Now prior to relative quantitation what we need to do? We need to normalize there is a phenomena of normalization.

What is normalization? It is the process of properly quantifying gene expression by dividing the measured quantity of RNA from gene of interest by the amount of RNA from a housekeeping gene measured in the same sample. Mind it we are talking about RNA over here. We will be discussing the next class how RNA can participate in the PCR, but almost I mean most of you must have guessed we need to convert this RNA to DNA with the help of reverse transcriptase enzyme. So, reverse transcriptase enzyme will convert this RNA to DNA and all these things are incorporated in the PCR reaction which is actually known as RT PCR. So, whenever we are targeting any gene of interest we are isolating mRNA and in that sample we have to properly quantitate the amount of the target RNA and since all sample has got housekeeping gene we need to measure quantitate the number of RNA all right and we need to divide it to get the normalized ratio.

Why? Because it helps us to compare various samples. This is done to account any change in amount and the quality of RNA across successive samples. So, normalization is must which is dividing the amount of measured quantity of RNA from the gene of interest that is GOI divided by the RNA of housekeeping gene fine. So, what is done? So, in this case this method of relative quantitation is known by Leveque's method or comparative CT method or delta delta CT method all right. So, what is done? The CT values of the sample. So, there are two experimental situations one is a control sample right and one is a experimental sample.

Experimental sample means whether it is treated or any treatment drug treatment has been given or any condition that has been applied to it and there will be another sample which is a reference sample or a control sample with which we will compare the result of the treated sample right. Now, in both the samples we have housekeeping gene we have our gene of interest right. Mind it we are trying to know how much change has

happened. We are not trying to detect the absolute number relative we are trying to compare with respect to controls fine. So, the CT values of both the control samples normalized control sample and the treated sample are measured.

So, there will be difference in the CT value of a target gene of interest. So, CT GOI and there will be a CT housekeeping gene. So, if we subtract that that will give us the delta CT mind it that is in one single sample in one single sample we are calculating the CT value of our gene of interest and we are calculating the CT value of housekeeping gene because both will amplify right. Housekeeping gene will remain unchanged whereas, gene of interest will vary depending on the treatment. So, we will get a difference or the CT value from both sample and non treated sample alright from control and samples.

Next what we will do? We will subtract these two CT values. So, delta CT from experiment will be subtracted from delta or delta CT control is subtracted from the delta CT of experiment and this will give a change in delta CT mind it delta refers to a change. So, in this case when you are considering a single sample the change in CT value from gene of interest to a housekeeping gene is actually the change in CT that is delta CT right delta CT of experiment. Similarly, we will get a delta CT of control when we are subtracting these two we will get a delta of this which is delta delta CT and the efficiency or the fold change of the treated sample compared to the control sample is actually measured by this formula 2 to the power minus delta delta CT which gives us a relative change in gene expression mind it. So, hence this method is also known as 2 to the power delta delta CT method or Leibnack method alright.

So, this is an experimental scenario where we are seeing the plot of the delta delta CT method of relative quantitation. So, there are genes over here. So, referent genes are GAB DH right and TNF alpha tumor necrosis factor alpha that is our gene of interest. So, we have run both of them mind it GAB DH will behave similarly in both the samples. So, you see an overlapping scenario we are denoting control with an overlap right.

So, in any situation where TNF alpha has increased what will happen if the expression of TNF alpha has increased it means there are more number of TNF alpha RNA mRNA start with hence it will have a CT value much less than that of control exactly that has happened. So, we are getting an amplification plot of or CT value which is less earlier amplification that is shift to the left of TNF alpha in the experimental curve whereas, in case of control also we have got an amplification of TNF alpha, but the CT value shifted towards the right which is a higher CT value means low. So, how will we calculate the delta delta CT? We will calculate the delta CT that is subtract the CT value of TNF alpha treated sample minus TNF GAB DH of treated sample and we will also calculate the delta CT of control sample that is by subtracting the TNF alpha CT value from the GAB DH CT value of the control sample. And finally, the fold increase will be done by if this

is the value of x the fold increase will be 2 to the power minus x or 2 to the power minus delta delta CT all right. So, these so, not only with this we can compare multiple sample you know the beauty of having a real time PCR machine is you can color the lanes you can plot the lanes.

And in this case we see that multiple lanes the amplification have happened and depending on the initial copy number the fluorescence the cycles have been crossed ok this is the CT value. Now since these are plots from real samples the machine often have got a direct mode of floral cop I mean plotting fluorescence versus cycle number we can also convert it into a logarithmic scale. So, log of fluorescence so, that actually helps us to identify the exponential region and set up the CT values. Next the last part of our discussion is melt curve analysis. So, what is melt curve analysis? So, after the whole amplification is done right the machine has got an inbuilt program built in that is known as melt curve analysis it depends from machine to machine, but the general principle is we heat the entire mixture to 94 degree centigrade mind it this is done after we have reached the plate phase.

So, it is denatured the whole mixture is denatured and then it is subsequently again cooled to the desired annealing temperature. Now we note that there will be the machine actually notes and plots it for you do not need to note in real time you can actually come back later and see what has happened, but you can also track it in real time. So, what happens when we slowly heat the double standing DNA mind it we in the last class I discussed that cyber green one or cyber green dye is actually very much useful in detecting or in utilizing in this melt curve analysis. So, we all know cyber green dye binds to double stranded DNA to emits fluorescence. So, when we are melting or when we are denaturing the DNA there will be when the double stranded DNA is being converted to single stranded DNA the fluorescence will drop.

There is a significant drop when 50 percent DNA melts all right and it gives us a characteristic plot. There are other multiple factors, but this is the main reason that is the denaturation of DNA that leads to this type of figure and from this we can have multiple interpretation basically the fluorescence is plotted against temperature and rather the change of fluorescence with unit change in time that is a dF dt it is first derivative against temperature is also plotted against time to obtain a clear view of the melting dynamics. What happens so, mind it this is the direct plot this is the derivative plot. So, different PCR products or the different amplification product with different length of base pair different GC content will have will give peak in different zones depending on the change of temperature which will help us to identify whether we have got a single variety of amplicon or there are multiple non specific amplicons in the sample. Mind it this thing for example, primer dimers it will exhibit a very typical peak like this the primer dimers will give a peak which will might appear over here fine.

So, we can also get the similar information by agarose electrophoresis where the size difference in size will give bands in different regions, but it is time consuming. So, melt curve analysis actually helps us in one step to detect the purity of the product mind it melt curve analysis helps us detect the purity of the product any non specific amplification everything is detected in the machine without any extra effort or time. Not only that this method is actually very much sensitive if it is designed by a probe that it can also helps us to help us to detects very minute genetic alteration. For example, a change in allelic temperature will lead allelic structure will lead to shift in melting temperature between the peaks. So, we can actually detect single nucleotide polymorphism differentiate between homozygous and heterozygous mutant and wild type alleles using this.

Well it will be discussed in much details when we are discussing about cancer infectious disease and multiple inherited disorders. So, let us move on, but know this melt curve analysis is very very very useful. So, having said all that qualities of a good PCR run what is an ideal PCR run? It should be 100 percent efficient right template product should double in each cycle it should be highly sensitive. So, even a small amount of starting product should be able to be detected right. It should be very specific means after the reaction after this run it should actually amplify our single target of interest there should be no primer diamond there should be no specific amplification.

Of course, as I discussed there are ways to detect it via melt curve where one peak will correspond to one product. And if agarose gel electrophoresis that is time consuming will also be able to detect this. Next the dynamic range means ability to detect gene with varied expression levels. So, if it can detect a gene with different copy numbers even a small difference should be easily detected.

So, this is referred to as a dynamic range of PCR should also be very high. And lastly the reproducibility. So, mind it these are all the characteristic where I have for example, in I in my lab have defined a set of reaction situation and every lab member should be able to get the same result if they replicate the same cycle provided that we are not considering the technical or sample handling errors or demand to man differences. The changes are only due to biology if there are any changes at all. So, that is also known as the reproducibility of the run. So, all these factors should ideally be met, but we see in real world not all of these are met and often we have to compromise with what we get and what we have that is why we always run each sample in triplicate.

We have to run each sample in three mirror tubes. So, that we can I mean mean the collectively and average out the data from all the three samples because one sample can have some error. And lastly what is the application of this real time PCR and molecular

diagnostic each and every thing that is mentioned here will be covered in some way in the later part of the lecture series. So, we see when we are dealing with mRNA we can study gene expression profile analysis, microRNA expression profile analysis. When we are detecting DNA when we are dealing with DNA we can detect single nucleotide polymerism and allelic discrimination as I discussed. Somatic mutation analysis, copy number variation analysis, food well food microbiology where we can detect the number of genetically modified organization organisms, DNA methylation epigenetic study, detection of pathogen, viral quantification.

So, everything ultimately boils down to the final common pathway which is PCR and best done by real time PCR. So, the concluding part of our real time PCR class this is the part 2 where we have covered all these important concepts we have discussed the components of a q PCR that is real time PCR machine, how the amplification plot is done, how we can visualize the concept of CT value, how it is how easily we can quantify with the number of copies of DNA by relating it to the cycle number absolutely or we can compare two genes and the relative expression and relative CT values to have idea about the relative fold increase. Last we also discussed about melt curve analysis and in brief we discussed what is yet to come that is the applications of real time PCR in molecular diagnostics. So, these are my references for today I encourage you if you are interested you can go through each and every text book and article that is mentioned here and feel free to shoot any query using the email system and if you have any more queries will be definitely addressing them in the live session. So, I thank you for your patient hearing and that is all for today. Thanks a lot.