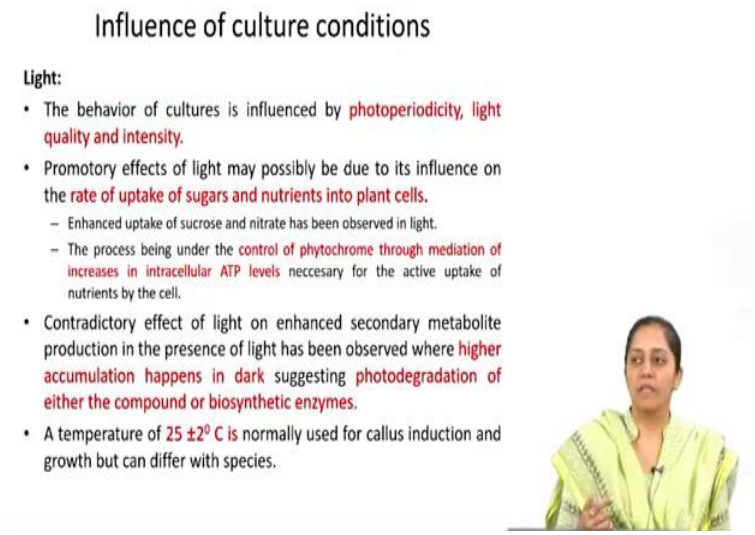
Plant Cell Bioprocessing
Dr. Smita Srivastava
Department of Biotechnology
Indian Institute of Technology, Madras

Lecture - 17
Optimization strategies - Part 2

(Refer Slide Time: 00:16)



We were discussing about screening and selection of the highest yielding cell line. So, what methods can be used for selecting the highest yielding cell line - active and passive techniques. So, now coming on to the effect of other parameters, like light.

Light is a crucial parameter, especially for plant cells. Light intensity means light quality and photo periodicity means light photoperiod cycle. Generally, in *in vitro* cultures we keep 16 to 8 hours light dark cycle, but it may vary and people optimize photoperiodicity. When I say light quality, what does it mean?

Student: Wavelength of the light.

So, sometimes secondary metabolism is also found to be a function of the kind of light which means the wavelength of light. Sometimes, in some secondary metabolites, blue light has been found to induce the production of a secondary metabolite. We will be discussing some examples. So, the behaviour of culture is influenced by three things: photoperiodicity, light quality and intensity.

Now promotory effects of light may possibly be that light might can play a role in activity and expression of enzymes. It can even lead to active transport of nutrients from the soil or the medium.

So, now this can be an indirect correlation because some of these enzymes which are responsible for active uptake are influenced by the light photoperiodicity or light intensity or the kind of light or wavelength. So, enhanced uptake of sucrose and nitrate has been reported in literature to be affected by light. Now I was saying an indirect effect - some of these might lead to enhanced production of ATP which in turn may facilitate the active transport of the nutrients. The process being under the control of phytochromes.

Now these are certain kind of proteins called PR proteins which are impacted by the light which in turn may help in the transcription of certain enzymes which may be responsible in the biosynthesis or induction of a particular secondary metabolite pathway.

So, the process is under the control of phytochrome and through increase in the intracellular level ATP, they might be directly impacting the transcription process. So, enhanced secondary metabolite production in the presence of dark has also been observed.

Now, if in nature sometimes a particular secondary metabolite is known to be synthesized in the underground parts. So, this could mean the synthesis of secondary metabolite regulation might be related to light. So, under *in vitro* conditions when you try to enhance the production, you may observe that under complete dark conditions in *in vitro* cultures, the secondary metabolite biosynthesis is found to be affected or regulated. So, some cues have to be obtained by how and where the secondary metabolite is getting synthesized under natural conditions.

Coming onto temperature, generally 25℃ ± 2 is found to be suitable for generally most of the species but sometimes even a change in 2℃ or 3℃ makes a difference. This is what we observed with *Viola odorata*. So, there are plant growth chambers which are maintained at 23℃ and a plant growth chamber which is maintained at 25℃. Under 25℃, we could find that their growth rate gets affected. So, how do you think the temperature can play a role in the growth as well as secondary metabolism; how does the temperature affect?

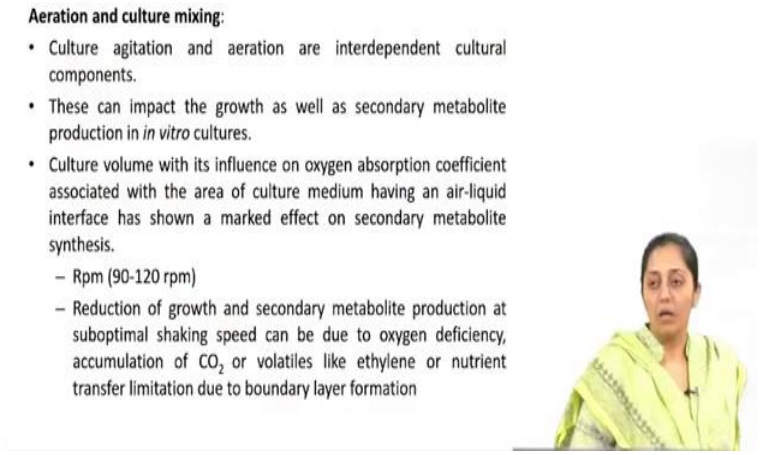
Student: Ma’am, presence of light is directly proportional to the growth

Light is one; Temperature?

Student: enzyme activity.

Enzyme activity can get impacted or membrane composition, membrane fluidity and permeability can get impacted. So, there are different ways in which the temperature affects.

(Refer Slide Time: 05:37)



Now, coming onto aeration and culture mixing. This is particularly important under *in vitro* conditions when you are working with *in vitro* cultures like hairy root cultures or cell cultures. It is more useful or more crucial when it is under liquid culture conditions.

Now culture agitation and aeration, these are set to be independent parameters. What will agitation facilitate?

Mixing is related to mass transfer homogeneity. So, not only will it allow the cells to be suspended, but it also does not allow it to settle down which again is affecting the mass transfer. Mixing should be properly done to have uniform conditions throughout.

The cells have to be kept in suspension without settling down.

How is the aeration taking place in a shake flask? It is through surface aeration. So, continuous circulation will facilitate the contact of the media with the air above. So, culture agitation and aeration are independent cultural components which can impact the growth as well as secondary metabolite biosynthesis.

Now culture volume may have influence in oxygen absorption coefficient which is associated with the area of the culture medium. Culture volume, how does it affect? The media to flask volume ratio will impact surface to volume ratio; where surface is the one which is exposed to the air.

RPM generally ranges between 90 to 120 rpm. In the case of microbial fermentations, it goes above 250 rpm, but in plant cell fermentations, nearly less than 100 rpm or around 150. How will you find that what rpm is the optimum? If I ask you to plan an experiment to find which agitation speed should be optimum, how will you decide?

All the culture conditions and experimental set up must be kept same and we can experiment on different rpms and then check for our desired result.

Now, why do you think there would be a change? What will get impacted?

The effect on secondary metabolite biosynthesis is a response of cells getting damaged.

There can be detrimental shear forces which arise with increase in rpm. So, when you talk about shear forces or shear sensitivity damage how will you quantify that?

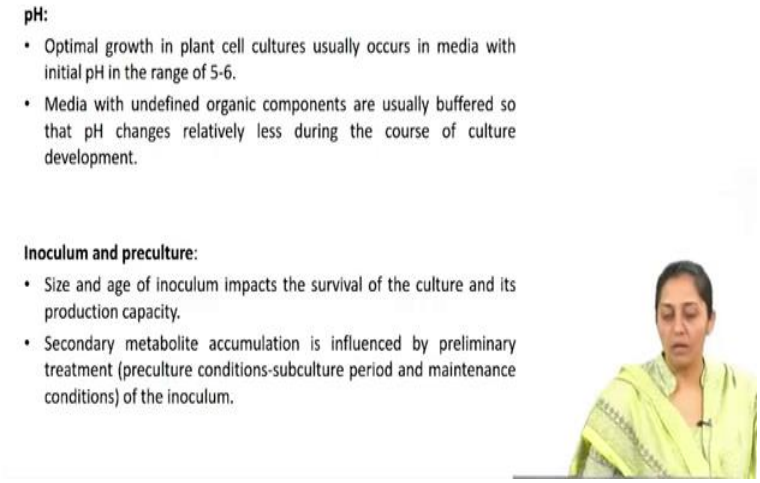
Cell viability. What are the methods did we study earlier for plant cell viability?

TTC assay. So, you can use these assays to check the viability. So, percentage viability drop with rpm can be assessed and its effect on biomass productivity and thereby product productivity can be checked.

What else do you think apart from oxygen transfer can also get impacted by media to flask volume ratio or rpm? There are other gaseous components like CO_2 or volatile components like ethylene which act as growth regulators specially in plant cell fermentations.

So, now if there is too less or too much of a rotation, then this may lead to escape of volatile components which may be needed for the plant growth regulation.

(Refer Slide Time: 11:05)



Plant cells generally prefer a little slight acidic pH which ranges between 5 to 6. One has to see that it is between 5 to 6 or which is the optimum pH.

Now how do you think the pH impacts the growth? Enzyme activity is crucially impacted by the pH of the medium. Media with undefined organic components are usually buffered. When we use complex substrates, there is batch to batch variation. So, for example, yeast extract or peptone or casein hydrolysate, these are complex substrates. So, now, when you use such kind of complex substances, then your pH might get affected. Then you generally have a buffering agent. So, what are the most common buffering salts which are used in media? Potassium dihydrogen phosphate.

So, your inoculum and preculture is also one of the significant factors which impacts the growth and secondary metabolite productivity. Now generally you optimize inoculum age and size.

So, when I say inoculum age what does it mean - Which phase the cells belongs to. Now when I say inoculum size what does it mean – Concentration of inoculum. So, that has to be optimized.

Now, secondary metabolite accumulation has been found to be influenced by preliminary treatment. Now what preliminary treatment if you remember I talked about. it is very crucial specially in plant cell technology which is the quality of your inoculum..

Do you do you agree that quality of inoculum is one of the factors which can be crucial?

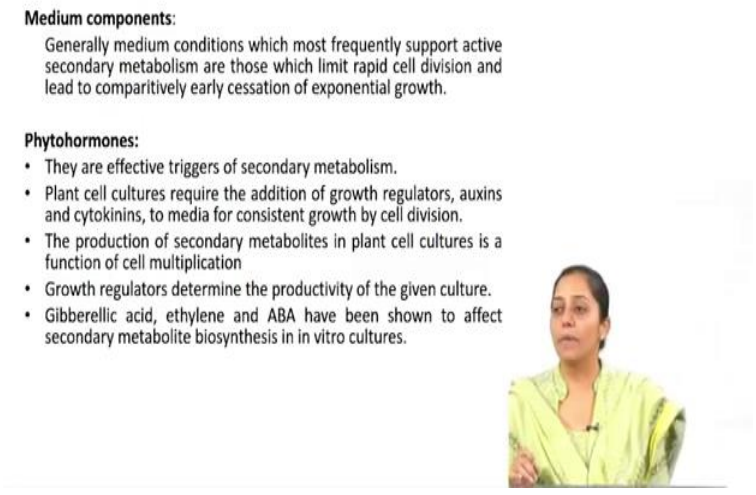
So, when I say quality of the inoculum what does it mean irrespective of any kind of fermentation.

Student: How fresh the inoculum is. Which phase the inoculum is?

So, which means age of the inoculum. What else. Where do you think your bioprocess efficiency in terms of inoculum quality can be impacted? What is the ratio of the most productive cells in your inoculum? So, that means that when you say age of the inoculum it refers to the percentage of productive cells in the inoculum.

Because always these are mixture of cells. If you take plant cell fermentations, they are coming from callus which is a heterogeneous mass. So, you need to make sure of a synchronous culture. So, when you are preparing the inoculum you need to make sure that you end up in an inoculum which is of high quality which means it has the right age. Larger number of cells which are required for the production of that secondary metabolite is higher other cells some cells which may be of different size, shape and therefore, metabolic activity.

(Refer Slide Time: 14:59)



Now, medium components we have already spoken about the different types of medium components which are required in plant cell fermentations. Generally what is observed is

that the media components which will support the growth may not be working well for the secondary metabolite yield.

So, if you need to optimize the medium composition, then the medium composition has to be separately optimized for biomass and for the production of the secondary metabolite. Now we will see how the media composition is optimized.

When the objective is to optimize for maximum biomass, then how will you find using single factor, a range in which you should be optimizing? If suppose the range is not right, what kind of trend will you observe?

Student: Low production of metabolite.

Trend; you are working in a range. So, what kind of trend can you observe if the range is not right?

Continuously decreasing or continuously increasing. So, it will be kind of a linear model. So, is that the right model for optimization?

No what kind of models will end up in giving you optimum values.

Student: Bell shaped curve.

So, what kind of models, now convert them into mathematical form. What kind of equations end up in convergence,

So, 2 degree, 3 degree polynomial, they will have a convergence point.

Generally you do not have to go beyond that because then it becomes too complicated for optimization and the effect is also not that significant to work with and spend time and cost. So, you go up till 3 factor level cubic models, most crucially impacting your product yield. Now this is what is done in your statistical optimization. Once you have done your single factor, you get the suitable range.

So, make sure that you end up seeing increasing and decreasing trend, bell shaped curve and then if you want to optimize it further then you select that range and then use designs which are called as screening designs using design of experiments.

Specially in bioprocess this should be taken up. This is a very useful tool which will be used in getting to the optimum parameters in minimum number of experiments possible and when you do single factor you do not end up in the right optima why. Why do you think varying one factor at a time may not give you the right optima although you will be able to find a bell?

Student: Because you not it does not account for interaction.

Interaction of?

Student: Of other components in the media.

So, now how do you think your statistical optimization tools will take into account the interactive effect?

Student: well we give high and lows of each of them via component and then.

Ok.

Student: Gives us a set of experiments varying different components.

So, a number of recipes can be designed because every factor is varying. So, there can be so many permutations. So, now, what is the advantage. I said the number of experiments are minimized. So, more the number of factors it will keep on exponentially increasing. So, now, these are called as fractional factorial designs. What does that mean? There are full factorial designs as well, but generally to minimize the number of experiments and get to the optimum value you use fractional factorial designs as well. What does it mean?

Student: Taking that value. High and Low value. Getting that trend

No no coming to the design right now, we are designing experiments. So, now, you pick and choose in the design of experiment tool only some of these design points. Now some of these design points will be then given to you and you will be asked to carry out the experiments.

Now some of these designs in the range which you have given, will be choosen depending on the interval there can be. So, many different recipes/design can be used

which chooses only 3 points for a particular variable or design can be which chooses 5 points for a particular variable in this range.

So, there are different types of designs depending on the number of data points, it chooses for every variable and then from there it chooses only set recipes and gives you a design of experiments which the tool asks you to carry out.

Now suppose coming back to a secondary metabolite which is non-growth associated you will carry out the experiment what should be your response for every experiment. Your media for biomass, will it be same for the media for your product?

Not necessarily.

So, then how would you know which component is for biomass and which component is for product such that ultimately you end up in maximum product productivity. So, therefore, you would like to optimize in one go how are your different media components are affecting the biomass and how are your different media components are affecting your product because these may not be directly correlated.

Depending on whether it is growth associated or non growth associated product formation and also whether there is a need to do biomass separately, and this also depends on whether it is extracellular intracellular product formation. Ultimately the goal is productivity.

Now, first thing is I have some so many different media components, first I need to know before even I come to the optimization part which are the most significant ones. So, that I was talking about in the earlier classes that screening designs like Plackett–Burman design is most frequently used, you must have heard, you must have seen many papers in literature. What it does connecting it with mathematical modelling. What does it do?

It depends on certain assumptions, let us talk about the Plackett Burman which is most frequently used. It is based on the assumption that although there is interaction among the factors, but if there is an interaction between any 2 factors or 3 factors you cannot negate that at least one of them will be having a very significant effect then only their interaction factors are having significant effect isn’t it.

So, and what else that there is factor sparcity which means that the factors which you have chosen are independent factors. They are not compounded which means the one of the components, is not affecting, you are not matching pen with pencils here. So, these are independent factors. So, now, with this it takes into account because it assumes that the main effects are the most significant effects and because of these significant effects, the interactive effects will gain importance.

So, it neglects the interactive effect and fits it into a linear model. Linear model means only taking into account the main effects, main effects are the effect of each of your parameters a b c d, and it will ask you to carry out the experiments in the range which you have given the lower level and the higher level. It will select some of the design points and then you will do the experiment based on that experiment it will try to fit your response. Now your response becomes in that y = mx +c, your response is y.

Isn't it and you are factors is x_1, x_2, x_3. Now based on whatever experiments you have carried out you will get a number of y's for number of x_1, x_2, x_3 function. So, you will fit all this in a linear polynomial model based on the fitting and the closeness of the fitting it will do a simple regression analysis and it will fit and it will give you the coefficients of that mx + c was there. Now this mx can be m_1x_1 +, m_2x_2 + and so on.

So, it will give you after fitting; fitting is what? modelling is what? You began with a simple equation y = mx +c, you give the data of y and x you will end up after best fitting it will give you the value of m and c similarly in the same model. It will give you the values of different m's for different x1 and x2 isn't it; now depending on the value of these m's and their signs what can you make out?

Suppose you see a positive value of m what does it mean. Let us talk about the sign first we will not talk about the magnitude.

Let us talk about the sign, first a positive coefficient means what?

Student: Growth rate increases

All the time? Be more specific.

It is dependent on what? It can only say that in the range which you have selected, as you increase from the lower end to the higher end, it is positively impacting the y. A negative

would mean what that in the range selected as you are increasing it is negatively impacting. Now why this is important? In the range selected this may be a major nutrient like for example, nitrogen. Negative dosen't mean that you take away nitrogen and make it to 0.

It would mean that the amount which you are currently adding or in this range which you have chosen it can go below the minus one which you have chosen or the lowest limit which you have chosen. Ideally it should be now checked below this minimum range. Now talking about the magnitude what does the magnitude signify? How will it help in your media optimization.? All the factors will be given some signs of magnitude isn't it?

So, now this will tell you what? Suppose there were 3 factors, one is -5, +5, +3. What can you do with these values? What can you do with these values, with magnitude can you not rank them?

Rank them for what? What was our aim before we began with Plackett–Burman?

Student: optimization.

Screening design means what; not optimization.

I said screening design what does that mean?

To choose.

Which one is significant and which one is least significant. Which means ranking can be done. Now for optimization if there were 16 components in the media all 16 will be ranked.

So, optimization means all 16 have to be optimized- ideally on paper, but do you think that is worth spending time, money and consumables? Depending on this magnitude and the sign you can pick the top ranked which are most significantly affecting your response for further optimization in the range. Now why do you think further optimization is crucial like the values which I gave you -5 which means that you need to now see - you would rather pick up -5 and +5 and leave the +3 because -5 magnitude wise it is more significant, which means the reduction of this component on optimization in the lower

range can affect your response much more than with the same change which you would make. Can affect your response much more than your +3 change.

So, that's what is useful in screening designs, then you rank them then you come on to optimization designs. You must have read papers the most frequently used is your response surface methodology. Response surface methodology, there are a number of tools there. One of the tools most frequently used is central composite design - you must have heard about Box–Behnken design also.

So, central composite design, what does it do? Or these response surface methodology- in general what does it do? It will have a number of tools which I was now talking about, that between 3 to 5 in that same range which you had given, it will pick and choose, there can be any number of xs.

You have suppose x_1, x_2, x_3.

Now, this x1 between -1 and +1 can also have a number of data points. So, now, this would choose from 3 to 5, those data points and will give you a recipe. Again, you will carry out experiments and it will try to fit the methodology or the technology is the same - it will fit the data in a model. Now what kind of a model? Preferably, a linear model?

If you end up in a linear model what does it mean?

That the range which you selected was not right. So, it has to be a converging point. Now we do not go beyond with an assumption. You will generally see people do not go beyond quadratic and cubic. No 4 factor or 4 degree polynomial or 5 degree polynomial. Why do you think? If you have selected 4 factors from your screening design which were crucial, ultimately the interaction can go up to 4 degrees. There are a b c d isn't it? But what is generally done is that only quadratic effects are taken into account, means up to only 2 degrees. The coefficient which is associated with these in a 3 degree polynomial, what all different factors you would have?

For a quadratic equation, for example, you will have $y = mx + mx^2 + mx^3$. If there are 2 factors can be $mx_1x_2 + mx_1x_2x_3 + mx_1^2x_2$ and so on. Can be such a long polynomial. But generally the higher order terms are neglected because their coefficient associated will not be significantly impacting your response. If so then one has to look into the model

equation. So, then thereafter once you fit, ANOVA (analysis of variance) is used. This will tell you that how is the fitting good or not. The error, once you did the experiment is acceptable or not.

So, that is where your statistics comes into the picture, whether to what confidence level you can accept this model. If the confidence level is good, the model is able to predict the confidence of the prediction by the model is accepted. Now suppose now model is ready which means now values to the coefficients have been assigned in this polynomial equation.

This is a model now. How do you use this equation to determine the optimum values? All this we were doing to generate the model isn't it? Now coming on to our ultimate objective. I need to optimize. Now how this model is going to help you to optimize?

Where your in silico approaches help? Manually it is difficult because it is a polynomial which means now this model can predict for different x_1, x_2, x_3, it can predict a number of y's?

So, now, in silico you can generate a plot with a change in x's how is y changing and if the objective function can be kept as y maxima, and solve this equation to give me the value of x_1, x_2 and x_3 for y maxima. We can use a number of mathematical methods. Numerical methods are available isn't it? So, this is what goes on in the background of these tools. You generate contour plots because these are converging. So, in a 3D surface which is called as response surface. You must have a seen surfaces coming out 3D surfaces. So, generally when you see contours it will show you the effect of any 2 factors it is a 2D plot.

So, now, it is the user defined 2 factors. Now how would you define which 2 factors. You can go back to your screening design and see in your model which factor coefficient was maximum. Suppose carbon and nitrogen were coming in the top 2 rank. So, you would try to make a contour plot between carbon and nitrogen, which is a 2D plot keeping rest of the factors are there average values and you see how is your response changing and you keep generating these data points.

So, in silico what it does is, all these data points because this is a converging plot. So, around a circumference whatever is the value of x_1, x_2, x_3 will give you the same response.

So, when you keep maximizing, it will converge to a single point where your x_1 and x_2 will be the value - a single value which is giving you the maximum y value. That is how it is done. So, this is how you will get to the optimum value. Now when you will see tools, depending on the confidence of your model they sometimes differentiate the areas because it is a 3D surface. How will you get to know although I am getting a maxima here but how do I know that how much is the confidence interval? Although I am getting a maxima here, but my x_1, x_2 predicted here are falling in a lesser confidence value.

So, then I would not pick that. I would rather pick a lower response value where the confidence interval of the model is high. So, that is where you use these 3D surfaces and 3D surfaces they have a colour demarcation through which the user can know where the model how is the model confidence varying in different regions of your design space. So, that is how media optimization or any kind of optimization is done and is useful because in minimum number of experiments you can get to the optima.