and $s = 0.50$. Abraham[51] achieved similar results modeling literature data on logPS for 30 molecules using a linear regression model fit to five Abraham descriptors, with $r^2 = 0.87$ and $s = 0.52$.

P-GLYCOPROTEIN EFFLUX

P-glycoprotein is an ABC cassette transporter encoded by the *MDR1* gene in humans that is responsible for the efflux of drugs from cells. It plays a significant role in limiting brain penetration and to a lesser extent limits intestinal absorption of drugs. For oral drugs dosed in quantities greater than 50 mg with reasonable dissolution rates, P-glycoprotein transport will be saturated and thus unable to limit absorption. It should be noted that drugs with poor solubility effectively have a "low dose" and may have limited absorption due to P-glycoprotein efflux (e.g., paclitaxel). Unfortunately, the blood concentrations of drugs at the BBB do not achieve the levels found for most drugs in the intestines. Thus, P-glycoprotein transporters at the BBB cannot be saturated and will decrease the brain penetration of substrates.[52,53]

In a study of P-glycoprotein substrates versus nonsubstrates, Varma et al.[54] concluded that substrate molecules with high passive permeability overwhelmed the transporter while substrate molecules with moderate passive permeability were more affected by P-glycoprotein. Approximately half of 63 P-glycoprotein substrates studied had MW > 400 and PSA > 75, indicating that larger, more polar molecules are more likely to be P-glycoprotein substrates.

Several QSAR models have been used to predict whether a molecule is a P-glycoprotein substrate. Gombar et al.[55] modeled a set of 95 P-glycoprotein substrates and nonsubstrates using stepwise linear discriminant analysis. Class assignment was based on efflux ratios measured by an in vitro Madin–Darby canine kidney cell assay run at GlaxoSmithKline. The initial 254 descriptors were trimmed down to a set of 27 descriptors with an accuracy of 98.9%. Performance on a test set was also good, with 50/58 (86.2%) correctly predicted. A single E-state descriptor, MolES, representing molecular bulk, was particularly good at discriminating substrates. For MolES > 110, eighteen of nineteen molecules were substrates, and for MolES < 49, eleven of thirteen molecules were nonsubstrates.

Cabrera et al.[56] modeled a set of 163 drugs using topological substructural molecular design (TOPS-MODE) descriptors with a linear discriminant model to predict P-glycoprotein efflux. Model accuracy was 81% for the training set and 77.5% for a validation set of 40 molecules. A "combinatorial QSAR" approach was used by de Lima et al.[57] to test multiple model types (kNN, decision tree, binary QSAR, SVM) with multiple descriptor sets from various software packages (MolconnZ, Atom Pair, VoSurf, MOE) for the prediction of P-glycoprotein substrates for a data set of 192 molecules. Best overall performance on a test set of 51 molecules was achieved with an SVM and AP or VolSurf descriptors (81% accuracy each).

Analyses of molecules that are P-glycoprotein substrates have suggested a number of possible pharmacophores. For example, based on an analysis of 100 molecules, Seelig[58] proposed that molecules containing at least one Type I or Type II unit would be P-glycoprotein substrates, and their binding increases with the strength and number of these groups. Type I units contain two electron donor groups 2.5 ± 0.3Å apart, and Type II units contain two or three electron donor groups whose maximum distance apart is 4.6 ± 0.6Å. Pajeva and Wiese[59] proposed a pharmacophore containing two hydrophobic groups, three HBA groups, and one HBD group. They conclude that binding depends on the number of these pharmacophore points present and that different drugs interact with varied groups with multiple possible binding modes. This pharmacophore hypothesis was shown to agree with a homology model of P-glycoprotein created using *Escherichia coli MsbA* as the template.[60]

Two 3D QSAR models were built using GRIND descriptors for P-glycoprotein substrate recognition. Cianchetta et al.[61] selected 100 proprietary molecules and 29 publicly available molecules having Caco-2 A-B/B-A ratios > 1 and screened them for inhibition of P-glycoprotein activity in a calcein-AM assay. The inhibition values were modeled using GRIND and VolSurf descriptors. The 3D alignment independent GRIND descriptors fit the data well, with $r^2 = 0.83$. VolSurf descriptors produced a model that was slightly better than random. The pharmacophoric GRIND features suggested the following features were important for P-glycoprotein substrate recognition: two hydrophobic groups 16.5Å apart, two HBA groups 11.5Å apart, plus the size of the molecule (21.5Å distance required between edges of the molecule). Crivori et al.[62] similarly compared VolSurf and GRIND descriptors for the prediction of P-glycoprotein substrates. Fifty-three drugs were classified as substrates or nonsubstrates by a cutoff of two for their Caco-2 efflux ratio and modeled using VolSurf descriptors; the model was 89% accurate. When tested on a proprietary data set of 272 molecules, the VolSurf model correctly classified 72% of the data set. Thirty of the 53 drugs were assayed in a calcein-AM assay and the data were used to select nine substrates and fourteen nonsubstrates for modeling with GRIND descriptors. The model was tested on a set of 125 drugs from the literature and accurately predicted 82% of them. Two GRIND features were important in the model: two hydrophobic regions 11.5Å apart and two HBA groups 8Å apart.

The effect of P-glycoprotein efflux limiting brain penetration has been examined by two analyses. A bagged recursive-partitioning model was built using the R software on 190 compounds with literature logBB data and three sets of descriptors.[63] The literature-based model was tested on 250 Pfizer compounds, of which approximately 60% showed significant P-glycoprotein mediated efflux based