DRUG SYNTHESIS

OPTIMIZATION OF ARIPIPRAZOLE SYNTHESIS

ANDRZEJ LEŚ ^{1,2*}, KATARZYNA BADOWSKA-ROSŁONEK¹, MARTA ŁASZCZ¹, AGATA KAMIEŃSKA-DUDA¹, PIOTR BARAN¹ and ŁUKASZ KACZMAREK¹

¹ Pharmaceutical Research Institute, Rydygiera 8, 01-793 Warszawa, Poland, ² Faculty of Chemistry, University of Warsaw, Pasteura 1, 02-093 Warszawa, Poland

Abstract: A series of aripiprazole (AR3) syntheses were performed at laboratory scale (10 mmol of the AR1 substrate) in order to optimize the amount of another substrate AR2, as well as Na_2CO_3 , ethanol and varying the reaction time. The reaction parameters were chosen according to the D-optimal plans. A high conversion ratio, about 90-99 %, was obtained. Purity of crude product (AR3) was determined by HPLC. Molar content of crude reaction product was predicted theoretically with the use of the mass balance and the corresponding HPLC parameters. The theoretical predictions were verified with the potentiometric and thermogravimetric analysis of selected samples. Based on the predicted molar content of reaction mixtures, a series of reaction response surfaces was calculated and optimal set of reaction parameters for aripiprazole synthesis was determined.

Keywords: aripiprazole, optimization, synthesis, reaction response surface

Aripiprazole (7-{4-[4-(2,3-dichlorophenyl)-1piperazinyl]butoxy}-3,4-dihydrocarbostyril) belongs to the group of new antipsychotics (like, for example, olanzapine, quetiapine, risperidone, amisulpride). It is used in the treatment of schizophrenia, bipolar mania and some dementia related psychosis symptoms (1). The drug was invented initially by the Otsuka Pharmaceutical Co., Ltd., Tokyo [OPC-14597 (2, 3)] and co-marketed with Bristol-Myers Squibb [Abilify[™], BMS-337039 (4)].

Aripiprazole can form several inclusion compounds containing polar and protic species (5). In a search for optimal conditions to obtain crystals of aripiprazole solvate with ethanol suitable for pharmaceutical technology, a series of experiments has been recently performed (6, 7). In order to get the product of pharmaceutical quality it was necessary to optimize synthetic process, maximizing reaction yield and minimizing the level of potential impurities and unconsumed substrates. This lead to the application of theoretical methodology supporting the reaction response hypersurface (8-10). Theoretical approach using statistical methods was shown to be very useful in analytical methods validation (11) as well as in prediction of biological activity (12).

The present paper reports optimization of ariprazole syntesis with the use of a combined

experimental and theoretical approach. The chemical formulas of reagents are presented in Scheme 1.

EXPERIMENTAL

Synthesis

Representative procedure for the synthesis of 7-{4-[4-(2,3-dichlorophenyl)piperazin-1yl]butoxy}-1,2,3,4-tetrahydroquinolin-2-one (**AR3**)

To a suspension of 7-(4-bromobutoxy)-1,2,3,4tetrahydrochinolin-2-one (1) (29.8 g; 0.1 M) and 1-(2,3-dichlorophenyl)piperazine hydrochloride (2) (29.4 g; 0.1 M) in 300 mL of technical ethanol, powdered anhydrous sodium carbonate (23.3 g; 0.2 M) was added. The mixture was refluxed for 12 h. The resulting solid was filtered, taken up in technical ethanol (50 mL) and refluxed for 10 min. The insoluble inorganic residue was filtered off and the two obtained filtrates were combined, refluxed and then left at the room temperature for crystallization for 12 h. The crystalline aripiprazole (**3**) was filtered and dried to give 42.2 g (yield 85 %) of final product (HPLC purity 99.32 %).

Analytical methods

Reagents and solvents were used without additional purification. Purity of 7-(4-bromobutoxy)-1,2,3,4-tetrahydrochinolin-2-one (1) – substrate

^{*} Corresponding author: e-mail: a.les@ifarm.waw.pl



Scheme 1.

AR1 (Topharman Shanghai Co. Ltd.) used for all optimization experiments was 98.89 % (determined by HPLC method). Purity of 1-(2,3-dichlorophenyl) piperazine hydrochloride (2) - substrate AR2 (Xiamen Mchem. Pharma Ltd.) was 99.88 % (determined by HPLC method). Ethyl alcohol (Linegal Chemicals), which was used during all optimization experiments, was of technical purity (96% of ethyl alcohol and acetone not more than 6%). Sodium carbonate anhydrous (POCh) was of technical purity. HPLC was carried out on Shimadzu LC-2010A (Kyoto, Japan) chromatograph equipped with quaternary pump, UV-VIS detector, autosampler and column oven. The chromatographic separation was accomplished using the reversed-phase column -Waters Symmetry[®] C8 4.6 \times 250 mm, 5 μ m. As a mobile phase, the mixture of water : methanol (POCh, pure p.a.) : triethylamine anhydrous (MERCK, 99 %) : acetic acid (MERCK, 100 %) (300:700:12:5, v/v/v) was used. Eluent pH 7.4 was obtained using acetic acid. The flow rate was 1.5 mL/min, detection was at 220 nm. The analyzed AR samples were dissolved in methanol, the concentration of the sample was about 0.5 mg/mL. The injection of 10 mL was used.

Potentiometric titration has been performed with the 721 NET Titrino instrument (Metrohm). The samples were dissolved in acetic acid and titrated with 0.1 M perchloric acid (HClO₄).

Thermal analyses were carried out by means of the TGA/SDTA 851 cell (Mettler Toledo) under the nitrogen atmosphere. Accurately weighed samples (5-7 mg) were packed in aluminium pans, hermetically sealed then perforated. The **AR3** samples were heated from 25 to 180°C, with the scanning rate of 5°C/min, whereas the **AR2** samples were heated from 25 to 300°C, with the scanning rate of 10°C/min. For the calibration of the TGA instrument, measurements with indium and aluminium were performed. TGA measurements were blank curve corrected.

Optimization

Experiments (n = 16) were performed varying four reaction parameters (amounts of AR2, Na₂CO₃, EtOH and reaction time) in a way close to the Doptimal plan (13, 14) and starting from 10.0 mmoL of AR1 (2.98 g), according to Scheme 1.

The results are presented in Table 1, columns

Table 1. Kesults	of $n = 16 exp$	eriments for c	optimization (of synthesis o	of aripiprazole (columns 1-1	I) and predicti	on of the crude	product cont	ent estimated	based on the p	presented theor	etical method (c	.(CI-21 sumulo:
	p1	p2	p3	p4	Synthesis	HPLC	HPLC	HPLC	HPLC	HPLC	Theory	Theory	Theory	Theory
Experiment	Na_2CO_3	AR2	time	EtOH	AR3 ¹	AR3	AR1	AR2	AR3	IMP	AR1	AR2	AR3	IMP ²
no.	[equiv.]	[mmoL]	[hours]	[mL]	[g]	[mg]	[%]	[%]	[%]	[%]	x [mmoL]	y [mmoL]	z [mmoL]	t [mmoL]
1	1.1	10	3	20.6	4.206	11.26	11.46	0.485	86.3	1.752	1.37	0.13	8.23	0.17
2	1.1	12	12	20.6	4.368	11.20	0.010	0.140	99.30	0.550	0.00	0.04	9.67	0.05
3	3.3	11	3	40.2	4.578	10.11	1.110	1.269	96.69	0.927	0.14	0.35	9.84	0.09
4	3.3	12	3	20.6	4.651	12.68	8.576	0.343	89.6	1.482	1.13	0.10	9.41	0.16
S	1.1	12	ю	20.6	4.216	10.79	0.607	3.270	95.36	0.761	0.07	0.83	8.86	0.07
9	3.3	10	12	20.6	4.100	10.77	0.09	0.033	98.92	0.96	0.01	0.01	9.05	0.09
7	3.3	12	12	40.2	4.398	10.09	0.125	0.475	98.74	0.637	0.02	0.13	9.67	0.06
8	3.3	10	3	20.6	4.482	11.81	0.794	1.190	96.85	1.170	0.10	0.32	9.65	0.12
6	2.2	10	12	40.2	4.221	10.38	0.937	0.113	96.01	2.942	0.11	0.03	9.05	0.28
10	1.1	10	12	20.6	4.238	9.87	2.000	0.315	95.46	2.223	0.24	0.08	9.04	0.21
11	1.1	10	3	40.2	4.198	9.36	17.741	0.960	79.45	1.851	2.14	0.25	7.63	0.18
12	3.3	10	7.5	33.56	4.450	11.40	0.170	0.250	98.61	0.970	0.02	0.07	9.78	0.10
13	2.2	11	7.5	20.6	4.654	9.72	0.044	1.656	97.41	0.894	0.01	0.46	10.04	0.09
14	1.1	12	7.5	40.2	4.375	11.2	0.040	0.340	99.03	0.590	0.01	0.09	9.65	0.06
15	1.1	11	12	32.74	4.099	9.64	3.644	0.476	93.17	2.693	0.42	0.12	8.55	0.25
16	2.2	12	3	31.14	4.819	12.54	0.438	6.893	91.82	1.289	0.04	1.93	9.60	0.12
¹ AR3 product molecular mas	is contamin s M = 450 [lated with un g]	nconsumed	AR1, AR2	and unidenti	fied impuri	ties (IMP). ²	under assum]	ption that all	impurities	are represent	ed by a hypo	thetical single	e molecule of

Optimization of aripiprazole synthesis

1-11. The available experimental techniques (mass weighting, HPLC) are viable for qualitative and semiquantitative determination of the content of crude reaction product. However, for optimization purposes a more precise determination of the molar content of crude product is necessary. For this reason, a complementary theoretical methodology was applied in order to include mass balances in the estimation of crude product content. A similar idea was used previously for semiguantitative estimation of the content of polyester resins (15). In the present work, it was assumed that crude product is composed of pure AR3 substance, unconsumed AR1 and AR2 substrates and unidentified impurities (IMP). The relative content of crude product components was determined by HPLC using a common linear relation between the mass of a given substance, or equivalently, the molar content (*x*) and the area under the corresponding peak (S) on the chromatogram.

 $S = \varepsilon x$

where ε is an empirical factor that can be determined from earlier calibration measurements.

Based on our laboratory experience, the following ratios were suggested:

 ϵ (AR1) : ϵ (AR2) : ϵ (AR3) : ϵ (IMP) =

0.80:0.37:1.00:1.00

For mass balances the following molar masses were used:

AR1
$$M_1 = 298.18$$
 g
AR2 $M_2 = 231.12$ g
AR3 $M_3 = 448.39$ g

In the subsequent analysis, it was assumed that the impurites (IMP) can be treated as a single, hypothetical average ,,molecule" with the molar mass of $M_4 = 450$ g. It was assumed also that the ε (IMP) factor is equal to the ε (AR3) of the AR3 pure product.

A detailed theoretical procedure is described below. At the *i*-th experiment one obtains m_i grams of the crude product containing X_i mol of **AR1**, Y_i mol of **AR2**, Z_i mol of **AR3** × H₂O and T_i mol of IMP, i = 1,...,n; n = 16. The mass balances are expressed with the use of the following equations:

$$M_1 X_i + M_2 Y_i + M_3 Z_i + M_4 T_i = m_i$$

The X_i , Y_i , Z_i , and T_i values are estimated from the HPLC measurements knowing mass samples μ_i (usually about 0.20-0.25% of crude product mass m_i) and the corresponding peak areas.

The ratio of molar amounts can be obtained from the normalized peaks on chromatograms and expressed in another set of equations:

$$\begin{aligned} & \epsilon_{1} x_{i} - (\mathbf{AR1\%-i}) (\epsilon_{1} x_{i} + \epsilon_{2} y_{i} + \epsilon_{3} z_{i} + \epsilon_{4} t_{i}) = 0 \\ & \epsilon_{2} y_{i} - (\mathbf{AR2\%-i}) (\epsilon_{1} x_{i} + \epsilon_{2} y_{i} + \epsilon_{3} z_{i} + \epsilon_{4} t_{i}) = 0 \\ & \epsilon_{3} z_{i} - (\mathbf{AR3\%-i}) (\epsilon_{1} x_{i} + \epsilon_{2} y_{i} + \epsilon_{3} z_{i} + \epsilon_{4} t_{i}) = 0 \\ & \epsilon_{z} t_{i} - (\mathrm{IMP\%-i}) (\epsilon_{1} x_{i} + \epsilon_{2} y_{i} + \epsilon_{3} z_{i} + \epsilon_{4} t_{i}) = 0 \end{aligned}$$

where AR1%, AR2%, AR3% and IMP% denote the percent contribution of a given reagent peak to the integral area of all peaks. The mass balances of HPLC samples are now expressed with the use of equations:

$$M_1 x_i + M_2 y_i + M_3 z_i + M_4 t_i = \mu_i$$

The X_i and x_i quantities are linearly dependent via a relation $X_i = x_i m_i / \mu_i$ (a similar relation holds for Y, Z and T). In total, we have $n \times 4 = 64$ unknowns x_i , y_i , z_i , and t_i . Fortunately, for each *i*-th experiment they can be estimated independently by means of solving simultaneously five equations given above. The predicted molar content (X, Y, Z

Experiment no. (AR1/AR2 substrate ratio in mmoL)	Titration sample mass [g]	Volume of acid used [mL]	Content of H_2O and other solvents based on TG analysis $[\%]$	Content of AR2 + AR3 based on titration [mmoL]	Content of AR2 + AR3 based on theoretical calculations [mmoL]
2 (10/12)	0.3503 0.3570	7.386 7.523	3.99 3.99	9.63 9.62	9.71
6 (10/10)	0.3988	8.674	0.15	8.93	9.05
9 (10/10)	0.3836	8.069	1.66	9.03	9.08
12 (10/10)	0.3572 0.3572	7.483 7.482	4.27 4.27	9.78 9.77	9.85
14 (10/12)	0.3506 0.3596	7.521 7.713	2.10 2.10	9.62 9.62	9.74

Table 2. Comparison of the AR2 + AR3 content in selected experiments estimated based on the titration and TG analysis and calculated theoretically.

Na ₂ CO ₃ [equiv.]	AR2 [equiv.]	Time [hours]	EtOH [mL]	AR1 [mmoL]	AR2 [mmoL]	AR3 [mmoL]	IMP [mmoL]
p_1	<i>p</i> ₂	p_3	p_4	<i>y</i> ₁	<i>y</i> ₂	<i>y</i> ₃	<i>y</i> ₄
3.03	1.20	8.63	40.20	0.16	0.20	9.93	0.05
3.30	1.15	9.75	40.20	0.08	0.06	9.91	0.05
3.30	1.18	8.63	40.20	0.17	0.09	9.94	0.05
3.30	1.18	9.75	40.20	0.13	0.07	9.91	0.05
3.30	1.20	8.63	40.20	0.29	0.10	9.94	0.05
3.30	1.20	9.75	40.20	0.22	0.08	9.91	0.04

Table 3. The sets of reaction parameters corresponding to the high yield (over 99%) of pure aripiprazole (**AR3**) in the reaction started with 10 mmoL of **AR1**, as predicted based on the theoretical methods.

Table 4. The *b*-parameters entering the z(p, b) formula (Eq. 1 and 2)¹ for the calculations within the R-2/R-1 combined theoretical methods.

Parameter	AR1	AR2	AR3	IMP
b_0	-5,3062	-3,2416	4,3543	-4,1975
b_I	0	0	0,6603	-0,2593
b_2	-0,2682	0,3524	0	-0,4306
b_3	-1,1514	-1,2375	0	0
b_4	0	0	0	0
<i>b</i> ₁₂	2,3430	0	0	0,3067
<i>b</i> ₁₃	0	0	-0,3106	-0,1248
<i>b</i> ₁₄	0	0	0,3690	-0,2489
<i>b</i> ₂₃	0	0	0	-0,2739
<i>b</i> ₂₄	0	0	0	0
<i>b</i> ₃₄	0	0	0	1,0932
<i>b</i> ₁₁	0	-1,5550	0	0
<i>b</i> ₂₂	0	0	0	0
<i>b</i> ₃₃	0	0	-1,9050	0,2645
b_{44}	0	0	0	-0,4349

¹ $y_1^{max} = 10 \text{ mmoL}, y_2^{max} = 12 \text{ mmoL}, y_3^{max} = 10 \text{ mmoL}, y_4^{max} = 10 \text{ mmoL}$

and *T*) of crude product for *i*-th experiment is given in Table 1 (columns 12-15). It can be seen that the predicted molar content of the **AR3** pure product (Zvalues) is poorly correlated with the mass of the crude product given in Table 1, column 6. The theoretical predictions were verified for five experiments (no. 2, 6, 9, 12 and 14) by means of potentiometric and thermogravimetric analysis (see Table 2). A good agreement between theoretical and empirical values of the **AR2** + **AR3** content validates our theoretical approach. Based on the HPLC measurements, a high molar content of **AR3** (99% and higher) is expected in experiments labeled as no. 2 and 14. It suggests that one can use 20% excess of **AR2** substrate, about 1.1 equivalents of sodium bicarbonate, and, perhaps, shorter time of the reaction.

Reaction response surface

In the next step, the predicted molar contents of **AR1**, **AR2**, **AR3** and IMP reagents were used in the calculation of the reaction response surface, in order to predict higher reaction yield and lower content of impurities. We followed an approach elaborated previously (8-10) and based on the representation of the reaction output by means of the logistic curve dependent on a quadratic form of reaction parame-

ters. Using such a curve one can fulfill natural constraints of non-negative reaction outputs that must be bound also from above due to mass balances. In particular, the following reaction outputs were considered: y_1 – the molar content of **AR1**, y_2 – the molar content of **AR2**, y_3 – the molar content of **AR3** and y_4 – the (hypothetical) molar content of IMP. An explicit form of the logistic curve is given below:

$$\hat{y}(\boldsymbol{p}) = \hat{y}(\boldsymbol{p}; \boldsymbol{b}) = \frac{y^{max}}{1 + e^{-z(\boldsymbol{p}; \boldsymbol{b})}}$$
(1)
where

$$z(\boldsymbol{p};\boldsymbol{b}) = b_0 + \Sigma b_i u_i + \Sigma \Sigma b_{ij} u_i u_j$$
(2)

 $u_i = (p_i - M_i) / R_i$, i = 1,...,m (3) $M_i = (b_i + a_i)/2$ is the midpoint of the range of the x_i parameter with a_i and b_i endpoints,

 $R_i = (b_i - a_i)/2$ is the half of the x_i range. y^{max} is the maximum value of the *y*-response estimated from the experimental conditions. The *b*-parameters are determined with the least-squares fitting. The suggested logistic curve is a nonlinear function of both *b*- and *p*-parameters. For the estimation of the *b*-parameters one can apply two methods:

R-1 method

This is a method for direct minimization of the mean squared deviation between the reference *y*-values and the model (using logistic curve) values:

$$\boldsymbol{d}^* = \arg\min \| \boldsymbol{y}(\boldsymbol{p}) - \boldsymbol{y}(\boldsymbol{p}; \boldsymbol{d}) \|$$
(4)

where y(p,d) = y(z(p;d)). d^* can be calculated with routine numerical minimization methods such as, for example, the modified Gauss-Newton method (16).

R-2 method

This is an indirect method where initially the reference y-data are logit-transformed and then approximated with a quadratic form of p-parameters using the least squares method. In particular, we calculate such b-parameters (by analogy to d-parameters) that minimize mean squared deviation of the logit-transformed reference y-data from the quadratic form of p-parameters:

$$\boldsymbol{b}^* = \arg\min \parallel \log (y(\boldsymbol{p}) / (y^{\max} - y(\boldsymbol{p}))) - z(\boldsymbol{p}; \boldsymbol{b}) \parallel (5)$$

Having the calculated *z*-function one can predict the *y*-values using the formula in Eq. (1). The \boldsymbol{b}^* linear coefficients can be calculated form the following expression:

$$\boldsymbol{b}^{*} = (X'X)^{-1} X' \log (y(\boldsymbol{p}) / (y^{max} - y(\boldsymbol{p})))$$
(6)

where *X* is a matrix with a series of columns:

$$\boldsymbol{X} = [I \ u_1 \ u_2 \ \dots \ u_m \ u_1 u_2 \ u_1 u_3 \ \dots \ u_{m-1} \ u_m] \tag{7}$$

The calculated reaction response surface can be expressed in a more compact form as:

$$\hat{y}(\boldsymbol{p};\boldsymbol{b^*}) = \hat{y}(\boldsymbol{p}, y(\boldsymbol{x})) = \frac{y^{max}}{1 + e^{-X(X'X)^T X' \log (y(\boldsymbol{p}) / (ymax - y(\boldsymbol{p})))}}$$
(8)

Both R-1 and R-2 methods are not equivalent and may, in general, lead to different estimations of linear (d or b) coefficients. In practice, however, the values of d and b coefficients are very close and do not affect final conclusions. In the present approach we used both methods sequentially, i.e. the R-2 method was used first in order to obtain a reasonable estimation of linear coefficients that were refined subsequently with the R-1 method. In the process of linear coefficients determination a stepwise multiple linear regression was used in order to eliminate nonsignificant terms. The optimized linear coefficients are presented in Table 4.

Exploration of reaction response surface

Due to rather small dimension of reaction parameter space $(\dim = 4)$ it was possible to scan the calculated response surface using 9 points per dimension ($9^4 = 6561$ points in total). As a result, we localized 6 sets of *p*-parameters corresponding to the molar content of AR3 above 99%. These sets are shown in Table 3. It is seen that the optimal set of reaction parameters corresponds to 3.3 equiv. of Na₂CO₃, 12 mmoL of AR2, 8-9 hours for reaction duration and about 40 mL of EtOH per 10 mmoL of **AR1** (7% of molar addition). The predicted shortening of the reaction timing becomes rather unexpected result, because the available prescriptions in the literature recommend about 12 hours or longer timing. From the present work it appears also that under a suitable choice of reaction parameters the yield of AR3 can exceed 99%. The unconsumed AR1 and AR2 substrates are probably the major components of impurities.

CONCLUSIONS

An optimization of aripiprazole synthesis has been performed at the laboratory scale based on the n = 16 experiments with n = 4 reaction parameters and supported by analysis of theoretically estimated reaction response surface. It appeared that high molar content (about 99%) of the final product can be obtained within 8-9 hours instead of about 12 hours recommended previously. The reaction should be conducted with some 20% excess of the **AR2** substrate, a 3-fold molar excess of sodium bicarbonate and 7% molar addition of ethanol. The calculated formula for the response surface can be used for prediction of molar content of crude reaction product at the conditions slightly perturbed from the ones determined in the present work.

REFERENCES

- Davies M. A., Sheffler D.J., Roth B.L.: CNS Drug Rev. 4, 317 (2004).
- 2. http://otsuka-europe.com
- Oshiro Y., Sato S., Kurahashi N.: U.S. Pat. 5006528 (April 9, 1991)
- 4. http://www.bms.com
- Tessler L, Goldberg I.: J. Inclusion Phenom. Macrocycl. Chem. 55, 255 (2006).
- Kaczmarek Ł., Łaszcz M., Pietraszko A., Szelejewski W.: Polish patent application P-378363 (December 12, 2005)
- Łaszcz M., Pietraszko A., Glice M., Kaczmarek Ł., Badowska-Rosłonek K., Korczak K., Maruszak W., Beczkowicz H.: Abstract on the Fifth Multidisciplinary Conference on Drug Research, Poland (2006); http://www.science24.com/paper/7006.

- Pucko W., Leś A., Szelejewski W.: Org. Proc. Res. Dev. 8, 157 (2004).
- Wojciechowska W., Kutner A., Leś A., Szelejewski W.: Pol. J. Appl. Chem. 3-4, 63 (2004).
- Martynow J., Krupa M., Leś A., Kutner A., Szelejewski A.: Org. Proc. Res. Dev. 8, 846 (2004).
- Rudzki P. J., Leś A.: Acta Pol. Pharm. Drug Res. 65, 743 (2008).
- Brzezińska E. O., Stolarska J. D.: Acta Pol. Pharm. Drug Res. 66, 25 (2009).
- 13. R. Carlson: Design and Optimization in Organic Synthesis; Elsevier, Amsterdam 2000.
- 14. Modde 6.0, Umetrics AB, Umea, Sweden; http://www.umetrics.com
- Skarżyński M., Cynkowska G., Abis L., Wawer I., Spera S.: J. Appl. Polym. Sci. 70, 675 (1998).
- 16. SAS Institute Inc., Cary, U.S.A.; http://www.sas.com.

Received: 24.08.2009