QSAR Study and VolSurf Characterization of Human Intestinal Absorption of Drugs

HU, Gui-Xiang(胡桂香) SHANG, Zhi-Cai*(商志才) ZOU, Jian-Wei(邹建卫)
YANG, Guo-Ming(杨郭明) YU, Qing-Sen(俞庆森)
Department of Chemistry, Zhejiang University, Hangzhou, Zhejiang 310027, China

The prediction of human intestinal absorption is a major goal in the design, optimization, and selection of candidates for the development of oral drugs. In this study, a computerized method (VolSurf with GRID) was used as a novel tool for predicting human intestinal absorption of test compound, and for determining the critical molecular properties needed for human intestinal absorption. The tested molecules consisted of 20 diverse drug-like compounds. Partial least squares (PLS) discriminant analysis was used to correlate the experimental data with the theoretical molecular properties of human intestinal absorption. A good correlation ($r^2 = 0.95$, $q^2 = 0.86$) between the molecular modeling results and the experimental data demonstrated that human intestinal absorption could be predicted from the three-dimensional (3D) molecular structure of a compound. Favorable structural properties identified for the potent intestinal absorption of drugs included strong imbalance between the center of mass of a molecule and the barycentre of its hydrophilic and hydrophobic regions and a definitive hydrophobic region as well as less hydrogen bonding donors and acceptors in the molecule.

Keywords human intestinal absorption, QSAR, VolSurf, principal component analysis (PCA), partial least squares (PLS)

Introduction

Oral administration is the most convenient and cost-effective way to administer drugs. Apart from dose and transit time, the main factors determining the extent of absorption are solubility and membrane permeability. Drug-discovery programs have traditionally focused on optimization of molecules with regard to potency and selectivity for the target receptor. But the failure rate of drug candidates by this way is significant in late stage of development for ignoring absorption, distribution, metabolism and excretion (ADME) properties of drugs. So we need predictive tools that can eliminate inappropriate compounds before substantial time and money are invested in testing. Physicochemical properties such as lipophilicity and hydrogen bonding are generally used to predict oral absorption.[1,2] However, the good model for structurally homogeneous data sets with these properties is not suitable for structural diversity.[3,4] So there is a great need for rapid and efficient computational methods capable of differentiating drugs with acceptable absorption at early stage in the drug development.

In 1997, Palm *et al*.[5] reported that calculated dynamic polar molecular surface area gave an excellent sigmoidal relationship with the fraction absorbed following oral administration of structurally diverse 20 drugs to humans. In 1999, Norinder *et al*.[6] reported the theoretical calculation and prediction of intestinal absorption of drugs in humans using MolSurf parametrization and PLS statistics. The data set consisted of 20 structurally diverse molecules and the relationship between experimental and calculated absorption gave an r^2 of 0.92 with a q^2 of 0.80. MolSurf calculates molecular descriptors related to physicochemical properties such as lipophilicity, polarity, polarizability, hydrogen bonding, *etc*. But the computational requirements are prohibitive for even medium-sized data sets, as it uses electronic wavefunction to calculate physicochemically relevant descriptors.[7]

Recently, a novel computational method, called VolSurf, has been developed for the modeling and prediction of physicochemical and pharmacodynamic properties of a compound.[7-12] VolSurf reads or computes 3D molecular interaction fields and uses image processing methods to convert them into simple molecular descriptors that are easy to understand and interpret. These descriptors quantitatively characterize size, shape, polarity and hydrophobicity of molecules, and the balance between them. In this paper we describe the use of this approach for modeling human intestinal absorption following oral administration of drugs.

Computational methods

Human intestinal absorption data

The experimental values of the human intestinal absorption data for the 20 data set compounds were taken from Palm *et al*.[5] and are given in Table 1. The data set covers a wide range of absorption as well as physicochemical properties. The selected drugs are predominantly absorbed by a passive process and complication factors such as low solubility and first-pass metabolism were either negligible or taken into con-

* E-mail: shangzc@mail.hz.zj.cn
Received May 31, 2002; revised August 23, 2002; accepted November 15, 2002.
Project supported by the National Natural Science Foundation of China (No. 20173050).

sideration in the calculation of the percentage absorption. The structures of drugs were gained from Merck index.

Table 1 Name, experimental, calculated and predicted human intestinal absorption of drugs

No.	Compound	MW	Exp.[a]	PLS model[b]		
				1		2
				Calc.[c]	Pred.[d]	Calc.[c]
1	Diazepam	285	97	96.2		99.5
2	Metoprolol	267	102	91.6		92.5
3	Lactulose	342	0.6		0.1	3.5
4	Nordiazepam	271	99		95.3	100.6
5	Oxprenolol	265	97		93.4	91.4
6	Practolol	266	95	89.3		85.3
7	Alprenolol	249	96	110.5		110.7
8	Oxazepam	287	97	88.0		93.2
9	Pindolol	248	92	79.7		80.4
10	Metolazone	366	64		53.2	53.5
11	Atenolol	266	54		69.7	62.1
12	Sulpiride	341	36	43.1		45.2
13	Foscarnet	126	17	16.5		20.4
14	Mannitol	182	26		11.5	19.9
15	Raffinose	504	0.3	-0.3		-1.6
16	Tranexamic acid	157	55	65.2		67.1
17	Sulfasalazine	398	12	4.4		1.9
18	Olsalazine	302	2.3	6.9		11.4
19	Ciprofloxacin	331	69	78.0		80.2
20	Phenazone	188	97		80.7	88.9

[a] The percentage absorbed after oral administration to humans taken from Palm *et al*.[5] [b] Two PLS models: model 1 and model 2. [c] Calculated values for two models. [d] Predicted values for the test set on the base of the training set.

Computational approach

The models calculated by VolSurf (version 3.04) are relatively insensitive to conformational sampling and do not require molecular alignment.[7] So conformational analysis was not performed on all the drugs. The structures were modeled in their neutral forms and further minimized with TRIPOS force field in Sybyl 6.8. The molecular descriptors were derived by using the VolSurf (version 3.04)/GRID program, which is a computational procedure for producing and exploring the physicochemical property space of a molecule, starting from 3D interaction energies grid maps between the target molecule and different chemical probes. In our study 3D-maps generated by probes of water OH2, hydrophobic DRY and carbonyl oxygen O were used. As a result, 72 chemical descriptors were produced. These molecular descriptors had clear chemical meanings, referring to, for example, molecular volume (V), shape (S), surface rugosity (R), size of the hydrophilic (W1—W8) and hydrophobic (D1—D8) regions, and hydrogen-bonding properties (HB1—HB8). Amphiphilic moments (A) are defined as a vector pointing from the center of the hydrophobic domain to the center of the hydrophilic domain. The vector length is proportional to the strength of the amphiphilic moment. A critical packing parameter (CP) described the ratio between the D and W part of a molecule. Other useful descriptors were integy (INTEraction enerGY) moments, measuring the imbalance between the center of mass of a molecule and the barycentre of its hydrophilic (Iw1—Iw8) and hydrophobic (ID1—ID8) regions, and capacity factors (Cw1—Cw8) representing the ratio between the hydrophilic regions and the molecular surface. Additionally, the best three local minima of interaction energies (Emin1, Emin2, Emin3) and the distances (d12, d13, d23) between them were determined by a water probe and were calculated by VolSurf. The number (1—8) after the abbreviation of the specific descriptor describes the energy level used for statistical analysis and the strength of the interaction energy between molecule and probe. A more detailed representation of VolSurf descriptors has been presented by Cruciani *et al*.[9]

Chemometric tools, principal component analysis (PCA) and partial least squares (PLS) analysis, were used to compile the information obtained from VolSurf. PCA summarized information contained in the X-matrix (*i.e.*, the molecular properties of the tested compounds) by a few principal components (PCs). PLS analysis, on the other hand, was used to correlate the experimental data (Y-matrix; human intestinal absorption of the compound) with the X-matrix. In PLS modeling, latent variables (LVs), which are quite equivalent to the PC in PCA, were used as linear combinations of the original X-variables. The significant LVs' number was determined by cross-validation. PLS modeling gives valuable information about molecular properties that are critical for high human intestinal absorption. Two analyses were performed: one of the 13 training set compounds (model 1) and one of the whole data set of 20 structures (model 2). The quality (*i.e.*, the explanatory power, r^2) and the predictive ability (*i.e.*, q^2) of the PLS model were evaluated by leave-one-out cross-validation.

Results and discussion

PCA

PCA on the autoscaled matrix resulted in three principal components which explained 68.3% of the variance of the data for all drugs (model 2). The first to the third components explained 39.6%, 19.3%, 9.4%, respectively. The first two PCs were the most informative. In fact, the first PC of PCA distinguished between the poor absorption (%Abs lower than 60) and the efficient absorption (%Abs higher than 60) except for **10**, **11** and **16** compounds that have the medium absorption 64, 54 and 55, respectively (Fig. 1).

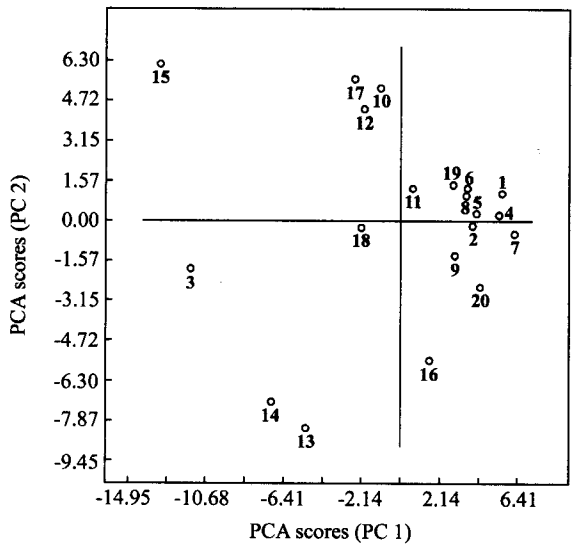

Fig. 1 PCA score plot for the compounds reported in Table 1. Numbers are those in Table 1. The first PC clusters the compounds according to human intestinal absorption of drugs.

PLS

First, the quality and predictive power of the PLS model was tested using 20 compounds (model 2). The relationship between experimental absorption and calculated absorption gave an r^2 of 0.95 with a q^2 of 0.86 (Fig. 2). The root mean squared error (RMSE) was 8.4% in the percent absorption.

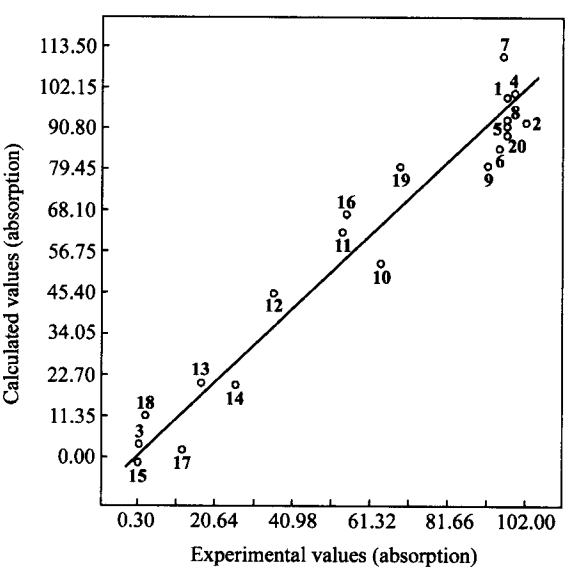

Fig. 2 Relationship between experimental and calculated human intestinal absorption. Compounds are numbered as in Table 1.

In order to explain the predictivity of the model, a training set consisting of 13 molecules was selected using the Most Descriptive Compounds (MDC)[13] method by PLS model of model 2. The quality and predictive power of the PLS model (model 1) was tested by the training set. The result showed that r^2 was 0.95 and q^2 was 0.82. RMSE was 8.3%. The model satisfactorily predicted the human intestinal absorption of the remaining seven compounds. RMSE of prediction was 11.1% (Fig. 3). The distribution of training set in 2D PLS score plot was given in Fig. 4.

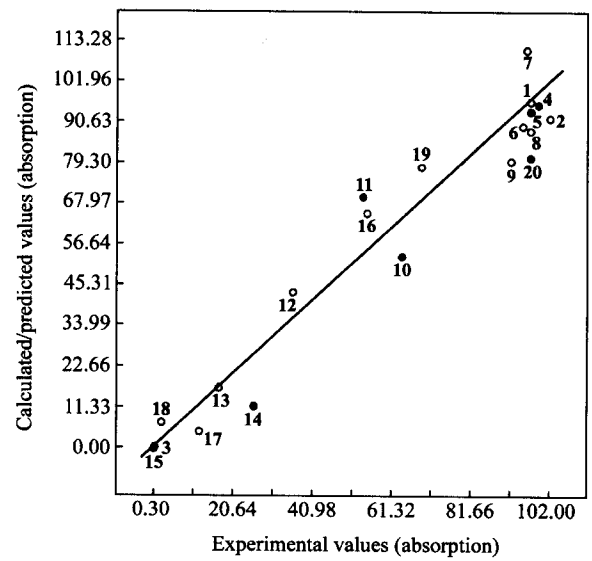

Fig. 3 Relationship between experimental and calculated (training set: ○) or predicted (test set: ●) human intestinal absorption (PLS model 1). Compounds are numbered as in Table 1.

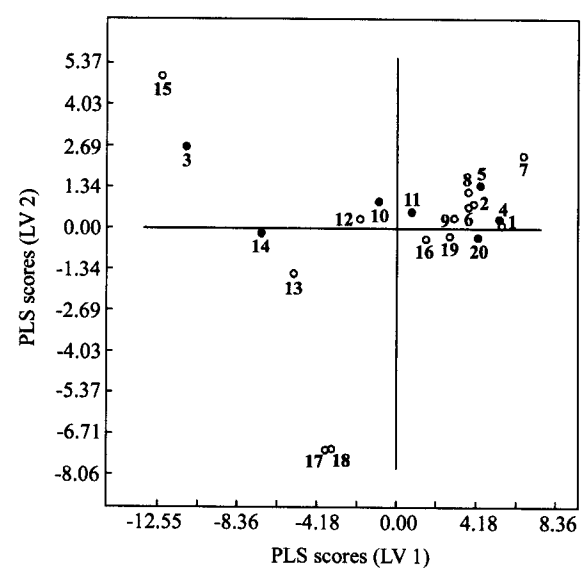

Fig. 4 PLS score plot with two LVs (LV 1 vs. LV 2) for the compounds reported in Table 1. ○ represents training set compounds, and ● represents test set compounds.

The two PLS analyses resulted in three significant PLS components according to cross-validation. Both models seem to be predictive from both an internal and an external point of view with respect to predictions of the test set in the latter case (model 1). The training set model is well balanced with respect to internal and external predictivity, which is indicated by similar RMSE values of training set (8.3%) and test

set (11.1%).

Most drugs are ionized at physiological pH. In our present study, twelve molecules are ionized at intestinal physiological pH 7.4, seven positive ions, three negative ions and two zwitterions. Abraham *et al*.[14] have shown that for strong Brϕnsted acids and bases, the rate constant for absorption of ionic species is close to that for absorption of the corresponding neutral species, so that the intestinal absorption can be calculated approximately from properties of the neutral species. Zhao *et al*.[15] have concluded that any ionization of Brϕnsted acids and based has a very small effect indeed (3%) on the percent absorption. Furthermore, these molecules must translate into neutral forms before transport across the membrane since it is generally accepted that it is the nonionized and desolvated molecules that penetrate the membrane. The deprotonation of amine salts and the protonation of carboxylates both are related to the hydrogen bonds. Water is the optimal hydrogen bond acceptor and donor and water probe OH2 was used in our study. Based on all of the cases, all chemicals were modeled in their neutral forms.

Figs. 4 and 5 show the score plot of the second versus the first LV and the corresponding loading plot of model 2.

It is important to recall that all the descriptors used here have positive numerical values and therefore, a positive loading means that an increase of the variable determines an increase in the considered LV toward positive score values, while a negative loading means that an increase of the variable determines a decrease in the corresponding LV. From the Fig. 4, it can be concluded that the first LV also distinguished between the poor absorption and the efficient absorption. In the PLS loading plot (Fig. 5), variables contributing very little to the LVs have small loading values and are plotted around the center of the plot. On the other hand, the variables that contribute most are plotted at the borders of the plot. The efficient absorption drugs are located in positive region of score plot and are influenced most by Iw1—Iw8, ID4—ID8, D1—D8, A and CP descriptors. Iw1—Iw8 and ID4—ID8 measure the imbalance between the center of mass of a molecule and the position of hydrophilic and hydrophobic regions around it. Positive Iw1—Iw8 and ID4—ID8 values show that there is a clear concentration of hydrophilic and hydrophobic regions in only part of the molecular surface. Positive D1—D8, A and CP values demonstrate that large hydrophobic regions are beneficial to the human intestinal absorption. In the present study, compounds **1**, **2**, **4**, **5**, **7**, **8** and **20** possess the similar molecular properties. They all have large and more hydrophobic groups (*e.g.* aromatic ring and alkyl) and less hydrophilic groups. The polar atoms (*e.g.* N, O) in these molecules are less than 5. At the same time, the hydrophilic groups are unsymmetrical in one molecule. So they have strong hydrophobic properties and high imbalance between the center of mass and the barycentre of hydrophilic and hydrophobic regions, which, further lead to their high human intestinal absorption (>95%).

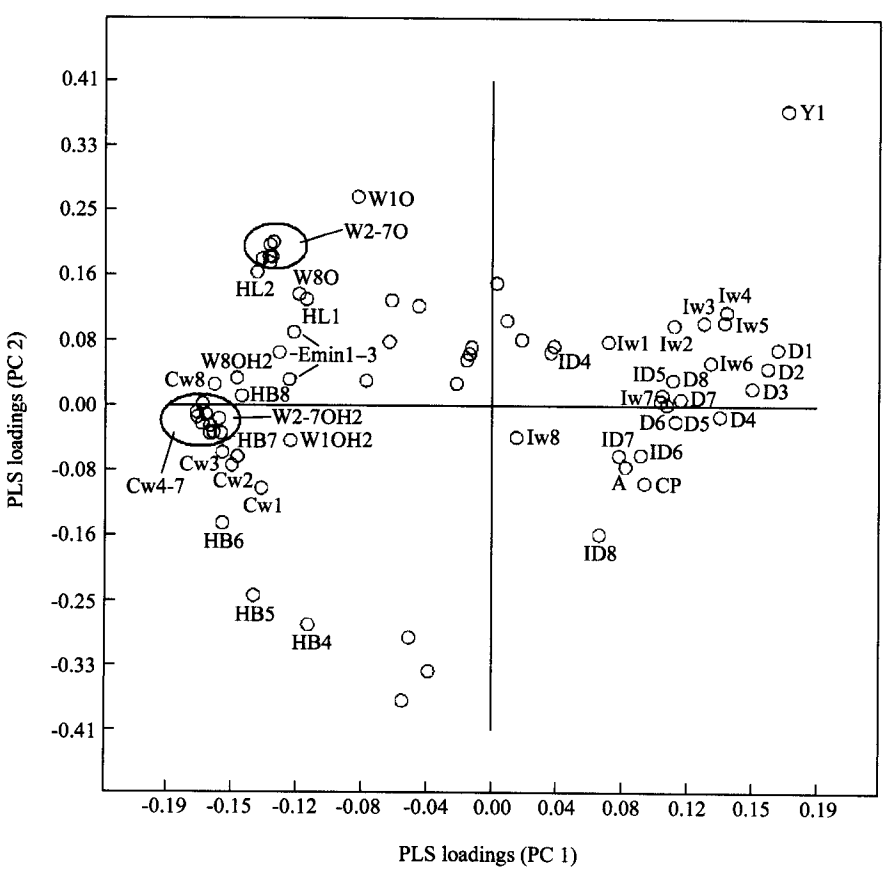

Fig. 5 PLS loading plot with two LVs (LV 1 vs. LV 2) for the compounds reported in Table 1.

At the negative extremity, the poor absorption drugs are influenced most by Cw1—Cw8, W1—W8 (OH2), W1—W8 (O) and HB4—HB8 (O) descriptors as well as HL1, HL2 and three energy minima. Capacity factors (Cw) represent the ratio between the hydrophilic regions and the molecular surface. WOH2 descriptors measure the size of hydrophilic regions with a water probe. Negative Cw1—Cw8, W1—W8 (OH2) values suggest that high polar and large hydrophilic regions of molecule lead to the poor human intestinal absorption. Carbonyl oxygen is able to accept two hydrogen bonds. So it examined the H-bonding donor capacity. Negative W1—W8 (O) values mean that more H-bonding donors are unfavorable to human intestinal absorption. The HB descriptors are defined as the difference between the water and carbonyl oxygen probe's interactions. So they refer mainly to the amount of the H-bonding acceptor atoms in a ligand. Negative HB4—HB8 (O) values represent that more hydrogen bond acceptors in the molecule are detrimental for intestinal absorption in humans. Hydrophilic-lipophilic balance (HL) is the ratio between the hydrophilic regions and the hydrophobic regions measured at two energy levels. Negative HL1 and HL2 values demonstrate that hydrophilic effect dominates in the molecule is not favorable to the intestinal absorption. For example, there are 11 polar atoms, 8 H-bonding donors in compound **3** and 16 polar atoms, 11 H-bonding donors in compound **15**, respectively. So hydrophilic regions dominate in their molecules and this is the most key role causing their low intestinal absorption (< 1%). Comparatively, molecule **18**, with high symmetry, has less polar atoms (8) and H-bonding donors (4) than compounds **3** and **15**. So its intestinal absorption (2.3%) is higher than that of molecules **3** (0.6%) and **15** (0.3%), but lower than that of other high absorption drugs. It is important to note that local interaction energy minima (Emin1—Emin3) are negatively correlated with the human intestinal absorption, which implies that low energy minima are beneficial to intestinal absorption.

Molecule **7** is located in the positive region and molecule **15** in the negative region. Fig. 6 shows the distribution of the hydrophilic (OH2) (a), hydrophobic (DRY) (b) and polar (O) (c) regions for molecule **7** and molecule **15**, respectively. Molecule **7** has high hydrophilic and hydrophobic integy moments as well as large hydrophobic regions and less hydrogen bond donors, so it possesses high intestinal absorption. But molecule **15** is converse, which results in its poor absorption.

It is useful to compare the statistics of the various equations put forward for the correlation of percent absorption. Palm *et al*.[5] and Norinder *et al*.[6] both listed statistics as shown in Table 2. In their analysis, Palm *et al*. gave the root mean squared error (RMSE) in the percent absorption 9.2%. In the study of Norinder *et al*., a logit transformation was performed on the experimental values and RMSE was 0.49. By the comparison with their study results, we get the best relationship between experimental absorption and calculated absorption.

Table 2 Comparison of statistics of equations of human intestinal absorption

Source	*n*	r^2	q^2	RMSE
Palm *et al*.	20	0.94	—	9.2%
Norinder *et al*.	20	0.92	0.80	0.49
Present study	20	0.95	0.86	8.4%

Norinder *et al*. chose 13 compounds as training set and the other 7 compounds as test set. The result was $r^2 = 0.90$ and $q^2 = 0.696$. So the cross-validation in our study was also done by randomly dividing the compounds into three groups, and then models were built by keeping one of these groups out of the analysis. The results gave $r^2 = 0.94$ and $q^2 = 0.80$. Compared with Norinder *et al*., our study represents a better prediction, bearing in mind the experimental error in the data.

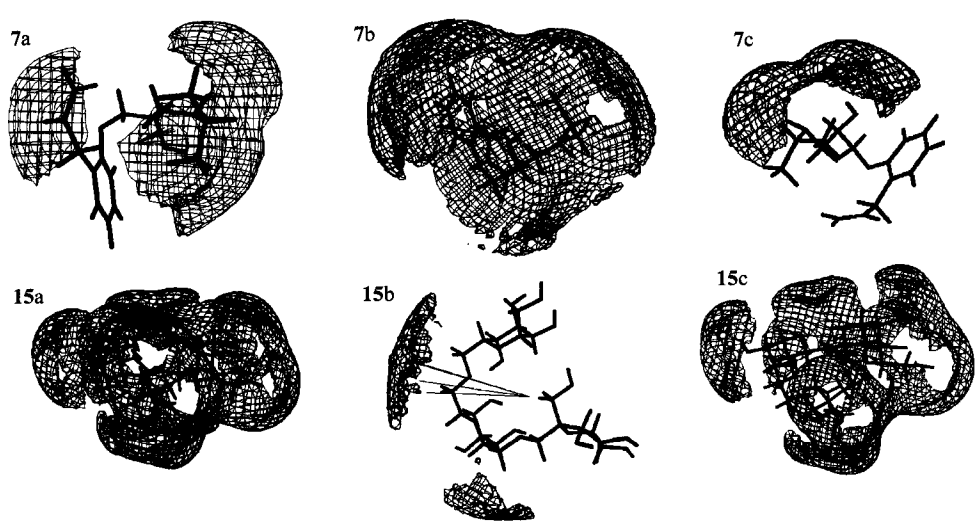

Fig. 6 (a) Distribution of hydrophilic regions for molecules (water probe). (b) Distribution of hydrophobic regions for molecules (DRY probe). (c) Distribution of polar regions for molecules (carbonyl oxygen probe).

Conclusions

The VolSurf method is a valuable tool for predicting the human intestinal absorption, because there is a clear correlation between the 3D structure of a compound and its absorption. The correlation between experimental and calculated absorption verifies that physicochemical properties are primary factors influencing the intestinal absorption of a compound. Strong hydrophilic and hydrophobic integy moments, that is, high imbalance between the center of mass of a molecule and the barycentre of its hydrophilic and hydrophobic regions, as well as large hydrophobic regions and less hydrogen bond donors and acceptors are favorable to human intestinal absorption of drugs.

References

1 Martin, Y. C. *J. Med. Chem.* **1981**, *24*, 229.

2 Testa, B.; Carrupt, P.-A.; Gaillard, P.; Billois, F.; Weber, P. *Pharm. Res.* **1996**, *11*, 335.

3 Rubas, W.; Cromwell, M. E. M. *Adv. Drug Deliv. Rev.* **1997**, *23*, 157.

4 Lee, C.-P.; de Vrueh, R. L. A.; Smith, P. L. *Adv. Drug Deliv. Rev.* **1997**, *23*, 47.

5 Palm, K.; Stenberg, P.; Luthman, K.; Artursson, P. *Pharm. Res.* **1997**, *14*, 568.

6 Norinder, U.; Osterberg, T.; Artursson, P. *Eur. J. Pharm. Sci.* **1999**, *8*, 49.

7 Cruciani, G.; Pastor, M.; Guba, W. *Eur. J. Pharm. Sci.* **2000**, *11* (suppl. 2), S29.

8 Tarvainen, M.; Sutinen, R.; Somppi, M.; Paronen, P.; Poso, A. *Pharm. Res.* **2001**, *18*, 1760.

9 Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. *J. Mol. Struct.* (*THEOCHEM*) **2000**, *503*, 17.

10 Filipponi, E.; Cruciani, G.; Tabarrini, O.; Cecchetti, V.; Fravolini, A. *J. Comput.-Aided Mol. Design* **2001**, *15*, 203.

11 Crivori, P.; Cruciani, G.; Carrupt, P.-A.; Testa, B. *J. Med. Chem.* **2000**, *43*, 2204.

12 Alifrangis, L. H.; Christensen, I. T.; Berglund, A.; Sandberg, M.; Hovgaard, L.; Frokjaer, S. *J. Med. Chem.* **2000**, *43*, 103.

13 Hudson, B. D.; Hyde, R. M.; Rahr, E.; Wood, J. *Quant. Struct.-Act. Relat.* **1996**, *15*, 285.

14 Abraham, M. H.; Zhao, Y. H.; Le, J.; Hersey, A.; Luscombe, C. N.; Reynolds, D. P.; Beck, G.; Sherborne, B.; Cooper, I. *Eur. J. Med. Chem.* **2002**, *37*, 595.

15 Zhao, Y. H.; Le, J.; Abraham, M. H.; Hersey, A.; Eddershaw, P. J.; Luscombe, C. N.; Boutina, D.; Beck, G.; Sherborne, B.; Cooper, I.; Platts, J. A. *J. Pharm. Sci.* **2001**, *90*, 749.

(E0205311 ZHAO, X. J.; DONG, H. Z.)