Available online at www.sciencedirect.com
SCIENCE DIRECT
Bioorganic & Medicinal Chemistry 14 (2006) 2225–2234
Bioorganic & Medicinal Chemistry

Effect of cholesterol on DMPC phospholipid membranes and QSAR model construction in membrane-interaction QSAR study through molecular dynamics simulation

Jianzhong Liu* and Liu Yang

Department of Chemistry and Biochemistry, University of Delaware, Newark, Delaware 19716, USA

Received 14 September 2005; revised 12 October 2005; accepted 2 November 2005
Available online 21 November 2005

Abstract—In this study, both pure DMPC and DMPC/cholesterol mixed membrane monolayer were built to compare the physical–chemical properties and dynamics properties through molecular dynamics simulation and normal-mode analysis. The results show that the addition of cholesterol decreases the area of per molecule of membrane, increases the lipid amplitude motion, and changes the solute diffusion coefficient. It is also found that the addition of cholesterol greatly changes the solute-membrane 1,4-nonbonded interaction energy (ΔE_{14}). MI-QSAR models were constructed based on solute–membrane interaction energy descriptors and other intramolecular descriptors. The results show that ΔE_{14} substitutes ΔE_{HB} as the second important descriptor compared with the previous study. Final results suggest that short range solute–membrane interaction energy changes due to the uptake of the solute may play an important decision on permeability in DMPC/cholesterol membrane. A test set was applied to evaluate the predictivity of MI-QSAR models. The result suggests that the combination of F_{H2O} and ΔE_{14} not only improves r^2 and q^2, but also greatly improves the model predictivity. Based on the combination of q^2 and r^2_{pre} values, a two-term model is better used to predict the solute permeability in this study ($r^2 = 0.859$, $q^2 = 0.803$, and $r^2_{pre} = 0.540$). Due to the small sample both in training set and test set, more datasets are necessary to make a final decision about the model construction and prediction.

1. Introduction

The principal computational approach to predict ADMET properties has been to use the quantitative structure–activity relationship, QSAR, methodology.[1–8] The QSAR methodology is based on the estimation of features and properties of chemical compounds, and subsequently in trying to relate these features and properties to biological activities. Unfortunately, QSAR methods are effectively limited to deal with analog structure–activity training sets,[9–11] while most ADMET datasets consist of structurally diverse compounds. A recently developed membrane-interaction QSAR (MI-QSAR) methodology[12–17] that combines structure-based design techniques with the QSAR paradigm permits structurally diverse training sets of some ADMET properties to be investigated. The operational key to this approach is to assume that the phospholipid-rich regions of cellular membranes constitute the effective 'receptor' for the compounds of the training set. The significance and utility of this approach rest upon the validity of the compound–membrane interaction in being a crucial event in the expression of the ADMET endpoint of interest. A comparison of MI-QSAR ADMET predictions to corresponding predictions made using 'traditional' intramolecular QSAR analyses clearly demonstrates that MI-QSAR analysis provides additional information, and the corresponding robustness, for estimating ADMET properties.[12–17] Moreover, MI-QSAR models permit mechanistic hypotheses to be made, which can be tested by experimental 'in vitro' methods. Thus, MI-QSAR analysis creates a bridge between experimental and computational ADMET approaches.

In MI-QSAR methodology, the dimyristoylphosphatidylcholine (DMPC) molecule usually is used to build a pure DMPC membrane monolayer and to explore the relationship of the permeability coefficients with the free energy force field, diffusion coefficient.[12–14] However, lipids cross the intestinal membrane by a diffusional process.[18,19] Therefore, it is reasonable to suggest that their permeability coefficients can be determined by physical

Keywords: MI-QSAR; DMPC; Cholesterol; Membrane; Normal-mode analysis.
* Corresponding author. Tel.: +1 302 831 3522; fax: +1 302 831 6335; e-mail: zhong@udel.edu

0968-0896/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2005.11.009

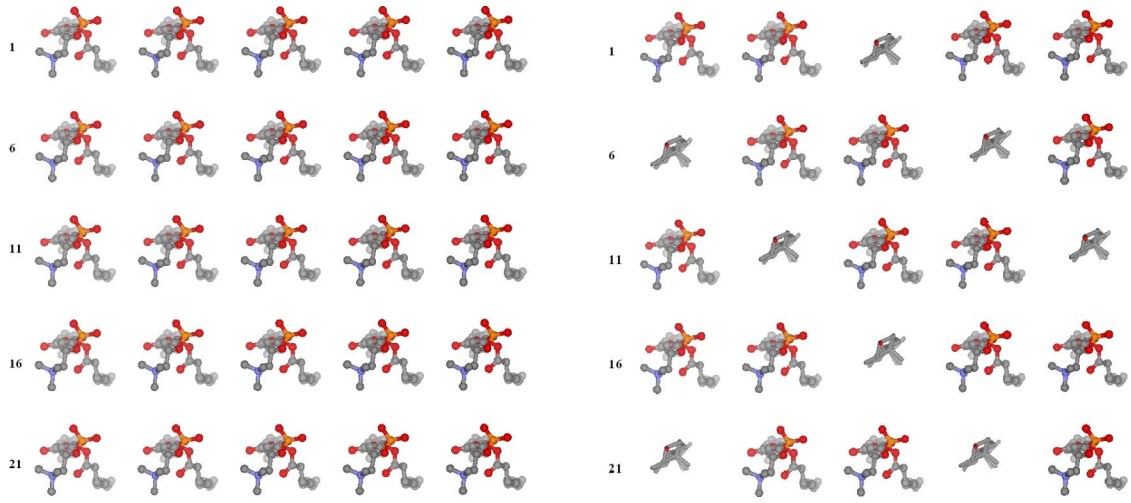

Figure 1. The top view of the initial structure of mandatory assembled pure DMPC monolayer and DMPC/cholesterol mixed monolayer according to the crystal structure of DMPC. The number on the left of each row means the position of that molecule.

properties of the membrane. In turn, the membrane physical properties are dependent upon the chemical composition of the layer. The experimental data supported that membrane fluidity can modulate lipid membrane permeability.[20] Cholesterol, one of the important components of the membrane, can increase membrane fluidity[21,22] which produces a significant effect on the membrane properties, subsequently, the calculation results of interaction among membrane molecules and solute.

In this paper, DMPC/cholesterol mixed membrane monolayer was built (see Fig. 1) based on the procedure of the MI-QSAR DMPC monolayer system[12–17] to explore the effect of cholesterol on the physical properties of the membrane, including membrane–solute interaction energy change, diffusion coefficient, and vibrational frequency of the membrane. MI-QSAR models were constructed based on intermolecular solute–membrane interaction descriptors, other solvation and dissolution intermolecular descriptors, and intramolecular solute descriptors. A test set was applied to evaluate the predictivity of these models. Final results suggest that solute–membrane 1,4-nonbonded interaction energy change plays an important role, while cholesterols were added into DMPC monolayer systems in this study.

2. Results and discussion

2.1. Area per molecule[23,24]

The averaged area per molecule is one of the most fundamental characteristics of lipid membrane. Although being one of the rather few structural quantities that can be measured accurately from model membranes via experiments, it also plays a major role in a number of quantities, including the ordering of acyl chains and the dynamics of lipids. It is highly useful as a means of monitoring the equilibration process. The averaged area per molecule of DMPC/cholesterol mixed monolayer is calculated according to the following formula:

$$A_{\mathrm{Mol}} = X_{\mathrm{DMPC}} A_{\mathrm{DMPC}} + X_{\mathrm{Chol}} A_{\mathrm{Chol}}, \quad (1)$$

where A_{Mol} is the overall area per molecule, X_{DMPC} and X_{Chol} are the ratios of DMPC and cholesterol in the membrane layer, which are 0.68 and 0.32, respectively. A_{DMPC} and A_{Chol} are the area per DMPC and the area per cholesterol.

The trajectory-dependent average area per molecule is presented in Figure 2, which shows that the obtained average area per molecule for the pure DMPC monolayer is 63.6 ± 0.8 Å^2. As for experimental data on the area per molecule, the results vary with the method used. The reported values are 62.9 Å^2,[25] 65.4 Å^2,[26] 65.6 Å^2,[27] and 70.3 Å^2,[28], respectively. The calculated result in our study is very close to these values, which certifies that the construction of pure DMPC monolayer system is reasonable. About the DMPC/cholesterol mixed monolayer, the average area per molecule is 44.8 ± 1.1 Å^2. This value is greatly smaller than pure DMPC monolayer. The previous paper[29] reported that the averaged area per molecule for DMPC/cholesterol bilayer with ratio 8:1 is 57.9 ± 0.9 Å^2. The averaged area per molecule for DPPC/cholesterol bilayer below the ratio 2:1 is about 41 Å^2.[30] The value in our study is obviously smaller than the former value, but is close to the latter value because our membrane system has a higher concentration of cholesterol, X_{Chol}, which greatly decreases the area per molecule. The DMPC and DPPC have the same head group and similar chains, which cause a slight difference in the area per molecule. Therefore, the difference between our calculations and DPPC/cholesterol is very small. It also supports our model construction about DMPC/cholesterol to be in agreement with that of the previous study. This result suggests that the addition of cholesterol will decrease the area of per molecule of membrane.

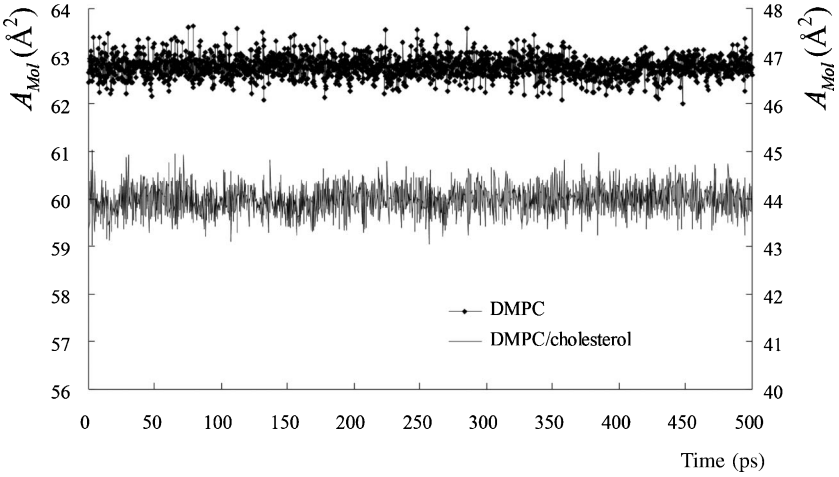

Figure 2. Time dependent area per molecule, A_{Mol}, for pure DMPC membrane monolayer (left Y axis) and DMPC/cholesterol mixed membrane monolayer (right Y axis).

2.2. Normal-mode[31,32] analysis

Normal-mode analysis is a time-honored technique in chemistry to obtain vibrational frequencies and the corresponding motions of molecules, molecular complexes, etc., in the harmonic approximation. It is finding renewed use in the area of biomolecular simulations, where, however, it can be very computationally demanding due to the large number of atoms. The standard normal-mode analysis is based on a calculation and subsequent diagonalization of the mass-weighted force constant matrix, also known as the Hessian. The normal-mode frequencies are directly related to the eigenvalues and the normal modes to the eigenvectors of this matrix. The order of the Hessian matrix is typically $3N$, where N is the number of atoms in the system of interest. The computational procedure to obtain the eigenvalues and eigenvectors of the Hessian matrix is an $O(N^3)$ process, i.e., the computational effort scales like N^3. For small and moderately sized molecules, N is of the order of 100 or less and the Hessian-based approach is easily carried out.

In this study, vibrational analysis was performed only on DMPC molecules for both monolayer systems. An examination of the density of states shows that the overall distribution is very similar for both systems (Fig. 3 inset). The differences between the systems are largely confined to the low-frequency motions region ($\leqslant$50 cm^{-1}) (Fig. 3). This plot shows that there are more distributions in this region for a mixed membrane system. The lowest vibrational frequencies of the pure DMPC and DMPC/cholesterol mixed system are presented in Table 1. The frequencies of the mixed membrane system, in the 1.9–2.9 cm^{-1} range, tend to be lower than the corresponding ones of the pure system, 2.5–3.3 cm^{-1}. Due to the lowest frequency modes making the largest contribution to configurational entropy and being associated with the largest amplitude

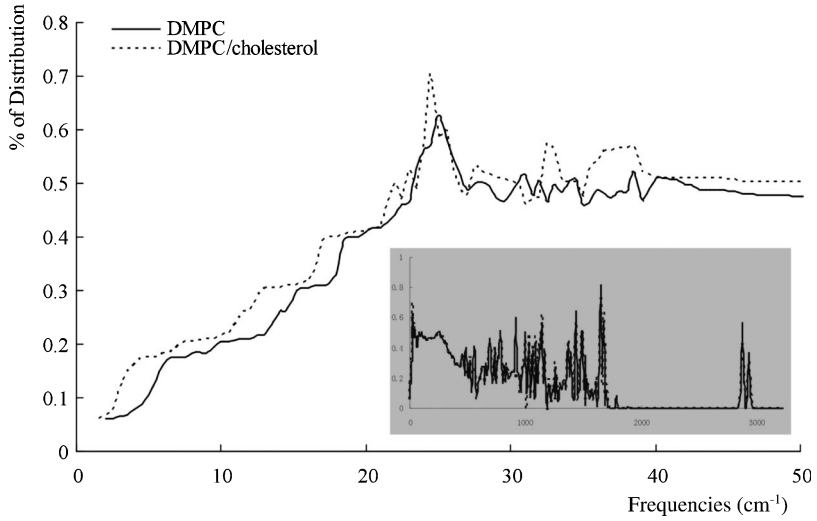

Figure 3. Density of states for the lowest-energy frequencies of vibration (main plot) and the full vibrational spectrum (inset) generated for the different trajectories for pure DMPC system (solid line) and DMPC/cholesterol system (dot line). These are generated from the distribution of vibrational frequencies resulting from the normal-mode analyses performed on multiple conformations taken from the different MD trajectories.

Table 1. The lowest five vibrational frequencies for 13th DMPC in pure DMPC membrane monolayer and DMPC/cholesterol mixed membrane monolayer, respectively

Membrane	1	2	3	4	5
Pure DMPC	2.48	2.84	2.91	3.00	3.25
DMPC/cholesterol mixed	1.91	2.12	2.45	2.67	2.94

The unit is cm^{-1}.

motions,[33] the calculations predict the fact that the mixed system should be more flexible than the pure DMPC membrane. This result is in agreement with previous experiment that cholesterol can increase the fluidity of membrane.[21,22]

2.3. Diffusion coefficient

In the MI-QSAR method, the lipid in 13th position was substituted by the solute to calculate a series of membrane–solute interaction descriptors.[12–17] One of them is diffusion coefficient, which is an important parameter used not only to characterize the position dependent mobility or local friction along the bilayer normal, but also to estimate the permeation ability. The variation of area of per molecule reminds us that the diffusion coefficient may be also linked. Therefore, the DMPC at 13th position was treated as the solute and its diffusion coefficient was calculated for both pure DMPC monolayer and DMPC/cholesterol mixed monolayer as comparison (see Fig. 4).

There are two different methods to estimate the diffusion coefficient, the mean square displacement (MSD)[34] and the force autocorrelation function (FACF).[35,36] MSD is used in this study. In a molecular system, a molecule moves in three dimensions and its motion of an individual molecule does not follow a simple path. If the path is examined in close detail, it will be seen to be a good approximation to a random walk. Mathematically, Einstein showed that mean square displacement (MSD) grows linearly with time, which is $\langle r(t)^2 \rangle = 6Dt + C$, where $\langle r(t)^2 \rangle$ is the mean square distance at time t. D and C are constants. The constant D is defined as the diffusion coefficient. Since all the atoms are considered to calculate a square displacement. This makes the mean square displacement scientifically significant. For each frame from the MDS trajectory, the average movement Δdi of the solute molecule, relative to the previous frame, was determined.

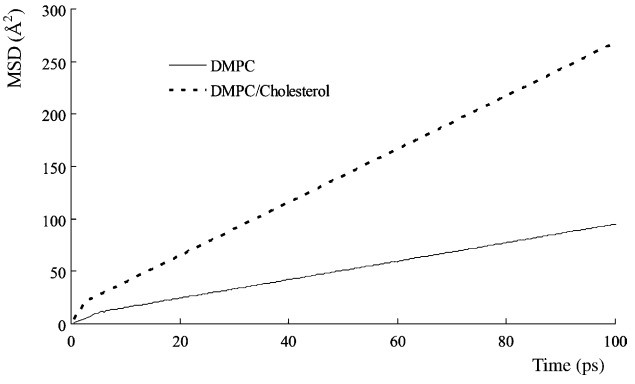

Figure 4. Mean-square displacement of the DMPC at the 13th position along simulation time (ps) for both pure DMPC system (solid line) and DMPC/cholesterol system (dot line). The diffusion coefficients calculated here are shown in Table 1.

Figure 4 shows the MSD of DMPC in both membrane systems. At the beginning of simulation, the movement of solute is irregular, but it reaches a regular movement short span of time. The calculated diffusion coefficients for pure DMPC monolayer and DMPC/cholesterol mixed monolayer are shown in Table 2. Both values are in the range of reported experimental values for organics in the membranes.[37] The value from a mixed membrane system is obviously higher than that from a pure system, which suggests that the the movement of 13th DMPC in a mixed membrane is quicker than that in a pure membrane. The result hints that cholesterol mixed membrane has better fluidity, which also caused us to investigate whether the mixed membrane system can increase the diffusion coefficient for those solutes in previous MI-SQAR studies.

Therefore, the DMPC molecule located at 13th position was substituted by 18 different solute molecules to calculate their diffusion coefficients for both membrane systems. The detailed methods to prepare these solute–membrane systems and to perform MDS refer to previous studies.[12–17] The results are shown in Table 1. For diazepam, which has the highest permeability capability, the addition of cholesterol decreases the diffusion coefficient, but opposite for acyclovir, which has the lowest permeability capability. There is no consistent tendency for other 16 solutes. The result suggests that the addition

Table 2. The calculated diffusion coefficients for DMPC, diazepam, acyclovir, and other 16 solutes in pure DMPC membrane monolayer and DMPC/cholesterol mixed membrane monolayer, respectively

Compound	Pure DMPC	DMPC/ cholesterol mixed	Permeability	V_m
DMPC	0.137	0.423	—	2179.03
Acyclovir	0.137	0.234	0.25	656.7
Diazepam	0.100	0.071	33.4	786.34
Alprenolol	0.188	0.164	25.3	851.72
Atenolol	0.235	0.213	0.53	875.35
Bremazocine	0.019	0.182	8.02	914.81
Caffeine	0.197	0.055	30.8	561.06
Clonidine	0.072	0.144	21.8	616.05
Dexamethasone	0.092	0.210	12.2	997.17
Dopamine	0.623	0.210	9.33	510.77
Ganciclovir	0.216	0.269	0.38	706.43
Metoprolol	0.218	0.182	23.7	909.22
Nadolol	0.070	0.253	3.88	934.34
Phenytoin	0.127	0.031	26.7	721.96
Pindolol	0.098	0.117	16.7	791.37
Salicylic acid	0.667	0.146	22.00	422.16
Sulfasalazine	0.042	0.250	0.30	1013.74
Terbutaline	0.284	0.231	0.47	719.43
Timolol	0.064	0.219	12.8	881.68

The unit is $Å^2/ps$. Last second column is the experimental permeability coefficient (10^{-6} cm/s). The last column is molecular volume ($Å^3$).

of cholesterol exactly causes a different impact on the diffusion behavior of solute although it does not produce the result we expected. The possible reason is that the addition of cholesterol changes the dynamic properties of the membrane system, subsequently changes the interaction behavior of the solute–membrane. To some solutes, this variation is strengthened; and for others, this variation may be weakened. From the table, it can be seen that for some molecules having a relatively small molecular volume, such as, salicylic acid, dopamine, and caffeine, the diffusion coefficients are greatly decreased. For some molecules having big molecular volume, such as bremazocine, sulfasalazine, nadolol, etc., the diffusion coefficients are increased. But not all solutes show this tendency. Therefore, further investigation about this change is necessary.

2.4. Interaction energy change (ΔE)

In order to explore the possible reason to explain the above results, the interaction energy changes due to the uptake of the solute to the total intermolecular system were explored based on stretch (ST), bend (BD), torsion (TOR), 1,4-nonbonded (14), general van der Waal (VDW), electrostatic (CHG), hydrogen bonding (HBD), and combinations thereof energies (TOT). The changes are calculated according to the following equation.[16] Also these energy changes are always used descriptors in MI-QSAR

$$\Delta E_{\mathrm{TT}} = \mathrm{ESM}_{\mathrm{TT}} - (\mathrm{ES}_{\mathrm{TT}} + \mathrm{EM}_{\mathrm{TT}}). \quad (2)$$

Here subscription TT means ST, BD, TOR, 14, VDW, CHG, HBD, and combinations thereof energies, TOT. ES means the energy of solute. EM means the energy of membrane system. ESM means the energy of the solute–membrane complex.

The calculated values are shown in Table 3. ΔE_{14} values in pure DMPC system are either negative or positive, which suggest that uptake of a solute either increases or decreases the 1,4-nonbonded interaction energy of a pure DMPC system. But in a mixed system, all ΔE_{14} values are negative. Since 1,4-nonbonded interaction energy considers 1,4-nonbonded Lennard-Jones interactions and 1,4 Coulombic contributions, which usually represent a repulsive/dispersion force, this suggests that the uptake of the solute should be helpful to decrease the repulsive forces among the solute–membrane complex, thereof to make this stable system. But this stability is not observed in other types of energy change or in the total energy change. Especially when only general van der Waals interaction change (ΔE_{VDW}) is considered, the uptake of solute will increase the energy of the system. The results suggest that the composition variation of phospholipids changes solute–membrane nonbonded interactions, which is in agreement with previous conclusion that nonbonded interactions play an important role in phospholipids due to the amphiphilic and zwitterionic (or ionic) nature of phospholipids.[38]

Our calculation results show that the values for $\mathrm{ESM}_{\mathrm{HBD}}$, $\mathrm{ES}_{\mathrm{HBD}}$, and $\mathrm{EM}_{\mathrm{HBD}}$ are zero for caffeine, diazepam, and phenytoin in a pure DMPC system, which means that there is no hydrogen bond energy formed between the solute and the membrane system. These results are also in agreement with previous publication.[16,17] In a mixed membrane system, the hydrogen bonding interaction energy change (ΔE_{HB}) for phenytoin is negative, but positive for caffeine and diazepam. Since $\mathrm{ES}_{\mathrm{HBD}}$ are zero for these three solutes, the results mean the $\mathrm{ESM}_{\mathrm{HBD}}$ is smaller than $\mathrm{EM}_{\mathrm{HBD}}$ for phenytoin, but for caffeine and diazepam $\mathrm{ESM}_{\mathrm{HBD}}$ is larger than $\mathrm{EM}_{\mathrm{HBD}}$. It suggests that in a mixed membrane system phenytoin has a stronger hydrogen bonding interactions with the membrane than caffeine and diazepam do. This is possible because phenytoin has a higher aqueous solvation free energy, which is −11.83 kcal/mol. These values are −5.47 and −6.87 kcal/mol for caffeine and diazepam according to a previous calculation method.[16] It is worth noting that we did not find any relationship between the diffusion coefficient and these solute–membrane interaction energy change descriptors as we expected.

2.5. MI-QSAR model construction

In order to investigate the effect of cholesterol on the MI-QSAR model, some intermolecular solute–membrane interaction descriptors, which can be highly affected by the addition of cholesterol, are extracted directly from the MDS trajectories and are listed in Part A of Table 4. The meaning and detail calculation methods of these particular intermolecular descriptors refer to previous studies.[16,17] These descriptors combined with other solvation and dissolution intermolecular descriptors (Part B in Table 4) and intramolecular solute descriptors (Part C in Table 4) are used to construct a MI-QSAR model by performing the genetic function approximation (GFA),[39] which is a multidimensional optimization method based on the genetic algorithm paradigm. The best MI-QSAR models for permeability construction realized by considering the combination of general intramolecular solute, intermolecular dissolution/solvation-solute, and intermolecular membrane–solute descriptors are presented as a function of the number of terms, that is descriptors, included in a given MI-QSAR model.

The models from one term to five term are listed below:

$$P = 35.515 + 0.898 * F_{\mathrm{H_2O}}, \quad n = 18, \quad r^2 = 0.822, \; q^2 = 0.777, \quad (3)$$

$$P = 44.292 + 0.908 * F_{\mathrm{H_2O}} + 0.125 * \Delta E_{14}, \quad = 18, \quad r^2 = 0.859, \; q^2 = 0.803, \quad (4)$$

$$P = 48.598 + 0.916 * F_{\mathrm{H_2O}} + 0.148 * \Delta E_{14} - 0.785 * \chi_{10}, \quad n = 18, \; r^2 = 0.876, \; q^2 = 0.810, \quad (5)$$

$$P = 49.998 + 0.932 * F_{\mathrm{H_2O}} + 0.159 * \Delta E_{14} - 0.996 * \chi_{10} - 0.088 * \Delta E_{\mathrm{TOR}}, \quad n = 18, \; r^2 = 0.895, \; q^2 = 0.828. \quad (6)$$

Table 3. The interaction energy changes for both pure DMPC and DMPC/cholesterol mixed monolayer systems due to the uptake of the solute to the total intermolecular system based on stretch (ST), bend (BD), torsion (TOR), 1,4-nonbonded (14), general van der Waals (VDW), electrostatic (CHG), hydrogen bonding (HBD), and combinations thereof energies (TOT)

Compounds	Pure DMPC monolayer system								DMPC/cholesterol mixed monolayer system							
	ΔE_{ST}	ΔE_{BD}	ΔE_{TOR}	ΔE_{14}	ΔE_{VDW}	ΔE_{CHG}	ΔE_{HBD}	ΔE_{TOT}	ΔE_{ST}	ΔE_{BD}	ΔE_{TOR}	ΔE_{14}	ΔE_{VDW}	ΔE_{CHG}	ΔE_{HBD}	ΔE_{TOT}
Acyclovir	−29.42	−0.28	−18.62	7.60	13.92	−39.39	−30.87	−97.06	−59.00	6.96	7.71	−86.15	59.52	−384.11	−45.87	−500.94
Diazepam	−74.73	−52.91	36.23	−31.13	9.87	110.62	0.00	−2.05	−93.92	−0.72	29.67	−46.19	48.48	58.19	62.44	57.95
Alprenolol	−80.28	−35.16	3.59	3.04	18.30	−112.48	−7.92	−210.91	−19.59	−11.24	−20.84	−54.96	136.40	−55.75	46.53	20.55
Atenolol	−44.24	−34.78	−2.19	−3.13	−1.34	93.39	−121.57	−113.86	36.81	−78.65	−11.97	−82.13	58.86	−292.13	19.94	−349.27
Bremazocine	−147.83	−110.11	33.49	19.71	27.19	29.24	−6.03	−154.34	−69.65	−6.50	−3.24	−80.62	52.47	−123.47	47.00	−184.01
Caffeine	−172.00	−76.28	22.80	18.76	33.65	136.20	0.00	−36.87	−53.57	32.90	−3.10	−90.32	58.43	−592.59	96.10	−552.15
Clonidine	−5.34	−53.70	−19.45	−3.73	−24.26	86.67	−19.71	−39.52	−69.61	−42.24	−1.26	−49.73	52.49	−264.79	−7.30	−382.44
Dexamethasone	−163.94	−16.28	103.15	67.58	28.21	1.41	−78.82	−58.69	−62.33	−52.51	−14.43	−43.01	10.83	−10.87	−25.38	−197.7
Dopamine	−7.38	−109.98	−20.65	76.97	22.56	82.30	−27.44	16.38	−101.45	−54.88	15.91	−52.13	124.18	−152.65	−7.95	−228.97
Ganciclovir	−52.91	−42.03	−9.89	36.21	15.85	31.82	−75.62	−96.57	−39.04	23.03	−42.41	−87.83	56.61	−98.74	27.30	−161.08
Metoprolol	−59.57	−60.23	−23.73	9.02	12.20	−16.10	−28.13	−166.54	4.74	−18.21	−17.53	−53.78	1.84	80.79	10.52	8.37
Nadolol	−160.11	−57.96	−2.02	−10.61	5.31	311.35	−50.19	35.77	−9.50	−24.28	4.38	−59.97	43.77	−15.80	−8.87	−70.27
Phenytoin	−53.01	−17.25	−9.39	31.80	15.87	20.37	0.00	−11.61	−6.25	−5.51	−26.87	−77.28	28.49	131.24	−36.38	7.44
Pindolol	−62.72	−45.03	9.99	−0.24	53.39	75.78	−86.31	−55.14	−24.31	−73.34	−12.34	−72.32	47.55	−317.73	56.85	−395.64
Salicylic acid	−69.24	−38.24	11.06	32.91	−1.07	−133.38	−22.04	−220.00	−17.35	−57.74	9.67	−104.87	49.75	−257.29	−64.47	−442.3
Sulfasalazine	−149.07	−25.83	22.59	5.51	3.49	33.68	−25.33	−134.96	−49.41	−7.40	1.60	−57.89	45.23	−246.44	−39.39	−353.7
Terbutaline	−193.16	−78.18	29.82	−0.19	52.78	85.44	−74.65	−178.14	−104.46	−51.15	27.33	−70.69	39.22	256.74	18.69	115.68
Timolol	16.55	−80.10	−18.46	−6.74	35.92	−1.48	−74.65	−128.96	−27.74	−82.30	−32.74	−52.54	62.24	277.40	123.32	267.64

The unit is kilocalories per mole.

Table 4. Descriptors used in the MI-QSAR model construction

Part A	
F	Average total free energy of interaction of the solute and membrane
E_{INTER}	Average interaction energy between the solute and the membrane
E_{TT}	1,4-Nonbonded, general van der Waals, electrostatic, hydrogen bonding, torsion and combinations thereof energies of the total (solute and membrane model) intermolecular minimum potential energy
ΔE_{TT}	Change in the TT = 1,4-nonbonded, general van der Waals, electrostatic, hydrogen bonding and combinations thereof of the total (solute and membrane model) intermolecular minimum potential energy
ΔS	Change in entropy of the membrane due to the uptake of the solute
S	Absolute entropy of the solute-membrane system
Part B	
F_{H_2O}	The aqueous solvation free energy
F_{OCT}	The 1-octanol solvation free energy
$\log P$	The 1-octanol/water partition coefficient
E_{coh}	The cohesive packing energy of the solute molecules
Part C	
HOMO	Highest occupied molecular orbital energy
LUMO	Lowest occupied molecular orbital energy
Dp	Dipole moment
V_m	Molecular volume
SA	Molecular surface area
MW	Molecular weight
MR	Molecular refractivity
N(hba)	Number of hydrogen bond acceptors
N(hbd)	Number of hydrogen bond donors
N(B)	Number of rotatable bonds
JSSA (X)	Jurs–Stanton surface area descriptors
Chi-N, Kappa-M	Kier and Hall topological descriptors
G_i	Gravitation index[44]
R_g	Radius of gyration[45]
PM	Principle moment of inertia
Se	Conformational entropy
Q(I)	Partial atomic charge densities
TIE	E-state topological parameter[46]
Eig1Z	Leading eigenvalue from Z-weighted distance matrix (Barysz matrix)[47]
Eig1m	Leading eigenvalue from mass-weighted distance matrix
Eig1v	Leading eigenvalue from van der Waals weighted distance matrix
Eig1e	Leading eigenvalue from electronegativity-weighted distance matrix
Eig1p	Leading eigenvalue from polarizability-weighted distance matrix
SEigZ	Eigenvalue sum from Z-weighted distance matrix (Barysz matrix)
SEigm	Eigenvalue sum from mass-weighted distance matrix
SEigv	Eigenvalue sum from van der Waals weighted distance matrix
SEige	Eigenvalue sum from electronegativity-weighted distance matrix
SEigp	Eigenvalue sum from polarizability-weighted distance matrix

Part A includes the membrane–solute interaction descriptors; Part B lists the intermolecular dissolution and solvation descriptors of the solute; Part C includes general intramolecular solute descriptors.

$$P = 53.242 + 0.926 * F_{H_2O} + 0.135 * \Delta E_{14} - 1.101 * \chi_{10} - 0.125 * \Delta E_{TOR} - 0.073 * \mathrm{LUMO},$$
$$n = 18,\ r^2 = 0.916,\ q^2 = 0.803. \quad (7)$$

Here, n is the number of compounds, r^2 is the coefficient of determination, and q^2 is the cross-validated coefficient of determination. The descriptors found in these MI-QSAR models are ranked as follows according to their contribution to the model.

1. F_{H_2O} is the aqueous solvation free energy.
2. ΔE_{14} is the change in the 1,4-nonbonded energy of the entire membrane–solute system for the solute relocated from free-space to the position corresponding to the lowest solute–membrane interaction energy state of the model system.
3. χ_{10} is a topological index measuring the size and shape of a molecule.[40]
4. ΔE_{TOR} is the change in the dihedral energy of the entire membrane–solute system for the solute relocated from free-space to the position corresponding to the lowest solute–membrane interaction energy state of the model system.
5. LUMO is the lowest occupied molecular orbital energy.

These descriptors are very similar to those of previous studies. Still, F_{H_2O} is considered as the most important descriptor and Kier–Hall topological descriptor, χ_{10}, is also found as before. LUMO can be excluded from the most significant descriptors in this study because q^2 value decreases in the five-term model, which hints that overfitting is being approached with the five or more term model. In the rest of the four descriptors,

Table 5. The test set with observed and predicted Caco-2 permeability coefficients, P (10^{-6} cm/s), for the 1–4 term MI-QSAR models

Test set	Observed	One term	Two term	Three term	Four term
Aminopyrine	36.5	27.695	33.226	33.661	28.783
Propranolol	21.8	16.781	15.575	25.121	32.430
Warfarin	21.1	19.28	26.741	25.554	15.101
Meloxicam	19.5	11.985	14.863	17.842	9.432
Zidovudine	6.93	12.122	11.856	8.399	0.320
Urea	4.56	18.83	−1.900	−1.991	−0.877
Sucrose	1.71	−39.444	5.192	7.831	11.252
Mannitol	0.38	−12.616	18.695	21.376	11.307
r^2_{pre}		−1.008	0.540	0.497	0.472

The last row is the validated correlation factor, r^2_{pre}.

F_{H_2O} and χ_{10} belong to Part B and Part C in Table 4, which are not affected by the addition of cholesterol into membrane. ΔE_{14} and ΔE_{TOR} are given in Part A of Table 4, whose values change greatly due to the addition of cholesterol. Especially for ΔE_{14}, all ΔE_{14} are negative as we discussed in previous section.

As the second important descriptor in previous MI-QSAR models, it is not surprising that ΔE_{HB}, which represents the hydrogen bonding energy of the entire membrane–solute system for the solute relocated from free-space to the position corresponding to the lowest solute–membrane interaction energy state of the model system, is substituted by other descriptors. Due to the effect of 3-hydroxy group, the addition of cholesterol positively changes the ΔE_{HB} value. In fact, different populations of hydrogen bonds between the cholesterol hydroxyl group and other molecules were found: between cholesterol and the phosphate oxygens of DMPC, between cholesterol and solute, and between cholesterol and carbonyl oxygens.[41]

The appearance of ΔE_{14} suggests that permeability depends on short distance solute–membrane nonbonded interaction energy change, which again supports the previous conclusion that nonbonded interactions play an important role in phospholipids. Worthy of mention is that ΔE_{TOR} appears as the fourth important descriptor. Usually E_{TOR} represents the energy due to rotation of a dihedral angle, described by a suitable cosine expansion. In particular, different forms for the van der Waals interactions and the dihedrals are in common use and the bonds are often constrained in simulations.[42] Therefore, ΔE_{TOR} represents short distance solute–membrane interaction energy change although it is a bonded interaction energy expression. The appearance of these two descriptors suggests that short range solute–membrane interaction energy changes due to the uptake of the solute playing an important decision on the permeability in DMPC cholesterol membrane.

A test set with eight solutes was used to evaluate the predictivity of the MI-QSAR models. The results are shown in Table 5. The last row is the validated correlation factor, r^2_{pre}.[43] Obviously, the two-term model has the 'better' predictivity than the other three models. For one-term model r^2_{pre} is negative, which means the prediction is even not better than random. The worst predictions are those of sucrose and mannitol because the aqueous solvation free energy for these two compounds is very strong, which is −53.67 and −83.58 kcal/mol, respectively. These two values are greatly lower than other solutes' F_{H_2O} value, causing them as outliers. Therefore, F_{H_2O} itself cannot predict the test set well although a one-term model, Eq. 3, has relatively high r^2 and q^2. For the two-term model, Eq. 4, the r^2_{pre} is 0.540. It suggests that the combination of F_{H_2O} and ΔE_{14} not only improves r^2 and q^2, but also greatly improves the model predictivity. Table 5 shows that the addition of term ΔE_{14} causes positive predictivity to the worst outlier, mannitol. The same result happens to sucrose too. Both of them are multiple hydroxyl compounds with very high hydrophilicity. The possible reason is that the mandatory docking of highly hydrophilic solutes into the lipophilic middle position of monolayer increases the unstable state of the solute–monolayer complex. This result again supports that ΔE_{14} plays an important role in determining the solute passing through the DMPC/cholesterol mixed membrane system. For three- or four-term model, although there are improvements on r^2 and q^2, there is no obvious improvement on r^2_{pre}. This suggests that these two models have a relatively worse predictivity than the two-term model based on the division of training set and test set in this study. Again the test set prediction suggests that the composition of the monolayer also plays an important role in predicting solute permeability. Due to both small training set and test set in this study, more datasets are necessary to further investigate the roles of these two descriptors on the effect of cholesterol on DMPC monolayer in a future study.

3. Conclusion

In this study, we explored the physical–chemical property and dynamic properties variations caused by the addition of cholesterol into the DMPC membrane and the effect on the MI-QSAR model construction. The results show that the addition of cholesterol decreases the area of per molecule of membrane, increases the lipid amplitude motion, and changes the solute diffusion coefficient. In our effort to explore the mechanism causing these variations, we found that the addition of cholesterol greatly decreases the solute–membrane 1,4-nonbonded interaction change. In the MI-QSAR model construction, this descriptor, ΔE_{14}, substitutes ΔE_{HB} as the second important descriptor. This result suggests the

building of membrane system may be a key step in MI-QSAR construction. Further investigation about the composition of cholesterol/DMPC monolayer in MIQ-SAR model construction and prediction or even about more sophisticated membrane composition is necessary in future study. This study is an effective supplement of MI-QSAR and supports that MI-QSAR methodology must be more mature and exert more functions in ADMET evaluation with the development of computer and knowledge of membrane structure and protein.

4. Materials and methods

4.1. Building phospholipid monolayer

Pure DMPC monolayer was built according to previous MI-QSAR methodologies.[12–17] DMPC/cholesterol monolayer was built according to the following procedure. The structure of cholesterol was constructed in Gaussian03[48] and was energy minimized using the quantum mechanical method MP2 and basis set 6-311+G(d). The pure 25 DMPC monolayer was used as the base. The DMPCs in position $3n$ ($n = 1$–8) were substituted by energy minimized cholesterol with its hydroxyl group toward the DMPC head group. In this way, the ratio of cholesterol/lipid is about 1:2, which is in agreement with the experimental composition of brush order in small intestine.[49,50] The assembled DMPC/cholesterol monolayer is shown in Figure 1.

4.2. Molecular dynamics simulation

The energy minimization and the following molecular dynamics simulation were performed on this assembled monolayer system using a Molsim package with an extended MM2 force field.[51] To remove unfavorable high-energy van der Waals interactions among molecules, the energy of the system was minimized by a series of steepest descent and conjugate gradient minimization steps. The energy convergence criterion was a gradient of less than 0.5 kcal $Å^{-1}$ mol^{-1}. Convergence was generally achieved within less than 10 ps. The model monolayer was first heated to 20 K and then to 50 K, and from that point in increments of 50 K to a final temperature of 311 K. At each temperature increment, MDS was carried out for 2 ps to allow for structural relaxation and distribution of kinetic energy throughout the system. After 311 K temperature was achieved, the whole system was held at this temperature and production trajectories were recorded every 0.1 ps over a 500 ps range for both DMPC and DMPC/cholesterol monolayer. The final temperature is above the phase transition temperature and therefore corresponds to stable systems in a liquid crystalline phase. All the calculations are based on this final 500 ps trajectory.

Acknowledgments

We gratefully acknowledge support from the Laboratory of Molecular Modeling & Design at UIC and from The Chem21 Group, Incorporated. Here we also express our grateful thanks to Dr. Anton J. Hopfinger for his contributions to MI-QSAR methodology.

References and notes

1. Akamatsu, M. *Curr. Top. Med. Chem.* **2002**, *2*, 1381–1394.
2. Anon. Classical QSAR: ADME and toxicology. *QSAR Comb. Sci.* **2004**, *23*, 686–691.
3. Anon. Classical QSAR: ADME toxicology. *QSAR Comb. Sci.* **2003**, *22*, 894–896.
4. Anon. Classical QSAR: ADME toxicology. *QSAR Comb. Sci.* **2003**, *22*, 778–781.
5. Klopman, G. S.; Liliana, R. *Eur. J. Pharm. Sci.* **2002**, *17*, 253–263.
6. Merlot, C.; Domine, D., et al. *Drug Discov. Today* **2003**, *8*, 594–602.
7. Cabrera, M. A.; Bermejo, M.; Gonzalez, M. P.; Ramos, R. *J. Pharm. Sci.* **2004**, *93*, 1701–1717.
8. Cabrera, M. A.; Bermejo, M.; Ramos, L.; Grau, R.; Gonzalez, M. P.; Gonzalez, H. *Eur. J. Med. Chem.* **2004**, *39*, 905–916.
9. Liu, J.; Pan, D.; Tseng, Y.; Hopfinger, A. J. *J. Chem. Inf. Model.* **2003**, *43*, 2170–2179.
10. Pan, D.; Liu, J.; Senese, C.; Hopfinger, A. J. *J. Med. Chem.* **2004**, *47*, 3075–3088.
11. Pan, D.; Iyer, M.; Liu, J.; Li, Y.; Iyer, M.; Hopfinger, A. J. *J. Chem. Inf. Model.* **2004**, *44*, 2083–2098.
12. Iyer, M.; Mishra, R.; Han, Y.; Hopfinger, A. J. *Pharm. Res.* **2002**, *19*, 1611–1621.
13. Kodithala, K.; Hopfinger, A. J.; Thompson, E. D.; Robinson, M. K. *Toxicol. Sci.* **2002**, *66*, 336–346.
14. Kulkarni, A.; Han, Y.; Hopfinger, A. J. *J. Chem. Inf. Model.* **2002**, *42*, 331–342.
15. Kulkarni, A.; Hopfinger, A. J. *Pharm. Res.* **1999**, *16*, 1245–1253.
16. Kulkarni, A.; Hopfinger, A. J. *Toxicol. Sci.* **2001**, *59*, 335–345.
17. Liu, J.; Li, Y.; Pan, D.; Hopfinger, A. J. *Int. J. Pharm.* **2005**, *304*, 115–123.
18. Sallee, V. L.; Dietschy, J. M. *J. Lipid Res.* **1973**, *14*, 475–484.
19. Winne, D.; Markgraf, I. *Naunyn Schmiedebergs Arch. Pharmacol.* **1979**, *309*, 271–279.
20. Meddings, J. B. *Biochim. Biophys. Acta—Biomembr.* **1989**, *984*, 158–166.
21. Brasitus, T. A.; Dahiya, R.; Dudeja, P. K. *Biochim. Biophys. Acta—Lipids Lipid Metab.* **1988**, *958*, 218–226.
22. Omodeo-Sale, F.; Lindi, C.; Marciani, P.; Cavatorta, P.; Sartor, G.; Masotti, L.; Esposito, G. *Comp. Biochem. Physiol. A* **1991**, *100*, 301–307.
23. Chaudhury, M. K.; Ohki, S. *Biochim. Biophys. Acta—Biomembr.* **1981**, *642*, 365–374.
24. Mombers, C.; De Gier, J.; Demel, R. A.; van Deenen, L. L. M. *Biochim. Biophys. Acta—Biomembr.* **1980**, *603*, 52–62.
25. Nagle, J. F.; Zhang, R.; Tristram-Nagle, S.; Sun, W.; Petrache, H. I.; Suter, R. M. *Biophys. J.* **1996**, *70*, 1419–1431.
26. Petrache, H. I.; Dodd, S. W.; Brown, M. F. *Biophys. J.* **2000**, *79*, 3172–3192.
27. Gurtovenko, A. A.; Patra, M.; Karttunen, M.; Vattulainen, I. *Biophys. J.* **2004**, *86*, 3461–3472.
28. Costigan, S. C.; Booth, P. J.; Templer, R. H. *Biochim. Biophys. Acta—Biomembr.* **2000**, *1468*, 41–54.
29. Smondyrev, A. M.; Berkowitz, M. L. *Biophys. J.* **2001**, *80*, 1649–1658.

30. Chiu, S. W.; Jakobsson, E.; Mashl, R. J.; Scott, H. L. *Biophys. J.* **2002**, *83*, 1842–1853.
31. Brooks, B.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 4995–4999.
32. Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423–447.
33. McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, 2003.
34. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, UK, 1987.
35. Marrink, S. J.; Berendsen, H. J. C. *J. Phys. Chem.* **1994**, *98*, 4155–4168.
36. Ulander, J.; Haymet, A. D. J. *Biophys. J.* **2003**, *85*, 3475–3484.
37. Stein, W. D. *Transprot and Diffusion Across Cell Membranes*; Academic press: NY, 1986.
38. Ledauphin, V.; Vergoten, G. *Biopolymers* **2000**, *57*, 373–382.
39. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Model.* **1994**, *34*, 854–866.
40. Kier, L. B. *Molecular Connectivity as a Descriptor of Structure for SAR Analysis*; Marcel Dekker: NY, 1980.
41. McMullen, T. P. W.; McElhaney, R. N. *Curr. Opin. Coll. Int. Sci.* **1996**, *1*, 83–90.
42. Tieleman, D. P.; Marrink, S. J.; Berendsen, H. J. C. *Biochim. Biophys. Acta—Rev. Biomembr.* **1997**, *1331*, 235–270.
43. Liu, J.; Yang, L.; Li, Y.; Pan, D.; Hopfinger, A. J. *J. Comp. Aided. Mol. Des.* **2005**, in press, doi:10.1007/s10822-005-9012-4.
44. Katritzky; Mu, L.; Karelson, M. *J. Chem. Inf. Model.* **1996**, *36*, 1162–1168.
45. Mattice, W. L. *Macromolecules* **1981**, *14*, 143–147.
46. Hall, L. M.; Hall, L. H.; Kier, L. B. *J. Chem. Inf. Model.* **2003**, *43*, 2120–2128.
47. Barysz, M.; Sadlej, A. J. *J. Chem. Phys.* **2002**, *116*, 2696–2704.
48. Frisch, M. J.; Trucks, G. W.; Schegel, H. B.; Gill, P. M. W.; Johnson, G. G.; Robb, M. A. *Guassian*, Guassian, Inc.: Pittsgurgh, PA, 2003.
49. Brasitus, T. A.; Dudeja, P. K.; Foster, E. S. *Biochim. Biophys. Acta—Biomembr.* **1988**, *938*, 483–488.
50. Roozemond, R. C.; Urli, D. C. *Biochim. Biophys. Acta—Biomembr.* **1981**, *643*, 327–338.
51. Hopfinger, A. J.; Pearlstein, R. A. *J. Comp. Chem.* **1984**, *5*, 486–499.