Department of Pharmacy[1], Zhejiang University City College, Hangzhou, First Affiliated Hospital of Zhejiang University[2], Hangzhou, Department of Chemistry[3], Zhejiang University, Hangzhou, College of Pharmaceutical Sciences[4], Zhejiang University, Hangzhou, Ningbo Institute of Technology[5], Zhejiang University, Ningbo, P.R. China

# Prediction of human intestinal absorption using an artificial neural network

X. C. Fu[1], C. X. Chen[2], G. P. Wang[3], W. Q. Liang[4], Q. S. Yu[5]

*Received October 25, 2004, accepted November 1, 2004*

*Prof. Xuchun Fu, Department of Pharmacy, Zhejiang University City College, Hangzhou, 310015, P.R. China*
*Fuxc@zucc.edu.cn*

*Pharmazie 60: 674–676 (2005)*

An artificial neural network model is developed to predict percent human intestinal absorption (%FA) of compounds from their molecular structural parameters. These parameters are the polar molecular surface area (PSA), the fraction of polar molecular surface area (FPSA, polar molecular surface area/molecular surface area), the sum of the net atomic charges of oxygen atoms ($Q_O$), the sum of the net atomic charges of nitrogen atoms with net negative atomic charges ($Q_N$), the sum of the net atomic charges of hydrogen atoms attached to oxygen or nitrogen atoms ($Q_H$), and the number of carboxyls ($n_{COOH}$). For a training set of 85 compounds and a test set of 10 compounds, root mean squared errors (RMSE) between experimental %FA values and calculated/predicted %FA values are 8.86% and 14.1%, respectively.

## 1. Introduction

The prediction of human intestinal absorption is important in the design, optimization, and selection of candidates for oral drugs. Researchers have therefore sought to delineate the physicochemical properties that favor intestinal absorption. Lipinski et al. (1997) put forward the famous "rule of five". The rule states that if a compound satisfies any two of the following rules, it is likely to exhibit poor intestinal absorption: Molecular weight >500, number of hydrogen bond donors >5, number of hydrogen bond acceptors >10, calculated partition coefficient (log P) >5. Zhao et al. (2001) utilized the general solvation equation developed by Abraham's group to predict the human intestinal absorption. Palm et al. (1997) and Clark (1999) derived a sigmoidal relationship between %FA and polar molecular surface area (PSA) from a data set of 20 compounds. In this paper, we report an artificial neural network model for the prediction of human intestinal absorption.

## 2. Investigations, results and discussion

The data set of 95 compounds (shown in the Table) consists of the Palm et al. set (1997), the Wessel et al. set (1998), and the Kansy et al. set (1998), in which methotrexate, zidovudine, cefatrizine, etoposide, cefuroxime axetil, lisinopril, and cephalexin are excluded because they can be absorbed by active transport (Clark 1999). These compounds are divided into a training set of 85 compounds and a test set of 10 compounds (Table).

The backpropagation algorithm with a modified learning rule, normalized cumulative delta is used to train the network. A tanh function is used as the transfer function. The neuronet model is a four-layer network that includes an input layer, two hidden layers, and an output layer. The initial learning coefficients are 0.3, 0.25, and 0.15 for the first hidden layer, the second hidden layer, and the output layer, respectively. The initial momentum is 0.4. Epoch size is 4. F′ offset is 0.1. Transition point is 10000 and learning coefficient ratio is 0.5. Inputs to the neural network consist of the polar molecular surface area (PSA), the fraction of polar molecular surface area (FPSA, polar molecular surface area/molecular surface area), the sum of the net atomic charges of oxygen atoms ($Q_O$), the sum of the net atomic charges of nitrogen atoms with net negative atomic charges ($Q_N$), the sum of the net atomic charges of hydrogen atoms attached to oxygen or nitrogen atoms ($Q_H$), and the number of carboxyls ($n_{COOH}$). The net atomic charges are obtained from the semiempirical self-consist-

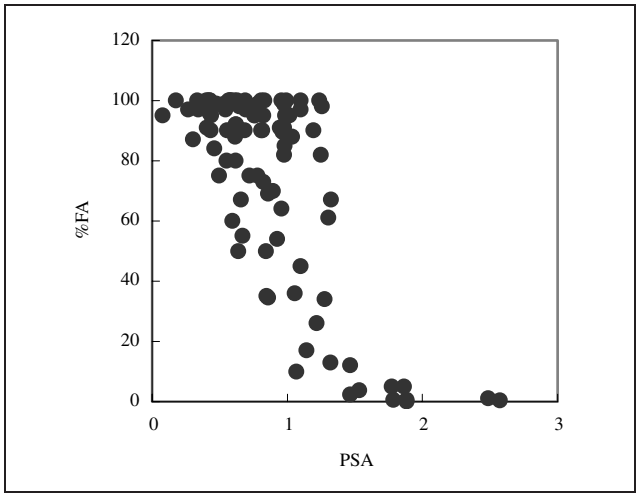


Fig.: Relationship between %FA of 95 compounds and PSA

**Table: %FA values of 95 compounds and their molecular structural parameters**

| Compound | PSA ($nm^2$) | FPSA | $Q_H$ | $Q_O$ | $Q_N$ | $n_{COOH}$ | %FA | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Exp.[a] | Calc.[b] |
| *Training set* | | | | | | | | |
| Metoprolol | 0.576 | 0.1460 | 0.3833 | −0.8356 | −0.2954 | 0 | 100 | 93.1 |
| Acetylsalicylic acid | 0.690 | 0.3064 | 0.2388 | −1.1719 | 0.0000 | 1 | 100 | 92.3 |
| Bumetanide | 1.236 | 0.2958 | 0.9632 | −2.6979 | −1.2712 | 1 | 100 | 99.4 |
| Corticosterone | 0.807 | 0.2039 | 0.4183 | −1.2366 | 0.0000 | 0 | 100 | 93.0 |
| Desipramine | 0.178 | 0.0499 | 0.1525 | 0.0000 | −0.5575 | 0 | 100 | 93.2 |
| Dexamethasone | 0.956 | 0.2355 | 0.6448 | −1.5633 | 0.0000 | 0 | 100 | 91.1 |
| Felodipine | 0.580 | 0.1368 | 0.0000 | −1.1341 | 0.0000 | 0 | 100 | 93.2 |
| Fluvastatin | 0.830 | 0.1717 | 0.7055 | −1.3648 | −0.1557 | 1 | 100 | 96.5 |
| Ibuprofen | 0.414 | 0.1410 | 0.2411 | −0.6740 | 0.0000 | 1 | 100 | 92.2 |
| Ketoprofen | 0.587 | 0.1869 | 0.2427 | −0.9911 | 0.0000 | 1 | 100 | 91.6 |
| Loracarbef | 1.100 | 0.2758 | 0.7729 | −0.8716 | −0.8546 | 1 | 100 | 101.5 |
| Lormetazepam | 0.568 | 0.1606 | 0.2285 | −0.6096 | −0.4526 | 0 | 100 | 93.1 |
| Ondansetron | 0.404 | 0.1135 | 0.0000 | −0.3191 | −0.4488 | 0 | 100 | 93.1 |
| Prazosin | 0.994 | 0.2255 | 0.4200 | −0.8964 | −1.4121 | 0 | 100 | 91.5 |
| Salicylic acid | 0.631 | 0.3716 | 0.5097 | −0.9579 | 0.0000 | 1 | 100 | 92.8 |
| Testosterone | 0.431 | 0.1225 | 0.1998 | −0.6200 | 0.0000 | 0 | 100 | 93.2 |
| Coumarin | 0.337 | 0.1812 | 0.0000 | −0.4760 | 0.0000 | 0 | 100 | 93.8 |
| Nordiazepam | 0.477 | 0.1550 | 0.2400 | −0.3318 | −0.4942 | 0 | 99 | 93.1 |
| Naproxen | 0.538 | 0.1848 | 0.2423 | −0.8825 | 0.0000 | 1 | 99 | 91.9 |
| Tiacrilast | 0.785 | 0.2661 | 0.2467 | −1.0112 | −0.4124 | 1 | 99 | 90.0 |
| Prednisolone | 0.978 | 0.2498 | 0.6345 | −1.5451 | 0.0000 | 0 | 98.8 | 91.9 |
| Cephalexin | 1.256 | 0.3210 | 0.7951 | −1.2051 | −0.9593 | 1 | 98 | 93.4 |
| Warfarin | 0.645 | 0.1792 | 0.2480 | −1.0475 | 0.0000 | 0 | 98 | 93.1 |
| Theophylline | 0.766 | 0.3598 | 0.2781 | −0.7049 | −0.8632 | 0 | 98 | 92.8 |
| Diazepam | 0.343 | 0.1049 | 0.0000 | −0.3280 | −0.4571 | 0 | 97 | 93.2 |
| Oxprenolol | 0.543 | 0.1419 | 0.3894 | −0.7316 | −0.3012 | 0 | 97 | 93.1 |
| Phenazone | 0.270 | 0.1102 | 0.0000 | −0.3443 | −0.3810 | 0 | 97 | 93.2 |
| Oxazepam | 0.695 | 0.2210 | 0.4735 | −0.6141 | −0.5234 | 0 | 97 | 92.9 |
| Trimethoprim | 1.099 | 0.3138 | 0.8823 | −0.5931 | −1.2116 | 0 | 97 | 81.6 |
| Practolol | 0.762 | 0.2064 | 0.6104 | −0.9085 | −0.6261 | 0 | 95 | 91.5 |
| Clonidine | 0.437 | 0.1675 | 0.4182 | 0.0000 | −0.7733 | 0 | 95 | 93.0 |
| Fluconazole | 0.825 | 0.2478 | 0.2102 | −0.3037 | −0.8090 | 0 | 95 | 92.7 |
| Imipramine | 0.078 | 0.0209 | 0.0000 | 0.0000 | −0.5047 | 0 | 95 | 93.2 |
| Labetalol | 1.017 | 0.2443 | 1.0570 | −0.9875 | −0.7218 | 0 | 95 | 99.8 |
| Sotalol | 0.984 | 0.2521 | 0.6261 | −2.4135 | −1.1249 | 0 | 95 | 83.9 |
| Verapamil | 0.756 | 0.1379 | 0.1525 | −0.8619 | −0.3583 | 0 | 95 | 93.0 |
| Diltiazem | 0.623 | 0.1254 | 0.0000 | −1.0825 | −0.6026 | 0 | 92 | 93.1 |
| Hydrocortisone | 0.945 | 0.2404 | 0.6289 | −1.5471 | 0.0000 | 0 | 91 | 91.7 |
| Progesterone | 0.406 | 0.1058 | 0.0000 | −0.5904 | 0.0000 | 0 | 91 | 93.2 |
| Terazosin | 0.976 | 0.2158 | 0.4444 | −1.0726 | −1.4688 | 0 | 91 | 91.3 |
| Betaxolol | 0.555 | 0.1251 | 0.3857 | −0.8376 | −0.2990 | 0 | 90 | 93.1 |
| Chloramphenicol | 1.193 | 0.3519 | 0.6516 | −1.6913 | −0.3390 | 0 | 90 | 89.9 |
| Phenytoin | 0.686 | 0.2343 | 0.5331 | −0.5396 | −0.6391 | 0 | 90 | 92.8 |
| Scopolamine | 0.614 | 0.1566 | 0.2163 | −1.2304 | −0.2592 | 0 | 90 | 93.1 |
| Tenidap | 0.817 | 0.2639 | 0.7726 | −0.8992 | −0.6443 | 0 | 90 | 74.3 |
| Timolol maleate | 0.807 | 0.2042 | 0.3552 | −0.7603 | −1.1324 | 0 | 90 | 92.7 |
| Propranolol | 0.433 | 0.1206 | 0.3845 | −0.5606 | −0.2963 | 0 | 90 | 93.1 |
| Acebutolol | 0.967 | 0.2087 | 0.6254 | −1.1617 | −0.6331 | 0 | 89.5 | 86.6 |
| Acrivastine | 0.616 | 0.1330 | 0.2454 | −0.6796 | −0.3505 | 1 | 88 | 89.3 |
| Trovafloxacin | 1.037 | 0.2412 | 0.5567 | −0.9639 | −1.0081 | 1 | 88 | 85.3 |
| Bupropion | 0.305 | 0.0954 | 0.1764 | −0.2813 | −0.2884 | 0 | 87 | 93.2 |
| Cimetidine | 0.981 | 0.2960 | 0.6672 | 0.0000 | −1.3045 | 0 | 85 | 86.1 |
| Bromazepam | 0.461 | 0.1182 | 0.2089 | −0.5473 | −0.2995 | 0 | 84 | 93.1 |
| Sorivudine | 1.251 | 0.3904 | 0.9487 | −1.9298 | −0.6329 | 0 | 82 | 84.4 |
| Acetaminophen | 0.553 | 0.2706 | 0.4448 | −0.6070 | −0.3259 | 0 | 80 | 93.1 |
| Quinidine | 0.617 | 0.1973 | 0.2395 | −0.3311 | −0.5862 | 0 | 80 | 93.0 |
| Guanabenz | 0.782 | 0.3136 | 0.8086 | 0.0000 | −1.0739 | 0 | 75 | 76.0 |
| Propylthiouracil | 0.498 | 0.2214 | 0.5544 | −0.3235 | −0.5900 | 0 | 75 | 92.9 |
| Terbutaline | 0.823 | 0.2721 | 0.7940 | −0.8180 | −0.2933 | 0 | 73 | 75.1 |
| Lamotrigine | 0.895 | 0.3572 | 0.9758 | 0.0000 | −1.1644 | 0 | 70 | 83.6 |
| Ciprofloxacin | 0.860 | 0.2339 | 0.3954 | −0.9672 | −0.6644 | 1 | 69 | 50.0 |
| Captopril | 0.659 | 0.2377 | 0.2426 | −1.0202 | −0.3313 | 1 | 67 | 91.4 |
| Hydrochlorothiazide | 1.325 | 0.5016 | 1.0001 | −3.7100 | −2.2012 | 0 | 67 | 66.5 |
| Furosemide | 1.304 | 0.4105 | 1.0110 | −2.6965 | −1.2693 | 1 | 61 | 60.9 |
| Ziprasidone | 0.594 | 0.1357 | 0.2557 | −0.3217 | −1.1537 | 0 | 60 | 92.9 |
| Tranexamic acid | 0.672 | 0.3144 | 0.5327 | −0.6860 | −0.3546 | 1 | 55 | 62.0 |

**Table: (continued)**

| Compound | PSA ($nm^2$) | FPSA | $Q_H$ | $Q_O$ | $Q_N$ | $n_{COOH}$ | %FA | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Exp.[a] | Calc.[b] |
| Atenolol | 0.926 | 0.2529 | 0.8311 | −0.9353 | −0.7338 | 0 | 54 | 48.3 |
| Gabapentin | 0.638 | 0.2762 | 0.5598 | −0.6918 | −0.3720 | 1 | 50 | 43.7 |
| Ranitidine | 0.843 | 0.2029 | 0.4482 | −0.9067 | −0.8886 | 0 | 50 | 92.6 |
| Phenoxymethylpenicillin | 1.101 | 0.2749 | 0.5136 | −1.4464 | −0.6051 | 1 | 45 | 50.6 |
| Sulpiride | 1.056 | 0.2642 | 0.7058 | −2.4474 | −1.5517 | 0 | 36 | 39.7 |
| Norfloxacin | 0.848 | 0.2368 | 0.3946 | −0.9719 | −0.7034 | 1 | 35 | 52.5 |
| Nadolol | 0.861 | 0.2111 | 0.8018 | −1.2005 | −0.2865 | 0 | 34.5 | 47.2 |
| Mannitol | 1.220 | 0.5469 | 1.3172 | −2.0080 | 0.0000 | 0 | 26 | 26.2 |
| Foscarnet | 1.142 | 0.8961 | 0.8232 | −3.2312 | 0.0000 | 1 | 17 | 16.5 |
| Sulfasalazine | 1.466 | 0.3424 | 0.7163 | −2.6862 | −1.1294 | 1 | 12 | 12.7 |
| Enalaprilat | 1.070 | 0.2513 | 0.7009 | −1.7234 | −0.6043 | 2 | 10 | 12.1 |
| Doxorubicin | 1.863 | 0.3481 | 1.4447 | −2.8787 | −0.3282 | 0 | 5 | 0.4 |
| Ganciclovir | 1.532 | 0.5200 | 1.1043 | −1.2776 | −1.1426 | 0 | 3.8 | 6.0 |
| Olsalazine | 1.465 | 0.4519 | 0.9276 | −1.7362 | −0.1319 | 2 | 2.3 | 0.6 |
| Ceftriaxone | 2.486 | 0.4475 | 1.1131 | −1.8219 | −1.6107 | 1 | 1 | 0.2 |
| Lactulose | 1.886 | 0.5240 | 1.7767 | −3.3840 | 0.0000 | 0 | 0.6 | 0.5 |
| Cromolyn | 1.786 | 0.3739 | 0.7222 | −2.7405 | 0.0000 | 2 | 0.5 | 0.5 |
| Raffinose | 2.574 | 0.5102 | 2.4057 | −4.8687 | 0.0000 | 0 | 0.3 | 0.2 |
| Gentamicin | 1.886 | 0.3379 | 1.9066 | −1.9603 | −1.6323 | 0 | 0 | 2.9 |
| *Test set* | | | | | | | | |
| Caffeine | 0.610 | 0.2553 | 0.0000 | −0.7128 | −0.8156 | 0 | 100 | 93.0 |
| Valproic acid | 0.427 | 0.1827 | 0.2417 | −0.6823 | 0.0000 | 1 | 100 | 92.4 |
| Alprenolol | 0.426 | 0.1153 | 0.3665 | −0.5566 | −0.3055 | 0 | 96 | 93.1 |
| Pindolol | 0.620 | 0.1811 | 0.6372 | −0.5542 | −0.5110 | 0 | 92 | 91.5 |
| Methylprednisolone | 0.977 | 0.2397 | 0.6334 | −1.5443 | 0.0000 | 0 | 82 | 91.6 |
| Sumatriptan | 0.723 | 0.1920 | 0.4847 | −1.8561 | −1.3873 | 0 | 75 | 92.3 |
| Metolazone | 0.958 | 0.2546 | 0.7243 | −2.1821 | −1.5757 | 0 | 64 | 44.8 |
| Pravastatin | 1.276 | 0.2401 | 0.8948 | −2.2986 | 0.0000 | 1 | 34 | 9.0 |
| Chlorothiazide | 1.321 | 0.5140 | 0.7593 | −3.6774 | −1.9979 | 0 | 13 | 11.0 |
| Cefuroxime | 1.775 | 0.4008 | 0.9902 | −2.1140 | −1.0577 | 1 | 5 | 0.3 |

a From reference (Clark 1999)
b From neural network model

ent field molecular orbital calculation AM1 method (Dewar et al. 1985). The atomic radii used to calculate molecular surface area are those used by Clark (1999). The two hidden layers consist of eight and three neurons, respectively. The output layer consists of a single neuron which is the percent human intestinal absorption (%FA). This network is a 6-8-3-1 architecture.

The calculated/predicted %FA values from the neural network model obtained after 100000 training cycles are listed in the Table. They are in good accordance with respective experimental %FA values. The root mean squared errors (RMSE) between experimental %FA values and calculated/predicted %FA values are 8.86% and 14.1% for the training set of 85 compounds and the test set of 10 compounds, respectively.

Although there is an excellent sigmoidal relationship between %FA and the polar molecular surface area for a data set of 20 compounds (Palm et al. 1997; Clark 1999), as shown in the Fig., the sigmoidal relationship is not obvious for the larger data set of 95 compounds. Our neural network model displays better predictive ability than the sigmoid model. Because the predictors in the neural network model can be easily calculated, it is convenient to predict human intestinal absorption of drug candidates using the model.

**References**


Clark DE (1999) Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 1. Prediction of intestinal absorption. J Pharm Sci 88: 807–814.

Dewar MJS, Zoebisch GE, Healy EF, Stewart JJP (1985) AM1: A new general purpose quamtum mechanical molecular model. J Am Chem Soc 107: 3902–3909.

Kansy M, Senner F, Gubernator K (1998) Physicochemical high throughput screening: parallel artificial membrane permeation assay in the description of passive absorption processes. J Med Chem 41: 1007–1010.

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv 23: 3–25.

Palm K, Stenberg P, Luthman K, Artursson P (1997) Polar molecular surface properties predict the intestinal absorption of drugs in humans. Pharm Res 14: 568–571.

Wessel MD, Jurs PC, Tolan JW, Muskal SM (1998) Predition of human intestinal of drug compounds from molecular structure. J Chem Inf Comput Sci 38: 726–735.

Zhao YH, Abraham MH, Hersey A, Eddershaw PJ, Luscombe CN, Boutina D, Beck G, Sherborne B, Cooper I, Platts J (2001) Evaluation of human intestinal absorption data and subsequent derivation of a quantitative structure–activity relationship (QSAR) with the Abraham descriptors. J Pharm Sci 90: 749–784.