RNA Crystal Improvement with Definitive Screening Designs

Barat Venkataramany[1], Francis Aqucah[1], Syed Aslam[1], Charles Charter Jr.[2], and Blaine Mooers[1]

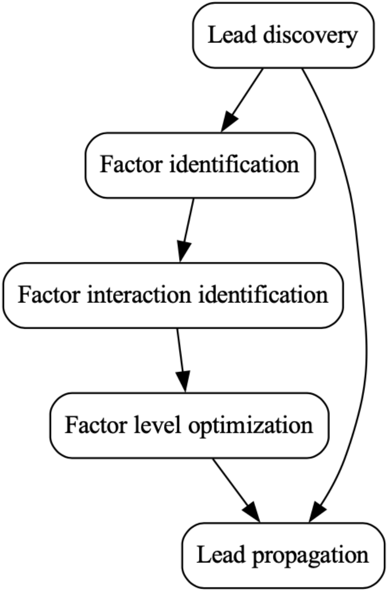
`MooersLab/dsd4xtals`

[1]Biochemistry & Molecular Biology, University of Oklahoma Health Sciences Center, Oklahoma City

[2]Biochemistry & Biophysics, University of North Carolina School of Medicine, Chapel Hill

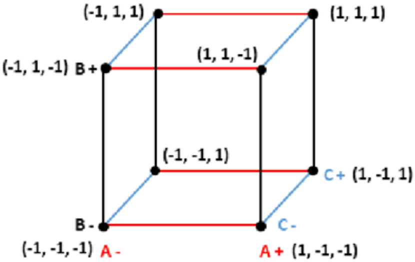
ACA Annual Meeting, Baltimore, MD
8 July 2023

Workflows


Lead discovery
Factor identification
Factor interaction identification
Factor level optimization
Lead propagation

2^3 Full factorial design

	Factor		
Run	A	B	C
1	-	-	-
2	+	-	-
3	-	+	-
4	+	+	-
5	-	-	+
6	+	-	+
7	-	+	+
8	+	+	+



https://www.afit.edu/stat/statcoe_files/Classical%20Designs-Full%20Factorial%20Designs_Final.pdf

No. drops for 3-level experiments

Design type	Number of factors				
	3	4	5	6	7
3-level Full Factorial	27	81	243	729	2187
Central Composite Design	15	25	27	45	79
Definitive Screening Design	9	9	13	13	17

Adapted from Table 2 from Errore et al. (2017) Using definitive screening designs to identify active first- and second-order factor effects. *J. Quality Technology* 49, 244-264.

Jones, B. and Nachtsheim, C.J. (2011) A class of three-level designs for definitive screening in the presence of second-order effects. *J. Quality Technology*, 43, 1–15.

DSD design matrix

$$\mathbf{D} = \begin{pmatrix} \mathbf{C}_4 \\ -\mathbf{C}_4 \\ \mathbf{0} \end{pmatrix} = \begin{array}{c} \begin{array}{rrrr} x_1 & x_2 & x_3 & x_4 \end{array} \\ \left(\begin{array}{rrrr} 0 & 1 & 1 & 1 \\ -1 & 0 & -1 & 1 \\ -1 & 1 & 0 & -1 \\ -1 & -1 & 1 & 0 \\ \hdashline 0 & -1 & -1 & -1 \\ 1 & 0 & 1 & -1 \\ 1 & -1 & 0 & 1 \\ 1 & 1 & -1 & 0 \\ \hdashline 0 & 0 & 0 & 0 \end{array}\right) \end{array}$$

Xiao,L., Lin,D.K.J. and Bai,F. (2012) *J. Quality Technology*, 44, 2–8.

The BigCrystalsFast library

Parameter	Values
Optimality criterion	J or D
Number of factors	1-5
Number of blocks	2-6
Tray sizes	9, 15, 24, 48, 72, 96
Total number of samples	9, 15, 30, 45, 60, 75, 90, 24, 48, 72, 96

User input

Factor Name	Mean	Delta	
[BaCl2]	50	30	
[Cobalt Hexammine]	7	5	
[% MPD v/v]	12.5	7.5	
number of wells * volume of drop * tip loss			
Total volume of protein or nucleic acid (uL):	26.4		
Reservoir:	60% MPD	Volume (uL):	350

Stocks	[Stock] (mM)	[Lead] (mM)
[BaCl2]	500	50
[Cobalt Hexammine]	50	7
[% MPD v/v]	50	12.5
NaCacodylate	500	40

Tube Vol. (uL)	100
Prot or NA Vol. (uL)	1
Xtal soln. Vol. (uL)	1
pH	6.7

Design matrix with three real factors

Wells	Runs	Random	Treatment	A Coding	B Coding	C Coding
Well 1, A1	1	0.767946518	21	0	0	0
Well 2, A2	2	0.017386382	8	1	1	-1
Well 3, A3	3	0.394998816	24	0	0	0
Well 4, A4	4	0.84510287	18	-1	-1	1
Well 5, A5	5	0.962814523	7	1	1	-1
Well 6, A6	6	0.492180712	20	-1	-1	-1
Well 7, B1	7	0.148571568	12	-1	0	1
Well 8, B2	8	0.651321708	11	0	-1	-1
Well 9, B3	9	0.722795032	15	-1	1	-1
Well 10, B4	10	0.65495096	22	0	0	0
Well 11, B5	11	0.513146964	17	-1	-1	1
Well 12, B6	12	0.322241229	10	1	1	1
Well 13, C1	13	0.014077323	6	1	-1	1
Well 14, C2	14	0.893487256	23	0	0	0
Well 15, C3	15	0.709750367	13	-1	1	0
Well 16, C4	16	0.943513685	9	1	1	1
Well 17, C5	17	0.192212226	4	1	-1	-1
Well 18, C6	18	0.835343853	3	1	-1	0
Well 19, D1	19	0.459329968	16	-1	1	-1
Well 20, D2	20	0.539832073	19	-1	-1	-1
Well 21, D3	21	0.403861778	1	0	1	1
Well 22, D4	22	0.389481159	14	-1	1	1
Well 23, D5	23	0.102614831	2	1	0	-1
Well 24, D6	24	0.025931854	5	1	-1	1

Design matrix with seven fake factors

C Coding	Fake1	Fake2	Fake3	Fake4	Fake5	Fake6	Fake7	A Level	B Level	C Level
0	0	0	0	0	0	0	0	50	7	12.5
-1	-1	1	1	-1	-1	0	1	80	12	5
0	0	0	0	0	0	0	0	50	7	12.5
1	1	-1	-1	1	1	0	-1	20	2	20
-1	1	1	-1	1	0	-1	-1	80	12	5
-1	0	1	-1	1	-1	1	1	20	2	5
1	-1	1	1	1	-1	-1	-1	20	7	20
-1	-1	-1	-1	-1	-1	-1	-1	50	2	5
-1	-1	-1	0	1	1	-1	1	20	12	5
0	0	0	0	0	0	0	0	50	7	12.5
1	-1	-1	1	-1	0	1	1	20	2	20

Example well

Well 6, A6	Treatment:	20	100
Reagent	[Final]	[Stock]	Vol. (uL)
[BaCl2]	20	500	4
[Cobalt Hexammine]	2	50	4
[% MPD v/v]	5	50	10
NaCacodylate	40	500	8
ddH2O	N/A	N/A	74
		Check:	100
Protein in drop (uL)	1		
Reservoir soln. in drop (uL)	1		
Reservoir (uL) in well	60% MPD	Volume (uL):	350
Response			

Treatments and scorecard

	Well 1, A1	Well 2, A2	Well 3, A3	Well 4, A4	Well 5, A5	Well 6, A6
Date:	21	8	24	18	7	20

	Well 7, B1	Well 8, B2	Well 9, B3	Well 10, B4	Well 11, B5	Well 12, B6
Date:	12	11	15	22	17	10

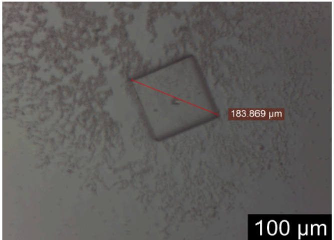
	Well 13, C1	Well 14, C2	Well 15, C3	Well 16, C4	Well 17, C5	Well 18, C6
Date:	6	23	13	9	4	3

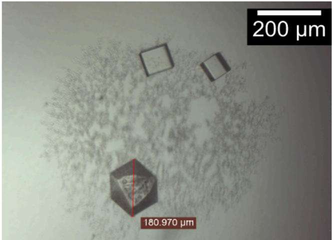
	Well 19, D1	Well 20, D2	Well 21, D3	Well 22, D4	Well 23, D5	Well 24, D6
Date:	16	19	1	14	2	5

Description	Description
Clear	0
Opaque precipitate or skin	1
Gelatinous precipitant	2
Phase separation (oil)	3
Granular (microcrystals)	4
Spherulates, urchins, dendrites	5
quasicrystals	6
1-D needles	7
2-D plates	8
Microxtals, 3-D xtals resolved	9
Fused 3-D xtals 20-100 micron	10
Fused 3-D xtals 100-200 micron	11
Fused 3-D xtals 200-400 micron	12
Fused 3-D xtals 400-600 micron	13
Fused 3-D xtals > 600 microns	14
3-D xtals 20-100 micron	15
3-D xtals 100-200 microns	16
3-D xtals 200-400 microns	17
3-D xtals 400-600 microns	18
3-D xtals 600-800 microns	19
3-D xtals > 800 microns	20

Response surface analysis (I of II)

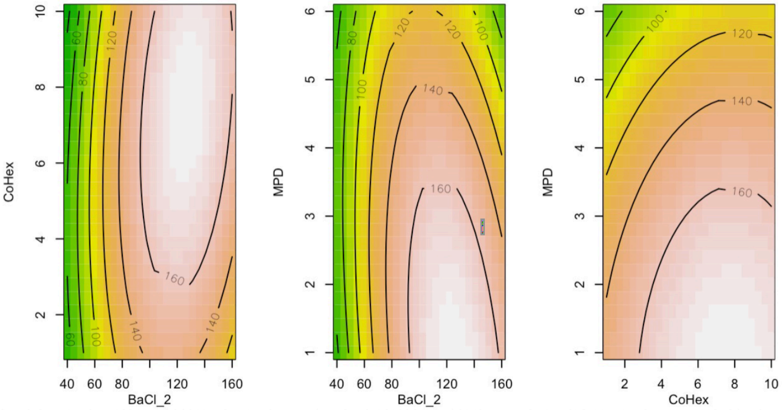
	Estimate	Std. Error	t-value	Pr(>|t|)	
(Intercept)	156.6	18.8	8.3	8.40E-07	***
x1	22.1	9.3	2.4	0.03	*
x2	10.2	9.3	1.1	0.29	
x3	-24.2	9.3	-2.6	0.02	*
x1*x2	15.5	10.5	1.5	0.16	
x1*x3	-21.2	10.5	-2	0.06	.
x2*x3	3.1	10.5	0.3	0.8	
x1^2	-59.4	25.3	-2.3	0.03	*
x2^2	-15.8	25.3	-0.6	0.54	
x3^2	-14.1	25.3	-0.6	0.59	
Significance Codes:			Multiple R-squared: 0.7414		
p < .001	***		Adjusted R-squared: .5751		
p < .01	**		F-statistic: 4.459 on 9 and 14 DF		
p < .05	*		p-value: .006436		
p < .1	.				





Response surface analysis (II of II)

$$y = x_1 + x_2 + x_3 + x_1 \cdot x_2 + x_1 \cdot x_3 + x_2 \cdot x_3 + x_1^2 + x_2^2 + x_3^3$$



Lenth, R.V. (2009) *J. Statistical Software*, 32(7), 1–17.

Conclusions

- sparse matrix $\rightarrow$ DSD $\rightarrow$ J-optimal $\rightarrow$ amplify
- Jones, B. and Montgomery, D.C. (2020) *Design of experiments: a modern approach.* Wiley and Sons.
  - DSDs are limited to 3 active factors.
  - DSD are limited to categorical factors with two levels
- dsd4xtals library of spreadsheets: `https://github.com/MooersLab/dsd4xtals`

Acknowledgements

Funding:

- NIH: R01 CA242845, R01 AI088011
- NIH: P30 CA225520 (PI: R. Mannel); P20GM103640 and P30GM145423 (PI: A. West)
- OCAST HR20-002;
- PHF Team Science Grant