Supporting Information

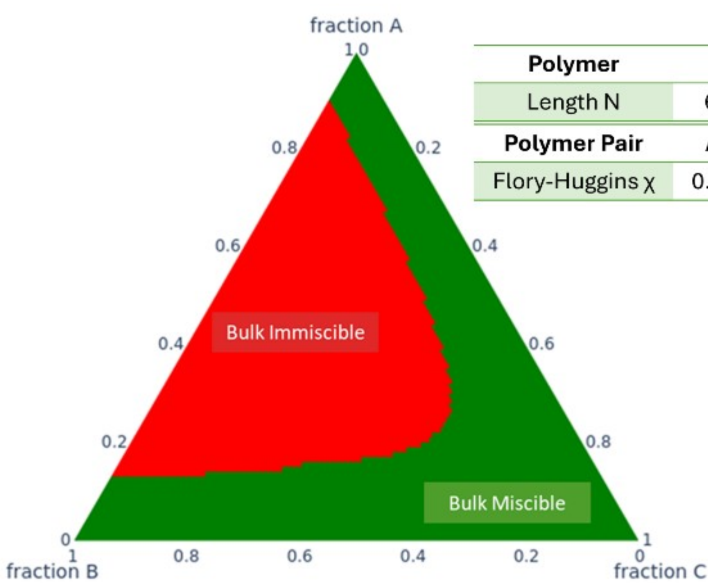
Efficient simulation of complex fluid phase diagrams with Bayesian optimization

Steven G. Arturo[a], Clyde Fare[d], Kaoru Aou[b], Dan Dermody[c], Will Edsall[c], Jillian Emerson[c], Kathryn Grzesiak[c], Arjita Kulshreshtha[b], Paul Mwasame[a], Edward O. Pyzer-Knapp[e] and Jed Pitera*[f]

[a] The Dow Chemical Company, Northeast Technology Center, 400 Arcola Road, Collegeville, PA 19426, USA.
[b] The Dow Chemical Company, Texas Innovation Center, 220 Abner Jackson Parkway, Lake Jackson, TX 77566, USA
[c] The Dow Chemical Company, Michigan Operations, 693 Washington Street, Midland, MI 48640, USA
[d] IBM Research UK, Hartree Centre, Keckwick Lane, Daresbury, Cheshire, Warrington, WA4 4AD, GB
[e] .Xyme, Botley Road, Oxford, England, OX2 0HA
[f] IBM Research Almaden, 650 Harry Road, San Jose, CA 95120, USA

The Supporting Information document details the methodology used to select the next point in a trajectory that efficiently finds the phase boundary of a ternary phase diagram.

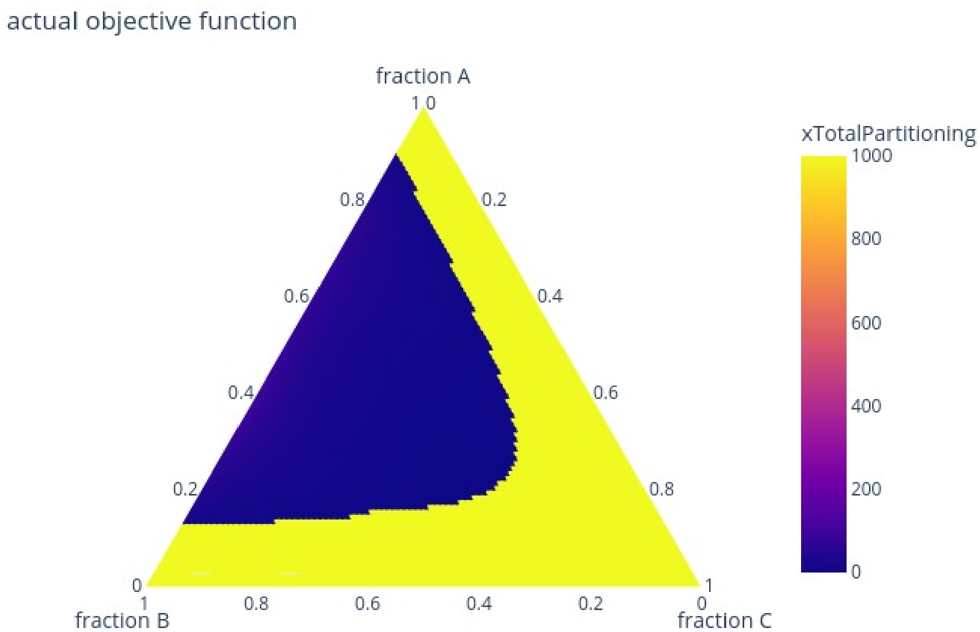
The ground truth phase diagram which serves as reference for our phase boundary search methodology is shown in Figure 1 of the Communication and here in Figure S1. The ternary phase diagram is produced by applying SCFT or Flory Huggins theory using the size and Flory-Huggins chi (χ) parameters shown in Figure S1. The polymers are modeled as incompressible unimodal linear homopolymers whose configurations are treated with Gaussian statistics. Points in the phase diagram are computed to be bulk immiscible or miscible at 1 wt% resolution. While PolyFTS (reference 14) was used to generate the phase diagram for this study, a correctly applied SCFT method or Flory-Huggins theory would generate the same phase diagram.



Polymer	A	B	C
Length N	6.4	10.0	1.0
Polymer Pair	**AB**	**AC**	**BC**
Flory-Huggins χ	0.570	0.167	0.303

Figure S1. The ground truth ternary phase diagram which serves as reference for our phase boundary search methodology.

The objective function is designed to drive a search for minima toward the boundary. The equations are given by eq 1-4 in the Communication but are summarized here equal 1; an artificially high value of 1000 is used at miscible points (points corresponding to the green points in Figure S1). Plots of the objective function are given in Figure S2, where the upper plot shows the true range of the objective function (called xTotalPartitioning), and the lower plot is presented with clipped range to show the minimum points at the phase boundary.



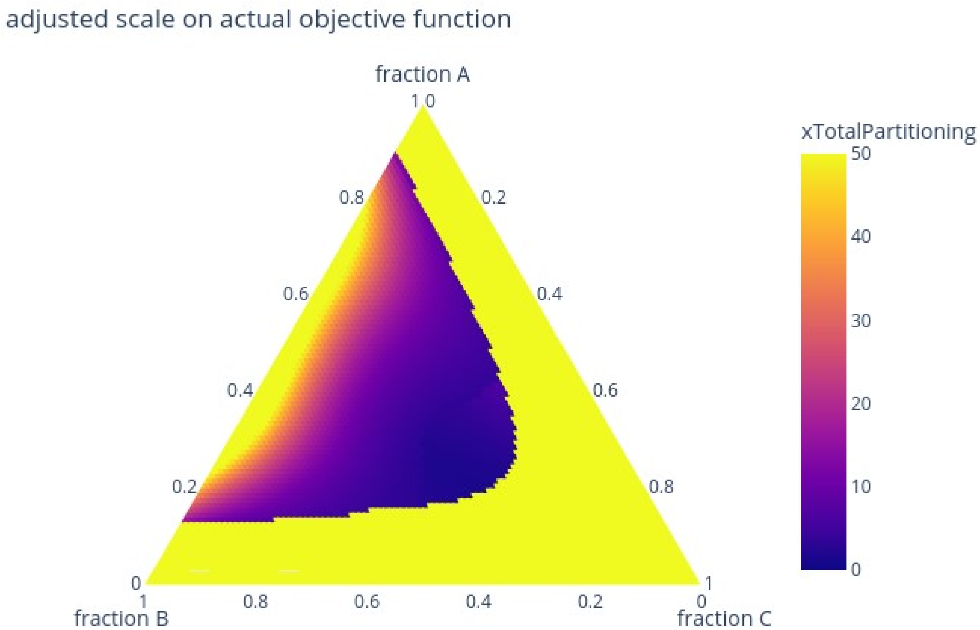


Figure S2. The values of the objective function for the ternary phase diagram. The upper figure shows the full range of values. The lower figure truncates all values that are greater than or equal to 50 to a value of 50.

The goal of the active learning effort is to determine the boundary of the phase diagram at 10 wt% resolution given three measured binary mixture points, one measured ternary mixture point, and the three pure species points at the corners of the phase diagram.

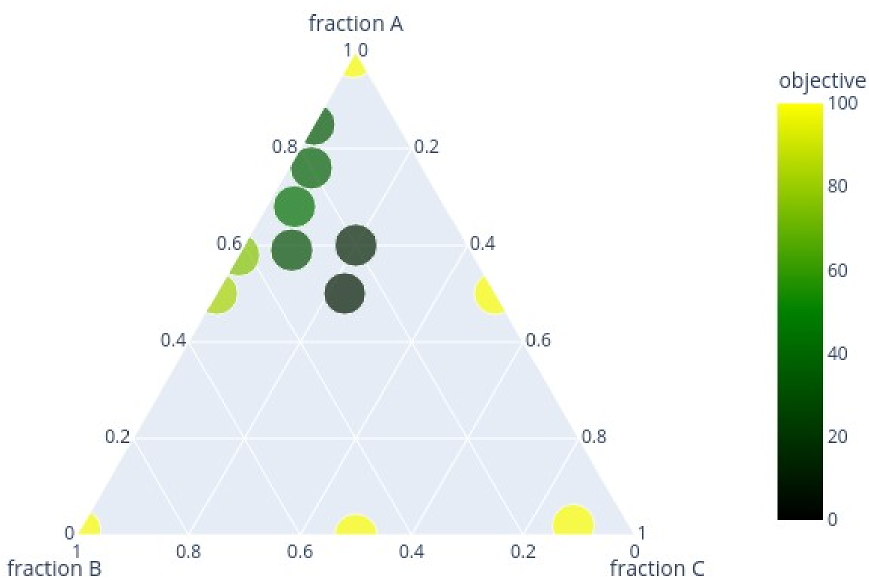
One iteration in the active learning trajectory is described in this Supporting Information. The state of the system at the beginning of the iteration (Table S1) is just after the ninth iteration, when the ninth experimental point is measured. Iteration 0 includes the initial four measured points (minus the pure species points). Each point has is categorized as miscibile or immiscibile. As mentioned above miscible points get assigned an objective function value of 1000. Immiscible points get a value based off the partition ratios of the three polymers shown in eqs 1-4 in the Communication.

Table S1. State of Active Learning Trajectory after Ninth Iteration

iteration	fraction A	fraction B	fraction C	miscible	objective function value
0	0.6	0.2	0.2	No	12.4
0	0.5	0	0.5	Yes	1000
0	0	0.5	0.5	Yes	1000
0	0.5	0.5	0	No	83.4
1	0.02	0.1	0.88	Yes	1000

2	0.59	0.32	0.09	No	30.0
3	0.68	0.27	0.05	No	42.6
4	0.5	0.27	0.23	No	9.57
5	0.01	0.99	0	Yes	1000
6	0.85	0.15	0	No	33.5
7	0.99	0.01	0	Yes	1000
8	0.58	0.42	0	No	76.9
9	0.76	0.2	0.04	No	36.9

Figure S3 depicts location of point and the predicted areas in ternary phase diagrams. The upper figure shows the points and their objective function values. Values of less than 1000 are assumed to be immiscible and are shades of green. The scale of Figure S3 is adjusted to show the values of the immiscible points. The predicted areas in the phase diagram after the ninth iteration are created by drawing every triangle between all possible sets of three immiscible points. The points within these triangles are assigned immiscible and are colored red. Points that are within 10 wt% of measured immiscible points and that are not within triangles made by immiscible points are assigned to hold the boundary and are colored orange. All other points at this iteration are assigned miscible and are colored green.



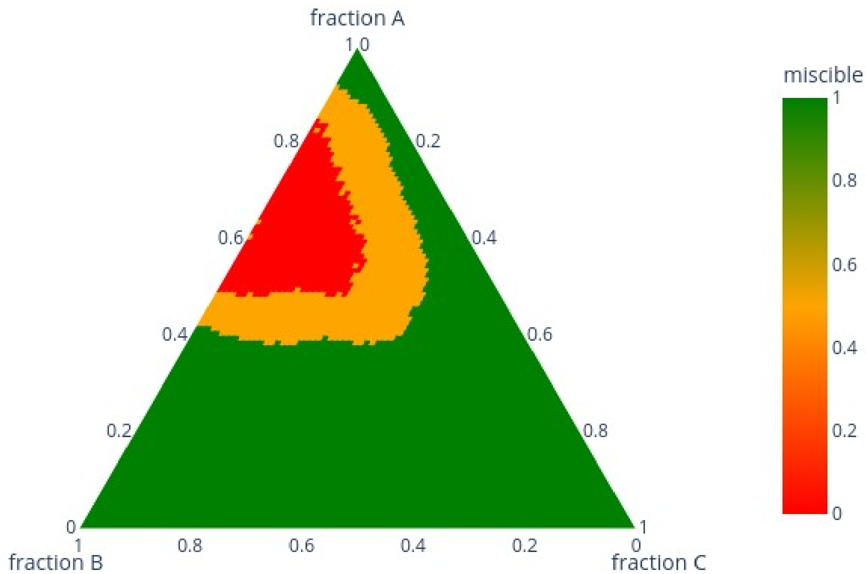


Figure S3. The points measured at the end of the ninth iteration, and the predicted areas assigned immiscible (red), miscible (green), and where the boundary lies (orange).

Figure S4 shows the predicted objective function (the negative of the objective function is shown) and the uncertainty of the prediction for a model generated with a Gaussian Process using a Matern(5/2) kernel with a single length scale optimized by minimizing the negative log marginal likelihood of the data. (reference 15). Points with a distance from a measured immiscible point that is less than the chosen resolution (10 wt%) are removed from the domain and are thus no longer possible choices for the next iteration.

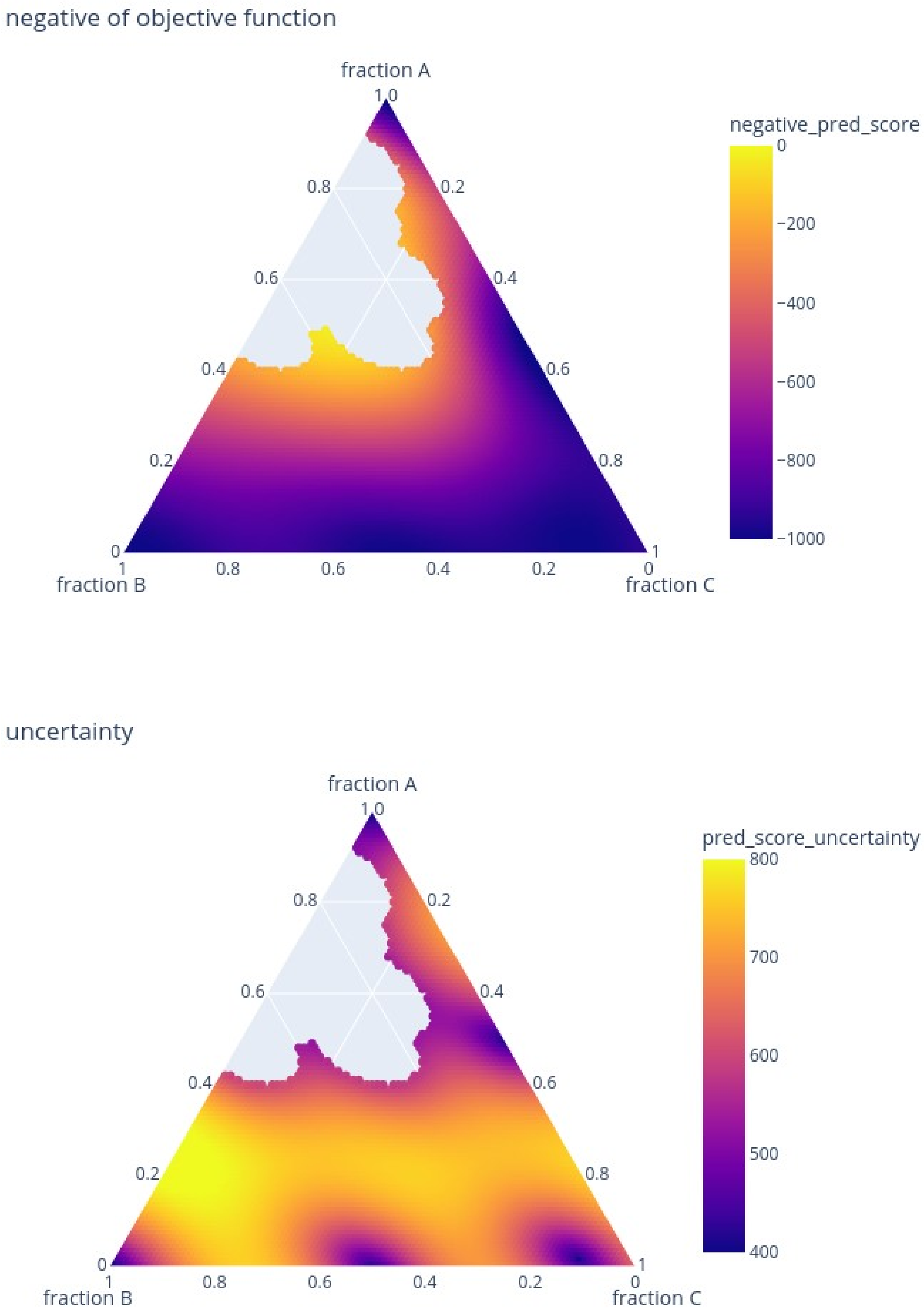


Figure S4. Predictions and uncertainties from a Gaussian Process model of the objective function over the ternary phase diagram.

The raw expected improvement acquisition function with use of the contextual improvement parameter of $c_v = 0.43$ for this iteration (reference 15 and 16) is shown in Figure S5. The acquisition function suggests the most information can be had by picking a point within one of the two yellow areas shown in the figure.

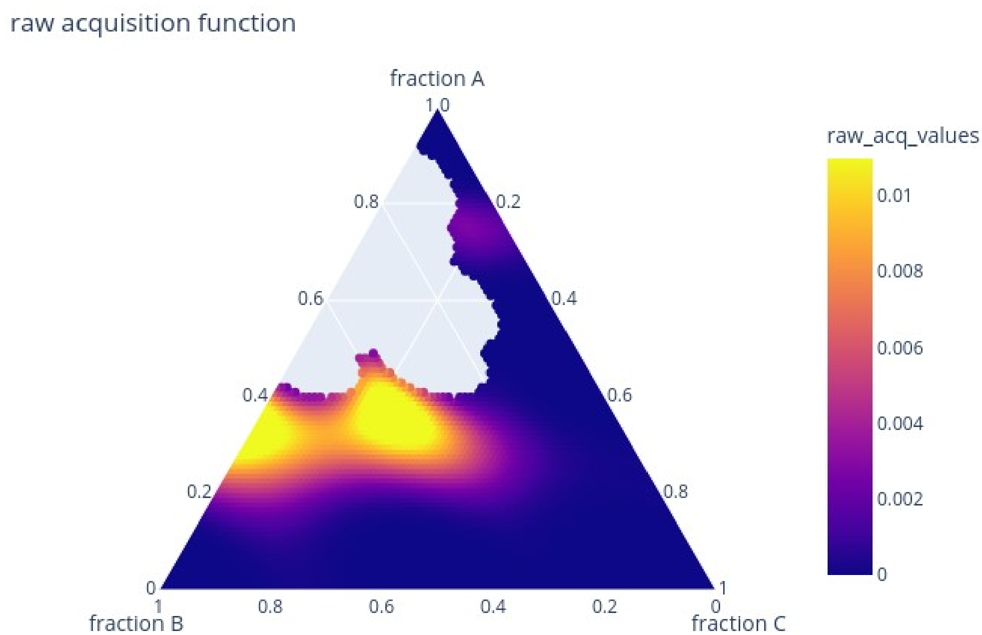


Figure S5. Raw acquisition function using expected improvement with contextual improvement.

An additional learned constraint is used to further reduce the probability that a point within the miscible space is chosen. This learned constraint is modelled via a Gaussian Process model using a radial basis function kernel with the length scale optimized via minimization of the negative log marginal likelihood. In this case the target value being modeled is a binary indicator variable defined as 0 if the point was classified miscible and 1 if the point was classified immiscible. The result of fitting the learned constraint to data is shown in Figure S6. The raw acquisition function is multiplied by the predicted mean values of this Gaussian Process thus reducing the acquisition function for points likely to be miscible.

Application of the constraint function to the raw acquisition function is shown in Figure S7. The two original areas suggested for the next experiment remain. However, the inclusion of the learned constraint function reduces the spread of proposed points. The maximum values are around the point (0.65, 0.35, 0.0), and that point is suggested to start iteration 10.

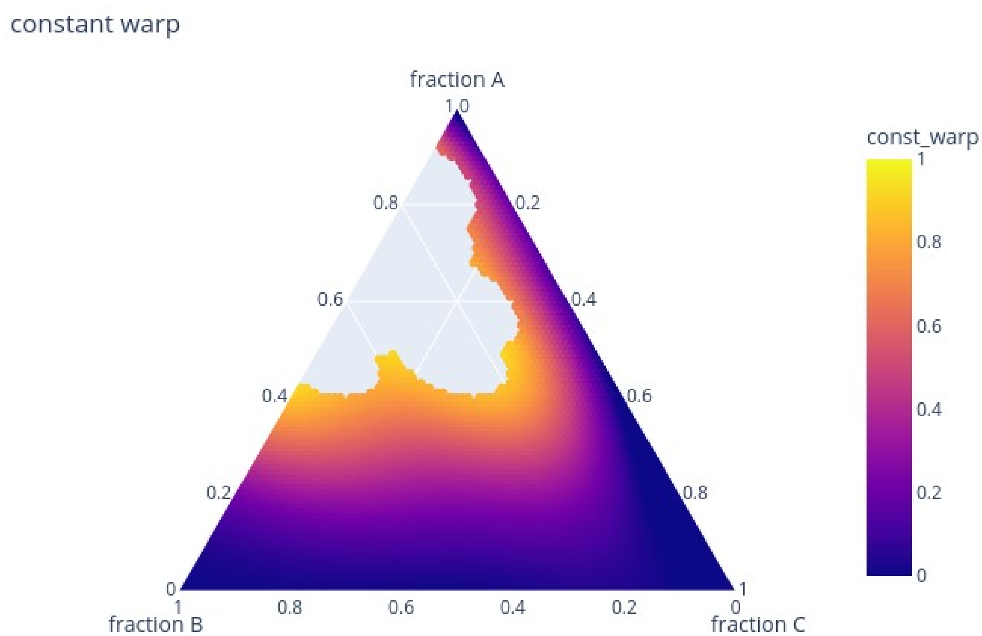


Figure S6. Learned constraint used to reduce the chance of selecting a point that is miscible.

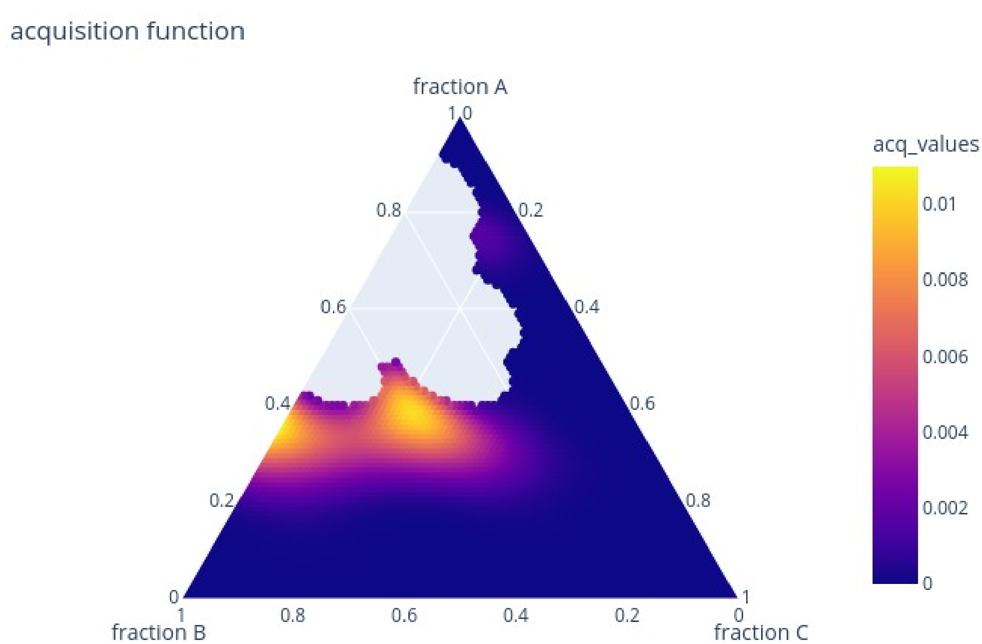


Figure S7. Acquisition function adjusted with the constraint function.

The computation of the (0.65, 0.35, 0.0) point shows it to be immiscible. The trajectory and the visualization of the ternary phase diagram are updated and shown in Figure S8. The area assigned immiscible is expanded along the A-B binary axis.

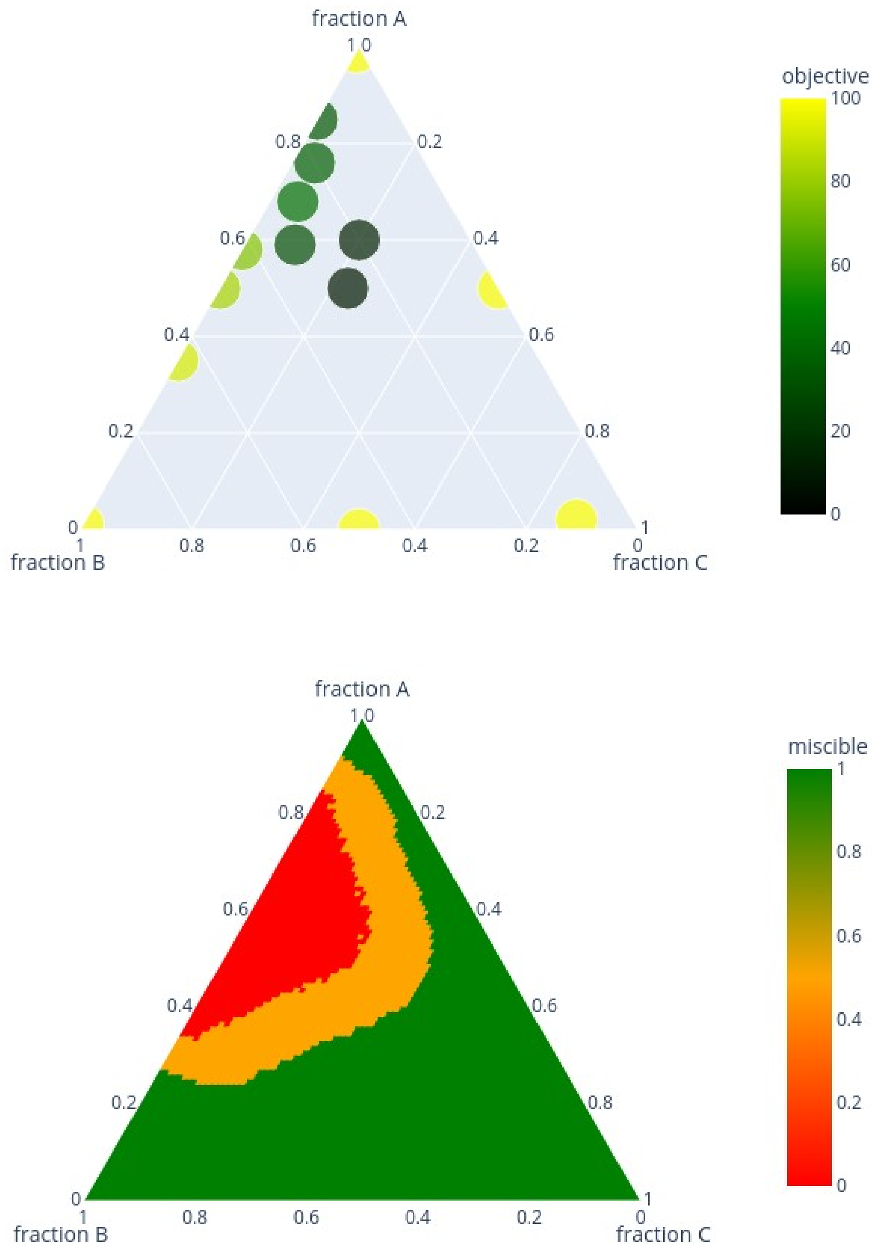


Figure S8. The points measured at the end of the tenth iteration, and the predicted areas assigned immiscible (red), miscible (green), and where the boundary lies (orange).