Quantitative Structure–Pharmacokinetic Relationships for Drug Distribution Properties by Using General Regression Neural Network

C.W. YAP, Y.Z. CHEN

Department of Computational Science, National University of Singapore, Blk SOC1, Level 7, 3 Science Drive 2, Singapore 117543

Received 28 January 2004; revised 16 August 2004; accepted 26 August 2004

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/jps.20232

ABSTRACT: Quantitative Structure–Pharmacokinetic Relationships (QSPkR) have increasingly been used for developing models for the prediction of the pharmacokinetic properties of drug leads. QSPkR models are primarily developed by means of statistical methods such as multiple linear regression (MLR). These methods often explore a linear relationship between the pharmacokinetic property of interest and the structural and physicochemical properties of the studied compounds, which are not applicable to those agents with nonlinear relationships. Hence, statistical methods capable of modeling nonlinear relationships need to be developed. In this work, a relatively new kind of nonlinear method, general regression neural network (GRNN), was explored for modeling three drug distribution properties based on diverse sets of drugs. The three properties are blood–brain barrier penetration, binding to human serum albumin, and milk–plasma distribution. The prediction capability of GRNN-developed models was compared to those developed using MLR and a nonlinear multilayer feedforward neural network (MLFN) method. For blood–brain barrier penetration, the computed r^2 and MSE values of the GRNN-, MLR-, and MLFN-developed models are 0.701 and 0.130, 0.649 and 0.154, and 0.662 and 0.147, respectively, by using an independent validation set. The corresponding values for human serum albumin binding are 0.851 and 0.041, 0.770 and 0.079, and 0.749 and 0.089, respectively, and that for milk–plasma distribution are 0.677 and 0.206, 0.224 and 0.647, and 0.201 and 0.587, respectively. These suggest that GRNN is potentially useful for predicting QSPkR properties of chemical agents.
 J Pharm Sci 94:153–168, 2005
Keywords: blood–brain barrier; computational ADME; distribution; general regression neural network; human serum albumin binding; milk–plasma distribution; pharmacokinetics

INTRODUCTION

Optimization of pharmacokinetic as well as the pharmacodynamic properties of a drug candidate is an important consideration in the drug design process.[1,2] One important aspect of pharmacokinetic properties of a drug candidate is its distribution in the human body. A drug is required to achieve sufficient concentration at the target site while possibly limiting its distribution elsewhere so as to produce desired therapeutic action with minimum side effects.[3] Traditionally, the distribution properties of a drug candidate are obtained via *in vivo* and *in vitro* studies, which tend to be time-consuming and costly. Therefore, an *in silico* method, Quantitative Structure–Pharmacokinetic Relationship (QSPkR) modeling, has recently been explored for predicting the distribution properties of drug candidates[4] in an effort to

Correspondence to: Y.Z. Chen (Telephone: 65-6874-6877; Fax: 65-6774-6756; E-mail: yzchen@cz3.nus.edu.sg)

Journal of Pharmaceutical Sciences, Vol. 94, 153–168 (2005)

eliminate undesirable agents in a fast and cost-effective manner.

The most common modeling methods for obtaining QSPkR models are linear methods such as multiple linear regression (MLR).[5] These methods can be easily used, and the derived models can be easily interpreted. However, multiple mechanisms may be involved in determining a particular pharmacokinetic property. A variety of factors may interact in complex ways to affect the pharmacokinetic property of a compound. Therefore, methods based only on linear relationships may not always be the most efficient approach for constructing a QSPkR model. Thus, nonlinear methods such as multilayer feed-forward neural networks (MLFN)[6] and general regression neural network (GRNN)[7] have increasingly been used for construction of QSPkR models.

GRNN has been explored for QSPkR modeling of human intestinal absorption[8] as well as for developing Quantitative Structure–Activity Relationships (QSAR) and Quantitative Structure–Property Relationships (QSPR) of chemical agents.[9] The prediction capability of GRNN has been found to be comparable to those of conventional nonlinear methods such as MLFN, but the former requires fewer descriptors.[9] Thus, GRNN is expected to be equally useful for developing QSPkR models of other pharmacokinetic properties. This work is intended to test this feasibility by applying GRNN for developing QSPkR models of three distribution properties, blood–brain barrier (BBB) penetration, binding to human serum albumin (HSA), and milk–plasma (M/P) distribution. The performances of the GRNN-developed models were compared with those developed by using MLR and MLFN to determine whether GRNN produces more predictive QSPkR models.

The BBB exists at the choroids plexus and at the tissue capillary membranes between the blood and brain fluid, and BBB penetration is necessary for CNS drugs.[10] Examples of these drugs are antipsychotics, antiepileptics, and antidepressants. For drugs not directed at targets in the brain, BBB penetration is undesirable because of potential CNS-related side effects. For example, the first-generation antihistamines are known to penetrate the BBB leading to drowsiness.[11] The second-generation antihistamines have a significantly reduced BBB penetration capability, and are thus less likely to cause drowsiness.[12] One method for assessing the effects of a compound in the brain is to determine its concentration in the brain. This concentration can be calculated from the brain–blood (BB) ratio, which is the concentration of this compound in the brain divided by that in the blood. Thus, the BB ratio is an important pharmacokinetic property and a number of QSPkR models of BB ratio have been developed,[13–27] the majority of which were developed by using MLR and the computed r^2 values are in the range between 0.723–0.941.

Most drugs bind to serum proteins, and such binding regulates drug distribution and subsequently its effect.[28] Albumin is the most abundant of all serum proteins, and is the most common drug-binding protein in the circulatory system. Because of the important role of albumin binding in regulation of drug distribution, QSPkR models for predicting the extent of albumin binding have been developed,[29–33] the majority of which were developed by using MLR and a congeneric series of compounds. In a study of a diverse set of 94 drugs and drug-like compounds, two QSPkR models developed by using MLR gave computed r^2 values of 0.88 and 0.82, respectively, on a separate testing set.[31]

Breast milk is the best form of nutrition available to a newborn infant. Certain drugs administered to a nursing mother may be distributed into breast milk and thus transferred into the infant. The concentration of drug present in the breast milk can be used as an indicator of breast feed risk. The ratio of drug concentration in milk and plasma (M/P ratio) is the most widely used quantity for describing drug concentration in breast milk.[34] However, the M/P ratio is seldom determined during clinical trials or after the drug has entered the market. In addition, M/P ratios were often obtained from studies involving a small number of women. This may lead to significant variations in the reported M/P ratio for a drug, and makes it difficult for clinicians to advise women on the safety of breast feeding. Methods for estimating the M/P ratios of drugs have been developed by using various modeling methods.[35–40] In a recent study,[40] MLFN was used to train and test on 123 diverse compounds. The computed r^2 and mean square error (MSE) values from this model are 0.61 and 0.814, respectively.

MATERIALS AND METHODS

GRNN Algorithm

GRNN was introduced by Specht in 1991,[7] and is a form of neural network designed for