doi:10.1017/S1431927622005190
Microsc. Microanal. 28 (Suppl 1), 2022

Expediting Cryo-EM Grid Optimization by Utilizing Statistical Analysis with JMP

Rose Marie Haynes[1,2*]

[1.] Pacific Northwest Center for Cryo EM, Oregon Health & Science University, Portland, OR, USA.
[2.] Pacific Northwest National Laboratory, Richland, WA, USA.
* Corresponding author: rose.haynes@pnnl.gov

The process of optimizing the freezing conditions for cryo-EM grids prior to collecting single particle analysis (SPA) data is generally iterative, requiring the researcher to perform multiple rounds of grid freezing and screening under the microscope to achieve the optimal ice quality and particle density needed for data collection. Even well-educated guesses about the initial sample concentration, blotting conditions, and need for additives may result in grids that are not only poor for data collection, but fail to provide useful information for improving future attempts. For researchers who rely on facilities such as the national cryo-EM centers for grid screening, the waiting period between screening sessions can be significant, which creates a major bottleneck when repeated iterations of optimization are needed [1]. For this reason, we propose a strategic method to expedite the iterative process of cryo-EM grid optimization by selectively testing a wide range of input conditions in batches for screening, and more efficiently optimizing the parameter space to produce data-collection-ready grids.

Our proposed method is rooted in the principles of Design of Experiments (DOE) which allow a researcher to determine how several input variables (factors) interact with one another to produce output variables (responses) [2]. Here, we use a common experimental design within DOE known as a fractional factorial design to probe a wide range of grid freezing parameters through a limited set of test conditions [3]. The fractional factorial design used here (Table 1) was designed to test the widest range of conditions possible while remaining within the 11-grid limit for each screening session at the Pacific Northwest Center for Cryo-EM [4]. The DOE-based software package JMP provides a user-friendly way to set up a table of test conditions using a fractional factorial design [5]. JMP furthermore provides the analysis and model-fitting capability needed to determine how each factor affects each response, how factors interact with each other, and predicts optimal conditions needed to achieve the desired response.

We have demonstrated a strategic method of cryo-EM grid optimization following DOE principles using apoferritin. Since DOE allows for the full range of a continuous input variable to be modeled by testing only a few values, we selected the starting conditions for sample concentration, blot time, and sample volume based on what were estimated to be the minimum and maximum reasonable values. Sets of test conditions were determined by creating a fractional factorial design (Table 1) and all other grid freezing parameters (temperature, humidity, blot force, glow discharge settings, and grid type) were kept constant. 11 grids were frozen on a Vitrobot and screened on PNCC's Arctica (Fig. 1). A rubric of consistent scoring criteria was developed for scoring grids during screening based on particle distribution and ice thickness. This was important for limiting subjectivity of values being used to create statistical models.

A least squares regression model was created in JMP using these responses, which was used to calculate the relative effects of each factor on each response and, more importantly, output an optimized set of freezing conditions (Fig. 2) [6]. Another set of grids was frozen using these optimized conditions. Screening results show that these grids have near-optimal ice thickness and particle distribution (Fig. 3),

with enough imageable area for multiple days of SPA data collection.

We conclude that applying DOE methods to cryo-EM grid preparation is a reliable way of expediting the grid optimization process by limiting the number of rounds of freezing and screening needed to obtain data-collection-ready grids. While apoferritin is well-characterized and robust, this method could reasonably be applied to more complicated samples by adjusting the factors being tested (i.e. by testing detergent type, detergent concentration, the presence of a continuous substrate, or different grid types in addition to sample concentration and blot time). This is particularly useful to those who rely on national cryo-EM facilities for screening and are subject to waiting periods between screening sessions, but it can provide benefit to anyone wishing to make the grid optimization process faster.

Concentration (mg/ml)	Blot Time (s)	Sample Volume (µl)	Particle Distribution Rating (-3 to 3)	Ice Thickness Rating (-3 to 3)
9.6	3	4	0.5	1
2.2	3	2.5	-1	-1
2.2	6	2.5	-2	-1.5
9.6	1.5	2.5	1	1.5
2.2	1.5	4	-2	1

Table 1. Conditions used for the first round of grid freezing, with resulting grid quality scores shown in the last two columns. The combination and scope of factors were determined by creating a fractional factorial design [2], which allowed for 6 sets of conditions to be tested with replicate grids for each. Grids were given scores between -3 and 3 with 0 representing optimal results.

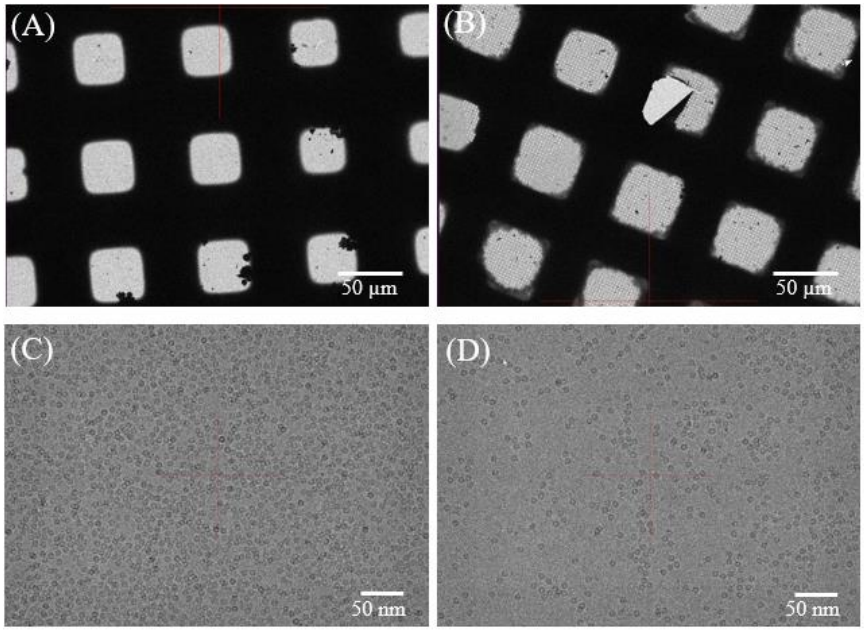


Figure 1. A selection of screening images from the first round of freezing and screening shows the range of resulting ice thickness and particle density. The ice thickness in (A) was high enough to have an undesirable effect on signal without entirely prohibiting data collection, thus this grid was given a score of 1.5 on a scale of -3 to 3. For comparison, the ice thickness in (B) was thin overall, with broken and dry holes throughout, and was given a score of -1.5. Examples of high and low particle distribution from this set of grids is shown in (C) and (D); the amount of particle overlap in (C) is high enough that it

https://doi.org/10.1017/S1431927622005190 Published online by Cambridge University Press

would likely present challenges with data processing, therefore it was given a score of 1 on a scale of -3 to 3. The sparseness of particles in (D) corresponds with a score of -2.

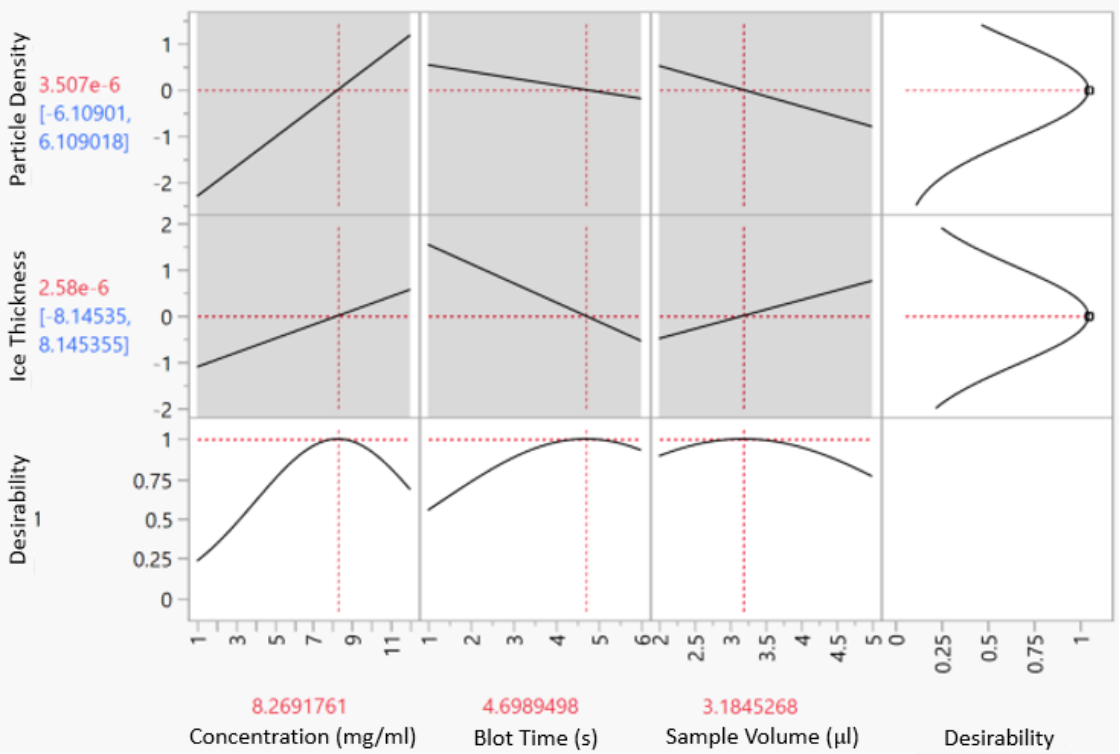
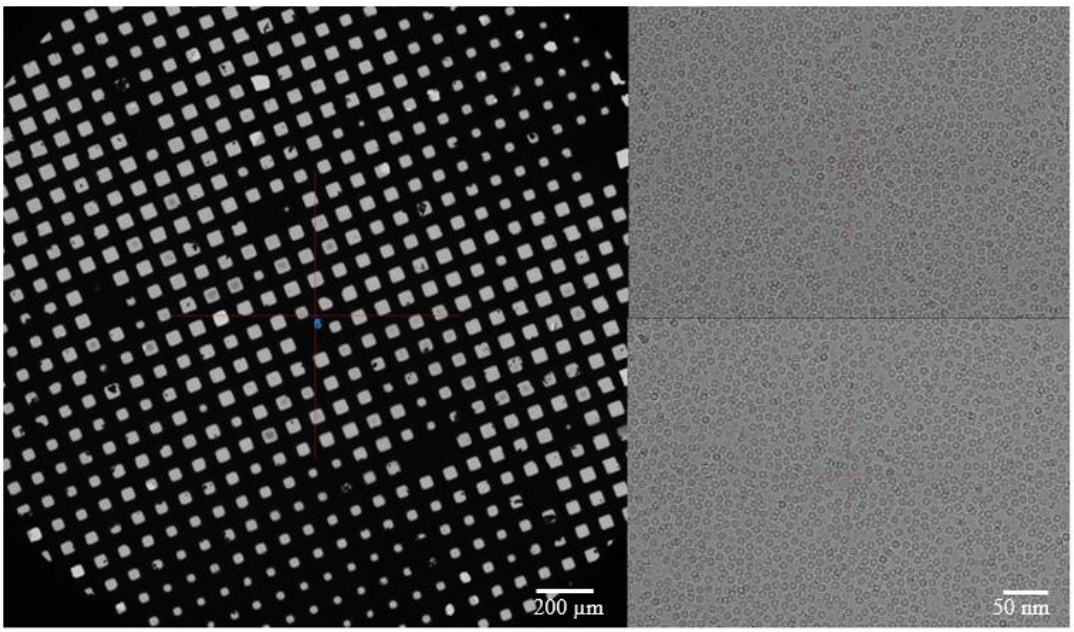


Figure 2. Results from JMP statistical model based on screening results, indicating the most likely optimal values for input conditions. This prediction profiler shows the model curves for which input values will produce the desired response; the values in red are optimized for maximum desirability and were used to freeze the final set of grids (concentration: 8.2 mg/ml, blot time: 4.7s, sample volume 3.2 µl).



https://doi.org/10.1017/S1431927622005190 Published online by Cambridge University Press

Figure 3. Screening results from a representative grid frozen with optimized conditions calculated using JMP, which is highly suitable for data collection. Thin ice covers the majority of the grid and particle distribution is a near-monolayer in all areas imaged, which makes it ideal for SPA data collection.
References:

[1] Cryo EM Centers, http://www.cryoemcenters.org (accessed February 16, 2022)
[2] M Barad, Applied Mathematics **5** (2014), p. 2120-2129. doi:10.4236/am.2014.514206
[3] C Natoli in "Classical Designs: Fractional Factorial Designs", STAT Center of Excellence, p.1-8.
[4] Pacific Northwest Cryo EM Center, http://pncc.labworks.org (accessed February 16, 2022)
[5] JMP®, Version *16.2.0*. SAS Institute Inc., Cary, NC, 1989–2021.
[6] G Watson, Annals of Mathematical Statistics **38(6)** (1967), p.1697-1699. doi:10.1214/aoms/1177698603
[7] This research was supported by NIH grant U24GM129547 and performed at the PNCC at OHSU and accessed through EMSL (grid.436923.9), a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research.