CHO Cell Culture Media Optimization

In This Tutorial You Will Learn How to

- Learn how to generate a mixture design using MODDE® and its Design Wizard;
- Analyze mixture data using the Analysis Wizard of the MODDE® software;
- Understand how changes in media composition correlate with the responses;
- Use trilinear contour plots for model interpretation;
- Use SweetSpot plots to propose media compositions optimized for different target responses.

Background and Objective

Biopharmaceutical companies today are challenged to develop high producing cell lines as quickly as possible. Commercially available media may fall short of performance expectations in order to meet targets. Design of experiments (DOE) is an efficient strategy to accomplish such targets.

In order to optimize a process for a CHO cell line producing an IgG1 antibody a design of experiments approach was used. In the first stage media optimization was studied by setting up a mixture design consisting of four mixture ingredients (i.e., four different base media). Twenty-four formulations were encoded by the mixture design to match the capacity of the Ambr®15 system. Cultures were run in a simple fed-batch mode and the batches lasted between 10 and 14 days. Peak viable cell density, doubling time and final IgG titer were three of the responses measured upon fermentation termination.

Responses

Three responses relating to titer, peak viable cell density and doubling time were defined. The first two responses should be maximized and the last minimized.

Responses

	Name	Abbr.	Units	Transform	Type	Min	Target	Max
1	**Peak titer**	Titer		None	Regular	240	268	
2	**Peak VCD**	VCD	x10^6 cells/mL	None	Regular	7.2	7.9	
3	**D3 DT**	D3_DT	hours	None	Regular		23.5	26
		Double-click here	*to add*	*a new response*				

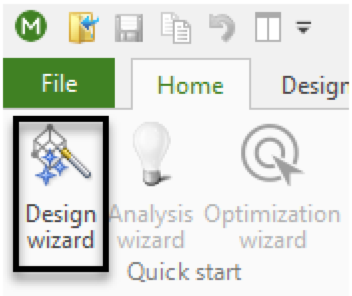
Factors

Four mixture factors ("base media") were mixed according to a mixture design to create media formulations that cover a wide range of mixture component compositions. Lower and upper bounds of the mixture factors ("base media") were [0 – 1] in all cases. There were no additional relational constraints.

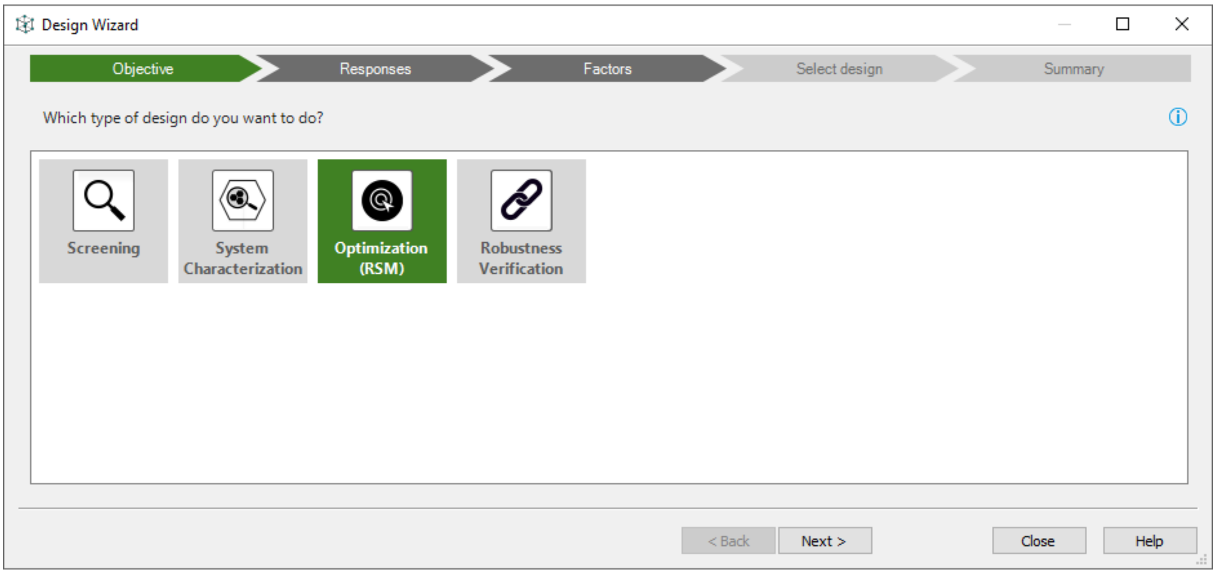
	Name	Abbr.	Units	Type	Settings	Transform	Precision
1	Media 1	M1	Fraction	Formulation	0 to 1	None	0.025
2	Media 2	M2	Fraction	Formulation	0 to 1	None	0.025
3	Media 3	M3	Fraction	Formulation	0 to 1	None	0.025
4	Media 4	M4	Fraction	Formulation	0 to 1	None	0.025
		Double-click here	*to add*	*a new factor*			

Using the Design Wizard to Generate the DOE WorksheetThe first step is to create a MODDE® project and reproduce the experimental design that was used. In this investigation, the Ambr®15 system was set up such that one standard was run thus leaving room for 23 true mixture combinations. This means the DOE protocol used in reality deviates a little from the mixture designs available in MODDE®'s list of eligible experimental designs (see screenshot below). We will therefore copy and paste the real data from an Excel-file. But first we need to specify the design framework.

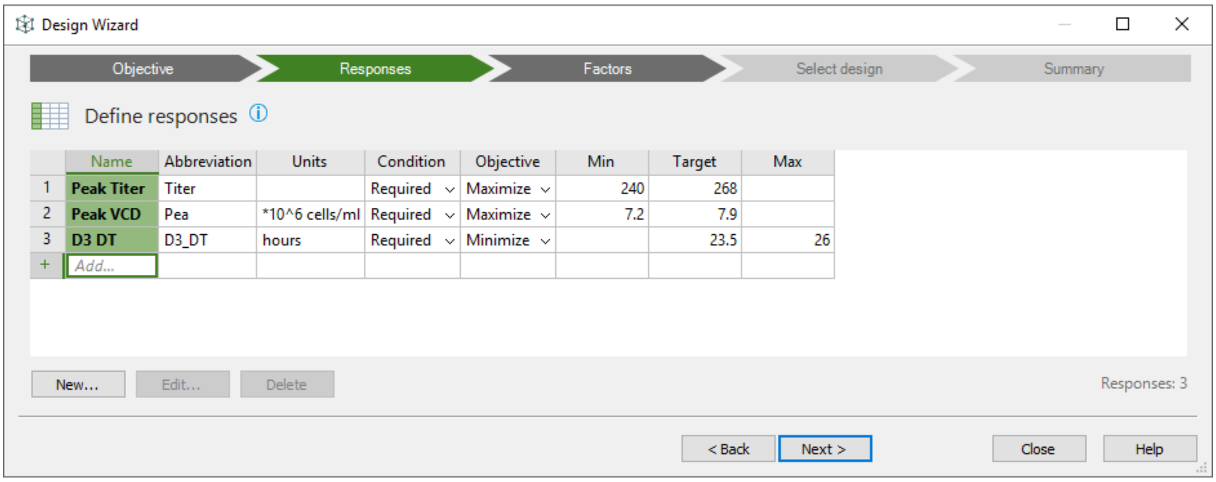
Use the Design Wizard to create a new MODDE® project:



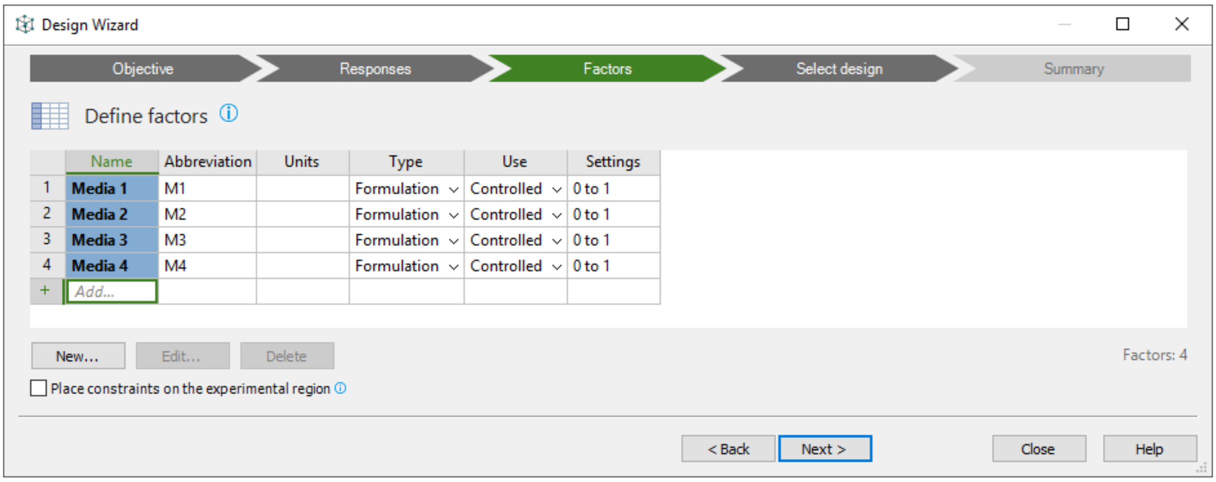
Select File/New/Experimental Design/Optimization (RSM) and step through the Design Wizard as shown below. Click Next.



On the responses page, Click New and enter the name of the response. Set Condition to Required and Objective Maximize. Enter the value Min = 240 and Target = 268. Click Add another and define the second response. Set Condition to Required and Objective Maximize. Enter the value Min = 7.2 and Target = 7.9. Click Add another and define the third response. Set Condition to Required and Objective Minimize. Enter the value Target = 23.5 and Max = 26. Click OK. Click Next.



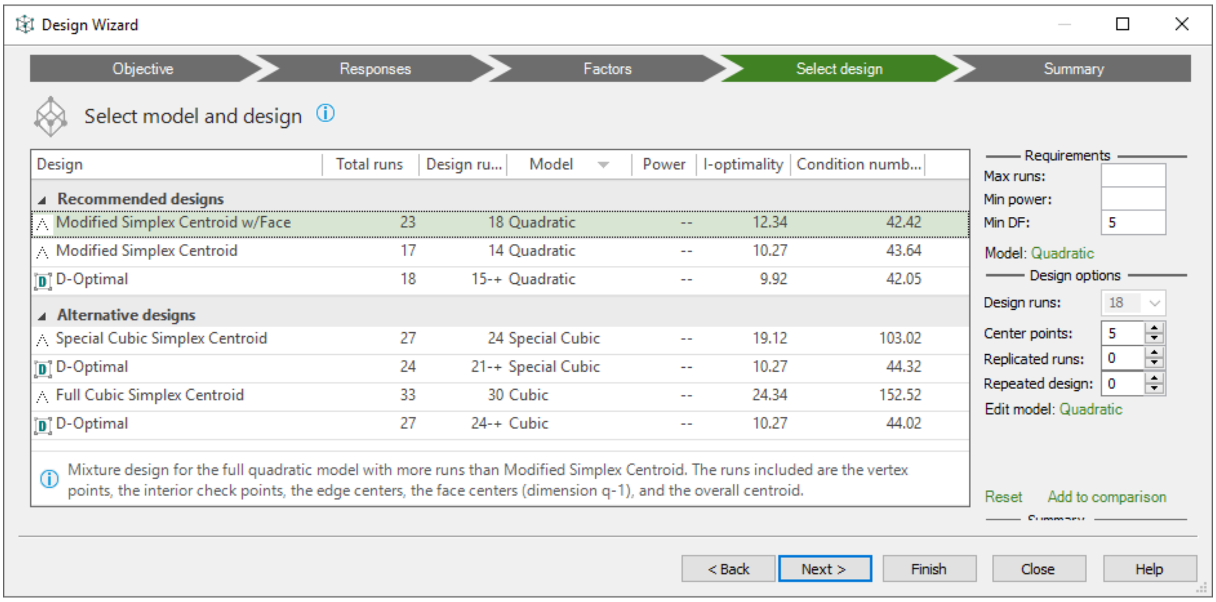
On the factors page, Click New and enter the name, abbreviation, and settings for the first factor. Note that each factor is of the Type Formulation. Click Add another and fill in the name, abbreviation, and settings for the second factor. Repeat to add the third and fourth formulation factors. Click on OK. The four factors have now been defined. Click Next.



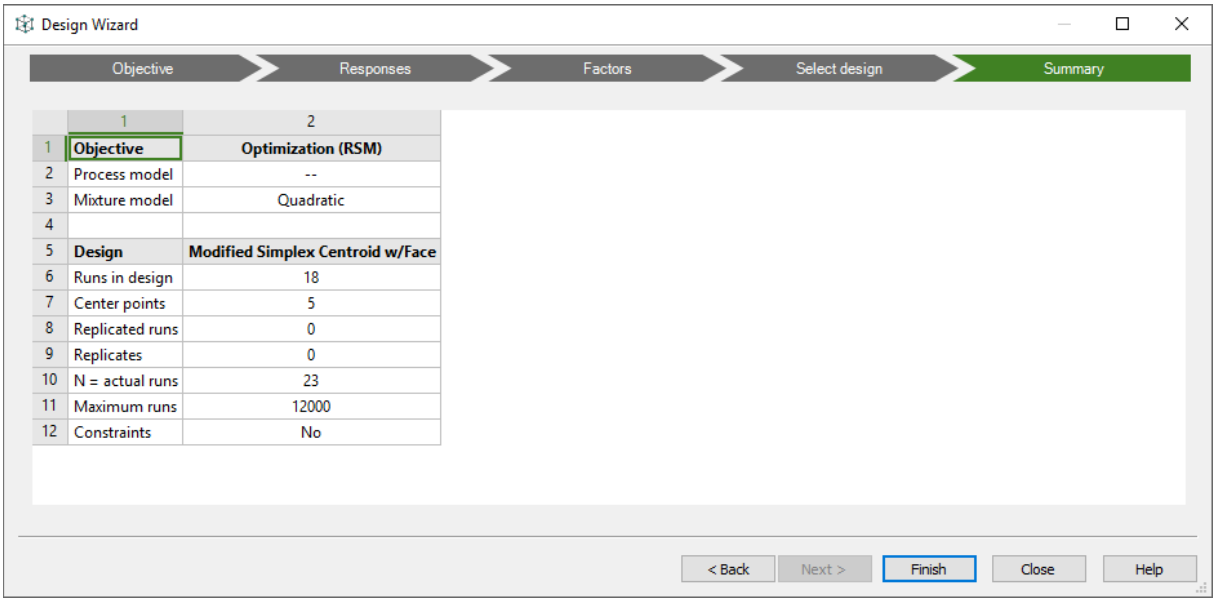
Select the Modified Simplex Centroid w/Face design and increase the number of center-points to 5. Verify that the number of center points = 5 and Total runs = 23.

Note: 5 replicates were not used in reality. It is here the easiest way to get to a worksheet including 23 runs. The real data will be copied into this worksheet.

Click Next.



On the final Summary page you can review your selections and settings, which should look like the screenshot below. Click Finish to exit the design wizard.



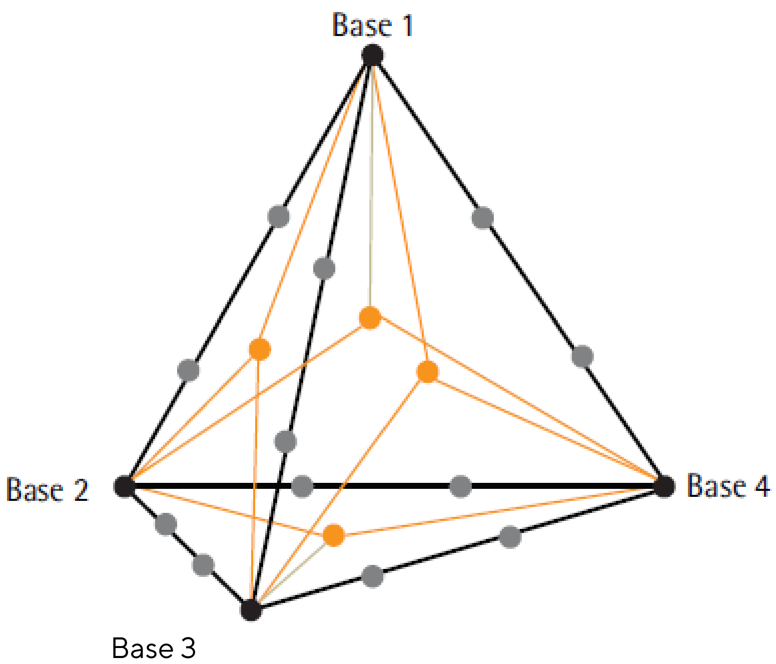
Upon exiting the Design Wizard a preliminary worksheet with 23 rows (23 experiments) is created. Open the file *Media Development CHO Example.xlsx* and copy/paste the real factor and response values into the worksheet. The real worksheet is seen below.

SARTORIUS

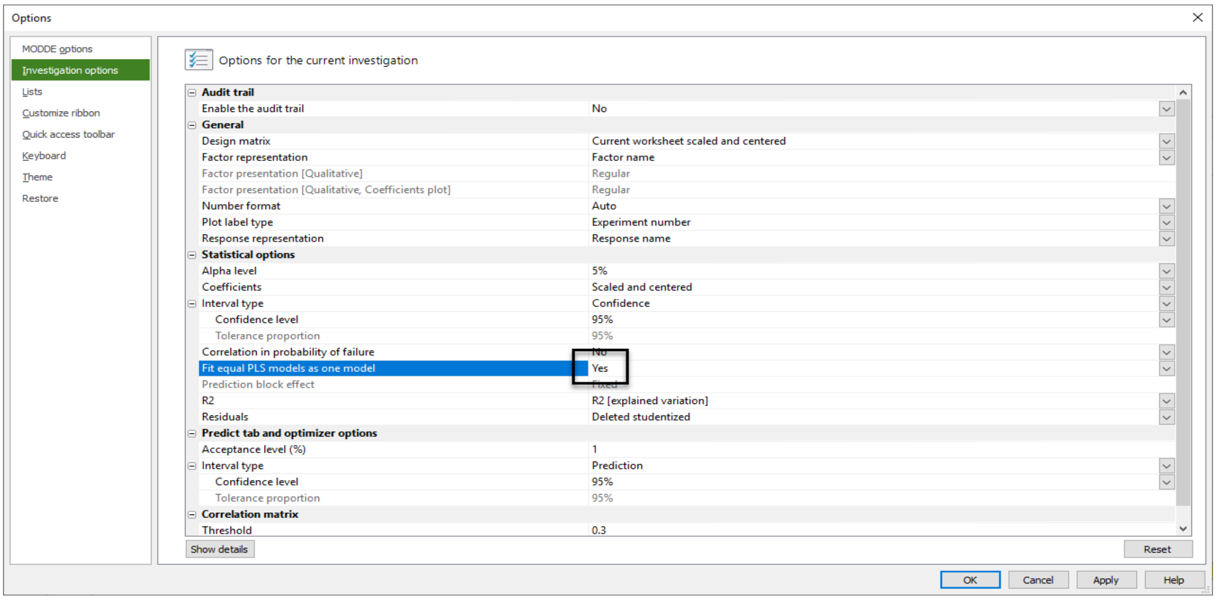
Worksheet

	1	2	3	4	5	6	7	8	9	10	11
	Exp No	Exp Name	Run Order	Incl/Excl	Media 1	Media 2	Media 3	Media 4	Peak Titer	Peak VCD	D3 DT
1	1	N1	5	Incl	1	0	0	0	265.3	6.3281	28.5321
2	2	N2	1	Incl	0	1	0	0	63.1	3.2721	29.3092
3	3	N3	8	Incl	0	0	1	0	81.3	4.4762	25.7874
4	4	N4	16	Incl	0	0	0	1	153.3	8.7821	20.1694
5	5	N5	15	Incl	0.666667	0.333333	0	0	213.2	7.8922	23.4292
6	6	N6	2	Incl	0.333333	0.666667	0	0	185.6	6.6423	22.3544
7	7	N7	12	Incl	0.666667	0	0.333333	0	227	7.6304	23.2045
8	8	N8	20	Incl	0.333333	0	0.666667	0	197.8	7.6304	23.337
9	9	N9	3	Incl	0.666667	0	0	0.333333	219.2	7.8202	22.6309
10	10	N10	23	Incl	0.333333	0	0	0.666667	187.2	8.396	21.5163
11	11	N11	11	Incl	0	0.666667	0.333333	0	68.7	3.7465	29.565
12	12	N12	21	Incl	0	0.333333	0.666667	0	79.7	4.0705	26.0905
13	13	N13	6	Incl	0	0.666667	0	0.333333	122.7	5.0521	26.1444
14	14	N14	13	Incl	0	0.333333	0	0.666667	147.2	7.7155	21.1443
15	15	N15	9	Incl	0	0	0.666667	0.333333	129.3	6.4328	30.4853
16	16	N16	4	Incl	0	0	0.333333	0.666667	158.9	8.9327	20.6944
17	17	N17	14	Incl	0.333333	0.333333	0.333333	0	181.9	7.4733	22.8998
18	18	N18	19	Incl	0.333333	0.333333	0	0.333333	197.8	7.2705	22.6432
19	19	N19	10	Incl	0.333333	0	0.333333	0.333333	188.5	7.7351	21.8119
20	20	N20	22	Incl	0	0.333333	0.333333	0.333333	123.1	5.5952	22.6635
21	21	N21	7	Incl	0.25	0.25	0.25	0.25	178.1	7.2901	21.1967
22	22	N22	17	Incl	0.25	0.25	0.25	0.25	171.7	7.421	21.6316
23	23	N23	18	Incl	0.25	0.25	0.25	0.25	172.9	7.3228	21.6686

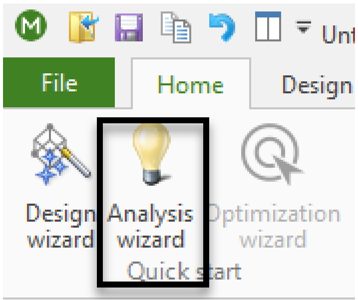
In the case of four base media in a mixture, the design region corresponds to a tetrahedron. The graph below is a visualization of the distribution of the experimental runs in the design region. Four base media are mixed in a systematic fashion to create media formulations that cover a wide range of component concentrations.



Using the Analysis Wizard to Model the Three ResponsesSince we work with mixture design it is recommended to fit similar PLS models to the responses. Goto File/Options/Investigation options. Set “Yes” as seen below. Click Apply. Click OK.

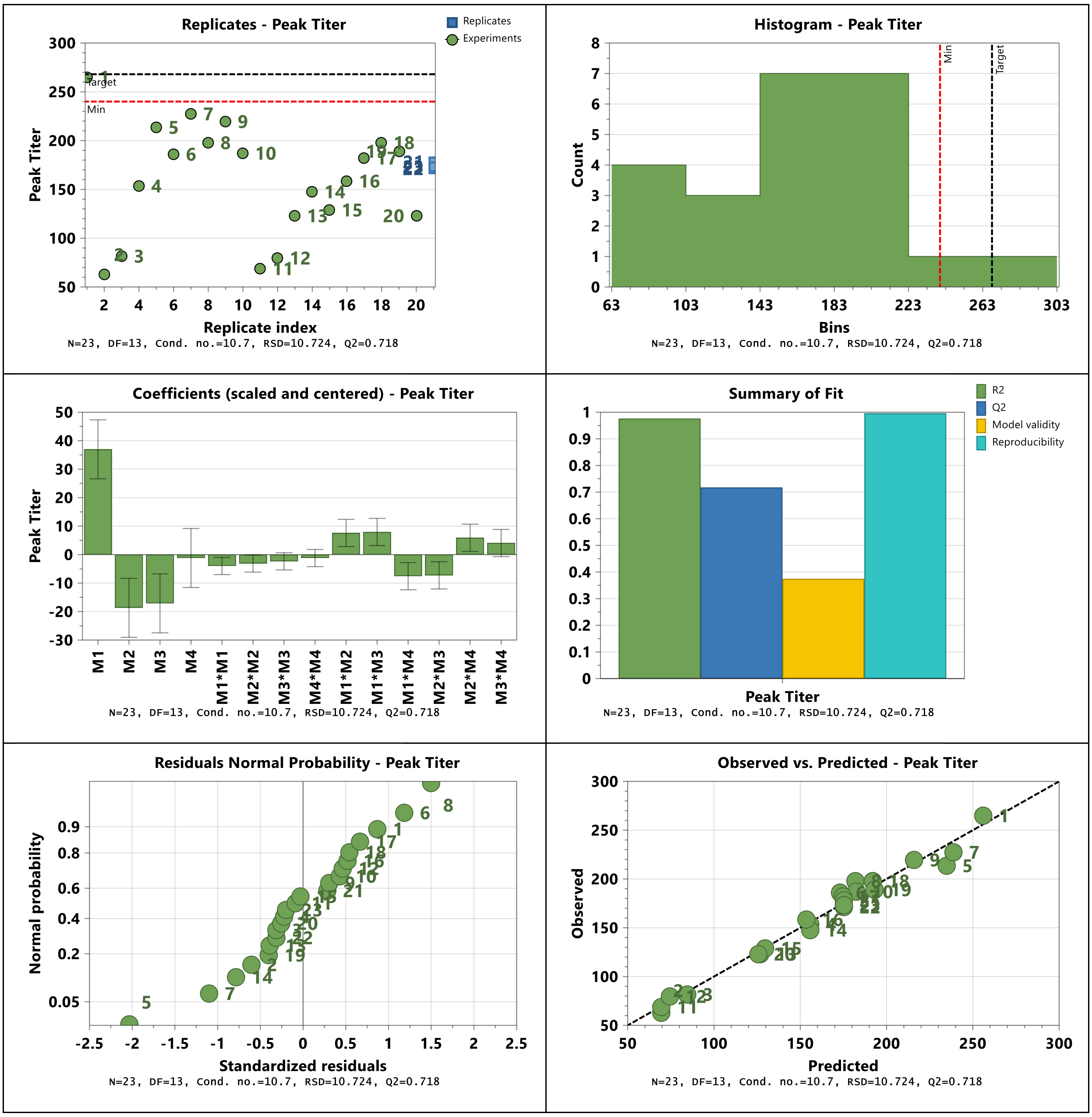


Use the Analysis Wizard to model the three responses.



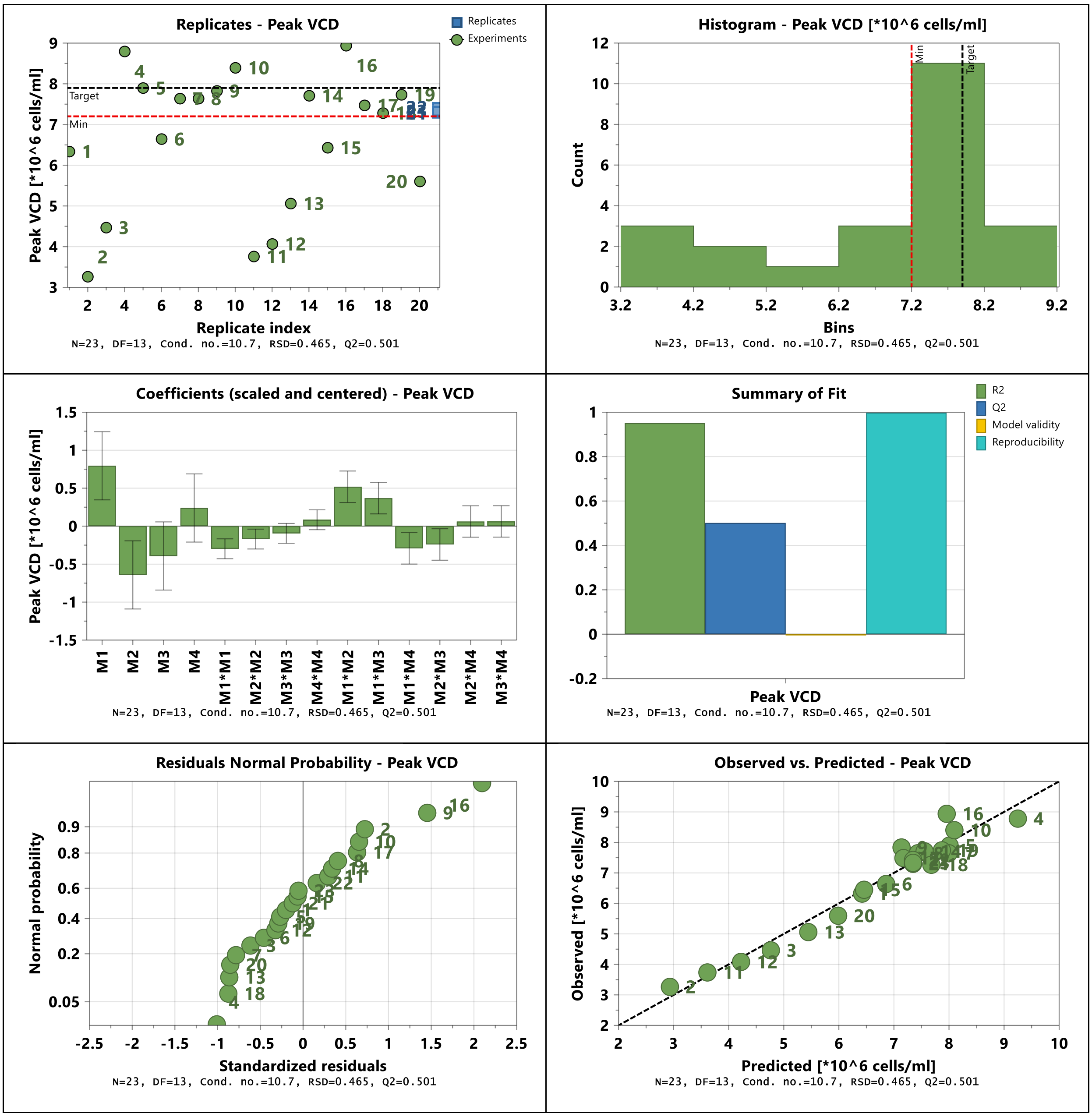
For each response judge the replicate experiments quality, response normal distribution, model quality etc. Are there any deviators? Which factors (“base media”) have the highest / lowest influence? Are the investigated factors (“base media”) influencing the three responses in the same way? Which mixture composition is favorable for maximizing titer? Maximizing peak VCD? Minimizing doubling time?

Peak Titer



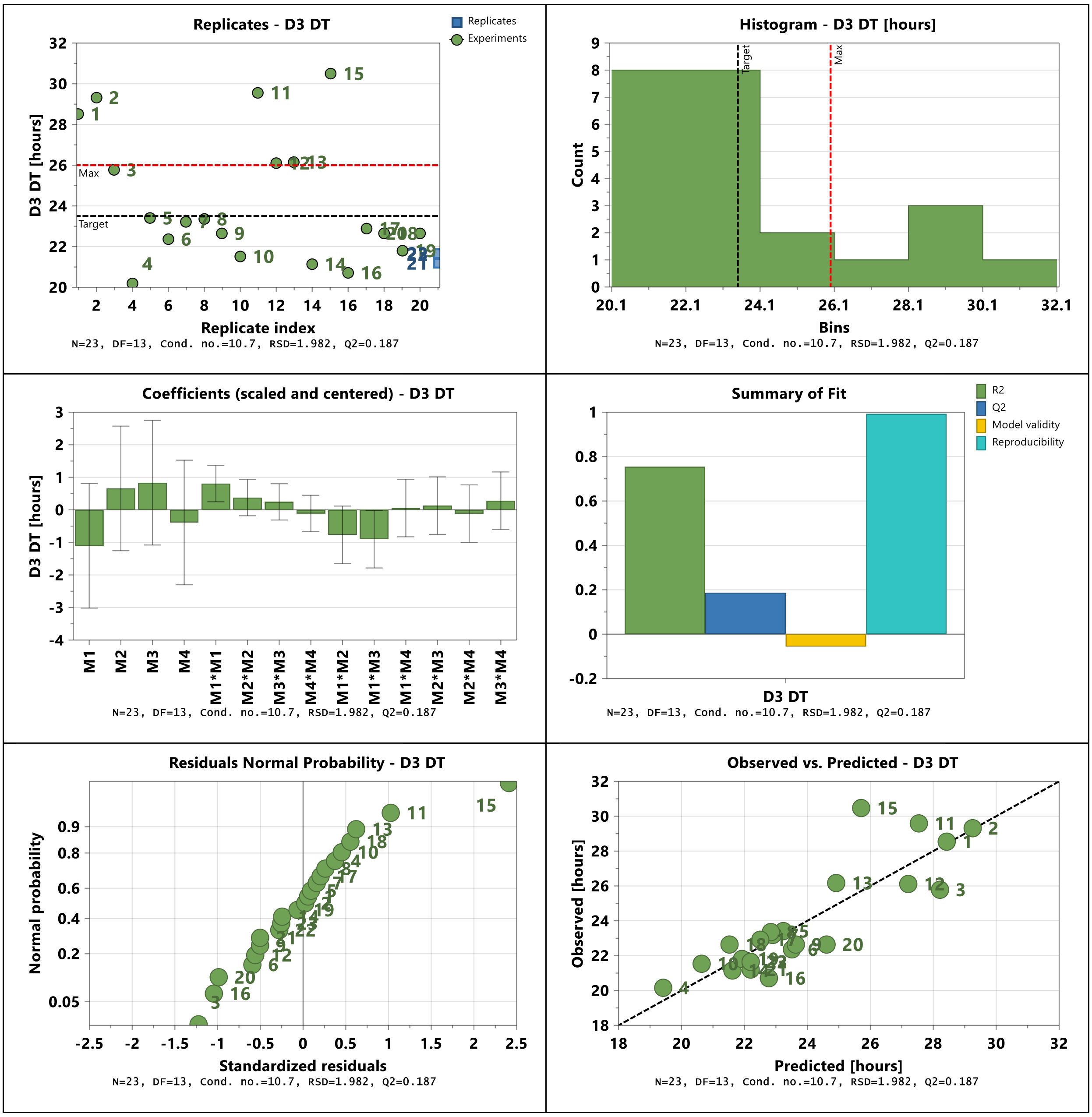
Comment: The replicate plot shows there is small variability among the replicates, which is encouraging and shows we have good data to work with. The histogram plot indicates that the response does not need a transformation and can be analyzed using the untransformed metric. According to the coefficient plot media M4 has a small impact on Peak titer; in order to get more titer a formulation with high amount of M1 and low amounts of M2 & M3 is desirable. The summary of fit plot, the normal probability plot of residuals and the observed versus predicted plot all point to a very good model for the titer response.

Peak VCD



Comment: The replicate plot shows there is small variability among the replicates. The histogram plot shows a slight tendency for tailing to the left; however, this tailing is not strong enough to justify a response transformation. The regression coefficient plot shows Peak VCD to depend similarly on the mixture factors as does Peak titer. The influence of the higher-order model terms is more pronounced for Peak VCD, however. In order to increase Peak VCD, the formulation media should contain high amount of M1 and low amounts of M2 & M3. As indicated by the Summary of fit plot, N-plot of residuals and the obs/pred plot we have obtained a rather good model, although not as strong as the model for Peak titer. The close-to-zero value for Model Validity is caused by the very small replicate error and is more of a "cosmetic" issue rather than a real model problem.

D3 DT ("Doubling Time")

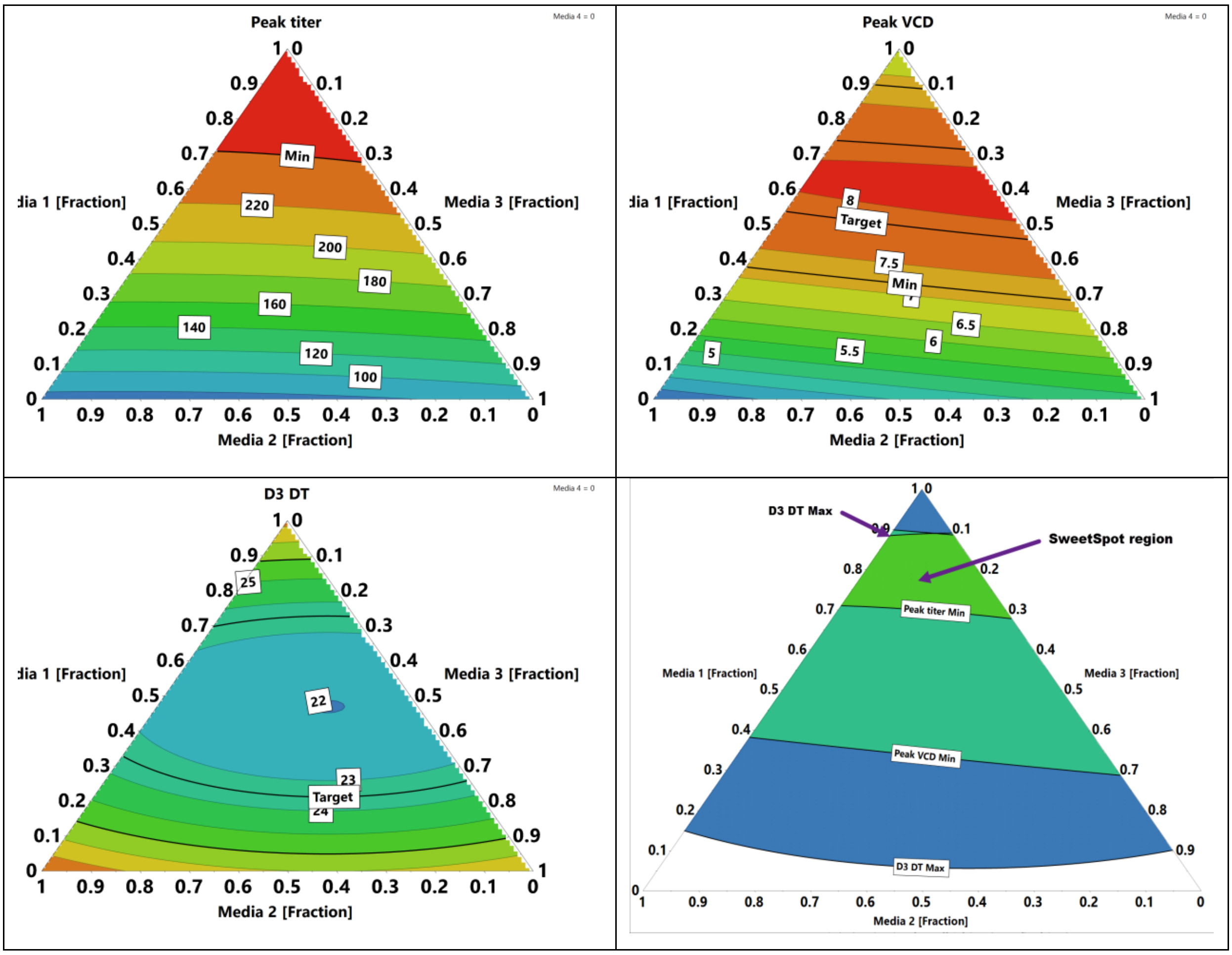


Comment: For the doubling time, the replicate plot suggests small replicate error. Additionally, we can see that there are a few runs with rather large numerical values for the doubling time, corresponding to a slight skewness to the right. This skewness is visualized in the histogram plot and indeed the skewness test is just triggered and is a borderline case whether to log-transform or not to transform the response. You may experiment analyzing data with and without a transformation and compare the results. In order to keep things simple our account is based on not transforming the response. Compared with the results for the two foregoing responses, the modelling results for the doubling time is a lot weaker. This is seen from the size of the confidence intervals in the coefficients plot and the comparatively low R2 and Q2. The low Q2 is primarily caused by a deviation for experiment #15, as seen in the observed versus predicted plot. Both linear and higher-order terms have some influence on the response.

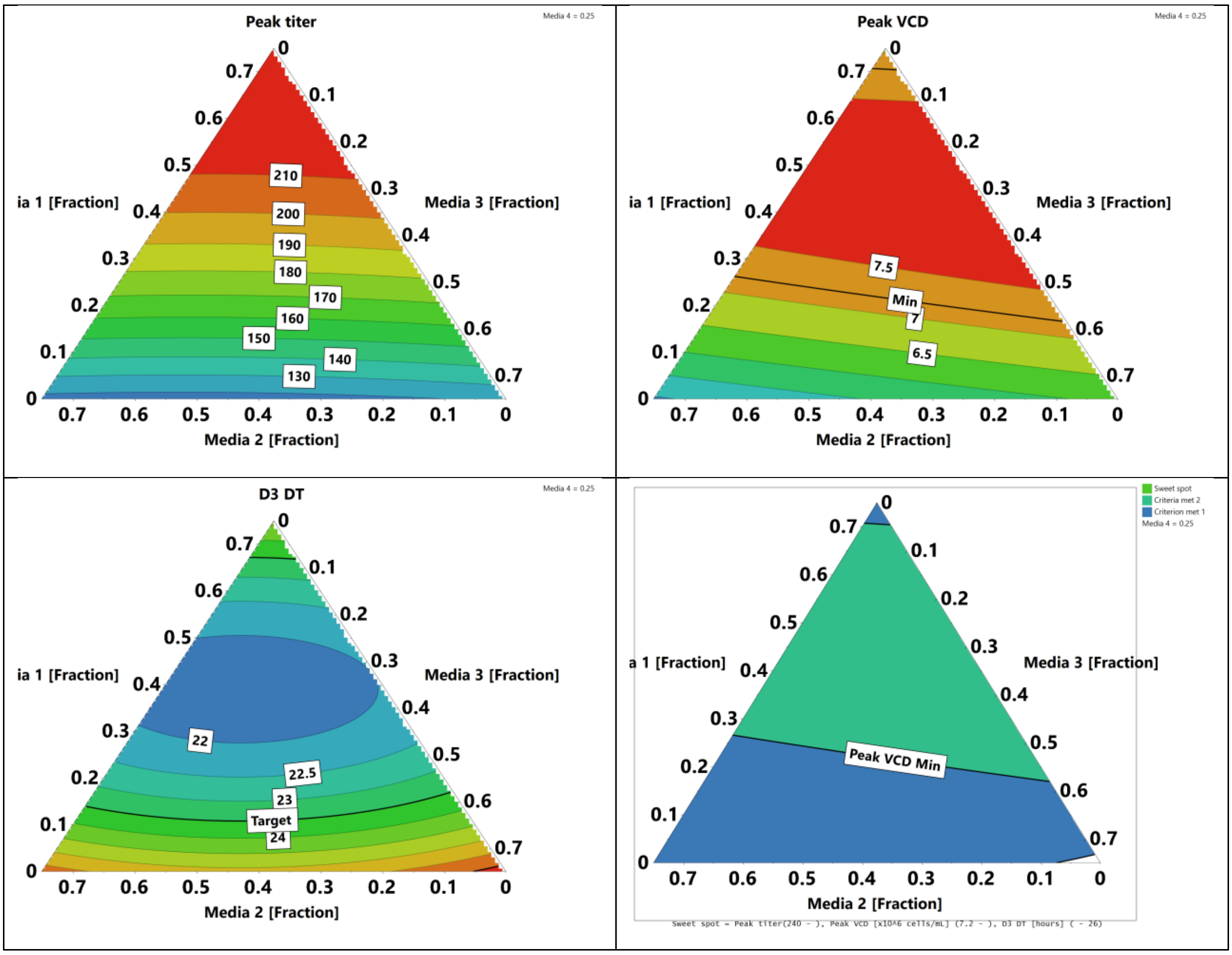
Using Trilinear Contour Plots to Visualize ResultsIn this step we will evaluate how changes in media composition affect the three responses using trilinear contour plots. This means one of the four mixture ingredients must be held constant. One way of identifying which mixture ingredient to use as a constant is through looking at the regression coefficients. Other means could be to consider other properties such as cost, purity, stability, half-life, toxicity or some other important property relevant for the application at hand.

The plots of regression coefficients used in the previous task jointly indicate that Media 4 is the least influential one. In the first set of trilinear contour plots given below the proportion of Media 4 = 0 and in the second set of contour plots Media 4 = 0.25. You can of course exploit other settings of Media 4 and alternative constant mixture factors, but that option is not pursued here.

To maximize Peak titer we would move towards the top vertex in the displayed contour plot for Peak titer (the upper, left plot). However, that co-ordinate would not satisfy the goals for Peak VCD and D3 DT (doubling time). For the latter two responses, we should stay in the interior top-to-mid-part of contour plots. The SweetSpot plot (lower, right) can be thought of as an overlay of the three trilinear contour plots that is colored according to how many response criteria are fulfilled. The SweetSpot plot shows there is a region where the demands for all three responses are met, i.e., we comply with the limits in each case but we are not necessarily at the ideal target setting. In the SweetSpot region we have an area where we achieve a compromise for all three responses.



In order to visualize the impact of increasing the proportion of Media 4 from 0 to 0,25 we can look at the next set of trilinear plots. Evidently, increasing the amount of Media 4 is beneficial for doubling time, but not the other two responses. For this setting of Media 4, no SweetSpot region is definable.



Conclusions

Design of experiments (DOE) is an efficient strategy for systematically and simultaneously changing proportions of mixture ingredients and exploring the consequences on critical quality attributes.

In the current example the ranking of the four base media is such that the overall influence of M1 > M2&M3 > M4.

Based on the presented methodology three media formulations were identified:

- One media was optimized for titer production
- One media was optimized for doubling time
- One media was optimized for all three responses weighted equally

The three chosen media were used in a second phase DOE looking into spent media analysis and process development. This dataset is reviewed in the sequel DOE investigation called *CHO Cell Culture Design Space Estimation*, which is part of our DOE training course material.