Supporting Information

Process development and optimization of linagliptin aided by Design of Experiments (DoE)

Hailong Wang[a,b,d], Kai Chen[a], Biyue Lin[b,d] , Jingping Kou[d], Lijun Li[d], Shuming Wu[b,d], Shouzhu Liao[b,d], Guodong Sun[b,d], Junwen Pu[d], Hua Yang[a,*] and Zhongqing Wang[b,c,d*]

[a] *College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, P. R. China*
[b] *State Key Laboratory of Anti-Infective Drug Development, Sunshine Lake Pharma Co., Ltd., Dongguan 523871, P. R. China*
[c] *School of Pharmacy, Xiangnan University, Chenzhou 423000 Hunan, China, Xiangnan University, Chenzhou 423000, China*
[d] *HEC Research and Development Center, HEC Pharm Group, Dongguan 523871, P. R. China*

Table of Content

* Corresponding author. e-mail: Hyangchem@csu.edu.cn

* Corresponding author. e-mail: Wangzhongqing@hec.cn

Table S1. The 1/4-fractional factorial DoE studies parameters and results of the residue of **2**

StdOrder	RunOrder	CenterPt	Blocks	Volumes of acetonitrile	BNB	Temp.	Time	DIPEA	Residual SM **2** (%)
1	1	1	1	10	1	50	20	1.5	12.9
6	2	1	1	16	1	80	8	1.5	10.7
4	3	1	1	16	1.3	50	20	1.1	1.7
9	4	0	1	13	1.15	65	14	1.3	1.7
7	5	1	1	10	1.3	80	8	1.1	1.2
3	6	1	1	10	1.3	50	8	1.5	2.1
10	7	0	1	13	1.15	65	14	1.3	2.0
11	8	0	1	13	1.15	65	14	1.3	2.2
8	9	1	1	16	1.3	80	20	1.5	1.0
5	10	1	1	10	1	80	20	1.1	3.3
2	11	1	1	16	1	50	8	1.1	7.5

Table S2. Analysis of Variance (ANOVA) for the residual starting material **2.**

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Model	3	142.276	47.425	9.99	0.006
Linear	2	122.051	61.026	12.86	0.005
BNB	1	100.848	100.848	21.25	0.002
DIPEA	1	21.203	21.203	4.47	0.072
2-Way Interactions	1	20.225	20.225	4.26	0.078
BNB*DIPEA	1	20.225	20.225	4.26	0.078
Error	7	33.219	4.746		
Curvature	1	20.936	20.936	10.23	0.019
Lack-of-Fit	4	12.172	3.043	54.58	0.018
Pure Error	2	0.112	0.056		
Total	10	175.496			

Table S3. Three process parameters were considered for the CCF

Symbol	Parameter	Unite	High	Low
1	Temperature	ºC	50	80
2	BNB	eq.	1.05	1.2
3	DIPEA	eq.	1.1	1.5

Table S4. The RSM DoE studies parameters and results of the residue of **2**

StdOrder	RunOrder	CenterPt	Blocks	Temperature	BNB	DIPEA	The residue of **2**
12	1	0	1	65	1.125	1.3	3.168
4	2	1	1	80	1.2	1.1	1.31
8	3	1	1	80	1.2	1.5	2.823
3	4	1	1	50	1.2	1.1	2.214
7	5	1	1	50	1.2	1.5	3.743
2	6	1	1	80	1.05	1.1	3.733
11	7	0	1	65	1.125	1.3	1.754
10	8	0	1	65	1.125	1.3	2.714
5	9	1	1	50	1.05	1.5	9.785
9	10	0	1	65	1.125	1.3	2.669
6	11	1	1	80	1.05	1.5	7.939
1	12	1	1	50	1.05	1.1	4.039
13	13	-1	2	50	1.125	1.3	2.361
15	14	-1	2	65	1.05	1.3	3.506
18	15	-1	2	65	1.125	1.5	3.383
17	16	-1	2	65	1.125	1.1	3.021
14	17	-1	2	80	1.125	1.3	1.578
16	18	-1	2	65	1.2	1.3	1.518

Table S5. ANOVA Analysis for the RSM DoE studies of the residue of **2**

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Model	5	70.575	14.1150	18.01	0.000
Linear	3	50.358	16.7861	21.42	0.000
Temperature	1	2.265	2.2648	2.89	0.115
BNB	1	30.255	30.2551	38.61	0.000
DIPEA	1	17.838	17.8383	22.76	0.000
Square	1	14.248	14.2484	18.18	0.001
DIPEA*DIPEA	1	14.248	14.2484	18.18	0.001
2-Way Interaction	1	5.969	5.9685	7.62	0.017
BNB*DIPEA	1	5.969	5.9685	7.62	0.017
Error	12	9.404	0.7837		
Lack-of-Fit	9	8.350	0.9278	2.64	0.229
Pure Error	3	1.054	0.3513		
Total	17	79.979			

Table S6. Four verification experiments for synthesizing N7-substituted xanthine **3**

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

[a] Determined by HPLC analysis

Table S7. The half-fractional factorial DoE studies parameters and results of the residue of **3**

StdOrder	RunOrder	CenterPt	Blocks	CMQ	K_2CO_3	Time	Temperature	Solvent	The residue of **3**
4	1	1	1	1.2	2	4	50	14	0.419
15	2	1	1	1	2	14	90	8	0.631
13	3	1	1	1	0.5	14	90	14	3.195
9	4	1	1	1	0.5	4	90	8	5.329
19	5	0	1	1.1	1.25	9	70	11	0.68
11	6	1	1	1	2	4	90	14	0.506
17	7	0	1	1.1	1.25	9	70	11	0.532
6	8	1	1	1.2	0.5	14	50	14	2.999
20	9	0	1	1.1	1.25	9	70	11	0.408
5	10	1	1	1	0.5	14	50	8	3.875
3	11	1	1	1	2	4	50	8	2.058
7	12	1	1	1	2	14	50	14	1.032
1	13	1	1	1	0.5	4	50	14	6.615
8	14	1	1	1.2	2	14	50	8	0.138
16	15	1	1	1.2	2	14	90	14	0.138
18	16	0	1	1.1	1.25	9	70	11	0.443
12	17	1	1	1.2	2	4	90	8	0.138
10	18	1	1	1.2	0.5	4	90	14	4.428
2	19	1	1	1.2	0.5	4	50	8	5.276
14	20	1	1	1.2	0.5	14	90	8	3.397

Table S8. ANOVA Analysis for the half-fractional factorial DoE studies of the residue of **3**

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
Error	15	15.9719	1.0648		
Curvature	1	12.7377	12.7377	55.14	0.000
Lack-of-Fit	11	3.1901	0.2900	19.71	0.016
Pure Error	3	0.0441	0.0147		
Total	19	83.4543			

Table S9. The RSM DoE studies parameters and results of the residue of **3**

StdOrder	RunOrder	CenterPt	Blocks	CMQ	K_2CO_3	Time	The residue of **3**
11	1	0	1	1.1	1.25	9	0.443
5	2	1	1	1	0.5	14	5.356
10	3	0	1	1.1	1.25	9	0.679
6	4	1	1	1.2	0.5	14	4.234
2	5	1	1	1.2	0.5	4	4.447
9	6	0	1	1.1	1.25	9	0.215
7	7	1	1	1	2	14	0.32
1	8	1	1	1	0.5	4	5.357
8	9	1	1	1.2	2	14	0.179
3	10	1	1	1	2	4	0.39
4	11	1	1	1.2	2	4	0.207
15	12	-1	2	1.1	2	9	0.242
16	13	-1	2	1.1	1.25	4	1.552
13	14	-1	2	1.2	1.25	9	0.71
17	15	-1	2	1.1	1.25	14	0.402
14	16	-1	2	1.1	0.5	9	5.066
12	17	-1	2	1	1.25	9	2.354

Table S10. The ANOVA for the RSM DoE studies of the residue of **3**

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	6	67.5510	11.2585	53.75	0.000
Linear	3	55.2764	18.4255	87.96	0.000
CMQ	1	1.6000	1.6000	7.64	0.020
K_2CO_3	1	53.4627	53.4627	255.23	0.000
Time	1	0.2137	0.2137	1.02	0.336
Square	2	11.9099	5.9549	28.43	0.000
CMQ*CMQ	1	0.3994	0.3994	1.91	0.197
K_2CO_3*K_2CO_3	1	6.6802	6.6802	31.89	0.000
2-Way Interaction	1	0.3647	0.3647	1.74	0.216
CMQ*K_2CO_3	1	0.3647	0.3647	1.74	0.216
Error	10	2.0947	0.2095		
Lack-of-Fit	8	1.9870	0.2484	4.61	0.190
Pure Error	2	0.1077	0.0538		
Total	16	69.6456			

Table S11. Four verification experiments of **4**

Entry	K_2CO_3 (eq.)	CMQ (eq.)	Time (h)	**3**(%)	95% CI
1	1.8	1.13	14	0.124	(-0.844,0.284)
2	1.8	1.13	14	0.181	
3	1.8	1.13	14	0.188	
4	1.8	1.13	14	0.110	
average	-	-	-	0.150	-

Table S12. The screen of the stoichiometry K_2CO_3 used in the preparation of **5**

Equev. of K_2CO_3	1.5	3	4
Allene impurity	0.03	0.13	0.21

Conditions:**4** (2 g) , NMP (16 ml) ,75-80ºC,16 h. Determined by HPLC

Table S13. The half-fractional factorial DoE studies parameters and results of the residue of **4** and allene impurity

StdOrder	RunOrder	CenterPt	Blocks	Temperature	Time	RBP	solvent	Na_2CO_3	The residue of **4**	allene impurity
1	1	1	1	70	10	1.00	6	2.0	6.556	0.000
3	2	1	1	70	16	1.00	6	0.5	4.346	0.000
13	3	1	1	70	10	1.12	12	2.0	2.757	0.000
8	4	1	1	100	16	1.12	6	0.5	0.01	0.063
19	5	0	1	85	13	1.06	9	1.25	1.744	0.000
7	6	1	1	70	16	1.12	6	2.0	0.23	0.000
11	7	1	1	70	16	1.00	12	2.0	4.267	0.000
17	8	0	1	85	13	1.06	9	1.25	1.005	0.000
2	9	1	1	100	10	1.00	6	0.5	2.946	0.048
14	10	1	1	100	10	1.12	12	0.5	0.028	0.000
4	11	1	1	100	16	1.00	6	2.0	2.285	0.037
15	12	1	1	70	16	1.12	12	0.5	1.915	0.000
12	13	1	1	100	16	1.00	12	0.5	3.966	0.057
9	14	1	1	70	10	1.00	12	0.5	10.601	0.000
10	15	1	1	100	10	1.00	12	2.0	4.109	0.050
6	16	1	1	100	10	1.12	6	2.0	0.007	0.00
5	17	1	1	70	10	1.12	6	0.5	2.964	0.000
18	18	0	1	85	13	1.06	9	1.25	0.794	0.000
16	19	1	1	100	16	1.12	12	2.0	0.04	0.000

Table S14. ANOVA Analysis for the half-fractional factorial DoE studies of the residue of **4**

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Model	4	104.604	26.1509	14.91	0.000
Linear	3	96.579	32.1931	18.35	0.000
Temperature	1	25.616	25.6163	14.60	0.002
Time	1	10.415	10.4151	5.94	0.029
RBP	1	60.548	60.5479	34.51	0.000
2-Way Interactions	1	8.024	8.0245	4.57	0.051
Temperature*Time	1	8.024	8.0245	4.57	0.051
Error	14	24.562	1.7544		
Curvature	1	7.809	7.8094	6.06	0.029
Lack-of-Fit	11	16.255	1.4777	5.94	0.153
Pure Error	2	0.498	0.2489		
Total	18	129.165			

Table S15. ANOVA Analysis for the half-fractional factorial DoE studies of the residue of allene impurity

Source	**DF**	**Adj SS**	**Adj MS**	**F-Value**	**P-Value**
Model	3	0.006144	0.002048	8.03	0.002
Linear	2	0.005104	0.002552	10.01	0.002
Temperature	1	0.004064	0.004064	15.94	0.001
RBP	1	0.001040	0.001040	4.08	0.062
2-Way Interactions	1	0.001040	0.001040	4.08	0.062
Temperature*RBP	1	0.001040	0.001040	4.08	0.062
Error	15	0.003824	0.000255		
Curvature	1	0.000642	0.000642	2.82	0.115
Lack-of-Fit	12	0.003183	0.000265	*	*
Pure Error	2	0.000000	0.000000		
Total	18	0.009969			

Table S16. The RSM DoE studies parameters and results of the residue of **4**, allene impurity and the assay of **5**

StdOrder	RunOrder	CenterPt	Blocks	Temperature	Time	RBP	The residue of **4** (%)	allene impurity	The assay of **5** (%)
4	1	1	1	90	16	1.00	3.218	0	95.452
3	2	1	1	70	16	1.00	5.1	0	94.831
1	3	1	1	70	10	1.00	6.128	0	92.565
10	4	0	1	80	13	1.08	1.175	0	97.674
6	5	1	1	90	10	1.16	0.000	0	98.674
7	6	1	1	70	16	1.16	0.233	0	98.779
11	7	0	1	80	13	1.08	0.612	0	98.27
8	8	1	1	90	16	1.16	0.000	0	98.775
9	9	0	1	80	13	1.08	0.517	0	98.338
5	10	1	1	70	10	1.16	0.936	0	98.13
2	11	1	1	90	10	1.00	3.534	0	95.067
16	12	-1	2	80	13	1.00	2.918	0	95.952
15	13	-1	2	80	16	1.08	0.203	0	98.662
13	14	-1	2	90	13	1.08	0.413	0	98.237
17	15	-1	2	80	13	1.16	0.252	0	98.883
14	16	-1	2	80	10	1.08	0.742	0	98.244
12	17	-1	2	70	13	1.08	2.698	0	96.144

Table S17. The ANOVA Analysis for the RSM DoE studies of the residue of **4**

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	6	55.6814	9.2802	48.41	0.000
Linear	3	44.8926	14.9642	78.07	0.000
Temperature	1	6.2885	6.2885	32.81	0.000
Time	1	0.6687	0.6687	3.49	0.091
RBP	1	37.9354	37.9354	197.91	0.000
Square	2	9.4218	4.7109	24.58	0.000
Temperature*Temperature	1	2.2109	2.2109	11.53	0.007
RBP*RBP	1	2.3662	2.3662	12.34	0.006
2-Way Interaction	1	1.3670	1.3670	7.13	0.023
Temperature*RBP	1	1.3670	1.3670	7.13	0.023
Error	10	1.9168	0.1917		
Lack-of-Fit	8	1.6639	0.2080	1.64	0.432
Pure Error	2	0.2530	0.1265		
Total	16	57.5982			

Table S18. The ANOVA Analysis for the RSM DoE studies of the assay of **5**

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Model	6	52.1410	8.6902	28.60	0.000
Linear	3	42.3068	14.1023	46.40	0.000
Temperature	1	3.3132	3.3132	10.90	0.008
Time	1	1.4585	1.4585	4.80	0.053
RBP	1	37.5352	37.5352	123.51	0.000
Square	2	9.0002	4.5001	14.81	0.001
Temperature*Temperature	1	2.7946	2.7946	9.20	0.013
RBP*RBP	1	1.6299	1.6299	5.36	0.043
2-Way Interaction	1	0.8340	0.8340	2.74	0.129
Temperature*RBP	1	0.8340	0.8340	2.74	0.129
Error	10	3.0390	0.3039		
Lack-of-Fit	8	2.7721	0.3465	2.60	0.308
Pure Error	2	0.2669	0.1335		
Total	16	55.1800			

Table S19. Four verification experiments for preparing N-Boc linagliptin **5**

Entry	Temperature (ºC)	RBP (equiv)	Time (h)	HPLC assay (%)		Allene impurity	95% CI
				4	**5**		
1	82	1.09	16	0.150	98.777	ND	
2	82	1.09	16	0.065	98.748	ND	(-0.440, 0.564)
3	82	1.09	16	0.257	98.599	ND	(98.229, 99.574)
4	82	1.09	16	0.088	98.741	ND	
average	-	-	-	0.140	98.716	-	

Table S20. Alternative deprotection methods to prepare linagliptin **1**

	Reaction temperature	Reaction time	The residue of **5**	**1**
HCl[a]	-10ºC	5 h	ND	62.5%
Zinc chloride[b]	30ºC	64 h	29.7%	69.1%
MeOH/H_2O[c]	Reflux	18 h	100%	

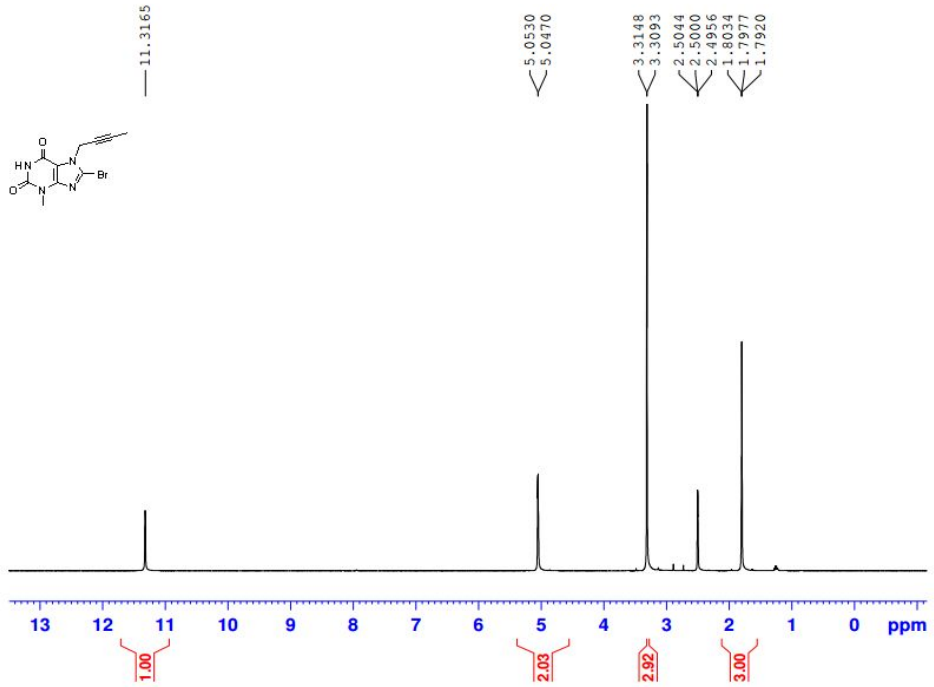
[a] **2** (1.5 g), DCM (7.5 ml), Con HCl (2.2 ml); [b] **2** (1.0 g), DCM (10 ml), Zinc chloride (1.3 g, 5 eq); [c] **2** (5.0 g), MeOH/H_2O (V/V=3:2, 15 ml)

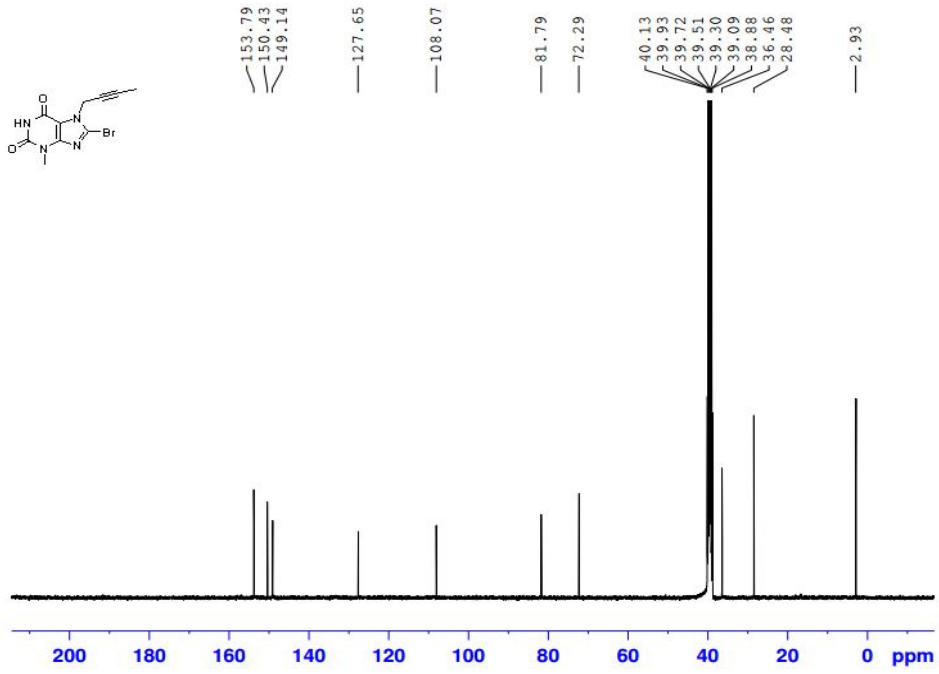
Table S21. The solubility screening of Linagliptin in different solvents

Solvent	Solubility (25 ºC)	Solubility (reflux)
Tertiary butanol	>200 mL/g	30 mL/g
EA	>200 mL/g	30 mL/g
THF	60 mL/g	15 mL/g
Toluene	>200 mL/g	10 mL/g
Isopropyl acetate	>200 mL/g	60 mL/g
2-Me-THF	>200 mL/g	35 mL/g
MTBE	>200 mL/g	>200 mL/g

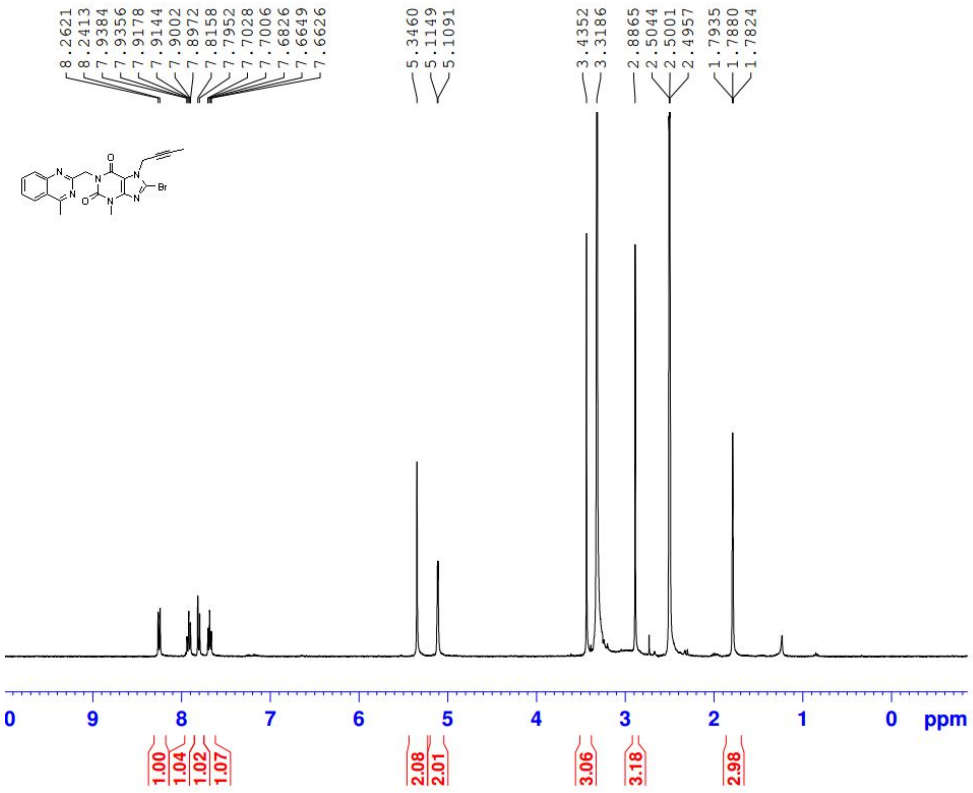
^{1}H NMR and ^{13}C NMR spectra

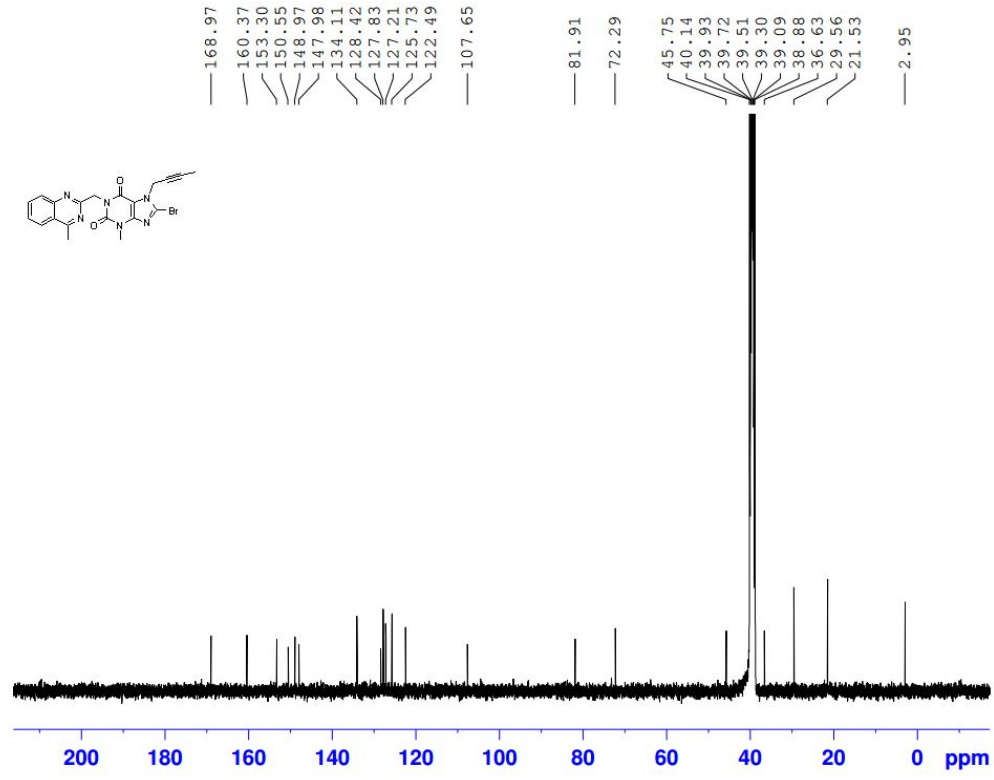
^{1}H NMR and ^{13}C NMR spectra for compound **3**



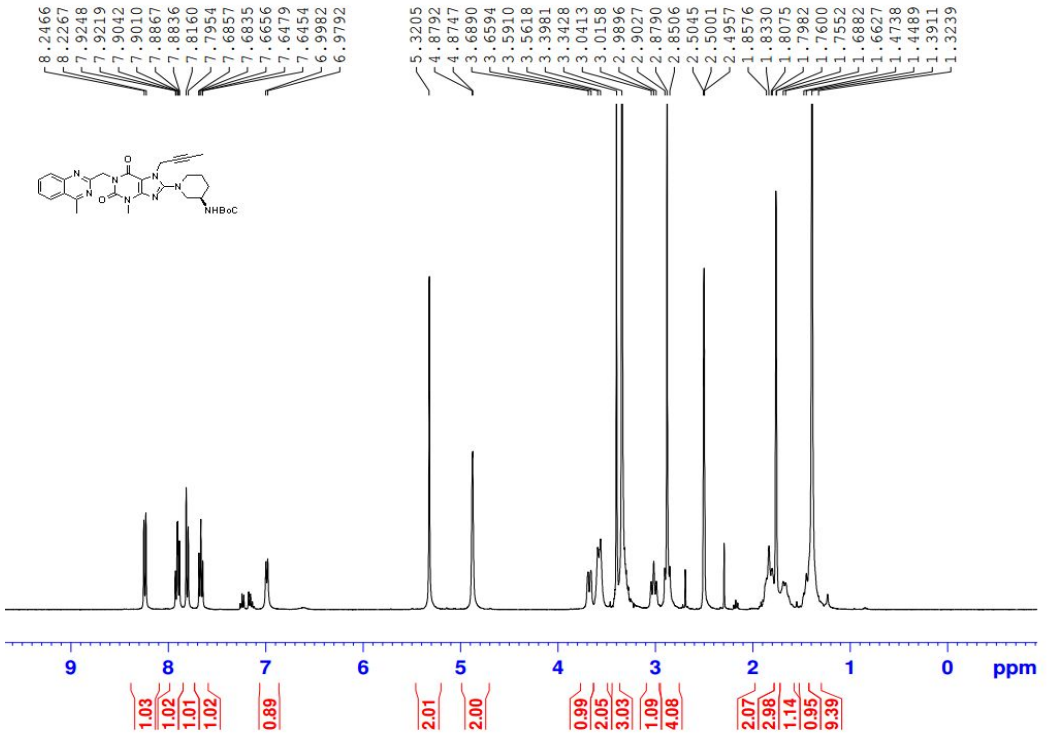


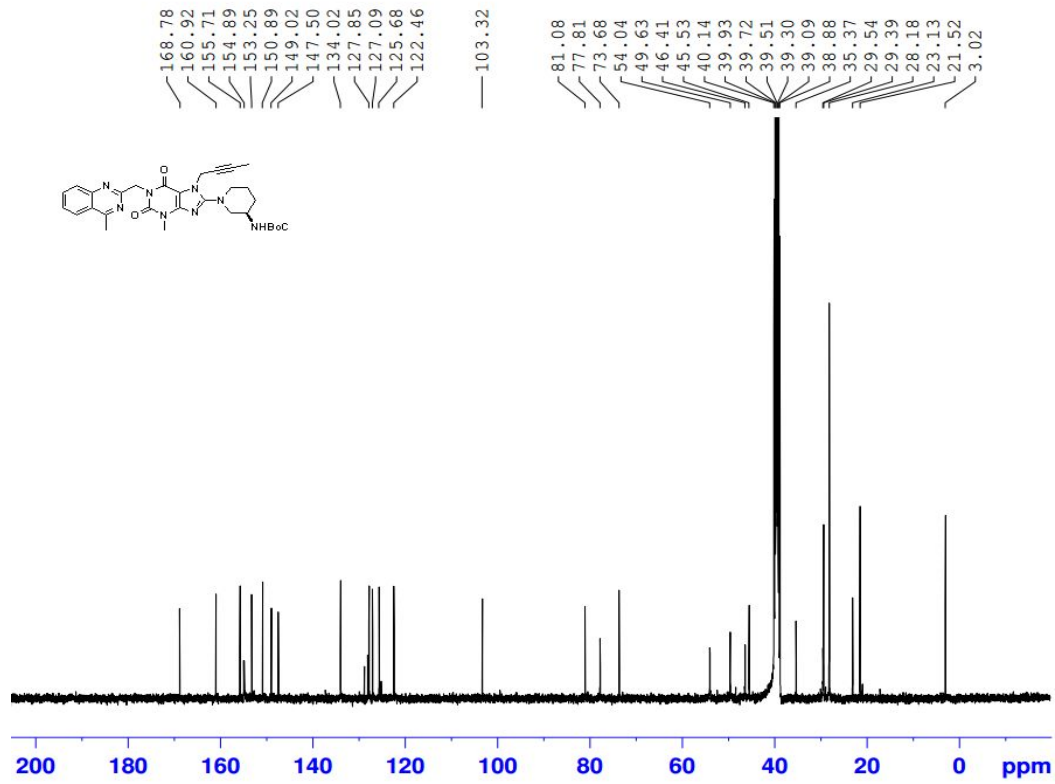
^{1}H NMR and ^{13}C NMR spectra for compound **4**



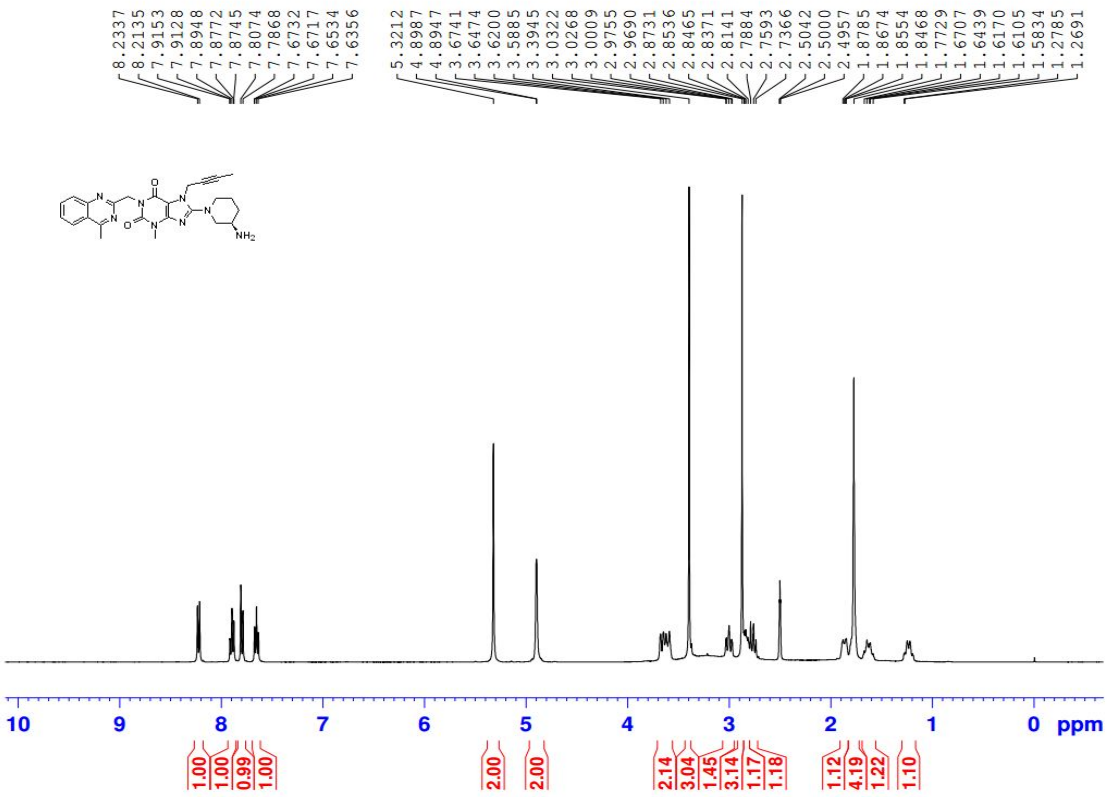


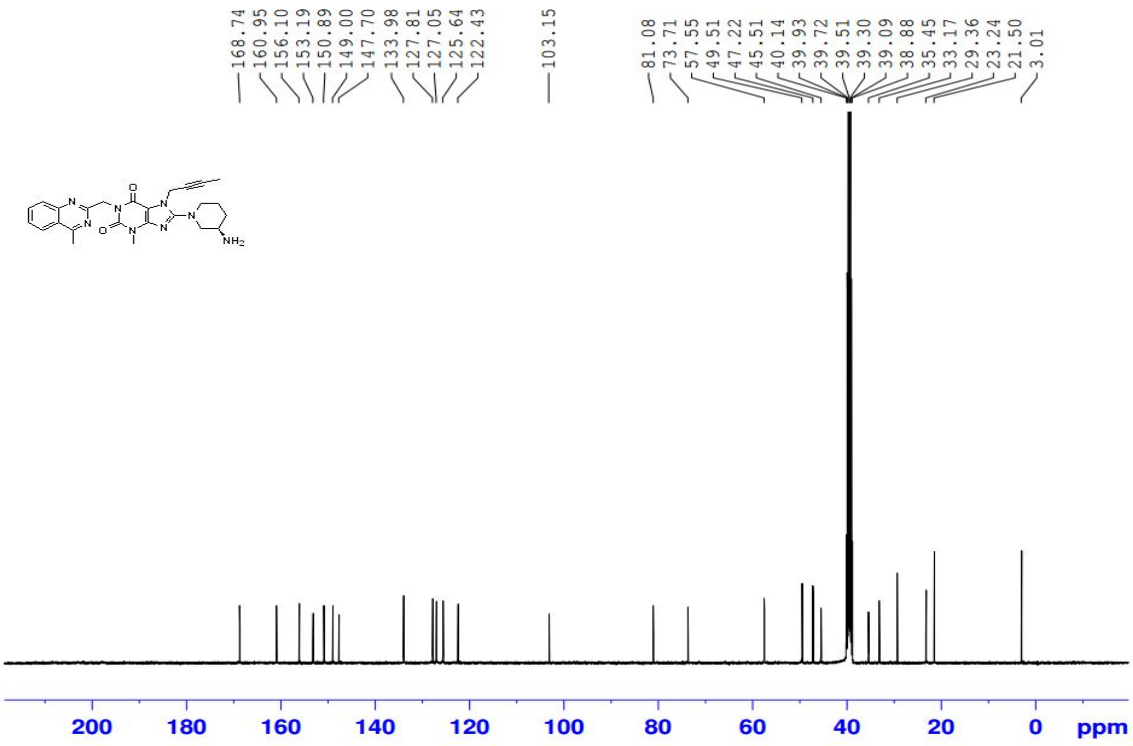
^{1}H NMR and ^{13}C NMR spectra for compound **5**



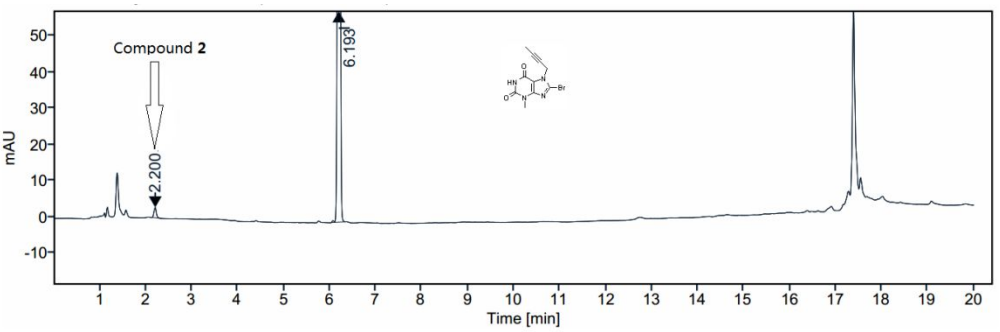


^{1}H NMR and ^{13}C NMR spectra for compound **1**





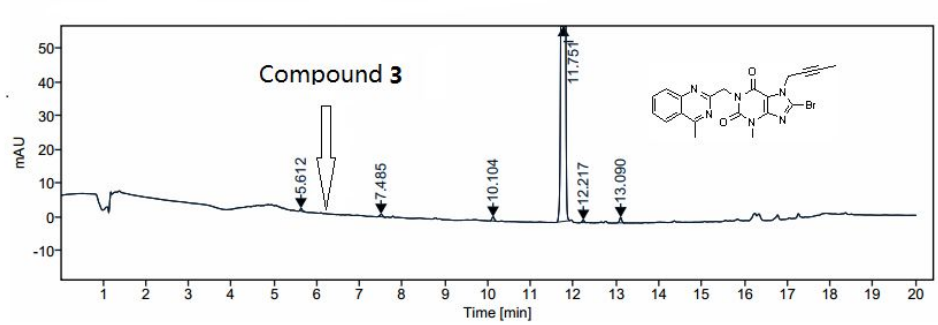
HPLC chromatography for Compound **3**



Signal: DAD1A,Sig=206,4 Ref=off

RT [min]	Height	Area	Area %	Resolution USP	Tail	Theoretical Plates USP	Peak Purity	S/N	RRT	Name
2.20	2.809	10.282	0.825		1.0	8195				
6.19	396.612	1236.516	99.175	44.3	1.0	88941				

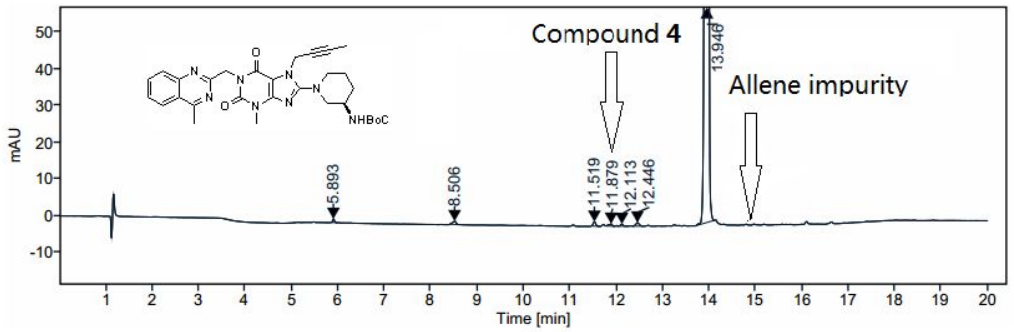
HPLC chromatography for Compound **4**



Signal: DAD1A,Sig=226,4 Ref=off

RT [min]	Height	Area	Area %	Resolution USP	Tail	Theoretical Plates USP	Peak Purity	S/N	RRT	Name
5.61	0.782	2.802	0.105		1.0	61563				
7.49	0.751	3.113	0.117	18.6	0.9	72847				
10.10	1.213	4.692	0.177	24.4	0.9	150879				
11.75	663.293	2638.020	99.311	15.7	1.0	200435				
12.22	0.697	2.259	0.085	4.8	1.0	299523				
13.09	1.571	5.430	0.204	9.6	1.0	313932				

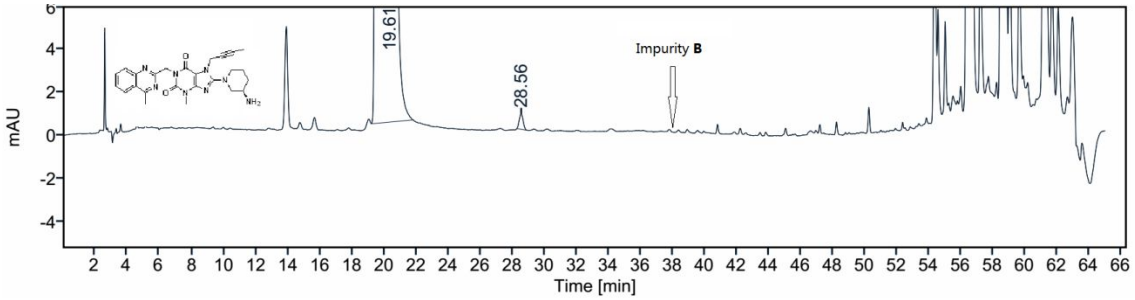
HPLC chromatography for Compound **5**



Signal: DAD1A,Sig=226,4 Ref=off

RT [min]	Height	Area	Area %	Resolution USP	Tail	Theoretical Plates USP	Peak Purity	S/N	RRT	Name
5.89	0.940	2.587	0.062		0.9	101756				
8.51	0.839	4.563	0.109	22.8	0.8	48220				
11.52	1.130	4.382	0.105	23.2	1.0	191825				
11.88	0.601	2.848	0.068	3.4	0.7	196716				
12.11	0.569	2.703	0.065	2.2	0.7	226335				
12.45	0.946	4.827	0.115	2.8	1.1	133193				
13.95	1175.626	4165.794	99.477	13.0	0.9	356061				

HPLC chromatography for Compound **1**



Signal: DAD1A,Sig=226,4 Ref=off

RT [min]	Height	Area	Peak Resolution USP	Peak Theoretical Plates USP	Area %
19.61	1403.4	52079.3		8498.4	99.98
28.56	0.7	12.3	13.63	60911.9	0.02