World Journal of Pharmaceutical and Life Sciences
WJPLS
www.wjpls.org
SJIF Impact Factor: 6.129

DEVELOPMENT AND OPTIMIZATION OF ERLOTINIB HYDROCHLORIDE ANALYTICAL METHOD BY USING QUALITY BY DESIGN APPROACH

Bandana Kumari*

Department of Pharmacy, Modern College of Pharmacy, Sector-21, Yamunanagar, Nigdi, Pune – 411044, India.

Corresponding Author: Bandana Kumari
Department of Pharmacy, Modern College of Pharmacy, Sector-21, Yamunanagar, Nigdi, Pune – 411044, India.

Article Received on 30/09/2020 **rticle Revised on 20/10/2020** **Article Accepted on 10/11/2020**

ABSTRACT

A reversed-phase high performance liquid chromatographic (HPLC) technique for the determination of Erlotinib Hydrochloride has been optimized using analytical quality by design (QbD) approach. All the compounds are monitored with the photodiode array (PDA) detector at 336 nm. The experiments were conducted by changing the three different factor, that are pH, concentration of methanol and flow rate by using of a 2^3 factorial design space. The whole technique is developed as per International Council for Harmonization (ICH) guidelines. The proposed method is robust, sensitive, rapid and successful and helpful in the regions where regulatory agencies recommend HPLC analytical method.

KEYWORD: Erlotinib Hydrochloride; Reversed-phase HPLC, Analytical Quality by Design.

INTRODUCTION

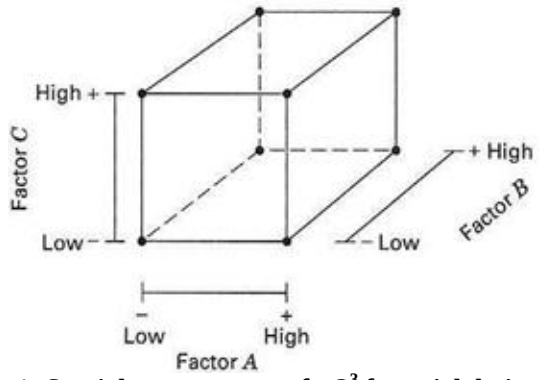
The US-FDA ICH Q8 (R2)[1,2] says that "Quality by Design is a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management." Not only in formulation optimization, but the QbD approach is also used in analytical method development and validation[3] Various analytical methods such as HPLC; capillary electrophoresis method; Karl Fischer titration methodology, the supercritical fluid chromatography method etc. are optimized using analytical QbD.[4] The various advantages that analytical QbD provides are- It requires less experiment time since it uses design of experiment (DOE) methods to obtain possible parameter combinations. The obtained design space obtained can ensure the robustness of the method. Here there is Flexible process within design space which allows continuous improvement. Any specifications are based on product performance requirements.[5]

Reversed Phase High-performance liquid chromatography particularly (RP-HPLC), is one of the most popular analytical technique in the pharmaceutical industry. This importance in turn makes the optimization process of analysis by HPLC more important and need of an hour.

According to the guidelines and the established sequences of an analytical QbD approach, method development consists of five parts. Firstly, the analytical QbD approach begins with the definition of the Analytical Target Profile (ATP), which mainly identifies the components to be analyzed and the required analytical technique based on the intended purpose of the method. In the meantime, the critical method attributes (CMAs) are determined. The chromatographic performance criteria that can be used as CMAs, are the resolutions of critical peaks; the tailing factors; the signal-to-noise ratio of target components; peak height; peak symmetry; peak width; analysis time and retention time etc. Secondly, the parameters that might pose a higher probability of affecting the analytical results are identified through a risk assessment approach, or DOE methods. These parameters are known as critical method parameters (CMPs). The CMPs can be mobile phase compositions, gradient conditions, column temperature, flow rate etc. Thirdly, the quantitative relationships between CMPs and CMAs are modeled with statistical models. Fourthly, the analytical design space can be established. Fifthly, the control strategy can be set up and employed to confirm the analytical attributes of the method. Finally, method validation is necessary to demonstrate the reliability of the method.

In our analytical QbD we are optimizing the method development of HPLC of Erlotinib. The drug is a as an epidermal growth factor receptor (EGFR) inhibitor - protein-tyrosine kinase inhibitor. It is used for the treatment of locally advanced or metastatic non-small cell lung cancer (NSCLC) or metastatic pancreatic cancer.

The design follows 2^3 Factorial design, where, the design consist of 3 factors at two levels (high and low level) the effects of three input variables can be evaluated in eight experimental conditions shown as the corners of a cube represented in following Figure 1.



Run	Factor x_1	Factor x_2	Factor x_3	Response Y
1	-1	-1	-1	Y_1
2	-1	-1	+1	Y_2
3	-1	+1	-1	Y_3
4	-1	+1	+1	Y_4
5	+1	-1	-1	Y_5
6	+1	-1	+1	Y_6
7	+1	+1	-1	Y_7
8	+1	+1	+1	Y_8

Figure 1: Spatial arrangement of a 2^3 factorial design space and the general run orders in a 2^3 factorial design.

In a design, if we have three factors (A, B and C) each at two levels (low and high) (i.e. if we have a 2^3 Factorial Design), then we need to perform a total of $2^3 = 8$ experimental runs in the design. In our experiment the three different factors that are used are:

1. Composition of buffer (Organic: Aqueous phase)
2. pH of the Buffer
3. Flow rate

The main effects that are being observed are as follows:

1. Retention time
2. AUC
3. No of theoretical plates
4. Tailing Factor

The experiments were conducted by changing one factor at a time (OFAT) approach.

MATERIALS AND METHODS

Materials and Instrument

Erlotinib Hydrochloride (ERL) was obtained as a gift sample from Cipla Ltd., (Goa, India). Remaining all solvents HPLC grade were purchased from Fisher Scientific (Mumbai, India). High-quality water (MilliQ) (pH 6.7±0.1) was obtained from water purification system (Millipore, MA, USA). Shimadzu Prominence HPLC system equipped with two LC-20AD pumps, SIL-20ACHT auto sampler, CTO-20AC column oven, SPD-M20A PDA detector, DGU-20A3R degasser was used for analysis (Shimadzu Corp., Kyoto, Japan). Data acquisition and integration was carried out using LC solutions software (version 1.25).

Methods

Preparation of buffer: Ammonium acetate buffer was used for the experimental runs. For the buffer preparation, 770 mg ammonium acetate was dissolved in deionized water. To the solution 2.5 ml glacial acetic acid was added. The pH of the buffer was adjusted using glacial acetic acid. The adjustments were done to pH 3.8 and 4.2 to obtain to different kinds of buffer.

Preparation of Drug solution: Weighed accurately 2 mg of Erlotinib Hydrochloride in a 10 mL volumetric flask and make up the volume with the methanol to obtain the concentration as 200 µg/ml which is named as solution S1. From the solution S1, 400 µl was withdrawn and added to 3600 µl of methanol to obtain the concentration as 20 µg/ml which is named as solution S2. From the solution S2, 50 µl was withdrawn and added to 950 µl of methanol to obtain a concentration of 1 µg/ml. The samples for 8 runs were prepared in triplicates for a better optimization.

Chromatographic conditions- Analysis of the samples were done by RP- HPLC connected to a UV detector operated at the following operating parameters

Mobile phase- pH 3.8 ammonium acetate buffer: methanol (40:60)
pH 3.8 ammonium acetate buffer: methanol (30:70)
pH 4.2 ammonium acetate buffer: methanol (40:60)
pH 4.2 ammonium acetate buffer: methanol (30:70)

Flow rate- 0.8 ml/min
1.2 ml/min
Column type – Phenomenex RP C18 250mm X 4.6 mm; 5 um particle size
Detection wavelength- 336 nm

Optimizing by design expert software- The optimization design was performed using 2^3 full factorial designs with a resolution of 4. The factors that were input were as shown in table 1.

Table 1: The inputs of 2^3 full factorial designs with a resolution of 4.

Name	Factor	Lower	Higher
pH	A (numeric)	3.8	4.2
Concentration of Methanol (%)	B (numeric)	60	70
Flow Rate	C (numeric)	0.8	1.2

The number of replicates was 3. The experiments were conducted as per the run order and the respective responses were entered in the columns provided. After executing these, the design summary and graphs were obtained which will be discussed in the results section as shown in table 2.

Table 2: 2^3 full factorial design.

Std	Run	Factor 1 A: pH	Factor 2 B: Methanol %	Factor 3 C: Flow rate ml/min	Response 1 Retention time	Response 2 AUC	Response 3 Theoretical plate number	Response 4 Tailing factor
16	1	4.2	70	1.2	5.899	63491	1098.37	0.7999
4	2	4.2	60	0.8	17.862	70085	1554.51	0.892
2	3	3.8	60	0.8	17.811	135466	1387.35	1.111
1	4	3.8	60	0.8	17.803	110769	1467.33	0.952
12	5	4.2	60	1.2	12.465	46558.3	1486.8	0.915
14	6	3.8	70	1.2	5.953	103753	1244.69	1.348
8	7	4.2	70	0.8	8.593	100283	1111.96	1.211
3	8	4.2	60	0.8	17.693	71702.3	1194.92	0.920
11	9	4.2	60	1.2	12.334	49784	1494.6	0.957
9	10	3.8	60	1.2	12.229	80983.7	1473.71	1.086
5	11	3.8	70	0.8	8.765	147575	1178.34	1.303
15	12	4.2	70	1.2	5.902	59914.3	1081.24	1.077
13	13	3.8	70	1.2	5.995	147487	1175.36	1.490
7	14	4.2	70	0.8	8.620	98212.7	985.013	1.273
10	15	3.8	60	1.2	12.227	119739	1427.03	1.140
6	16	3.8	70	0.8	8.781	226760	1183.99	1.355

RESULTS AND DISCUSSION

The analysis of the responses will be discussed individually.

Response type- Area Under the Curve (AUC)

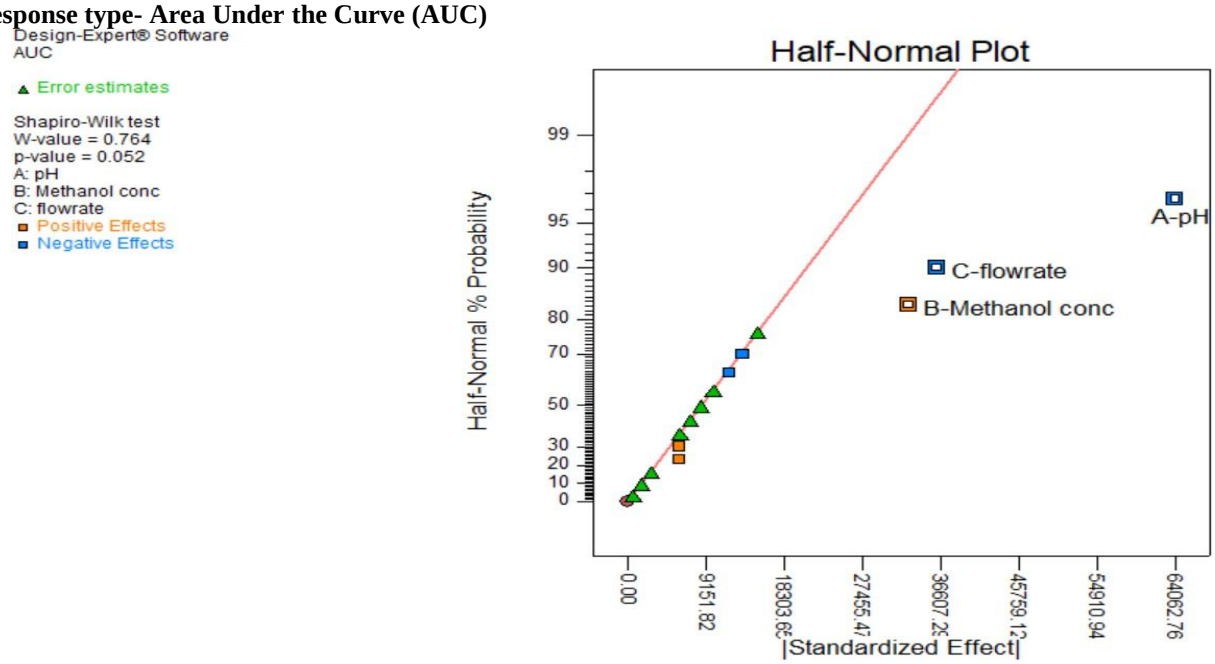


Figure 2: Half-Normal plot of Area under curve (AUC).

In the above figure 2 A, B, C factors are significant. The factors AB, BC, AC, ABC are not significant and are not considered in the model to maintain hierarchy. The ANOVA of the data as shown in table 3.

Table 3: ANOVA table for Area Under Curve (AUC).

ANOVA for Selected factorial method						
Analysis of variance table [Partial sum of squares – Type III]						
Source	**Sum of Squares**	**df**	**Mean Square**	**F value**	**p-value Prob > F**	
Model	2.594E+010	3	8.648E+009	15.42	0.0002	significant
A-pH	1.642E+010	1	1.642E+010	29.27	0.0002	
B-Methanol conc.	4.303E+009	1	4.303E+009	7.67	0.0170	
C-Flow rate	5.225E+009	1	5.225E+009	9.31	0.0100	
Residual	6.731E+009	12	5.609E+008			
Lack of Fit	1.569E+009	4	3.922E+008	0.61	0.6685	Not significant
Pure Error	5.162E+009	8	6.453E+008			
Cor Total	3.268E+010	15				

The Model F-value of 15.42 implies the model is significant. There is only a 0.02% chance that an F-value this large could occur due to noise. **"Values of ""Prob > F"" less than 0.0500 indicate model terms are significant. "** In this case A, B, C is significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model. "The **""Lack of Fit F-value""** of 0.61 implies the Lack of Fit is **not significant** relative to the pure error. There is a 66.85% chance that a ""Lack of Fit F-value"" this large could occur due to noise. Non-significant lack of fit is good -- we want the model to fit.

The "Pred R-Squared" of 0.6338 is in reasonable agreement with the "Adj R-Squared" of 0.7425; i.e. the difference is less than 0.2. "Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. Your ratio of 11.231 indicates an adequate signal. This model can be used to navigate the design space.

Final Equation in Terms of Coded Factors:
AUC = +1.02035206250000E + 005 -32031.38124999999900*A +16399.29374999999700 * B -18071.41875000000400 * C

The equation in terms of coded factors can be used to make predictions about the response for given levels of each factor. By default, the high levels of the factors are coded as +1 and the low levels of the factors are coded as -1. The coded equation is useful for identifying the relative impact of the factors by comparing the factor coefficients.

Final Equation in Terms of Actual Factors
AUC=+6.19829106249999E+005 -1.60156906250000E +005*pH+ 3279.8587500000003 * Methanol conc -90357.09374999997100* flowrate

The equation in terms of actual factors can be used to make predictions about the response for given levels of each factor. Here, the levels should be specified in the original units for each factor. This equation should not be used to determine the relative impact of each factor because the coefficients are scaled to accommodate the units of each factor and the intercept is not at the center of the design space.

Response type- Tailing Factor

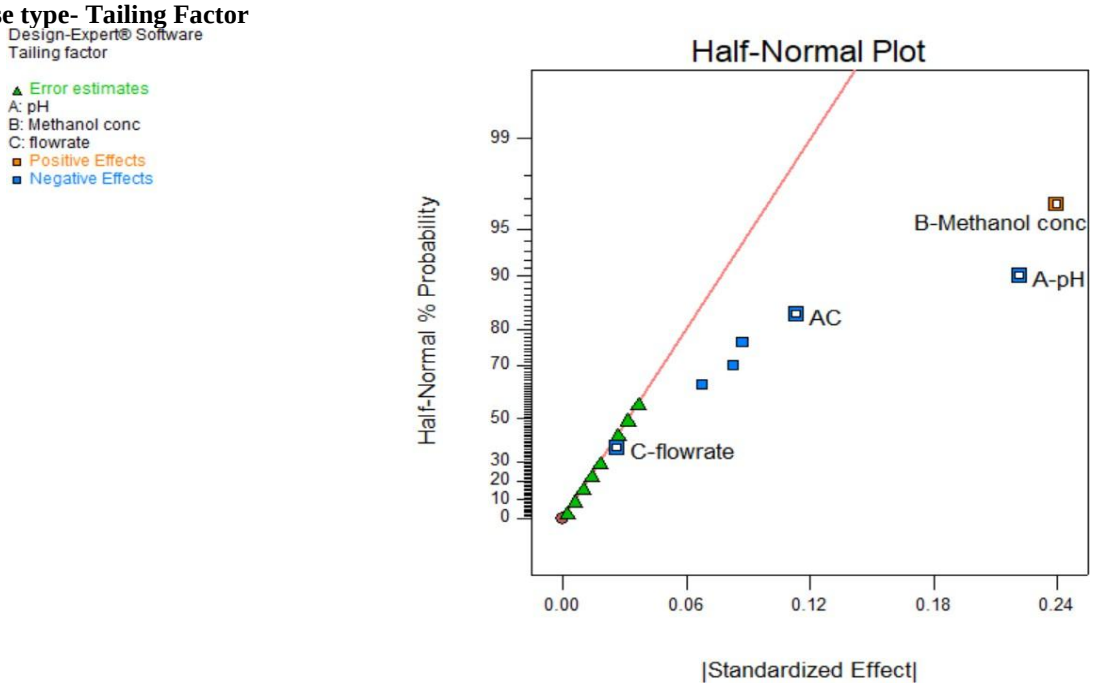


Figure 3: Half-Normal plot of Tailing Factor.

In the figure 3, A, B, AC factors are significant. The factors AB, BC, ABC are not significant and are not considered in the model to maintain hierarchy. The factor C is not significant but is considered in the model to maintain hierarchy. The ANOVA of the data as shown in table 4.

Table 4: ANOVA table for Area Under Curve (AUC).

ANOVA for selected factorial model						
Analysis of variance table [Partial sum of squares - Type III]						
Source	**Sum of Squares**	**df**	**Mean Square**	**F Value**	**p- value Prob > F**	
Model	0.46	4	0.12	9.06	0.0017	**significant**
A-pH	0.19	1	0.19	14.83	0.0027	
B-Methanol conc	0.22	1	0.22	17.34	0.0016	
C-flowrate	2.603E-003	1	2.603E- 003	0.20	0.6603	
AC	0.050	1	0.050	3.88	0.0746	
Residual	0.14	11	0.013			
Lack of Fit	0.073	3	0.024	2.89	0.1020	**not significant**
Pure Error	0.067	8	8.415E- 003			
Cor Total	**0.60**	**15**				

The Model F-value of 9.06 implies the model is significant. There is only a 0.17% chance that an F-value this large could occur due to noise. **"Values of ""Prob > F"" less than 0.0500 indicate model terms are significant."** In this case A, B is significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model. The **""Lack of Fit F-value""** of 2.89 implies the Lack of Fit is **not significant** relative to the pure error. There is a 10.20% chance that a ""Lack of Fit F-value"" this large could occur due to noise. Non-significant lack of fit is good -- we want the model to fit.

"The ""Pred R-Squared"" of 0.5075 is in reasonable agreement with the ""Adj R-Squared"" of 0.6826;"i.e. the difference is less than 0.2. "Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. Your ratio of 8.931 indicates an adequate signal. This model can be used to navigate the design space.

Final Equation in Terms of Coded Factors

Tailing factor = +1.11470412500000-0.10875425000000*A+0.11762000000000 * B-0.01275425000000* C-0.05562837500000* AC

The equation in terms of coded factors can be used to make predictions about the response for given levels of each factor. By default, the high levels of the factors are coded as +1 and the low levels of the factors are coded as -1. The coded equation is useful for identifying the relative impact of the factors by comparing the factor coefficients.

Final Equation in Terms of Actual Factors

Tailing factor = -3.73833712500003 +0.84693812500001 *pH +0.02352400000000 * Methanol conc +5.49906625000001 * flowrate-1.39070937500000 * pH * flowrate

The equation in terms of actual factors can be used to make predictions about the response for given levels of each factor. Here, the levels should be specified in the original units for each factor. This equation should not be used to determine the relative impact of each factor because the coefficients are scaled to accommodate the units of each factor and the intercept is not at the center of the design space.

Response type- Retention Time

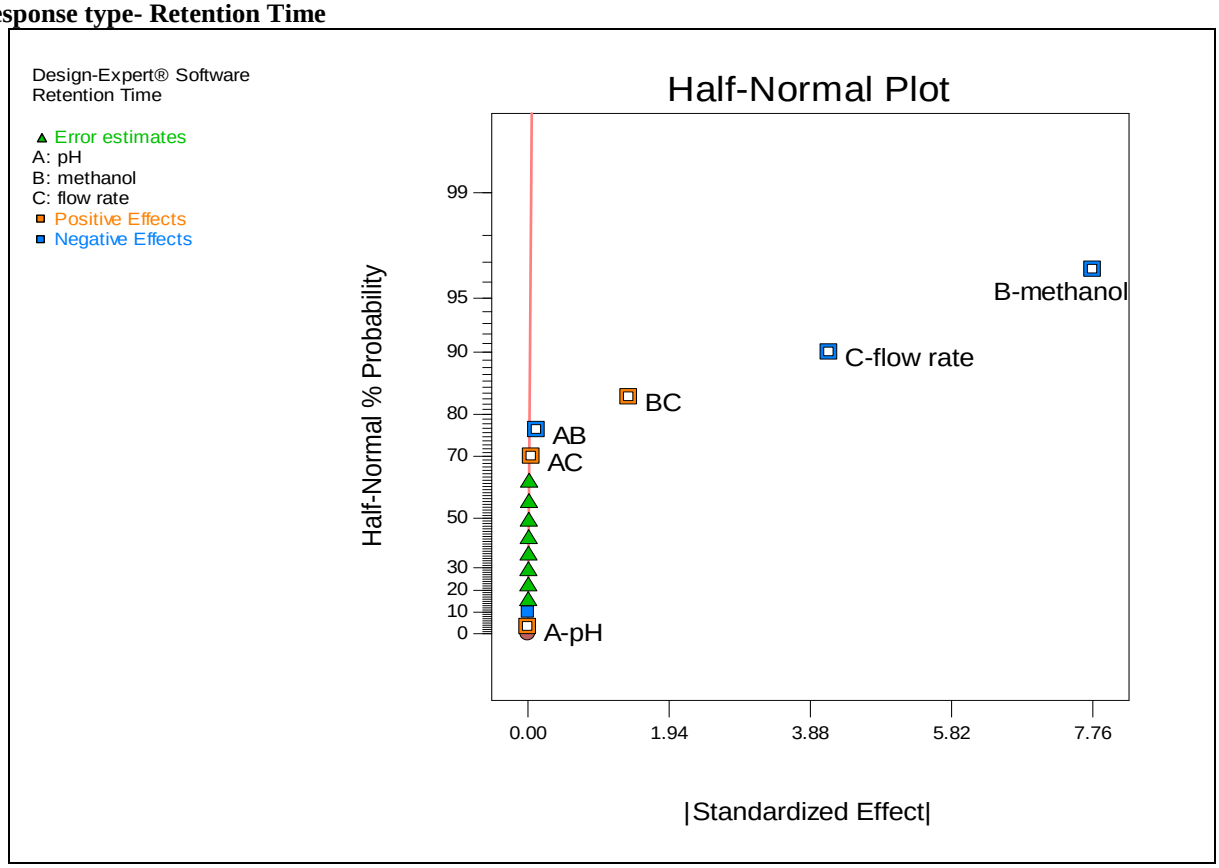


Figure 4: Half-Normal plot of Retention time.

In figure 4, the factors which are significant are Factor B, Factor C, Factor BC, Factor AB, and Factor AC are significant. The Factor A is not significant but to maintain the hierarchy of the model it is consider. The ANOVA of the data as shown in table 5.

Table 5: ANOVA table for Retention time.

ANOVA for selected factorial model						
Analysis of variance table [Partial sum of squares - Type III]						
Source	**Sum of Squares**	**df**	**Mean Square**	**F Value**	**p-value Prob > F**	
Model	317.50	6	52.92	45569.03	< 0.0001	**significant**
A-pH	1.000E-006	1	1.000E-006	8.612E-004	0.9772	
B-methanol	241.16	1	241.16	2.077E+005	< 0.0001	
C-flow rate	68.57	1	68.57	59048.59	< 0.0001	
AB	0.058	1	0.058	49.81	< 0.0001	
AC	9.474E-003	1	9.474E-003	8.16	0.0189	
BC	7.71	1	7.71	6636.97	< 0.0001	
Residual	0.010	9	1.161E-003			
Lack of Fit	1.736E-005	1	1.736E-005	0.013	0.9110	**not significant**
Pure Error	0.010	8	1.304E-003			
Cor Total	**317.51**	**15**				

The Model F-value of 45569.03 implies the model is significant. There is only a 0.01% chance that an F-value this large could occur due to noise. **Values of "Prob > F" less than 0.0500 indicate model terms are significant**. In this case B, C, AB, AC, BC are significant model terms. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model. The "**Lack of Fit F-value**" of 0.01 implies the Lack of Fit is **not significant** relative to the pure error. There is a 91.10% chance that a "Lack of Fit F-value" this large could occur due to noise. Non-significant lack of fit is good -- we want the model to fit.

The "Pred R-Squared" of 0.9999 is in reasonable agreement with the "Adj R-Squared" of 0.9999; i.e. the difference is less than 0.2. "Adeq Precision" measures the signal to noise ratio. A ratio greater than 4 is desirable. Your ratio of 531.350 indicates an adequate signal. This model can be used to navigate the design space.

Final Equation in Terms of Coded Factors

Retention Time = +11.20 + 2.49999999995441E-00 * A - 3.88229166666667 * B - 2.070166666667 * C - 0.06012499999999 * AB + 0.0243333333333 * AC + 0.69404166666667 * BC

The equation in terms of coded factors can be used to make predictions about the response for given levels of each factor. By default, the high levels of the factors are coded as +1 and the low levels of the factors are coded as -1. The coded equation is useful for identifying the relative impact of the factors by comparing the factor coefficients.

Final Equation in Terms of Actual Factors

Retention Time = +103.92533333333833+3.30104166666558 * pH 1.23000000000005 * methanol -57.89687500000171 * flow rate -0.06012499999999 * pH * methanol +0.60833333333362 * pH * flow rat +0.69404166666667 * Methanol * flow rate

The equation in terms of actual factors can be used to make predictions about the response for given levels of each factor. Here, the levels should be specified in the original units for each factor. This equation should not be used to determine the relative impact of each factor because the coefficients are scaled to accommodate the units of each factor and the intercept is not at the center of the design space.

Response type- No. of Theoretical Plates

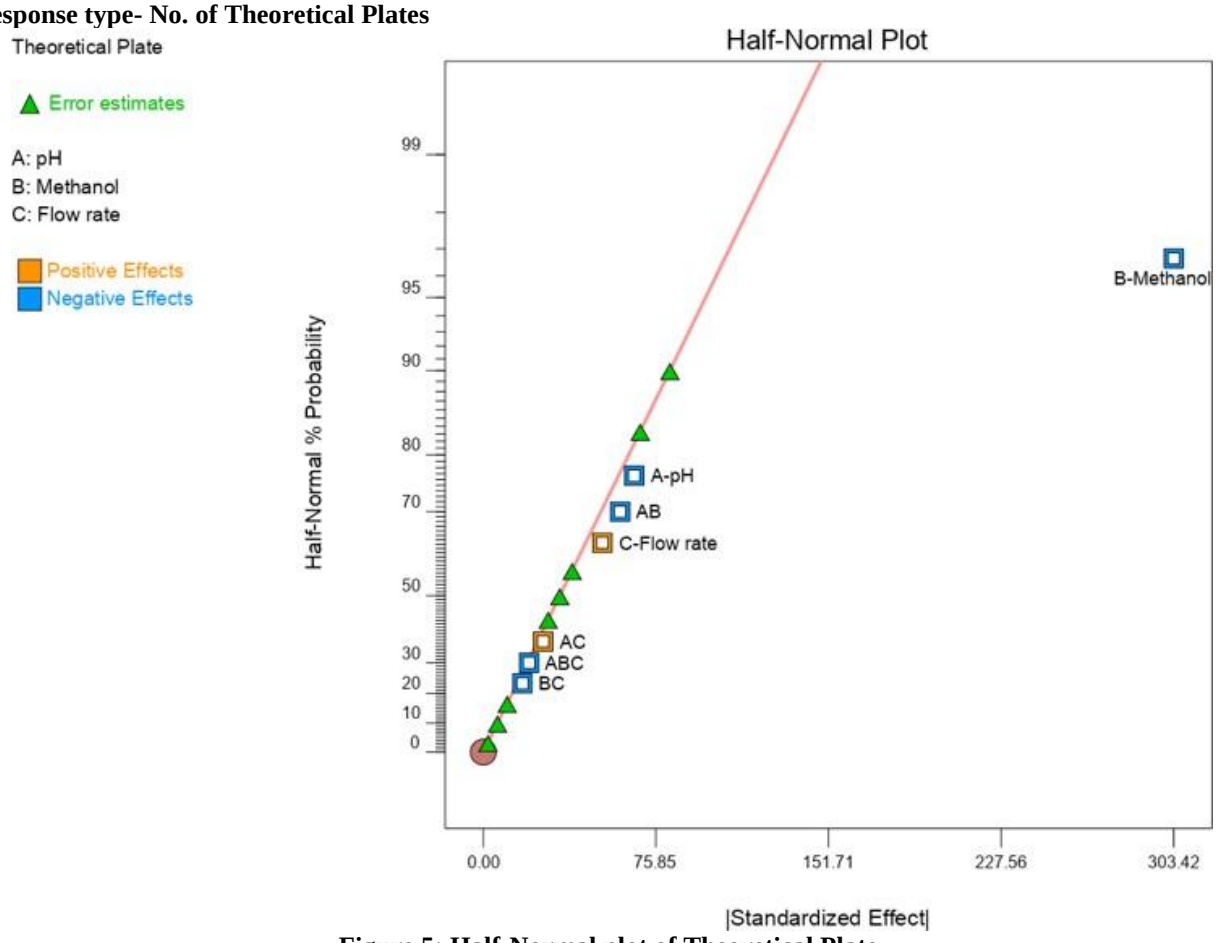


Figure 5: Half-Normal plot of Theoretical Plate.

As shown in figure 5, Only the factor B i.e methanol composition % is really significant for affecting the number of theoretical plates as its t_{cal} value is higher than the bonferroni limit , rest all factors are insignificant as their t_{cal} value is lesser than the t_{crit} limit. The ANOVA of the data as shown in table 6.

Table 6: ANOVA table for Retention time.

Source	Sum of Squares	df	Mean Square	F-value	p-value	
Model	3.682E+05	1	3.682E+05	40.22	< 0.0001	**significant**
B-Methanol	3.682E+05	1	3.682E+05	40.22	< 0.0001	
Residual	1.282E+05	14	9155.09			
Lack of Fit	48578.60	6	8096.43	0.8138	0.5878	**not significant**
Pure Error	79592.72	8	9949.09			
Cor Total	**4.964E+05**	**15**				

The **Model F-value** of 40.22 implies the model is significant. There is only a 0.01% chance that an F-value this large could occur due to noise. **P-values** less than 0.0500 indicate model terms are significant. In this case B is a significant model term. Values greater than 0.1000 indicate the model terms are not significant. If there are many insignificant model terms (not counting those required to support hierarchy), model reduction may improve your model. The **Lack of Fit F-value** of 0.81 implies the Lack of Fit is not significant relative to the pure error. There is a 58.78% chance that a Lack of Fit F-value this large could occur due to noise. Non-significant lack of fit is good -- we want the model to fit.

The **Predicted R^2** of 0.6628 is in reasonable agreement with the **Adjusted R^2** of 0.7234; i.e. the difference is less than 0.2. **Adeq Precision** measures the signal to noise ratio. A ratio greater than 4 is desirable. Your ratio of 8.969 indicates an adequate signal. This model can be used to navigate the design space. The coefficient estimate represents the expected change in response per unit change in factor value when all remaining factors are held constant. The intercept in an orthogonal design is the overall average response of all the runs. The coefficients are adjustments around that average based on the factor settings. When the factors are orthogonal the VIFs are 1; VIFs greater than 1 indicate multi-colinearity, the higher the VIF the more severe the correlation of factors. As a rough rule, VIFs less than 10 are tolerable.

Final Equation in Terms of Coded Factors
Theoretical plate = +1284.08 − 151.71 B
The equation in terms of coded factors can be used to make predictions about the response for given levels of each factor. By default, the high levels of the factors are coded as +1 and the low levels are coded as -1. The coded equation is useful for identifying the relative impact of the factors by comparing the factor coefficients.

Final Equation in Terms of Actual Factors
Theoretical Plate = +3256.28150 − 30.34159 Methanol
The equation in terms of actual factors can be used to make predictions about the response for given levels of each factor. Here, the levels should be specified in the original units for each factor. This equation should not be used to determine the relative impact of each factor because the coefficients are scaled to accommodate the units of each factor and the intercept is not at the center of the design space.

CONCLUSION

The experiments were conducted as per the run order and the respective responses were entered in the columns provided. After executing the design we conclude that the Factor A, B and C affect the response (AUC); Factor A, B and AC affect the response (Tailing Factor); Factor B, C, BC, AB and AC affect the response (Retention time) whereas the Factor B affect the response (Theoretical Plate)

ACKNOWLEDGEMENTS

This work was supported by the research funding from Modern College of Pharmacy, Nigdi.

REFERENCES

1. FDA: Bioanalytical Method Validation Guidance for Industry; Webpage URL: (Accessed on: 25-09-2018) https://www.fda.gov/downloads/drugs/guidances/ucm070107.pdf.
2. ICH: Validation Of Analytical Procedures:Text And Methodology Q2(R1); Webpage URL: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Quality/Q2_R1/Step4/Q2_R1__Guideline.pdf.
3. Pandey P, Dua K, Dureja H. Erlotinib loaded chitosan nanoparticles: Formulation, physicochemical characterization and cytotoxic potential. International journal of biological macromolecules, 2019 Oct 15; 139: 1304-16.
4. Marchetti S, de Vries NA, Buckle T, Bolijn MJ, van Eijndhoven MA, Beijnen JH, Mazzanti R, van Tellingen O, Schellens JH. Effect of the ATP-binding cassette drug transporters ABCB1, ABCG2, and ABCC2 on erlotinib hydrochloride (Tarceva) disposition in in vitro and in vivo pharmacokinetic studies employing Bcrp1−/−/Mdr1a/1b−/−(triple-knockout) and wild-type mice. Molecular cancer therapeutics, 2008 Aug 1; 7(8): 2280-7.

5. Sarwar Beg, MD Saquib Hasnain. Pharmaceutical Quality by Design: Principles and Application, 1 ed. 125 London Wall, London EC2Y 5AS, United Kingdom: Andre Gerhard Wolf, 2019; 22.