

# 基于人工神经网络的合成脱水长春碱工艺优化研究

李 硕 赵春芳 吴泽强 熊 茵 余龙江\*  
(华中科技大学生命科学与技术学院, 湖北 武汉 430074)

**摘 要:** 脱水长春碱是合成长春瑞滨、长春氟宁等新一代高价值抗癌药物最为重要的中间体。本文以硫酸长春碱为原料,进行了脱水长春碱合成的工艺研究。采用正交实验与人工神经网络对该工艺进行了优化,建立了三层改进的误差反向传播网络(BP-ANN神经网络),并验证了该模型的有效性。扩大正交实验的水平范围,通过模型的模拟及实验验证后,得到了合成脱水长春碱的最佳工艺条件:在 DMF 体系中,底物浓度为 15 mmol/L,采用  $C_2O_2Cl_2$  为脱水剂,其用量与底物的当量比为 50:1,在 4℃ 下进行脱水反应,反应 12 h 后,加入冰水,氨水调节 pH 为 10.0,乙醚与二氯甲烷混合溶液(体积比为 10)萃取,得到产物的总收率达到 70.88%。

**关键词:** 脱水长春碱; 人工神经网络; 合成; 正交实验

**中图分类号:** TQ460.31

## 引 言

脱水长春碱是第一个被人工合成的长春花属生物碱<sup>[1]</sup>,也是合成长春瑞滨及长春氟宁等其他新型长春花属抗肿瘤药物最重要的中间体之一。目前,合成脱水长春碱已有多种方法<sup>[2-4]</sup>,而长春碱通过定向脱去一个分子的  $H_2O$  得到脱水长春碱的方法,由于其合成工艺简便而受到广泛的关注<sup>[5-6]</sup>,但反应中控制条件较多而收率一般不高<sup>[6]</sup>。

人工神经网络是一类模拟人脑功能的信息加工处理系统,其理论基础是对真实的生理神经网络进行数学抽象而得到人工网络的数学模型<sup>[7-8]</sup>。误差反向传播网络(BP-ANN)<sup>[9-10]</sup>是一种典型的人工神经网络,它能够通过对样本的提炼与分析,获取数学模型,再通过对模型输出值与实际测量值之间的比对来不断更正模型本身误差的“训练”过程,获得准确的反应数据内在规律的数学模型。而利用增加了冲量项的改进 BP 算法得到的改进 BP 网络<sup>[11-12]</sup>,更加真实可靠。本文在对合成过程中多个因素初步分析后,选取 5 个主要因素,在正交实验基础上,利用改进的 BP-ANN 网络,对脱水长春碱的合成过程进行了优化研究,得到了最优合成条件。

## 1 实验部分

### 1.1 实验原料与仪器

#### 1.1.1 原料

硫酸长春碱, HPLC > 65%, 海南海宁化工厂; 标样, Sigma 公司; 其他所需的试剂均为分析纯, 上海试剂一厂, 使用前严格脱水。

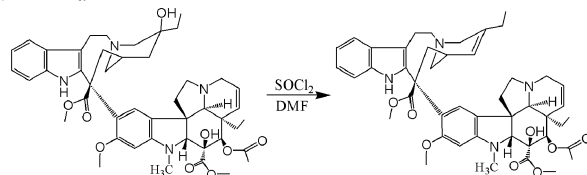
#### 1.1.2 仪器

高效色谱系统 Waters Millipore HPLC, 美国 Waters 公司。

### 1.2 实验方法

#### 1.2.1 脱水长春碱的合成

在反应瓶中加入一定量的脱水试剂, 在 Ar 气保护、冰浴下, 含有硫酸长春碱 (HPLC > 65%) 的 N,N-二甲基甲酰胺 (DMF) 溶液加入到反应瓶中。去除冰浴, 保持惰性环境, 在一定温度下搅拌反应, 采用薄层色谱 (TCL) 跟踪至原料点消失时终止反应 (硅胶 GF<sub>254</sub> 板, 展开条件甲醇与二氯甲烷的体积比为 1:4, 紫外显色)。加入冰水, 氨水调节 pH 值为 10.0, 有机溶剂萃取, 水反洗后, 无水  $Na_2SO_4$  干燥有机相, 减压浓缩至干, 残余物加入含有一定水的甲醇 50℃ 热溶解, 冷冻结晶后过滤得到脱水长春碱。反应式如下:



收稿日期: 2007-07-20

第一作者: 男, 1980 年生, 博士生

\* 通讯联系人

E-mail: yulongjiang@mail.hust.edu.cn

### 1.2.2 脱水长春碱的含量测定

产物采用 Waters Millipore HPLC 检测; 色谱柱: Dikma Diamonsil™ C18 5 $\mu$ ; 检测条件甲醇与乙腈和水的体积比为 25:35:40, pH6.0; 检测波长 267 nm, 流速 1.0 mL/min, 柱温 25℃。配制 1.0 mg/mL 的标准脱水长春碱溶液, 稀释至 3.0、6.0、10.0、15.0、20  $\mu$ g/mL。每次进样 10  $\mu$ L, 测量峰面积。以进样样品的浓度为横坐标 ( $X$ ), 峰面积为纵坐标 ( $Y$ ) 得到标准曲线:  $Y = 35237X + 7471.5$ ,  $R^2 = 0.9989$ 。计算平均回收率 100.78% ( $n = 5$ ), 相对标准偏差 RSD = 0.35%。

## 2 结果与讨论

### 2.1 影响因素的选择

影响脱水长春碱收率的因素有: 所用反应体系的选择、底物的浓度、脱水剂的种类、脱水剂的用量、反应的温度及时间、反应结束后萃取剂的选择。本文在大量实验的基础上对上述因素进行了考察。

DMF 是一种非离子弱碱性溶液, 被广泛应用在合成领域中。以 DMF 为该脱水反应的溶剂体系, 不仅能够起到很好的溶解样品的作用, 并且还能够与脱水剂  $C_2O_2Cl_2$  反应形成复合物, 增加脱水效率<sup>[13]</sup>, 因此本文选择 DMF 作为该反应的体系, 以  $C_2O_2Cl_2$  为反应的脱水剂。

反应结束后, 有机溶剂萃取时, 试剂的选择尤为重要, 传统的方法一般用乙醚萃取<sup>[5-6]</sup>。本文研究表明, 虽然乙醚萃取得到的产物脱水长春碱的含量较高, 但是当选用二氯甲烷时, 萃取得到的产物虽然纯度不及乙醚萃取物, 但其萃取产物的重量相对较高, 因此选用乙醚和二氯甲烷的混合溶液进行萃取分离。

### 2.2 利用正交实验的工艺优化研究

本文选择 DMF 为反应体系, 以  $C_2O_2Cl_2$  为脱水剂, 需要对底物的浓度 A (mmol/L), 脱水剂的用量 B (脱水剂与底物的摩尔比, mol/mol)、反应的温度 C (℃) 及时间 D (h) 及反应结束后萃取剂的选择 E (乙醚与二氯甲烷的体积比) 5 个因素进行考察。每个因素选取 4 个水平, 采用  $L_{16}(4^5)$  正交表进行实验。结果见表 1、2。

从表 2 中可以看出, 反应温度及反应底物的浓度对收率的影响最大, 其次是萃取用有机溶剂的选择及反应时间的长短, 而对脱水剂的用量对得率的影响最小。从正交实验得出的较好的工艺条件是:

表 1 各因素水平的选择

Table 1 Factors and levels

水平	因素				
	A	B	C	D	E
1	5	20	4	6	1
2	10	40	10	8	3
3	20	60	15	10	6
4	30	80	25	12	9

表 2 正交结果表

Table 2 The results of orthogonal tests

No.	A	B	C	D	E	收率/%
1	1	1	1	1	1	10.65
2	1	2	2	2	2	20.56
3	1	3	3	3	3	31.93
4	1	4	4	4	4	7.96
5	2	1	2	3	4	50.77
6	2	2	1	4	3	63.05
7	2	3	4	1	2	16.84
8	2	4	3	2	1	33.06
9	3	1	3	4	2	41.52
10	3	2	4	3	1	15.93
11	3	3	1	2	4	67.52
12	3	4	2	1	3	47.10
13	4	1	4	2	3	18.35
14	4	2	3	1	4	28.08
15	4	3	2	4	1	44.58
16	4	4	1	3	2	50.76
$k_1$	18.275	30.322	47.995	25.668	26.055	
$k_2$	40.930	32.405	41.252	35.373	32.920	
$k_3$	43.017	40.218	33.648	37.347	40.107	
$k_4$	35.442	34.720	14.770	39.275	38.582	
极差 R	24.742	9.896	33.225	13.610	14.052	

$A_3B_3C_1D_2E_4$ 。但是, 正交实验只是考虑到这几个水平之间的实验结果, 而不能从整个实验条件区间进行考虑, 因此尝试利用 BP-ANN 网络来进一步优化合成过程。

### 2.3 利用 BP-ANN 网络寻找最佳工艺条件

#### 2.3.1 神经网络的建立

人工神经网络有几十种之多, 而误差反向传播网络 (BP-ANN) 是目前应用最为广泛的一种。网络结构由三层组成, 即接收输入信号的单元输入层、输出信号的单元输出层以及不直接与输入输出发生联

系的单元隐藏层<sup>[10-11]</sup>。利用 BP-ANN 网络能够从实验数据中获取数学模型,它无需事先给出公式的形式,而是以实验数据为基础,经过“训练”后得到一个反应实验数据内部规律的数学模型,然后利用这一模型进行推理。

BP-ANN 网络采用 Matlab 软件进行模拟,经过对多个模型的比对,选择  $5 \times 10 \times 8 \times 1$  模型,即输入层共有 5 个神经元,包括有脱水剂的选择、脱水剂的用量、反应温度及时间、萃取剂的选择;输出层为脱水长春碱的收率;中间隐藏层包含有两层共 18 个神经元。利用上面 16 组正交实验数据对建立的 BP-ANN 网络模型进行“训练”,选取“训练”参数: Momentum 0.5, Learning Rate 0.2;停止训练条件: Iterations  $10^6$ , Tolerance 99%, Correct 100%。“训练”结束后得到神经网络模型,然后输入正交实验的各个参数,得到网络输出值  $Y_K$ ,通过与实验结果  $y_K$  比较,计算出相对误差 RD(表 3)及平均相对误差 ARD。

相对误差 RD 及平均相对误差 ARD 通过如下公式进行计算:

$$RD = [(Y_K - y_K) / y_K] \times 100$$

$$ARD = (1/n) |RD|$$

表 3 BP-ANN 网络对实验数据的模拟结果

Table 3 Simulation of experimental results with BP-ANN

No.	$Y_K$	$y_K$	RD/%	No.	$Y_K$	$y_K$	RD/%
1	16.64295	16.65	-0.042	9	41.52947	41.52	0.023
2	20.57632	20.56	0.079	10	15.92412	15.93	-0.037
3	21.9135	21.93	-0.075	11	66.8448	67.52	-1.000
4	7.979774	7.96	0.248	12	47.1278	47.10	0.059
5	50.75591	50.77	-0.027	13	20.34394	20.35	-0.030
6	63.12395	63.05	0.117	14	18.09163	18.08	0.064
7	16.84648	16.84	0.038	15	34.57224	34.58	-0.022
8	30.04618	30.06	-0.046	16	40.7618	40.76	0.004

从表 3 中可以看出,经过“训练”过的 BP-ANN 网络输出数据与实验数据相当吻合,平均相对误差 ASD 0.119%,说明建立的 BP-ANN 网络能够很好的逼近训练集样本变化规律。

本文选取模型样本外的几个条件进行验证实验,取 5 个因素的水平分别为 (7, 30, 6, 9, 5), (15, 50, 12, 9, 8), (25, 70, 20, 11, 8) 进行实验,实验收率 (%) 分别为: 46.15, 62.03, 31.66; 模型模拟结果为: 46.70, 62.63, 32.11, 平均相对误差 ASD 1.08%, 进

一步说明,建立的模型能够反应样本自身的内在规律。

### 2.3.2 利用 BP-ANN 网络预测结果

利用训练好的神经网络,在正交实验结果最好的 11 号结果的基础上,固定其他因素,研究每个因素对收率的影响。并且将正交实验中各个因素的水平拓宽,将底物的浓度 (mmol/L) 从 5~30 拓宽至 1~50,脱水剂的用量 (脱水剂与底物的摩尔比, mol/mol) 从 20~80 拓宽至 1~110、反应的温度 (°C) 从 4~25 拓宽至 1~50,反应时间 (h) 从 6~12 拓宽至 3~18,萃取剂的选择 E(乙醚与二氯甲烷的体积比) 从 1~9 拓宽至 0.5~20。通过得到的神经网络输出,得到的模拟结果如图 1~5 所示。

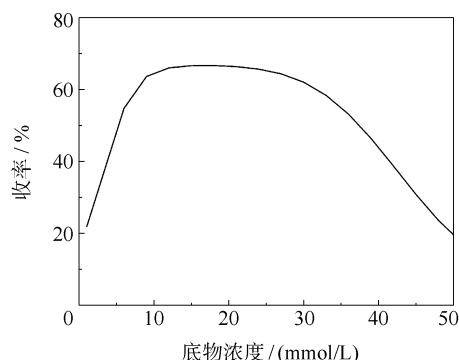


图 1 底物浓度对收率的影响

Fig. 1 Simulation curve of variation of yield with content of raw material

根据本文的实验表明,当底物浓度过低或过高时,副反应加剧导致收率下降明显,这和 BP 网络得到的模拟曲线是一致的。

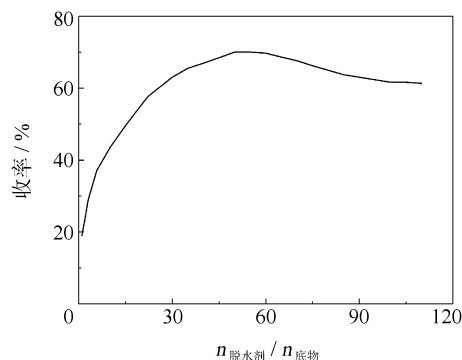


图 2 脱水剂用量对收率得影响

Fig. 2 Simulation curve of variation of yield with dosage of  $C_2O_2Cl_2$

脱水剂的用量对收率影响明显,其用量不足导致反应进行不彻底,用量太大,由于脱水剂的强脱水性而导致副产物增多。

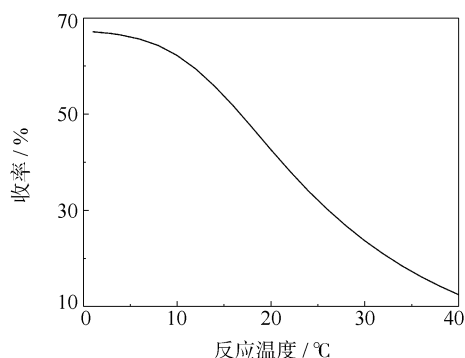


图 3 反应温度对收率的影响

Fig. 3 Simulation curve of variation of yield with reaction temperature

实验表明,温度越低,收率越高,因为由于底物具有多个可去的羟基,温度的升高导致副反应的加剧。实验结果与模拟曲线一致。

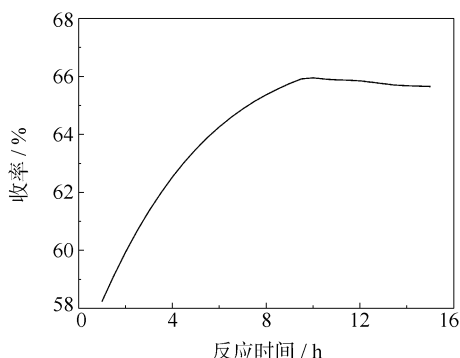


图 4 反应时间对收率的影响

Fig. 4 Simulation curve of variation of yield with the duration of reaction

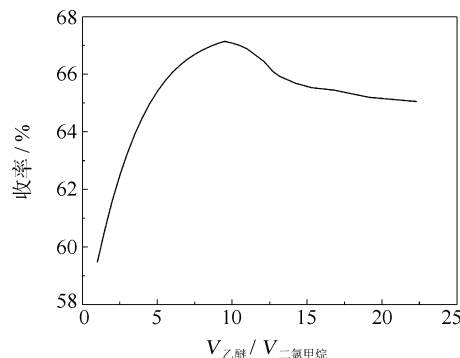


图 5 萃取剂的选择对收率的影响

Fig. 5 Simulation curve of variation of yield with the composition of extractant phase

实验还表明,虽然乙醚萃取得到的产物脱水长春碱的含量较高,但是当选用二氯甲烷时,萃取得到的产物虽然纯度不及乙醚萃取物,但其萃取产物的重量相对较高,模拟曲线表明在两者配比合适时,效

果最佳。

由图 1~5 得到最适合的反应条件为:底物浓度为 15 mmol/L,脱水剂用量与底物摩尔比为 50,反应温度 1~6 °C,反应时间 12 h,萃取剂乙醚与二氯甲烷的体积比 10。输入这一最佳组合条件到训练好的 BP-ANN 网络中,得到模拟输出收率 70.116%。

## 2.4 最佳反应条件的验证

为了验证得到的最佳条件,本文选用模拟得到的最佳条件,进行 5 次平行实验,实验收率为  $(70.884 \pm 0.12)\%$ ,与模拟结果偏差为 1.1%,表明神经网络模拟得到的最佳条件是很合适的。

## 3 结束语

通过建立人工神经网络 BP-ANN,在正交实验的基础上,对网络进行“训练”后得到能够反应数据相互关系的网络模型,能够对实验结果进行预测以得到最佳的工艺条件。在 DMF 为反应体系、 $\text{C}_2\text{O}_2\text{Cl}_2$  为脱水剂的脱水反应过程中,利用 BP-ANN 网络进行预测,得到了最优的反应条件。由得到的最优条件,通过 5 次平行实验进行验证,实验收率为  $(70.884 \pm 0.12)\%$ ,与 BP-ANN 网络得到的结果非常的接近,说明利用人工神经网络对合成过程进行优化与预测是行之有效的。

## 参考文献:

- [1] Fahy J. Modifications in the upper or velbenamine part of the *vinca* alkaloids have major implications for tubulin interacting activities[J]. Current Pharmaceutical Design, 2001, 34(7): 1181–1197.
- [2] Fahy J, Thillaye du Boullay V, Bigg D C H. New method of synthesis of *vinca* alkaloid derivatives[J]. Bioorganic & Medicinal Chemistry Letters, 2002, 12(3): 505–507.
- [3] Gueritte F, Langlois N, Langlois Y, et al. 20'-Deethyl-anhydrovinblastine[J]. Journal of Organic Chemistry, 1981, 46(1): 5393–5395.
- [4] Hardouin C, Doris E, Rousseau B. Selective deoxygenation of leurosine: concise access to anhydrovinblastine[J]. Journal of Organic Chemistry, 2002, 67(2): 6571–6574.
- [5] Langlois N, Gueritte F, Langlois Y, et al. Application of a modification of the polonovski reaction to the synthesis of vinblastine-type alkaloids[J]. Journal of the American Chemical Society, 1976, 98(1): 7017–7024.
- [6] 陈永江,陈洪明,李莉,等. 酒石酸长春瑞滨的合成

- [J]. 中国医药工业杂志, 1999, 30(1): 6-8.
- [7] 陈允平, 王旭蕊, 韩宝亮. 人工神经网络原理及其应用[M]. 北京: 中国电力出版社, 2002: 1-3.
- [8] 党建武. 神经网络技术及应用[M]. 哈尔滨工程大学出版社, 2000.
- [9] 吴启迪, 汪镭. 反馈式神经网络智能控制[M]. 上海: 上海科技教育出版社, 2004: 6-7.
- [10] Lacher R C. Back-propagation learning in expert networks[J]. IEEE Transaction on Neural Networks, 1992, 1(3): 62-72.
- [11] Liu Jiangto, Chang Hongbin, Xu Zuyao, et al. Prediction of the flow stress of high-speed steel during hot deformation using a BP artificial neural network[J]. Journal of Materials Processing Technology, 2000, 103(2): 200-205.
- [12] 胡德文, 王正志, 王耀男, 等. 神经网络自适应控制[M]. 长沙: 国防科技大学出版社, 2006.
- [13] 包俊敏, 夏德生, 陈云华. 合成长春瑞滨的方法: CN 1552716A[P]. 2004-12-08.

## Process optimization of the synthesis of anhydrovinblastine using artificial neural networks

LI Shuo ZHAO ChunFang WU ZeQiang XIONG Yin YU LongJiang

(Institute of Resource Biology and Biotechnology, College of Life Science and Technology,  
Huazhong University of Science and Technology, Wuhan Hubei 430074)

**Abstract:** Anhydrovinblastine, which is a crucial intermediate in the synthesis of vinorelbine and vinflunine, can be synthesized from vinblastine, but the reported yields are not sufficiently high for practical application. The synthetic conditions for this reaction have been optimized using orthogonal tests and artificial neural networks. A network model was established by improving back-propagation of an artificial neural network (BP-ANN) with 3 layers, and was shown to be correct. A wider range of factors was then introduced into the BP-ANN model in order to obtain the optimal conditions as follows: the vinblastine was dissolved in DMF with a concentration of 15 mmol/L; the reagent was oxalyl chloride,  $C_2O_2Cl_2$ , with the stoichiometric proportion between the reagent and vinblastine being 50:1; the reaction temperature was 4 °C and the duration was 12 hours; the extractant phase was the mixture of ether and methylene chloride (volumen ratio 10). The maximum yield of anhydrovinblastine obtained under these conditions was 70.88%.

**Key words:** anhydrovinblastine; artificial neural network; synthesis; orthogonal tests