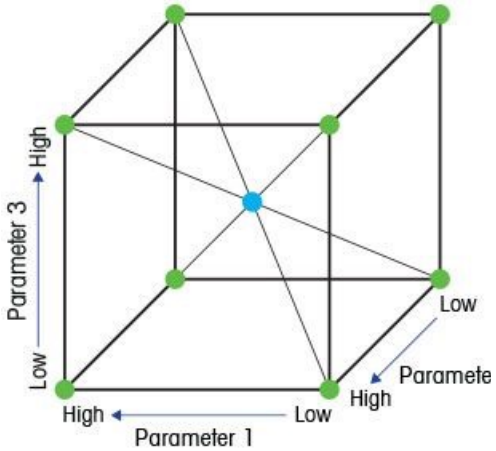
Statistical Design of Experiments



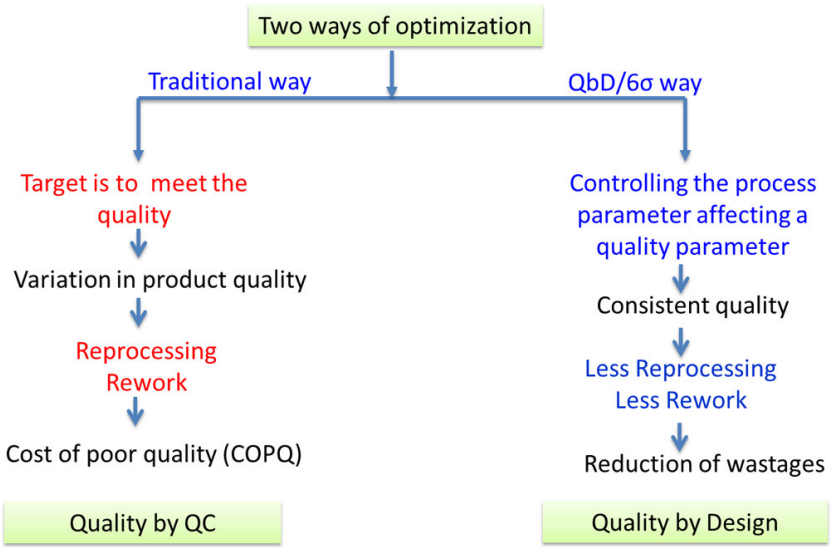
Travis Menard

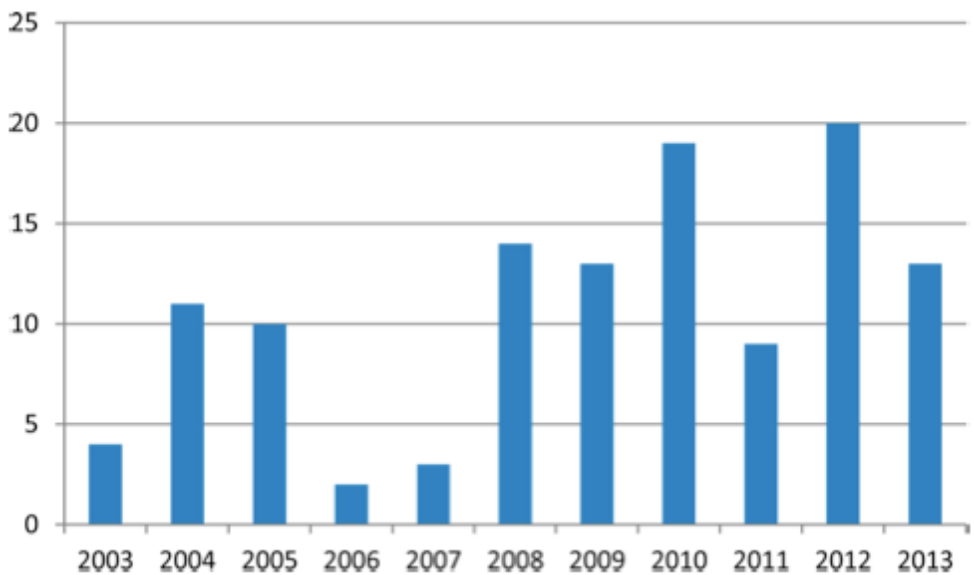
Group Meeting

January 23rd, 2023

Quality by Design

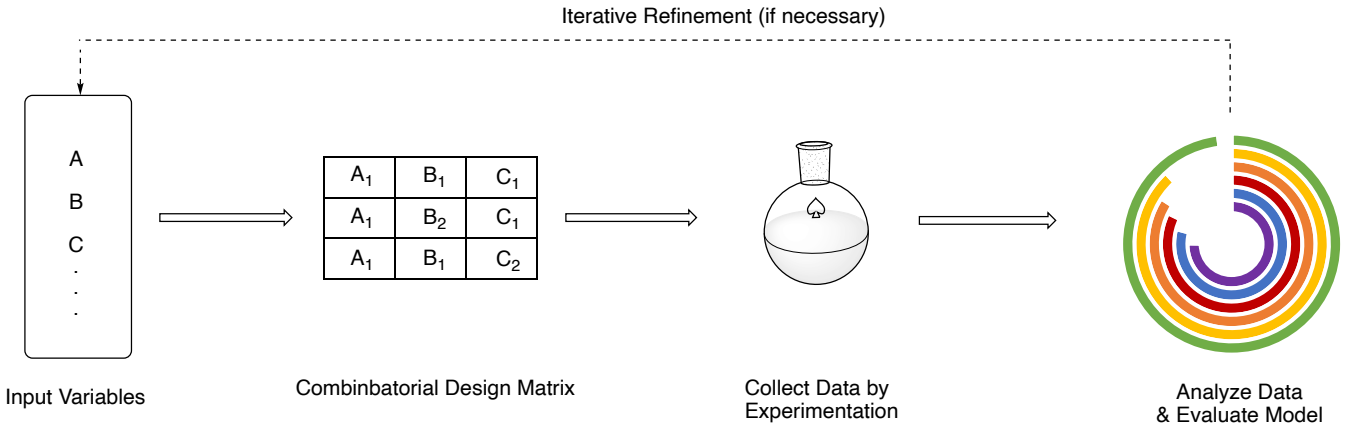
Quality by Design (QbD):
A systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management. - ICH

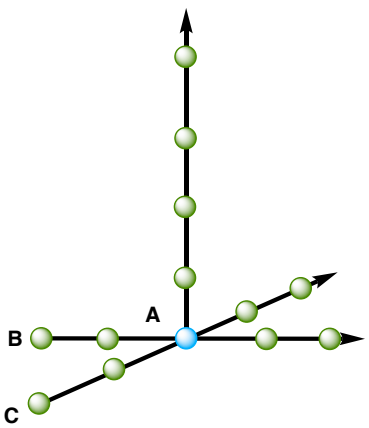




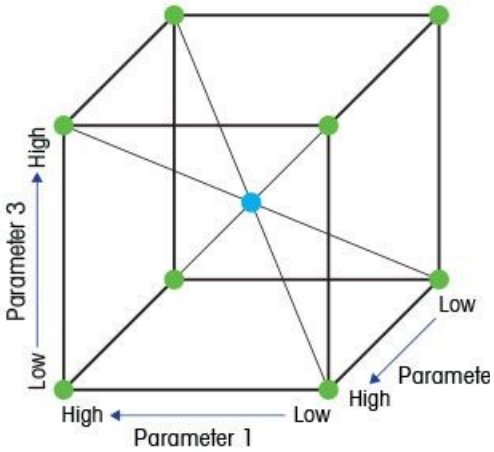
Org. Process Res. Dev. 2015, 19, 1634–1644
Org. Process Res. Dev. 2015, 19, 11, 1605–1633

Design of Experiments



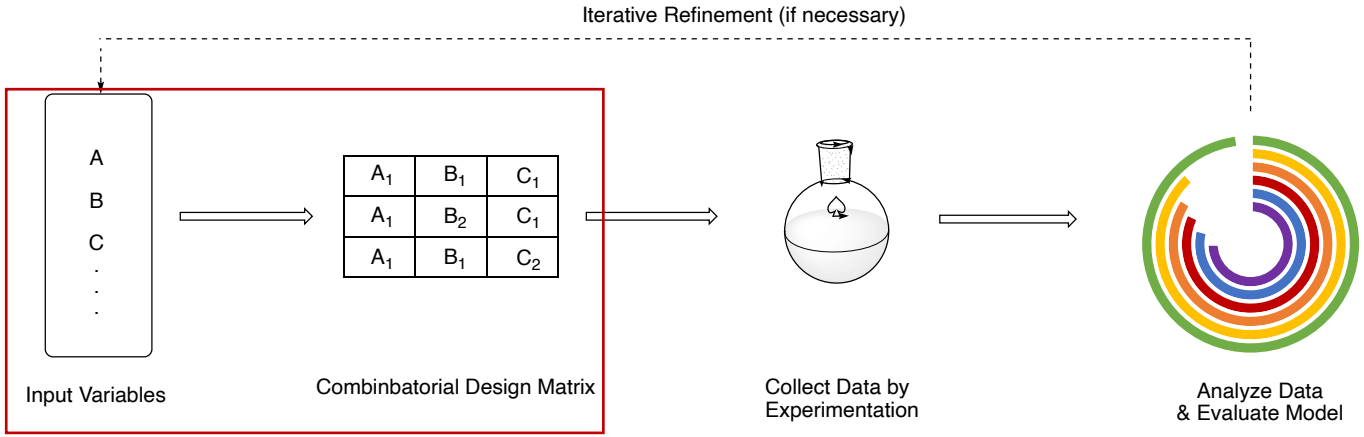


One Factor at a Time (OFAT)



https://www.mt.com/us/en/home/applications/L1_AutoChem_Applications/L2_ReactionAnalysis/design-of-experiments-doe.html

Design Matrixes



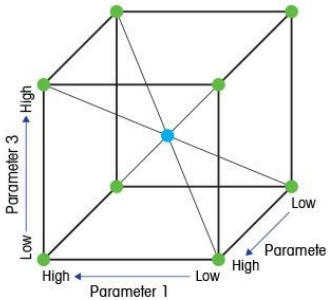
Temperature (25-50 ˚C)

Reagent X (2-5 equiv)

Concentration (0.1-0.5 M)

Input Variables

	Run Order	Temperature (˚ C)	Reagent X (Equiv.)	Concentration (M)
1	2	25 (-1)	2 (-1)	0.2 (-1)
2	6	50 (+1)	2 (-1)	0.2 (-1)
3	3	25 (-1)	5 (+1)	0.2 (-1)
4	1	25 (-1)	2 (-1)	0.5 (1)
5	8	50 (+1)	5 (+1)	0.2 (-1)
6	4	50 (+1)	2 (-1)	0.5 (+1)
7	10	25 (-1)	5 (+1)	0.2 (-1)
8	5	(50) +1	5 (+1)	0.5 (+1)
9	9	37.5 (0)	3.5 (0)	0.35 (0)
10	7	37.5 (0)	3.5 (0)	0.35 (0)

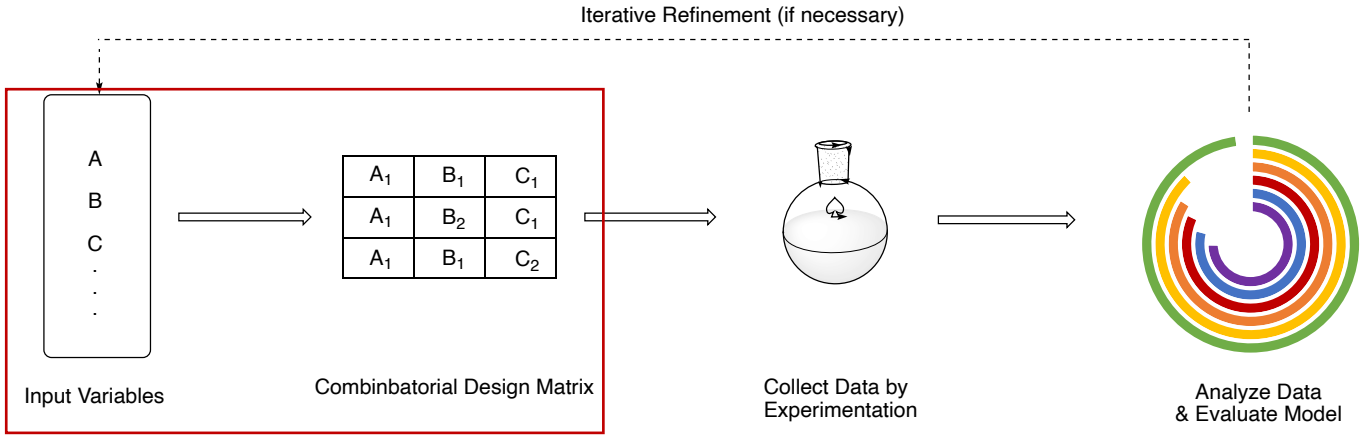


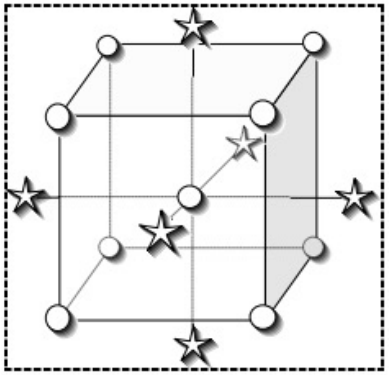
Full Factorial

(2^n runs + Center points)

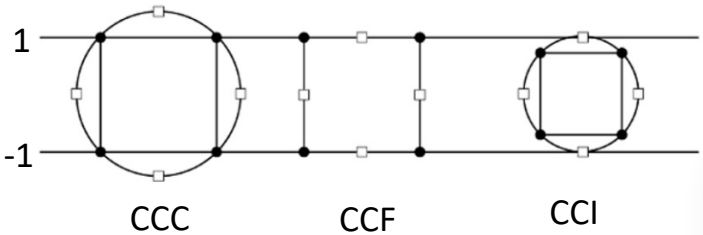
https://www.mt.com/us/en/home/applications/L1_AutoChem_Applications/L2_ReactionAnalysis/design-of-experiments-doe.html

Composite Centered Design



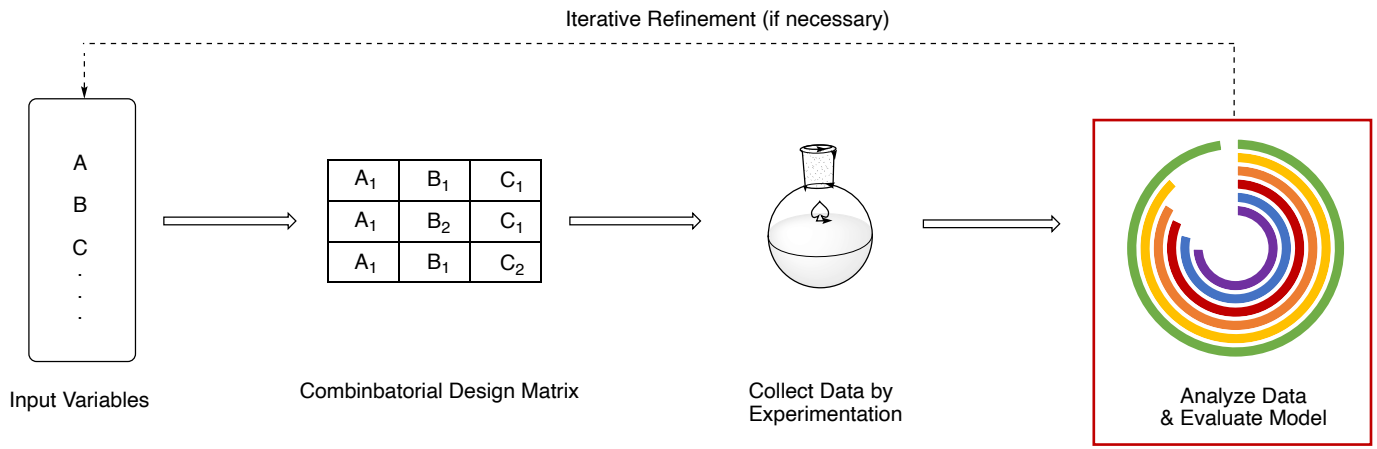


Composite Centered Design (CCD)



ICH PHARMACEUTICAL DEVELOPMENT Q8(R2) - *https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf*
Ait-Amir, B.; Pougnet, P.; El Hami, A. 6 - Meta-Model Development; El Hami, A., Pougnet, P. B. T.-E. M. S. 2, Eds.; Elsevier, 2015; pp 151–179.

ANOVA Basics



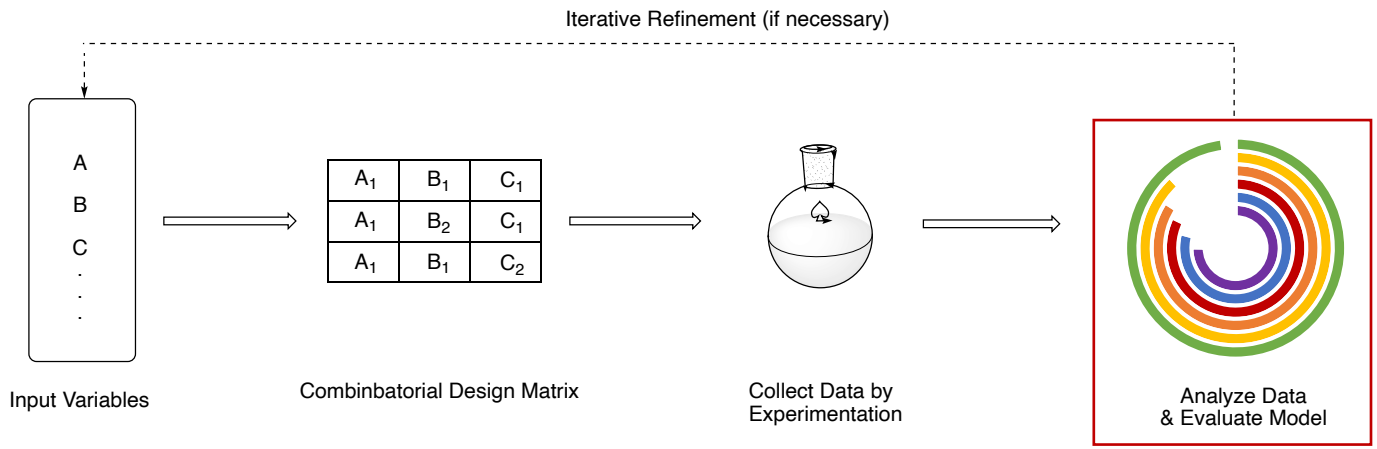
ANOVA

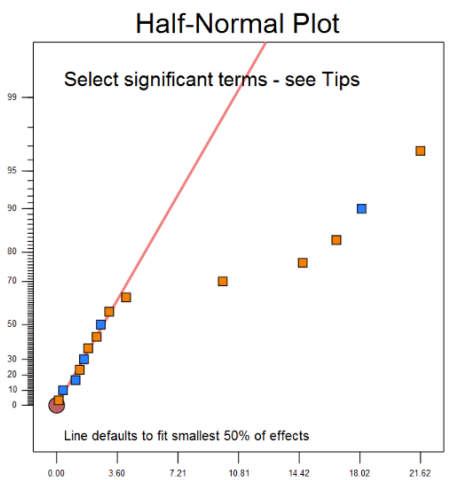
Sum of Squares	Mean Square	F-Value	P-Value
$\sum_{i=0}^{n} (X_i - \overline{X})^2$	$\frac{\text{Sum of Squares}}{\text{° of Freedom}}$	$\frac{\text{Varience Attributed to Investigated Factor (mean squared)}}{\text{Varience Atributed to Random Error}}$	Computed from a standard distribution of F-values ($P_{.05}$)

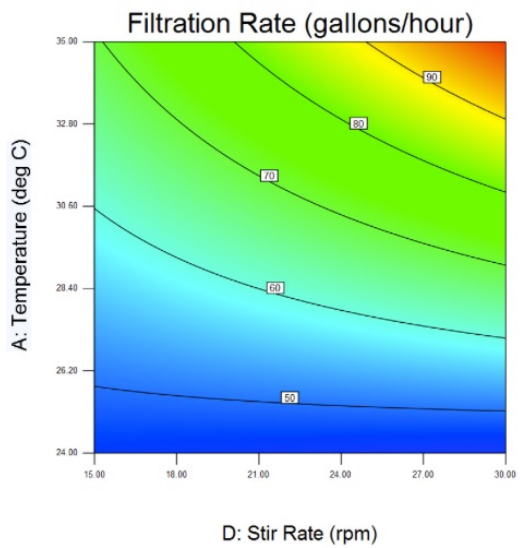
Residuals – Measure of random errors in a regression analysis (Should be normally distributed)

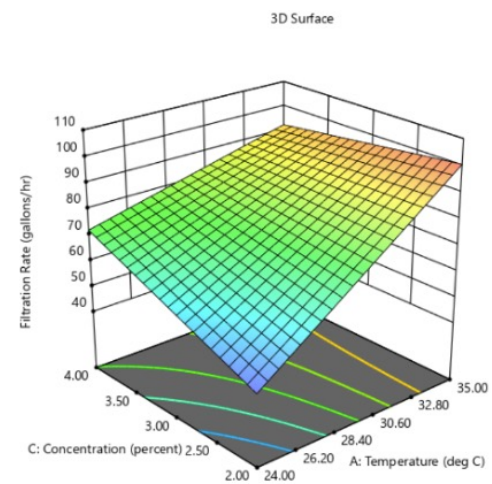
Curvature – Measure of linearity in the regression model (Statistically significant curvature Suggests quadratic terms are required)

Visualization Tools



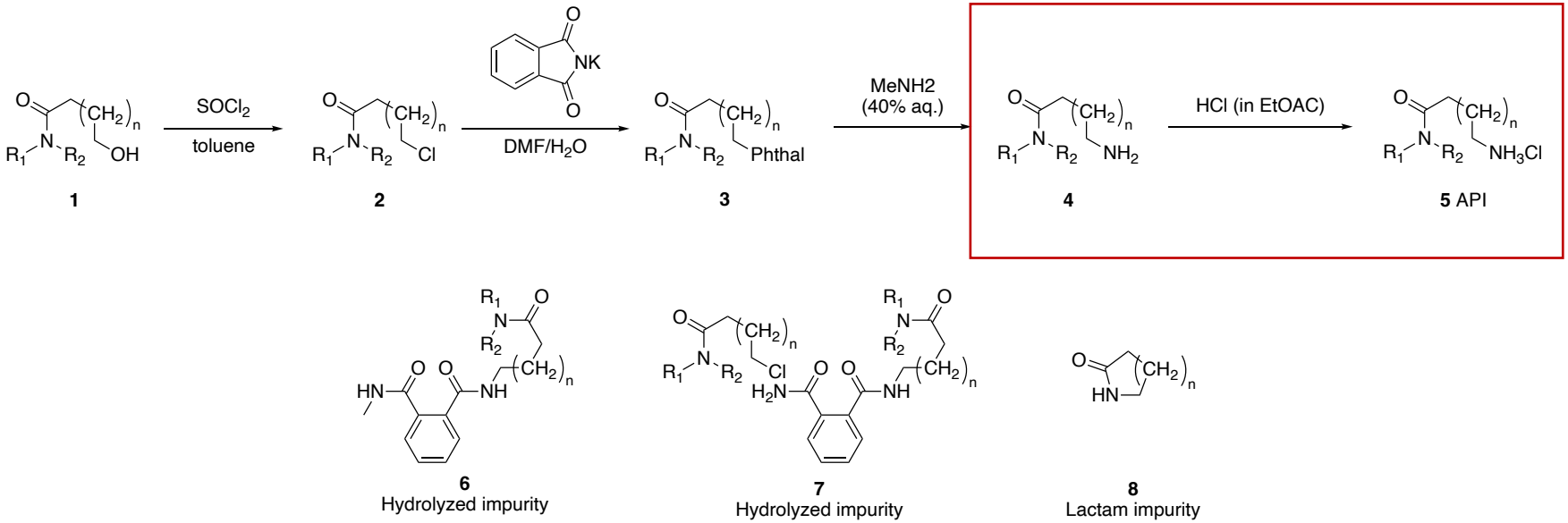


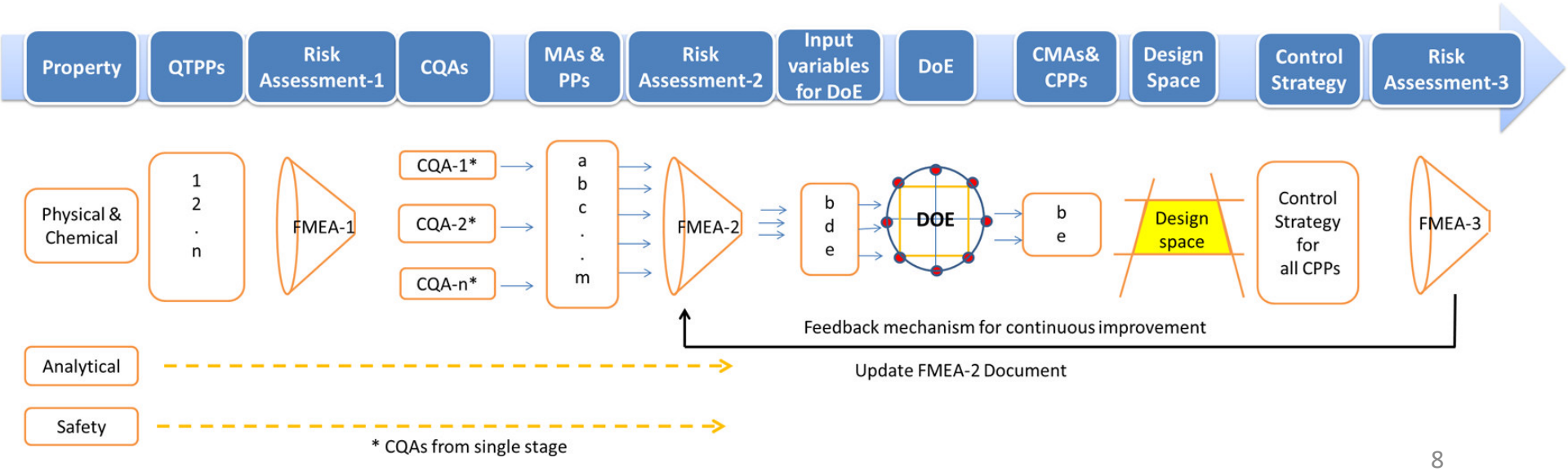




https://www.statease.com/docs/v13/tutorials/multifactor-rsm/

QBD Workflow





Org. Process Res. Dev. 2015, 19, 1634−1644

FMEA Tables

Ethyl acetate and isopropyl alcohol are added to a round- bottom flask. Crude compound 4 is added, and the mixture is stirred for 15 min to achieve complete dissolution. The solution is then cooled to 25 ± 10 °C, and ethyl acetate/HCl is added at the same temperature. The reaction mass is maintained for another 2 h at 25 ± 10 °C. The precipitated API hydrochloride salt is filtered under a nitrogen atmosphere, and the cake is washed with ethyl acetate. The material is then vacuum-dried, unloaded into a vacuum tray drier, and further dried at 47.5 ± 2.5°C until a constant weight is obtained.

			CQAs API (5)													
S. No	**activity**	**potential failure mode**	**yield**	**purity**	**hydrolyzed impurity 6**	**unreacted 4 impurity**	**HCl content**	**dimer impurity 7**	**lactam impurity 8**	**failure mode**	**present control**	**status of present control**	**occurrence**	**severity**	**detection (not)[13]**	**RPN**
1	charge 8.5-10 volumes of EtOAc into the reactor	more quantity of EtOAc	↕	↕	↕	↕	↕	↕	↕	measurement error	calibrated day tank	working	2	5	2	20
		low quantity of EtOAc	↓	↕	↕	↕	↕	↕	↕	measurement error	calibrated day tank	working	2	9	2	36
2	charge 1-1.5 Volumes of IPA into the reactor	more quantity of IPA	↓	↕	↕	↕	↕	↕	↕	measurement error	calibrated day tank	working but flow meter to be installed	5	9	7	315
		low quantity of IPA	↑	↓	↕	↕	↕	↕	↕	measurement error	calibrated day tank		5	9	7	315
3	charge crude stage **4** into the reactor	more quantity of 4	↓	↕	↕	↕	↕	↕	↕	analytical error	toluene correction to the assay	working	2	9	2	36
		less quantity of 4	↓	↕	↕	↕	↕	↕	↕	analytical error	toluene correction to the assay	working	2	9	2	36
4	stir the reaction mass till clear dissolution is observed	what if clear dissolution not observed	↓	↕	↕	↕	↕	↕	↕	manual error not critical	continue stirring till clear solution is obtained	working	2	9	2	36
5	add EtOAc.HCl into the reaction mass at 25±10 °C	charging at high temperature	↓	↓	↕	↕	↕	↕	↑	failure of chilled water line valve of RT water not closed	ensuring that RT water line is closed	working	5	9	5	225
		charging at low temperature	↓	↕	↕	↕	↕	↕	↕	RT water temperature fluctuation during winter	no effect	No control	5	7	7	245
		charging of more eq. of HCl	↓	↓	↕	↕	↕	↕	↑	manual error	releasing material on vendor's COA	working	5	7	9	315
		charging of less eq. of HCl	↓	↕	↕	↕	↕	↕	↕				5	5	9	225
		fast addition of EtOAc.HCl[14]	↓	↓	↕	↕	↕	↕	↑	manual error	addition at constant rate with flow meters in 45-60 minutes	working	5	5	5	125
		slow addtion of EtOAc.HCl	↕	↕	↕	↕	↕	↕	↕				5	5	5	125
6	maintain the reaction mass at 25±10 °C for 2 hour	more maintenance time	↕	↕	↕	↕	↕	↕	↕	no effect at 25±10 °C	not required	not required	5	2	3	30
		less maintenance time	↓	↓	↕	↕	↕	↕	↕	manual error	IPC for the absence of **4**	working	2	7	3	54
		maintenance at high temperature	↓	↓	↕	↕	↕	↕	↑	failure of chilled water plant	standby brine supply	working	2	7	3	54
		maintenance at low temperature	↕	↕	↕	↕	↕	↕	↕	No effect	Not required	Not required	2	5	3	30
7	filter the compound under nitrogen atmosphere.	filter under atmospheric conditions	↓	↕	↕	↕	↕	↕	↑	Failure of nitrogen plant	Standby nitrogen cylinders	working	3	9	3	81

↑ increase in **desired** CQA ↑ increase in **undesired** CQA ↕ no effect of CPPs on CQA

Org. Process Res. Dev. 2015, 19, 11, 1634-1644

FMEA Tables

2	charge 1-1.5 Volumes of IPA into the reactor	more quantity of IPA	↓	↕	↕	↕	↕	↕	↕	measurement error	calibrated day tank	working but flow meter to be installed	5	9	7	315
		low quantity of IPA	↑	↓	↕	↕	↕	↕	↕	measurement error	calibrated day tank		5	9	7	315
5	add EtOAc.HCl into the reaction mass at 25±10 °C	charging at high temperature	↓	↓	↕	↕	↕	↕	↑	failure of chilled water line valve of RT water not closed	ensuring that RT water line is closed	working	5	9	5	225
		charging at low temperature	↓	↕	↕	↕	↕	↕	↕	RT water temperature fluctuation during winter	no effect	No control	5	7	7	245
		charging of more eq. of HCl	↓	↓	↕	↕	↕	↕	↑	manual error	releasing material on vendor's COA	working	5	7	9	315
		charging of less eq. of HCl	↓	↕	↕	↕	↕	↕	↕				5	5	9	225
		fast addition of EtOAc.HCl[14]	↓	↓	↕	↕	↕	↕	↑	manual error	addition at constant rate with flow meters in 45-60 minutes	working	5	5	5	125
		slow addtion of EtOAc.HCl	↕	↕	↕	↕	↕	↕	↕				5	5	5	125

symbol	input variable for DoE	variable [RPN[b]] levels used for DoE	
		low (−1)	high (+1)
A	volume of IPA	0.8 volumes [315]	2 volumes [315]
B	addition temperature	5 °C [245]	35 °C [225]
C	amount of HCl	1 equiv [225]	1.2 equiv [315]

[a]A 2^3 full factorial design with four center points was planned to study the effect of the three important input PP_5 on the CQAs with all of the other PPs kept at predefined levels. [b]Risk priority numbers (RPNs) were taken from Table 2.

S. no.	PP	limit	justification
1	EtOAc volume	8–12 volumes	no effect on the responses at high volume
2	time the reaction mass is stirred until clear dissolution	not defined	not critical if stirred for a longer time, as it is just for dissolution
3	addition time of EtOAc/HCl	45–60 min	to control the exothermicity of reaction
4	cooling temperature before addition	10–15 °C	a low temperature was chosen to control the exothermicity
5	maintenance temperature	10–15 °C	same as the cooling temperature
6	maintenance time	2–2.5 h	no effect on the responses after 1 h
7	filtration	under nitrogen	dry atmosphere required

Org. Process Res. Dev. 2015, 19, 11, 1634-1644

ANOVA Evaluation

S. no.	IPA volume	cooling temp. (°C)	amount of HCl (equiv)	yield (%)	purity (%)	unreacted **4** (%)	impurity **6** (%)	impurity **7** (%)	impurity **8** (%)
1	2	5	1	78.37	99.96	0.01	0.02	0	0
2	2	5	1.2	80.54	99.89	0.01	0.01	0.03	0
3	2	35	1.2	74.01	99.98	0	0	0.04	0
4	0.8	5	1	82.72	99.85	0.01	0.05	0	0
5	0.8	35	1	82.72	99.94	0.01	0.01	0.03	0
6	2	35	1	80.54	99.92	0.01	0.01	0.01	0
7	0.8	5	1.2	87.07	99.96	0.01	0.03	0	0
8	0.8	35	1.2	87.07	99.99	0.01	0	0.02	0
9	1.4	20	1.1	87.07	99.97	0.01	0.02	0.01	0
10	1.4	20	1.1	84.90	99.97	0.01	0.02	0.02	0
11	1.4	20	1.1	87.07	99.93	0.01	0.01	0.03	0
12	1.4	20	1.1	84.90	99.95	0.01	0.01	0.02	0

source	sum of squares	degrees of freedom	mean square	F value	p value prob > F	
model	80.27	3	26.76	18.09	0.0011	significant
A (IPA volume)	61.28	1	61.28	41.42	0.0004	
C (equiv of HCl)	8.69	1	8.69	5.87	0.0459	
AC	10.30	1	10.30	6.96	0.0335	
curvature	39.65	1	39.65	26.80	0.0013	significant
residual	10.36	7	1.48			
lack of fit	5.62	4	1.40	0.89	0.5614	not significant
pure error	4.74	3	1.58			
cor total	130.28	11				

Org. Process Res. Dev. 2015, 19, 11, 1634-1644

Addition of CCD

point type	volumes of IPA	addition temp. (°C)	equiv of HCl	yield (%)
axial	1.4	20	1.2	78.67
axial	2	20	1.1	78.67
axial	0.8	20	1.1	83.04
centre	1.4	20	1.1	84.90
axial	1.4	35	1.1	80.85
axial	1.4	5	1.1	83.04
axial	1.4	20	1	80.85
centre	1.4	20	1.1	85.22

run	volumes of IPA	addition temp. (°C)	equiv of HCl	yield (%) actual	yield (%) predicted (95% CI)
1	1.5	10	1.1	83.9	82–85
2	1.5	15	1.1	84.3	82–85
3	1.5	20	1.1	84.7	82–85
4	1.5	15	1.1	83.9	82–85

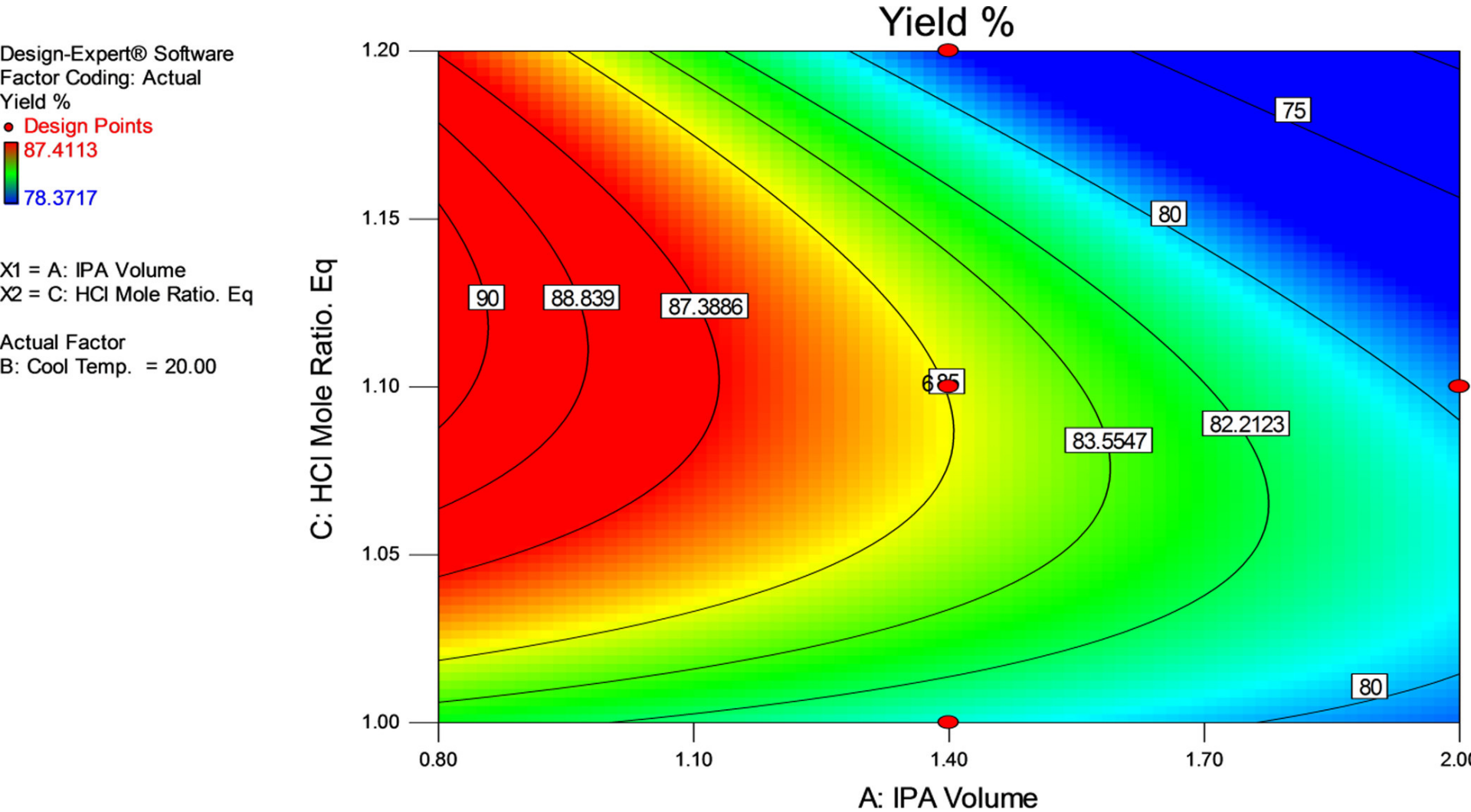
source	sum of squares	degrees of freedom	mean square	*F* value	p-value prob > *F*	
model	131.09	4	32.77	10.05	0.0008	significant
A	92.74	1	92.74	28.44	0.0002	
C	7.61	1	7.61	2.33	0.1524	
AC	35.77	1	35.77	10.97	0.0062	
C^2	68.05	1	68.05	20.87	0.0006	
residual	39.14	12	3.26			
lack of fit	31.87	7	4.55	3.13	0.1136	not significant
pure error	7.27	5	1.45			

$$\% \text{ yield} = -623.97 + 58A + 1240.5C - 61AC - 531C^2$$

Org. Process Res. Dev. 2015, 19, 11, 1634-1644

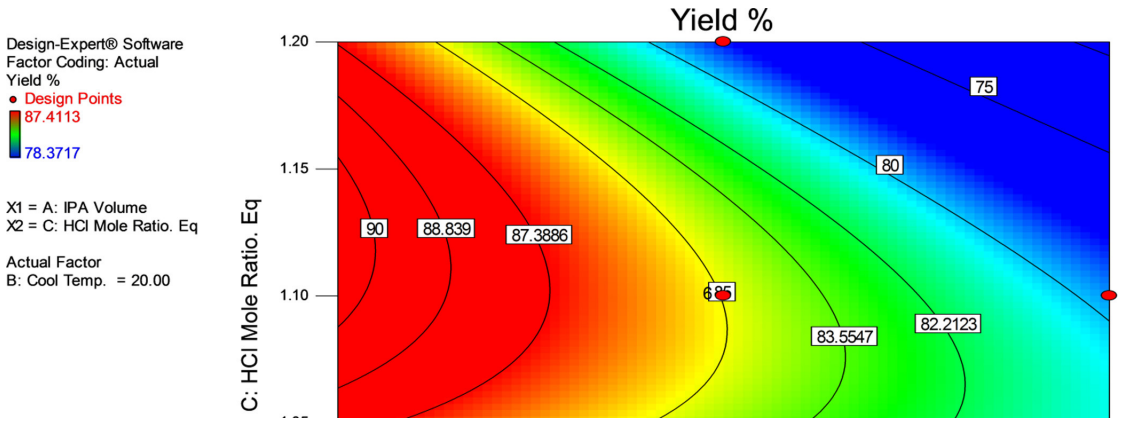
Contour Visualization

$$\% \text{ yield} = -623.97 + 58A + 1240.5C - 61AC - 531C^2$$



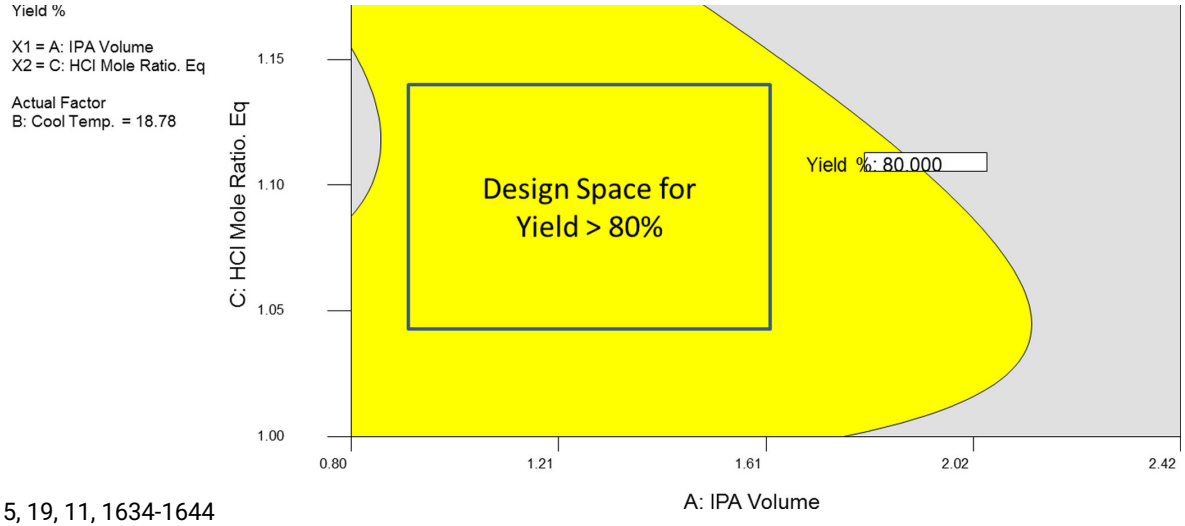
Org. Process Res. Dev. 2015, 19, 11, 1634-1644

Design Space



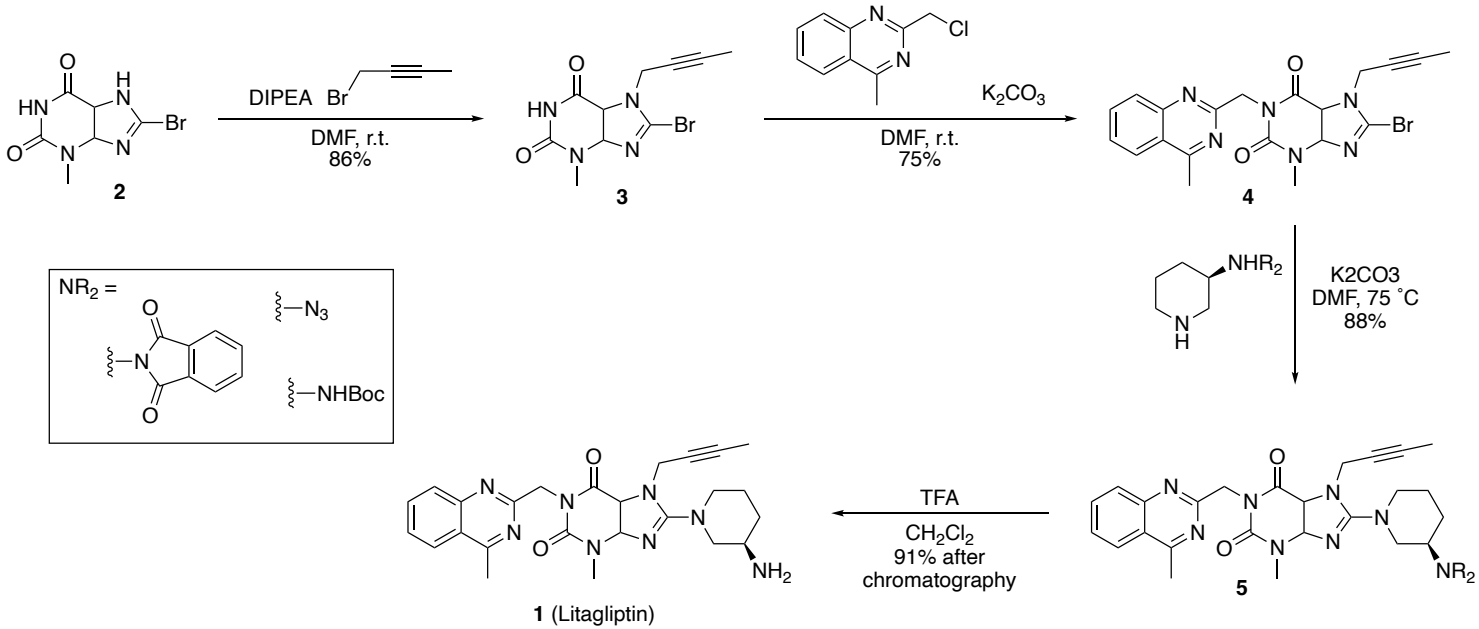
S. no.	factor	acceptable range	control strategy	risk assessment 3[a] O	S	D	RPN
1	IPA volume	0.8–1.6 volumes	only the calculated quantity of IPA is to be dispensed using a flow meter with additional calibration on the day storage tank	2	9	2	36
2	addition temperature[b]	10–30 °C	found to be noncritical from DoE, but the target was set at 20 °C	2	2	2	8
3	amount of HCl	1.04–1.14 equiv	assay of EtOAc/HCl solution to be done just before the batch is started to account for any HCl loss upon storage	2	5	2	20

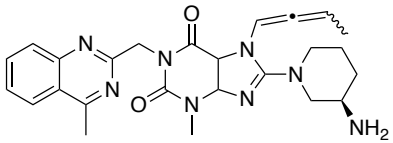
[a]O = occurrence, S = severity, D = detectability (not). [b]After DoE, temperature was found not to be a CPP.



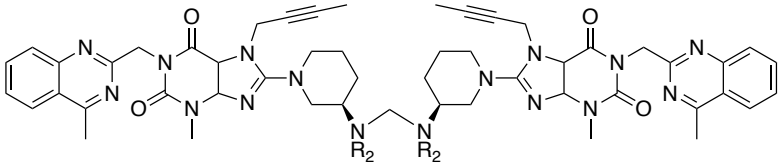
Org. Process Res. Dev. 2015, 19, 11, 1634-1644

Optimization of Litagliptin Synthesis





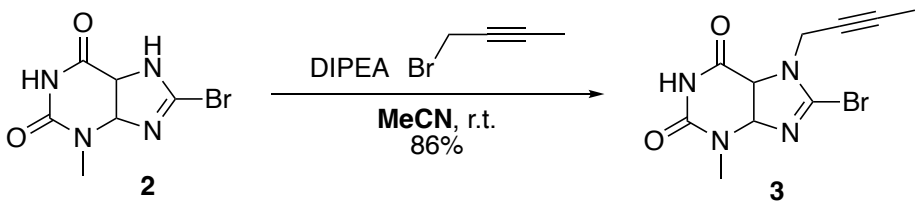
Impurity A



Impurity B

Org. Process Res. Dev. 2022, 26, 12, 3254−3264

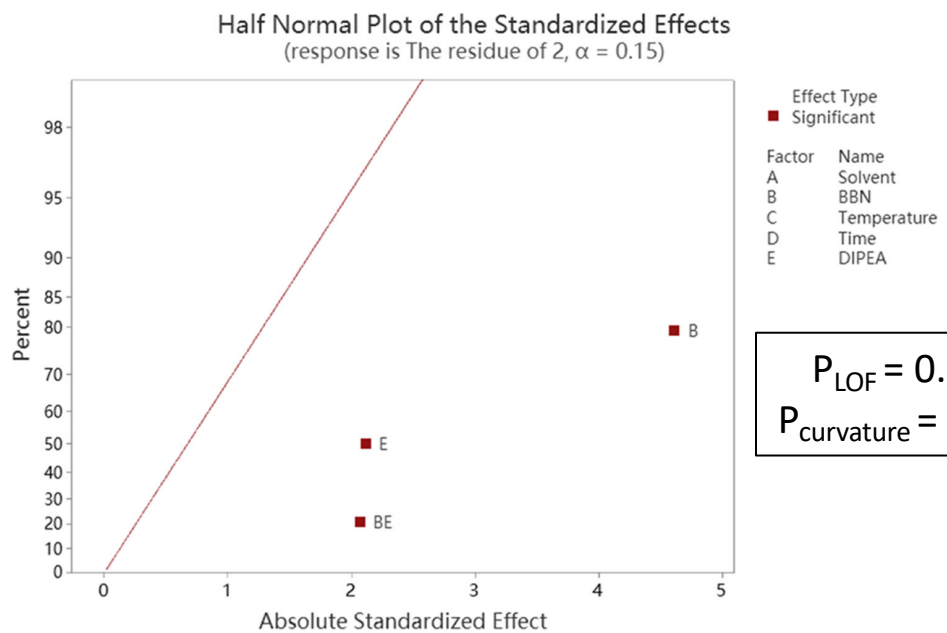
Optimization of Step 1



factor	unit	range
acetonitrile	mL/g	10–16
BNB	equiv	1.0–1.3
temperature	°C	50–80
time	h	8–20
DIPEA	equiv	1.1–1.5

[a]0.05 equiv of tetrabutylammonium bromide (TBAB) was added.

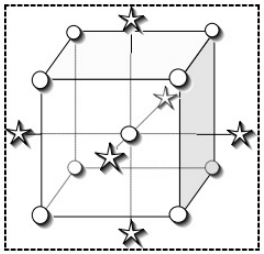
Full factorial: 2^5 = 32 experiments
¼ factorial: 8 experiments + 3 center points
Output = Residual 2



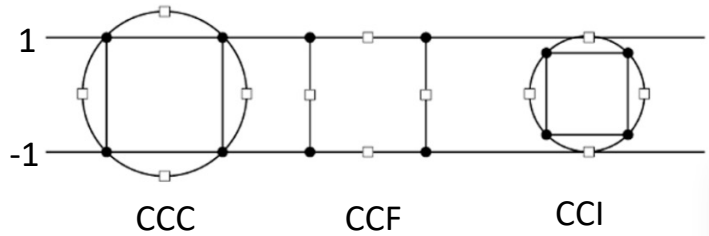
P_{LOF} = 0.018
$P_{curvature}$ = 0.019

Org. Process Res. Dev. 2022, 26, 12, 3254−3264
https://www.mt.com/sg/en/home/products/L1_AutochemProducts/chemical-synthesis-reactor-systems/EasyMax-102-Basic-Synthesis-Workstation.html

Step 1 – Addition of CCD

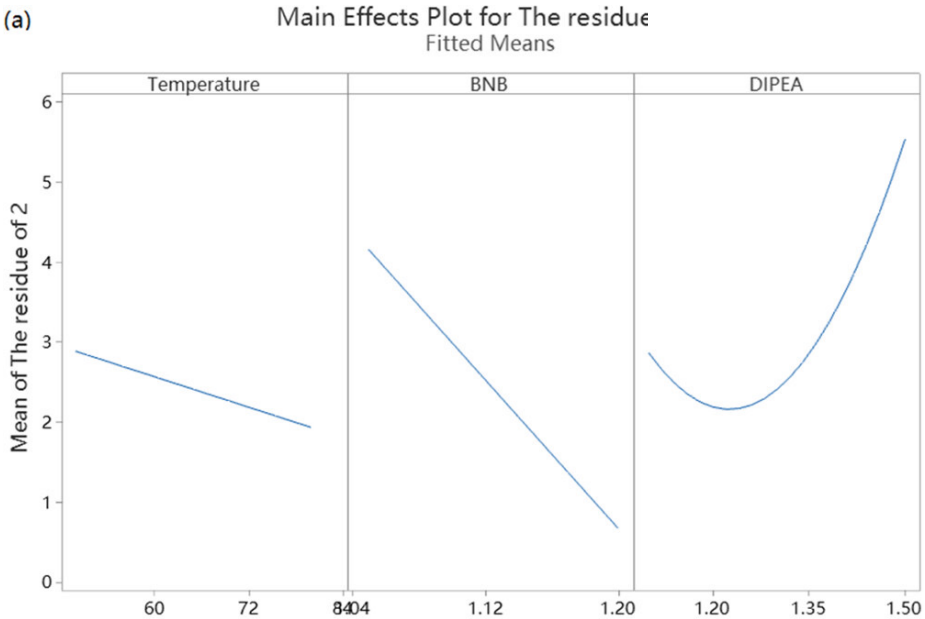


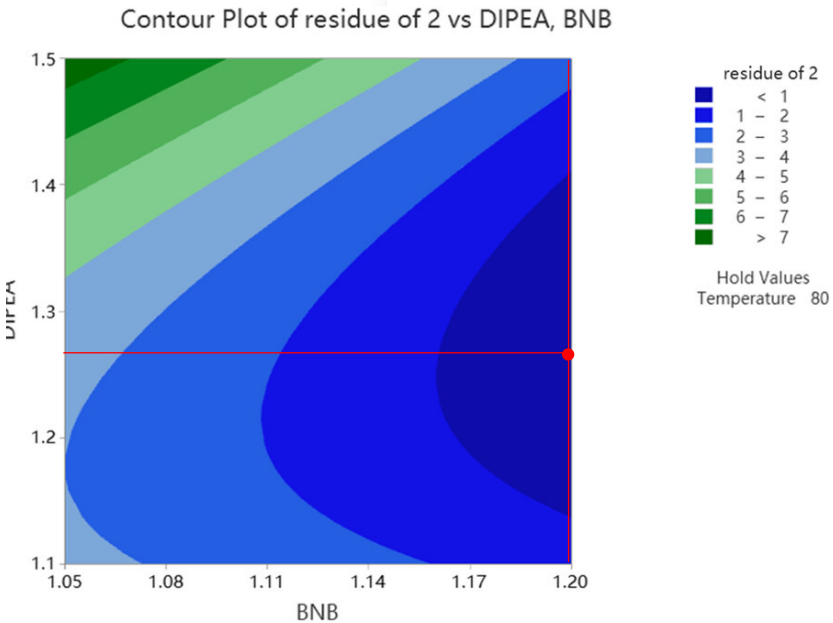
Composite Centered Design (CCD)



18 Additional CCF experiments carried out

(a)



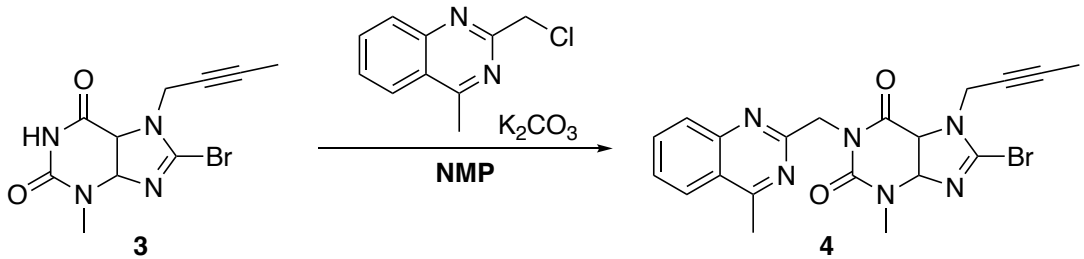


Entry	BNB (eq.)	DIPEA (eq.)	Temperature (°C)	**2** (%)[a]	95% CI
1	1.20	1.27	80	0.748	
2	1.20	1.27	80	0.810	(-0.937,1.261)
3	1.20	1.27	80	0.915	
4	1.20	1.27	80	0.909	
Average	-	-	-	0.845	-

[a] Determined by HPLC analysis

Org. Process Res. Dev. 2022, 26, 12, 3254–3264
ICH PHARMACEUTICAL DEVELOPMENT Q8(R2) - https://database.ich.org/sites/default/files/Q8_R2_Guideline.pdf
Ait-Amir, B.; Pougnet, P.; El Hami, A. 6 - Meta-Model Development; El Hami, A., Pougnet, P. B. T.-E. M. S. 2, Eds.; Elsevier, 2015; pp 151–179.

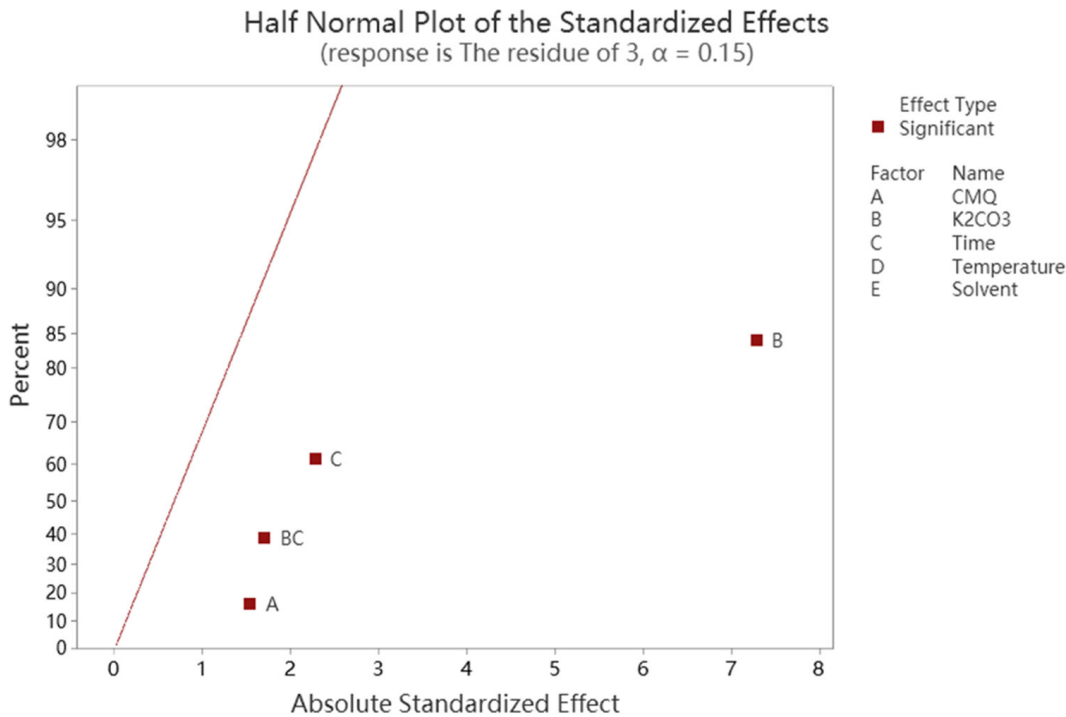
Optimization of Step 2



factor	unit	range
CMQ	equiv	1.0–1.2
K_2CO_3	equiv	0.5–2.0
time	h	4–14
temperature	°C	50–90
NMP	mL/g	8–14

[a]0.1 equiv of TBAB was added.

Full factorial: 2^5 = 32 experiments
½ factorial: 16 experiments + 4 center points
Output = Residual 3

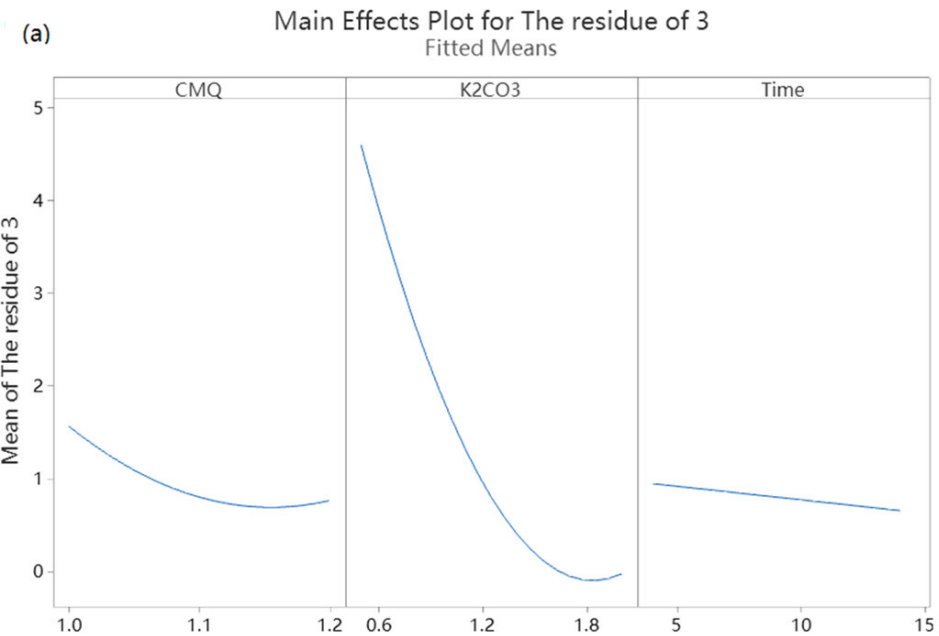


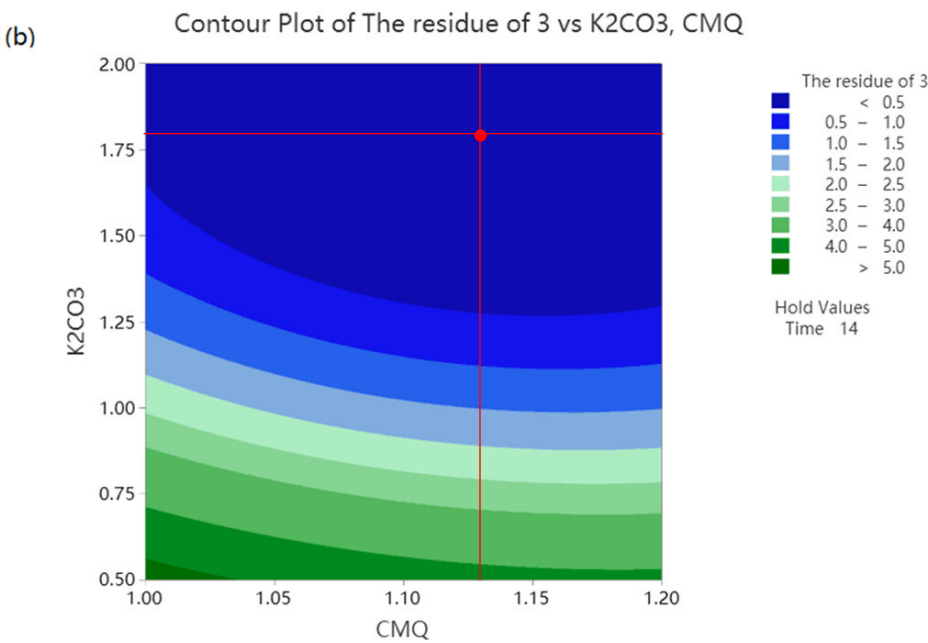
Org. Process Res. Dev. 2022, 26, 12, 3254–3264

Step 2 – Addition of CCD

17 Additional CCF experiments carried out

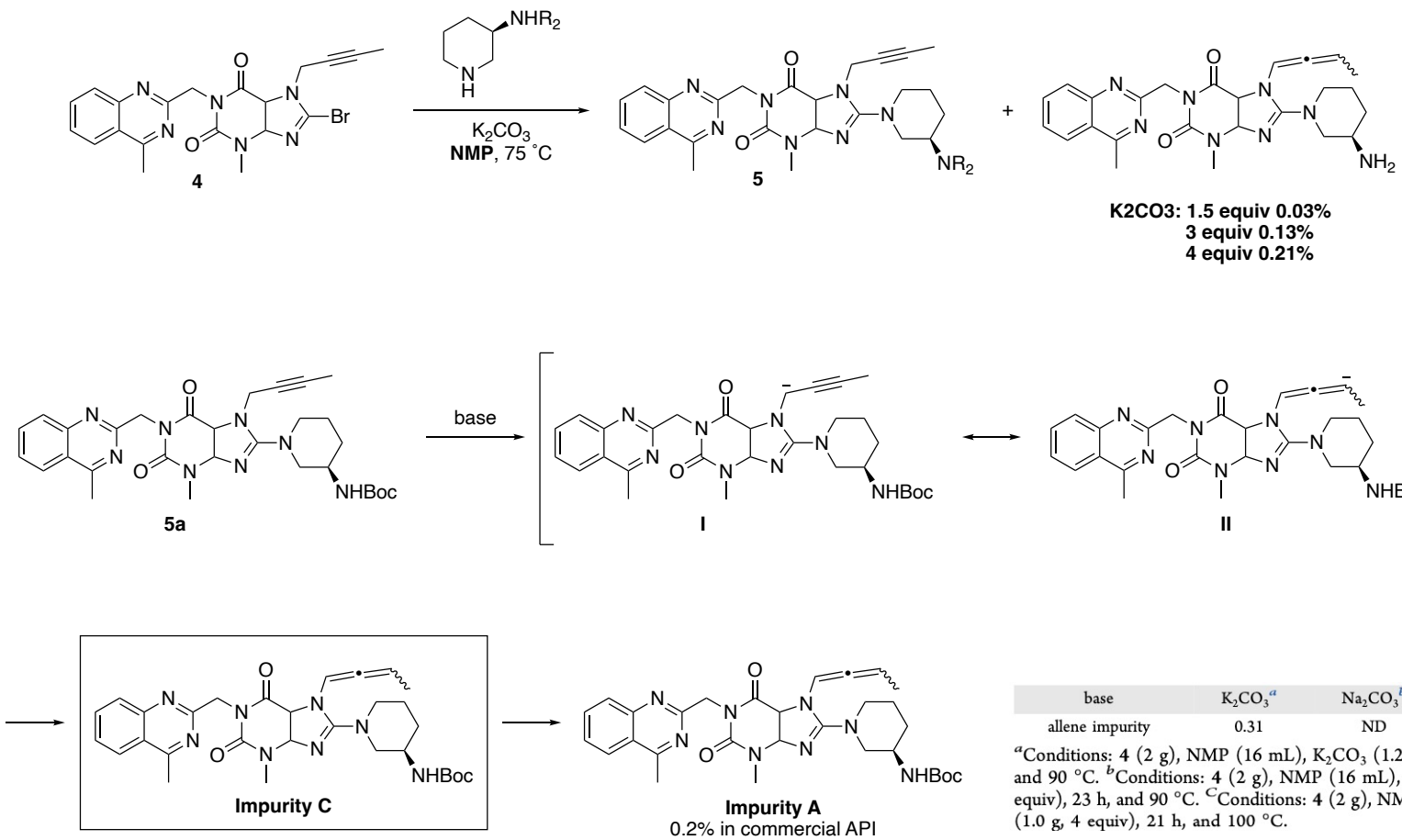
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	4	67.4824	16.8706	15.84	0.000
Linear	3	64.4199	21.4733	20.17	0.000
CMQ	1	2.4869	2.4869	2.34	0.147
K_2CO_3	1	56.4527	56.4527	53.02	0.000
Time	1	5.4803	5.4803	5.15	0.038
2-Way Interactions	1	3.0625	3.0625	2.88	0.111
K2CO3*Time	1	3.0625	3.0625	2.88	0.111





Org. Process Res. Dev. 2022, 26, 12, 3254–3264

Mechanism of Impurity C and A

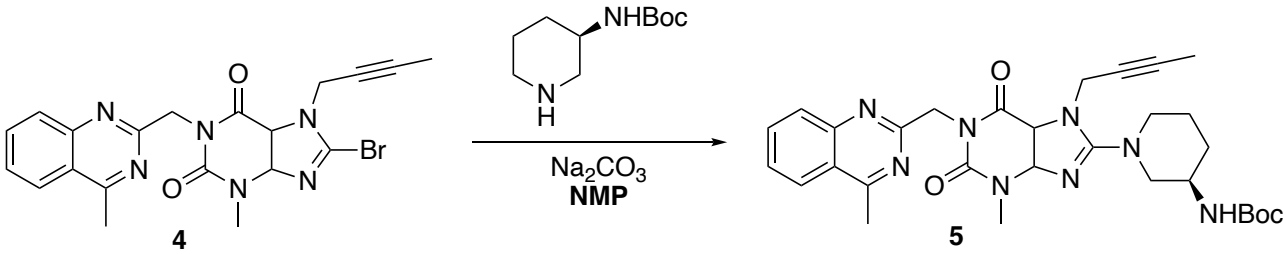


base	K_2CO_3[a]	Na_2CO_3[b]	KOH[C]
allene impurity	0.31	ND	0.70

[a]Conditions: 4 (2 g), NMP (16 mL), K_2CO_3 (1.2 g, 2 equiv), 23 h, and 90 °C. [b]Conditions: 4 (2 g), NMP (16 mL), Na_2CO_3 (0.9 g, 2 equiv), 23 h, and 90 °C. [C]Conditions: 4 (2 g), NMP (16 mL), KOH (1.0 g, 4 equiv), 21 h, and 100 °C.

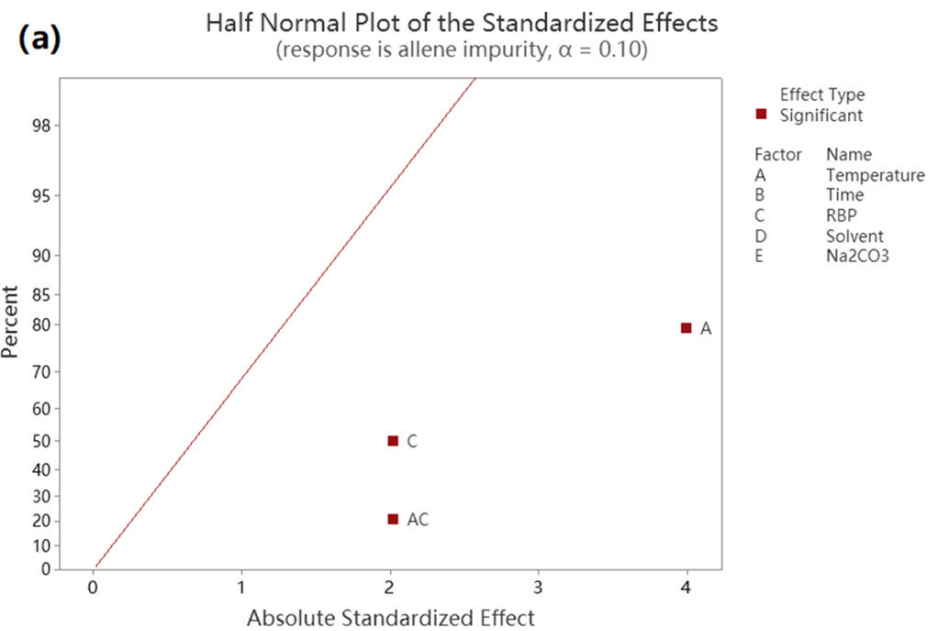
Org. Process Res. Dev. 2022, 26, 12, 3254–3264

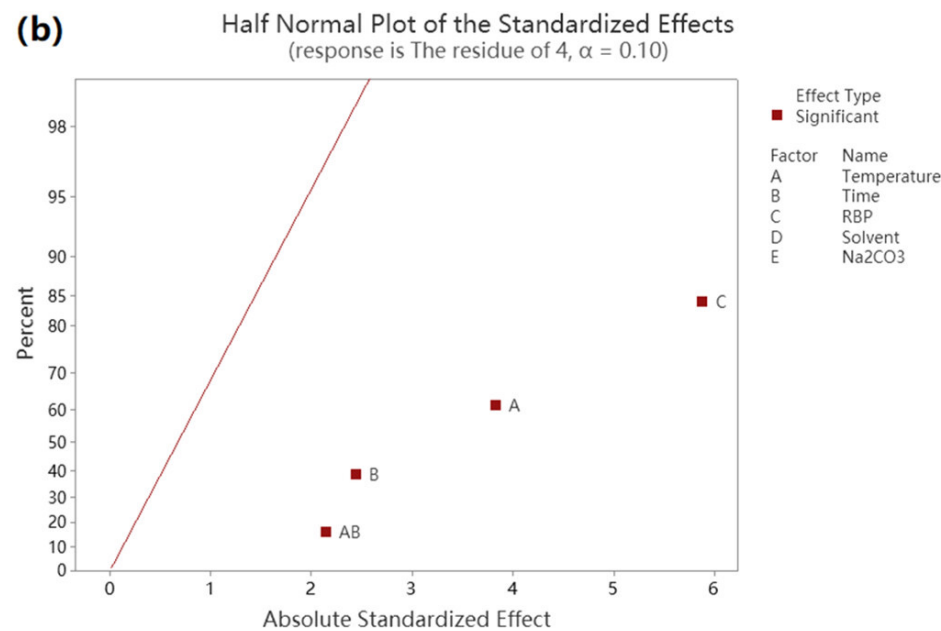
Optimization of Step 3



factor	unit	range
temperature	°C	70–100
time	h	12–16
Na_2CO_3	equiv	0.5–2.0
NMP	mL/g	6–12
RBP	equiv	1.00–1.12

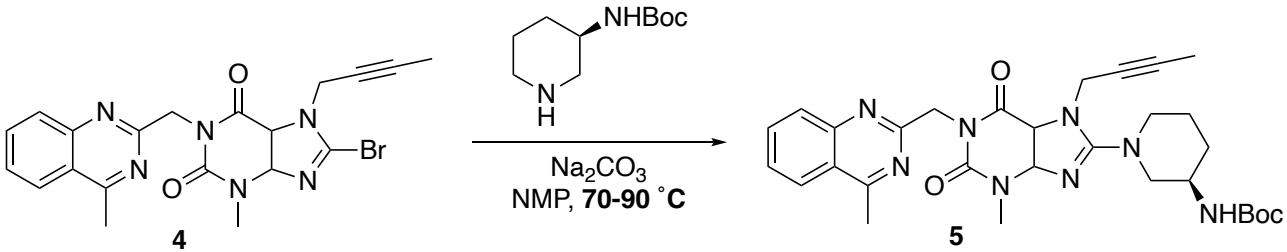
Full factorial: 2^5 = 32 experiments
½ factorial: 16 experiments + 3 center points
Output = Residual 4 + Impurity C



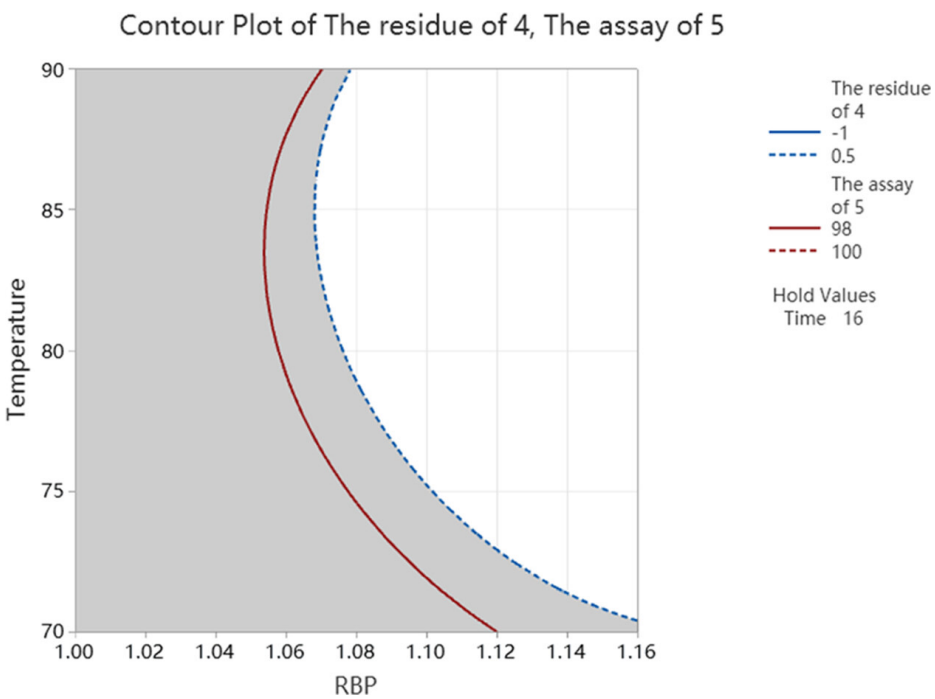


Org. Process Res. Dev. 2022, 26, 12, 3254–3264

Step 3 – Addition of CCD



17 experiment RSM generation.
Output : Residual 4 and assay yield of 5

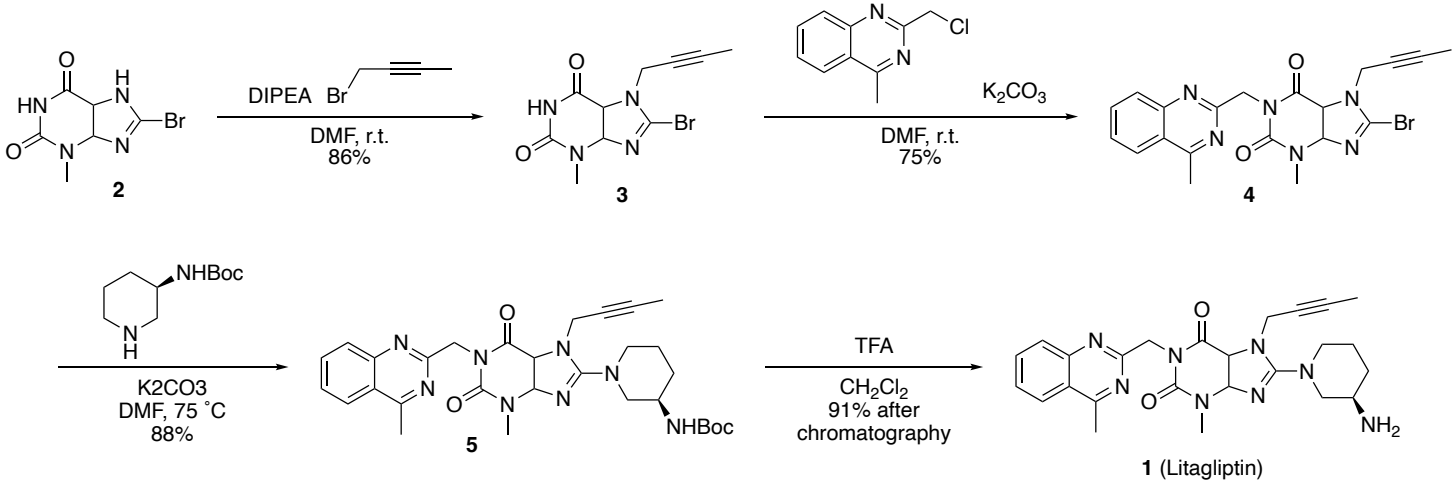


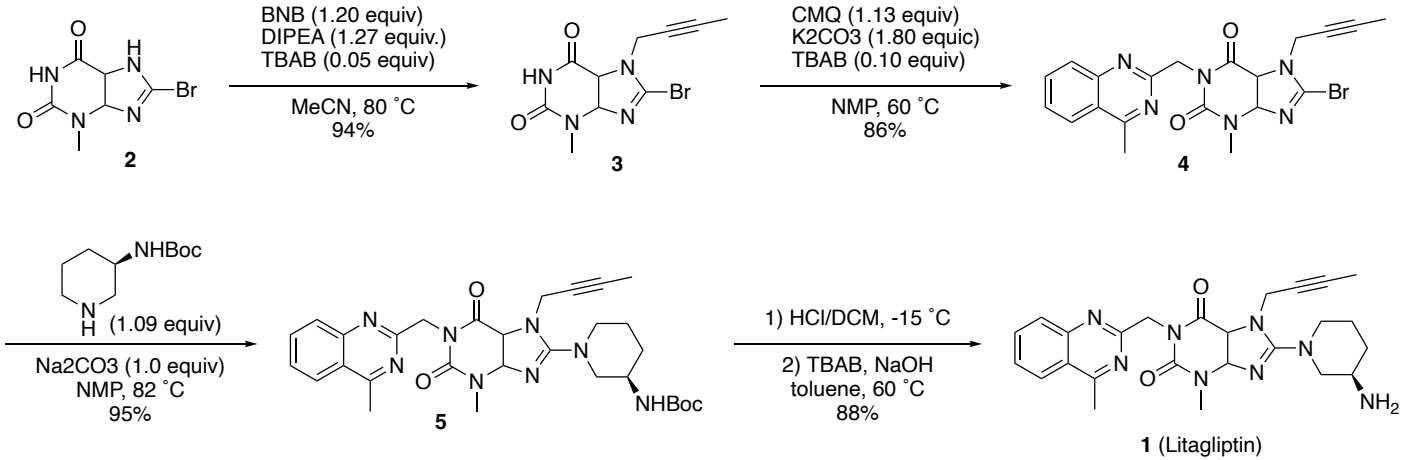
Entry	Temperature (°C)	RBP (equiv)	Time (h)	HPLC assay (%)		Allene impurity
				4	**5**	
1	82	1.09	16	0.150	98.777	ND
2	82	1.09	16	0.065	98.748	ND
3	82	1.09	16	0.257	98.599	ND
4	82	1.09	16	0.088	98.741	ND
average	-	-	-	0.140	98.716	-

100g scale: 96% isolated yield of 5 at 99.5% purity

Org. Process Res. Dev. 2022, 26, 12, 3254–3264

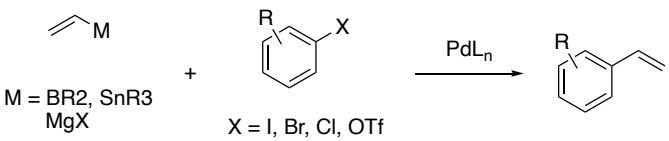
Summary of Litagliptin Synthesis





Org. Process Res. Dev. 2022, 26, 12, 3254–3264

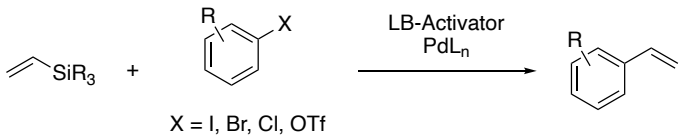
DoE on Discrete Variables



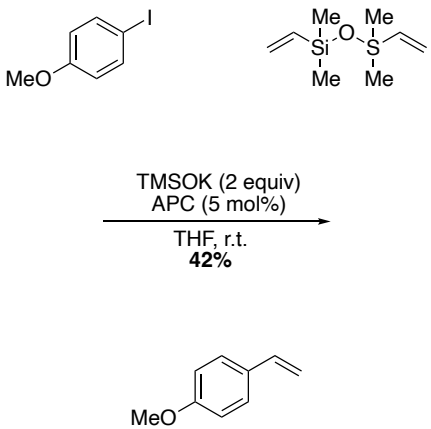
BR_2 - Must be freshly prepared, costly, require high temperatures, only work well with aryl-Br

SnR_3 - Toxic and costly

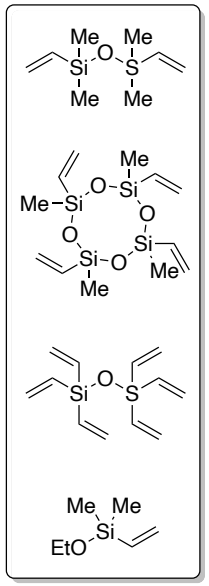
MgX - Not general resulting in a limited scope without tuning



Non-toxic, easy to purify, good substrate scope, and mild conditions



Vinyl Donor

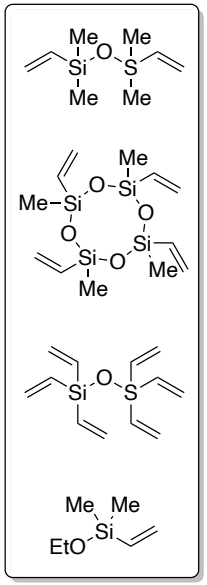


Activator	Pd Source	Ligand	Ligand mol%	Solvent
KOH	PdCl2	dppb	0	DMF
CsOH	Pd(dba)2	dppbO	5	dioxane
TMSOK	APC	dppbO2	10	toluene
		PPh3		THF
		AsPh3		
		10		

J. AM. CHEM. SOC. 2008, 130, 3690-3704

1st Round DoE

Vinyl Donor



Activator

KOH
CsOH
TMSOK

Pd Source

PdCl2
Pd(dba)2
APC

Ligand

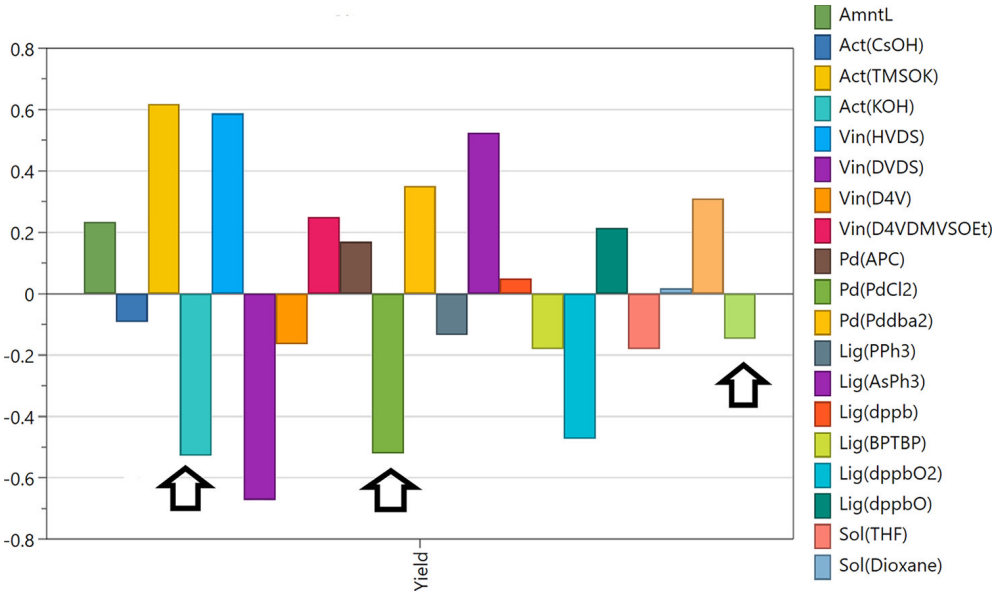
dppb
dppbO
dppbO2
PPh3
AsPh3
10

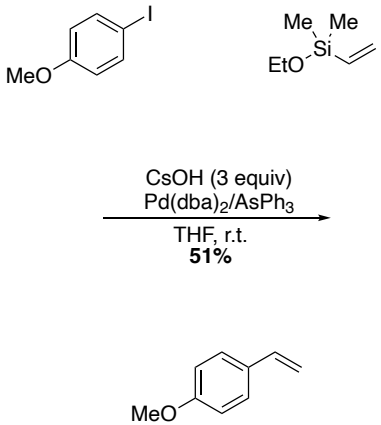
Ligand mol%

0
5
10

Solvent

DMF
dioxane
toluene
THF

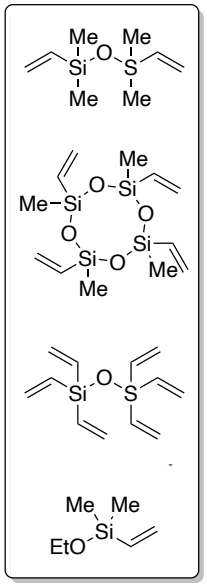




J. AM. CHEM. SOC. 2008, 130, 3690-3704
Org. Process Res. Dev. 2022, 26, 3, 899–914

2nd Round DoE

Vinyl Donor



Activator

~~KOH~~
CsOH
TMSOK

Pd Source

~~PdCl2~~
Pd(dba)2
APC

Ligand

dppb
dppbO
dppbO2
PPh3
AsPh3
10

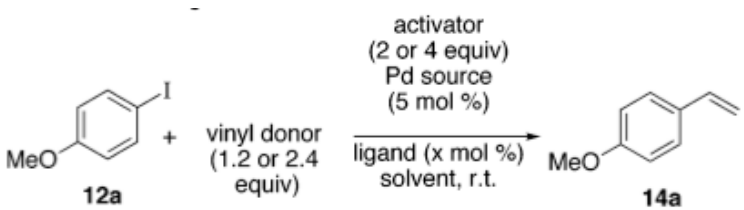
Ligand mol%

0
5
10

Solvent

DMF
dioxane
~~toluene~~
THF

100 reaction matrix selected for 2nd round DoE

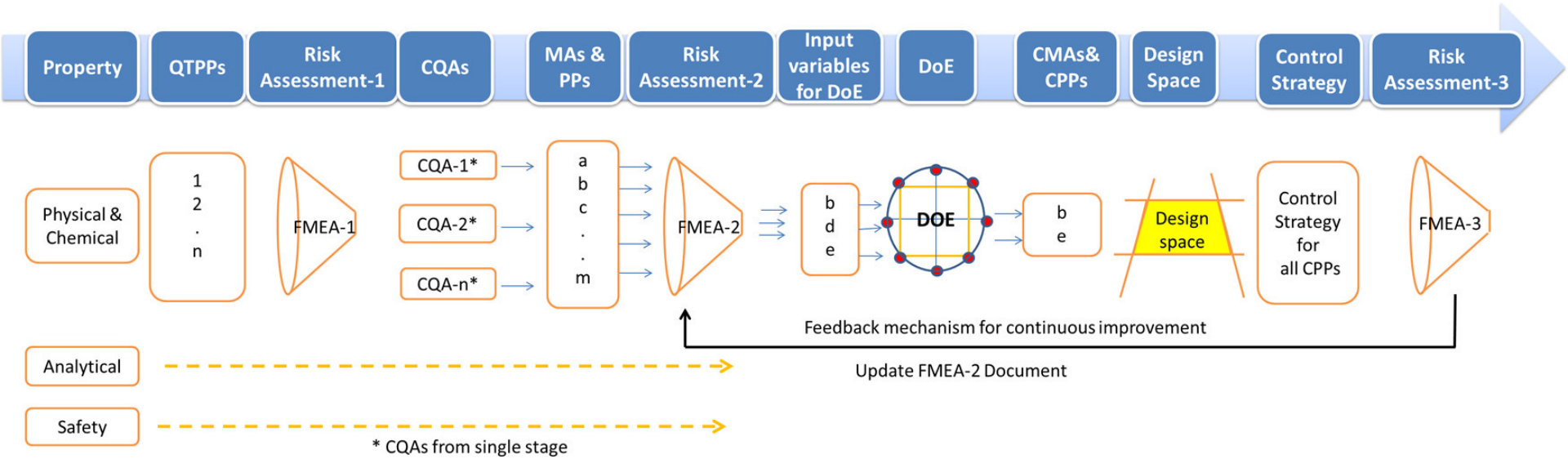


run	activator	vinyl donor	Pd source	ligand	solvent	ligand, mol %	vinyl, equiv[b]	12 h yield, %[c]
40	$KOSiMe_3$	HVDS	APC	$dppbO_2$	DMF	5	2	98
89	$KOSiMe_3$	DVDS	$Pd(dba)_2$	Ph_3P	dioxane[d]	0	1	93
7	$KOSiMe_3$	DMVSOEt	$Pd(dba)_2$	dppb	DMF	0	2	89
31	$KOSiMe_3$	DVDS	APC	dppbO	DMF	0	2	86
69	$KOSiMe_3$	DMVSOEt	APC	$dppbO_2$	DMF	0	1	85
64	$KOSiMe_3$	DMVSOEt	APC	$dppbO_2$	THF	0	2	85
26	$KOSiMe_3$	HVDS	$Pd(dba)_2$	dppbO	DMF	0	2	85
86	$KOSiMe_3$	DVDS	$Pd(dba)_2$	$dppbO_2$	DMF	0	1	83
96	$KOSiMe_3$	D_4^V	$Pd(dba)_2$	$AsPh_3$	DMF	1	2	82

[a] 0.5 mmol of aryl iodide used in the reaction. [b] The ratio of activator to vinyl equiv is always 2:1.2. [c] Yield determined by GC analysis versus an internal standard. [d] In this reaction, a 2:1 mixture of dioxane/DMF was inadvertently used.

J. AM. CHEM. SOC. 2008, 130, 3690-3704

Summary



Capabilities:

- Systematically and efficiently generate interpolative and predictive models
- Can evaluate any number of variables for multiple outputs simultaneously
- Easily provides powerful tools for analysis and visualization of data

Challenges:

- Number of runs scales exponentially with number of variables and levels
- Screening of discrete variables is difficult
- Models are only as good as the data they are given
- Incapable of accounting for lurking variables and scale up challenges

Org. Process Res. Dev. 2022, 26, 12, 3254–3264