Supplementary Information

Lyotropic Liquid Crystal Engineering – Ordered Nanostructured Small Molecule Amphiphile Self-Assembly Materials by Design

Celesta Fong,[a,*] Tu Le[b] & Calum J. Drummond[c,*]

[a] *CSIRO Materials and Science & Engineering (CMSE), Bag 10, Clayton South, VIC 3169, Australia; Fax: 61 3 9545 2515; Tel: 61 3 9545 2608; E-mail: Celesta.Fong@csiro.au*

[b]*CSIRO Materials and Science & Engineering (CMSE), Bag 10, Clayton South, VIC 3169, Australia; Tel: 61 3 9545 8113; E-mail: Tu.Le@csiro.au*

[c] *CSIRO Materials and Science & Engineering (CMSE), Bag 10, Clayton South, VIC 3169, Australia; Tel: 61 3 9545 2050; E-mail: Calum.Drummond@csiro.au*

[*] *Corresponding Authors*

QSPR modelling

Input descriptors

Descriptors are mathematical descriptions of molecular properties of the surfactant molecules that are used in developing QSPR models. We included descriptors calculated for the whole surfactant molecules as well as those computed for head groups and tails with methyl terminus.

The simplest of these are atomistic representations. Molecules were represented by counting the numbers of atoms of specific elemental type with specific numbers of connections as well as the number of rings of varying sizes. Although this representation is simple, it has been shown to be adequate to encode not only physicochemical parameters such as hydrophobicity and molar refractivity but also biological acitivity such as dihydrofolate reductase inhibition[1].

The Burden index (B) which encodes the connectivity of the molecules and nature of the valence electrons[2] was also calculated. These eigenvalue descriptors are provided by diagonalising the adjacency matrices derived from the molecular graphs. These matrices describe how atoms in a molecule are connected. Off-diagonal elements of the matrices are squareroots of the number of bonds between two atoms if these atoms are chemicalley bonded and 0 if they are not.

The binned charges index[3] contains descriptors that describe the charge properties of the compounds (and indirectly the dipolar and hydrogen bonding properties) which are

encoded by charge fingerprint descriptors. Atom charges were computed using electronegativity equalization methods for each structure and the charges for each element type were used to populate bins. For each of the elemental types represented in the data set, the value of the charge was compared to bin-boundaries. The vector of bin occupancies for all element types represented the charge fingerprint.

Functional group representations were also employed to deduce their contributions to the phase behaviour of the drug delivery carriers. Although there is some overlap with the atomistic representation, functional group counts such as number of primary, secondary or tertiary hydroxyl groups (nOHp, nOHs or nOHt), number of donor or acceptor atoms for hydrogen bonds (nHDon or nHAcc) and number of esters - aliphatic and aromatic - (nRCOOR and nArCOOR) are relatively informative. The functional group counts included in this study were calculated using the DRAGON software package[4].

We also used the DRAGON software package to calculate other molecular descriptors including the unsaturation index[5] Ui, hydrophilic factor[5] Hy, molar refractivity[6] AMR, topological polar surface areas[7] using N, O polar contributions TPSA(NO) or using N,O,S,P polar contributions TPSA(Tot) and different octanol-water partition coefficients[6, 8] MLOGP or ALOGP.

Multiple linear regression

The approach employed to derive the relationship between the formation of the inverse phase (indicator parameter has the value of 1 if inverse phases can be formed and 0 otherwise) and relevant descriptors was multiple linear regression with expectation maximization (MLREM)[9, 10]. This method pruned out the least informative descriptors by using the sparse Laplacian prior feature selection. The sparsity of the MLREM was tuned progressively until the quality of the derived models deteriorated drastically, indicating that some of the most relevant descriptors have been removed. The data set for 82 surfactants listed in Table 1 was separated into a training set (80%) and a test set (20%) using the *K*-means clustering algorithm. The best MLREM model is the one that can predict most accurately the formation of the inverse phase for both the training and test sets using the smallest number of descriptors. Using this approach, the most important descriptors that contribute to the formation of the inverse phase were be found, as shown in Table 2.

Table 1. List of surfactants considered for QSPR modelling

1	Phytantriol	42	Phytanyl biuret
2	Monopentadecenoin	43	Hexahydrofarnesyl biuret
3	Monomyristolein	44	1-(2-hydroxyethyl)-1-oleyl biuret
4	Monoolein	45	1-(2-hydroxyethyl)-1-phytanyl biuret
5	Oleyl glycerate	46	1-O-(3,7,11,15-tetramethylhexadecyl)erythritol
6	Phytanyl glycerate	47	mono-O-(3,7,11,15-tetramethylhexadecyl)pentaerythritol
7	Phytanyl ethylene oxide	48	1-(O-5,9,13-trimethyltetradecyl)erythritol
8	Phytanyl di ethylene oxide	49	mono-(O-5,9,13-trimethyltetradecyl)pentaerythritol
9	Phytanyl tri ethylene oxide	50	1-O-(3,7,11-trimethyldodecyl)erythritol
10	Phytanyl tetra ethylene oxide	51	mono-O-(3,7,11-trimethyldodecyl)pentaerythritol
11	Phytanyl penta ethylene oxide	52	1-O-(3,7,11,15-tetramethylhexadecanoyl)glycerol
12	Phytanyl hexa ethylene oxide	53	1-O-(3,7,11,15-tetramethylhexadecanoyl)erythritol
13	Phytanyl hepta ethylene oxide	54	mono-O-(3,7,11,15-tetramethylhexadecanoyl)pentaerythritol
14	Phytanyl octa ethylene oxide	55	1-(O-5,9,13-trimethyltetradecanoyl)glycerol
15	Hexahydrofarnesyl ethylene oxide	56	1-(O-5,9,13-trimethyltetradecanoyl)erythritol
16	Hexahydrofarnesyl di ethylene oxide	57	mono-(O-5,9,13-trimethyltetradecanoyl)pentaerythritol
17	Hexahydrofarnesyl tri ethylene oxide	58	2-monoolein
18	Hexahydrofarnesyl tetra ethylene oxide	59	Monoerucin
19	Hexahydrofarnesyl penta ethylene oxide	60	Monovaccenin
20	Hexahydrofarnesyl hexa ethylene oxide	61	Oleyl ethanolamide
21	Hexahydrofarnesyl hepta ethylene oxide	62	Linoleoyl ethanolamide
22	Hexahydrofarnesyl octa ethylene oxide	63	γ - linolenoyl ethanolamide
23	Glyceryl monophytanoate	64	3,7-dimethyl octanoyl monoethanolamide
24	Glyceryl monohexahydrafarnesoate	65	Hexahydrofarnesyl monoethanolamide
25	2-glyceryl monohexahydrafarnesoate	66	Phytanoyl monoethanolamide
26	Octadecyl glycerate	67	Phytanoyl amide
27	Hexahydrofarnesyl glycerate	68	Hexahydrofarnesyl amide
28	1-glyceryl oleyl ether	69	3,7-dimethyl octanoyl amide
29	1-dodecyl urea	70	Dodecaoxyethylene mono-*n*-dodecyl ether
30	1-octadecyl urea	71	1-octadecyl urea
31	*cis*-octadec-9-enyl urea	72	Linolenyl urea
32	*cis*-octadec-9-enyl biuret	73	1-decyl urea
33	*cis*,*cis*-octadec-9,12-dienyl urea	74	5-decyl urea
34	3,7,11,15-tetramethyl-hexadecyl urea	75	1-dodecyl urea
35	3,7,11-trimethyldodecyl urea	76	2-dodecyl urea
36	1-(2-hydroxyethyl)-1-oleyl urea	77	4-dodecyl urea
37	1-(2-hydroxyethyl)-1-phytanyl urea	78	6-dodecyl urea
38	3-(2-hydroxyethyl)-1-oleyl urea	79	1-O-(5,9,13,17-tetramethyloctadecyl)erythritol
39	3-(2-hydroxyethyl)-1-phytanyl urea	80	mono-O-(5,9,13,17-tetramethyloctadecyl)pentaerythritol
40	3-(2-hydroxyethyl)-1-hexahydrofarnesyl urea	81	1-O-(5,9,13,17-tetramethyloctadecanoyl)erythritol
41	1-(2,3-dihydroxypropyl)-1-oleyl urea	82	mono-O-(5,9,13,17-tetramethyloctadecanoyl)pentaerythritol

Table 2. Important descriptors selected by MLREM

1	Tail - molar refractivity	8	Head - number of ether groups
2	Tail - molecular weight	9	Tail - hydrophilic factor
3	Tail - number of double bond of carbons	10	Head - number of quaternary carbons
4	Tail - logarithm of octanol/water partition coefficient	11	Molecule - logarithm of octanol/water partition coefficient
5	Tail - unsaturation index	12	Head - number of imides
6	Head - number of primary hydroxyl groups	13	Head - logarithm of octanol/water partition coefficient
7	Head - hydrophilic factor	14	Molecule - unsaturation index

1 F. R. Burden, *Quantitative Structure-Activity Relationships*, 1996, **15**, 7.
2 D. A. Winkler and F. R. Burden, *Molecular Simulation*, 2000, **24**, 243.
3 F. R. Burden, M. J. Polley, and D. A. Winkler, *Journal of Chemical Information and Modeling*, 2009, **49**, 710.
4 TALETE srl, in 'Dragon for Windows (Software for Molecular Descriptor Calculations)', 2007.
5 R. Todeschini and V. Consonni, 'Handbook of Molecular Descriptors', Wiley-VCH, 2000.
6 V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, and R. K. Robins, *Journal of Chemical Information and Computer Sciences*, 1989, **29**, 163.
7 P. Ertl, B. Rohde, and P. Selzer, *Journal of Medicinal Chemistry*, 2000, **43**, 3714.
8 I. Moriguchi, S. Hirono, Q. Liu, I. Nakagome, and Y. Matsushita, *Chemical & Pharmaceutical Bulletin*, 1992, **40**, 127.
9 F. R. Burden and D. A. Winkler, *Qsar & Combinatorial Science*, 2009, **28**, 645.
10 F. R. Burden and D. A. Winkler, *Qsar & Combinatorial Science*, 2009, **28**, 1092.