Supporting Information

A QSPR study on liquid crystallinity of five-ring bent-core molecules using decision trees, MARS and artificial neural networks

Jelena Antanasijević,*[,#] Davor Antanasijević, [‡] Viktor Pocajt,[#] Nemanja Trišović,[#] and Katalin Fodor-Csorba[§]

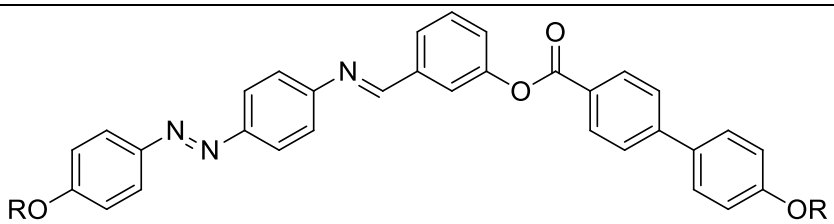
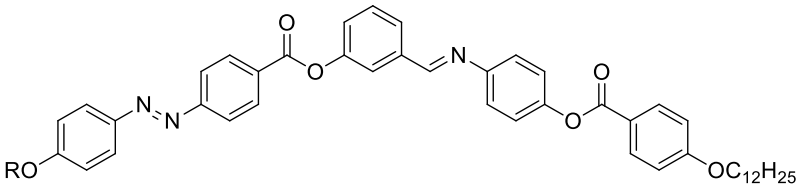
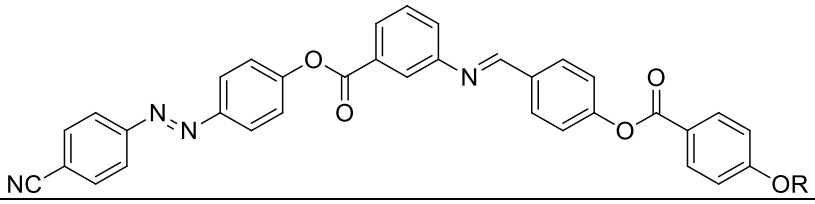
[#] *University of Belgrade, Faculty of Technology and Metallurgy, Karnegijeva 4, 11120 Belgrade, Serbia*

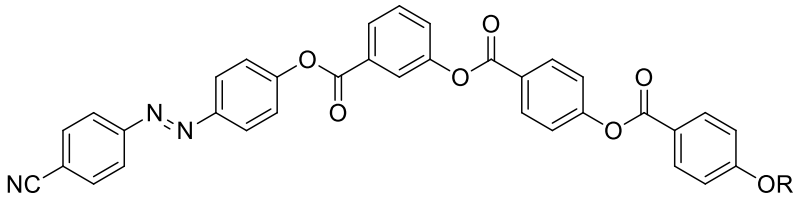
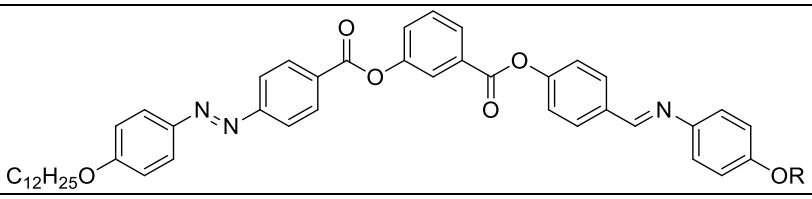
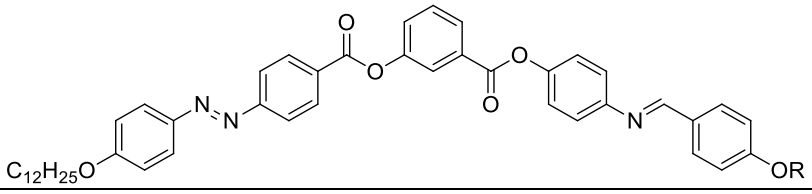
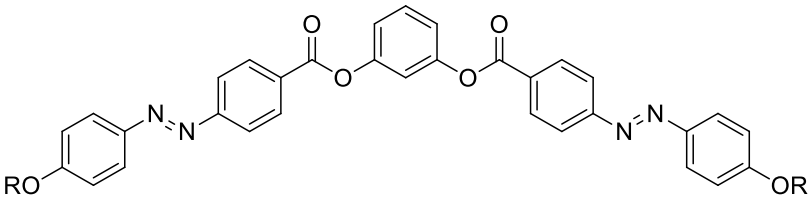
[‡] *University of Belgrade, Innovation Center of the Faculty of Technology and Metallurgy, Karnegijeva 4, 11120 Belgrade, Serbia*

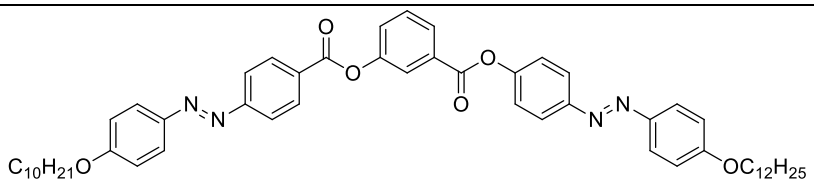
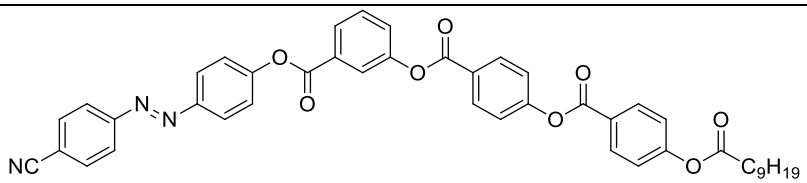
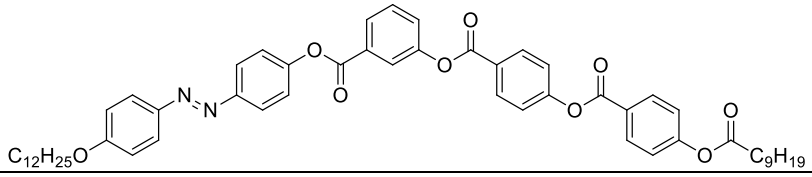
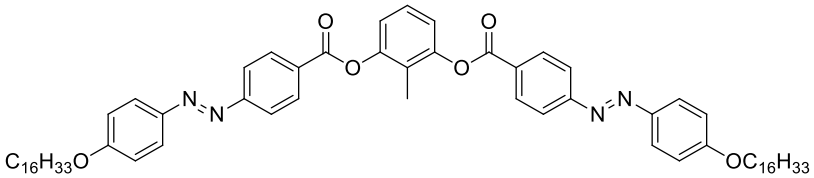
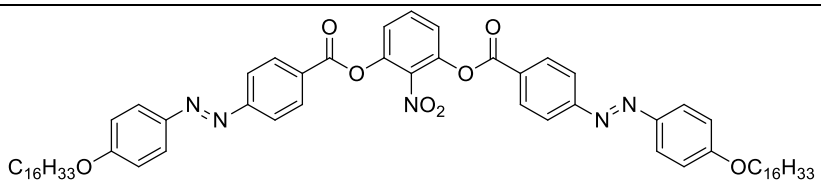
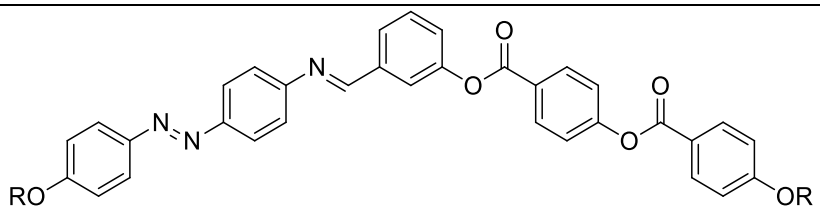
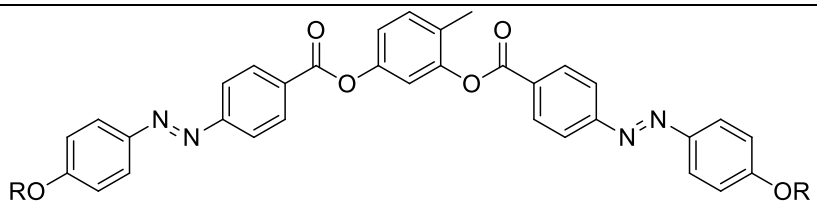
[§] *Wigner Research Centre for Physics, Institute for Solid State Physics and Optics of the Hungarian Academy of Sciences, H-1525 Budapest, P.O. Box 49, Hungary*

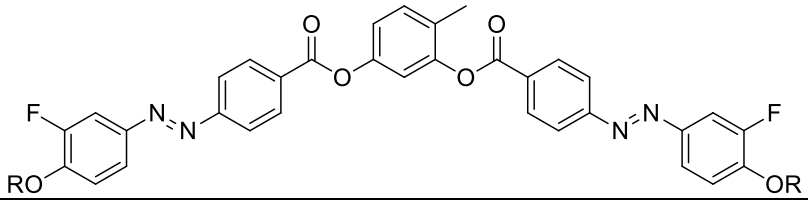
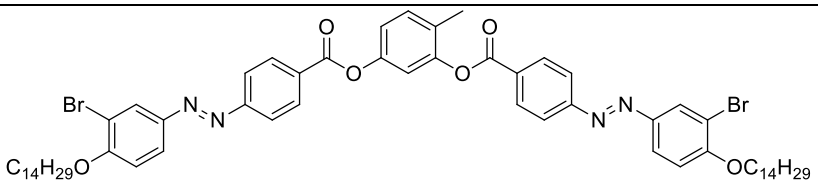
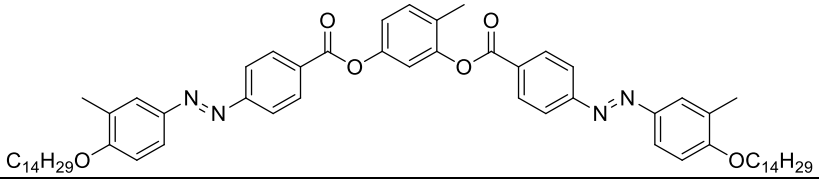
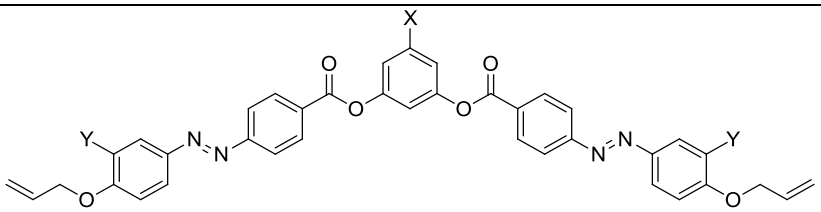
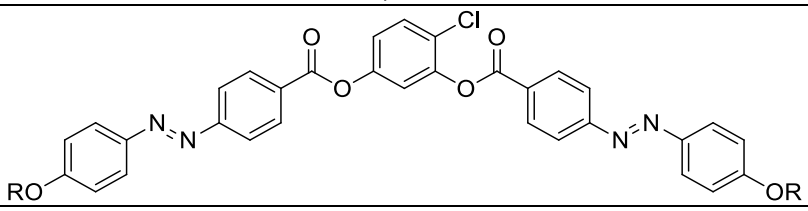
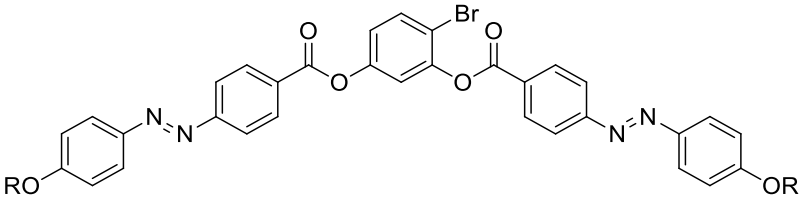
*Corresponding author e-mail: jantanasijevic@tmf.bg.ac.rs

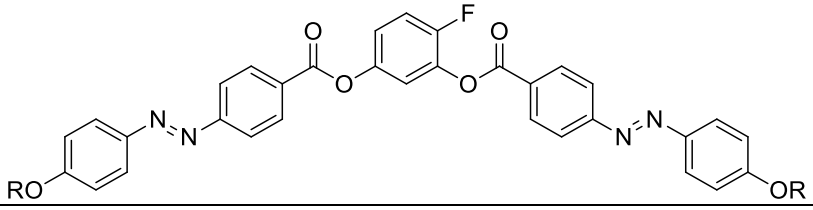
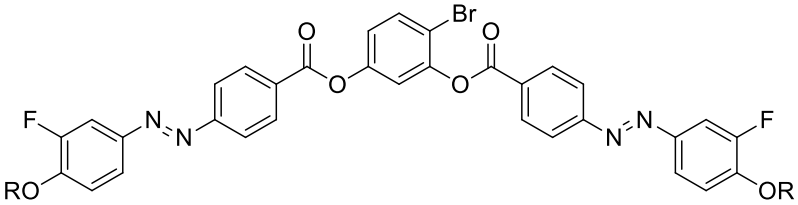
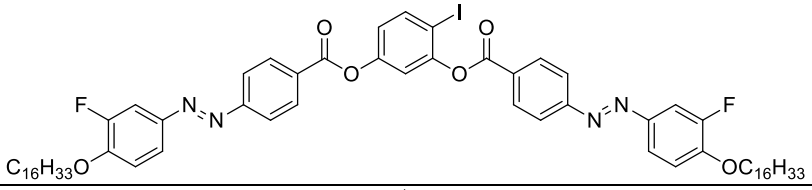
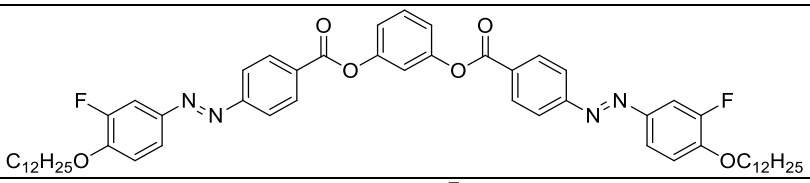
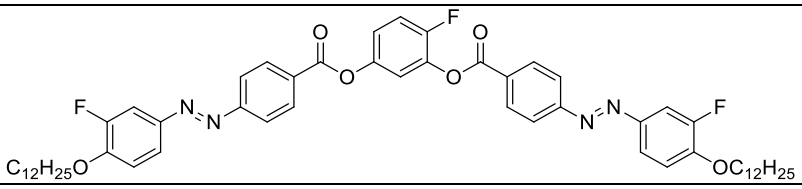
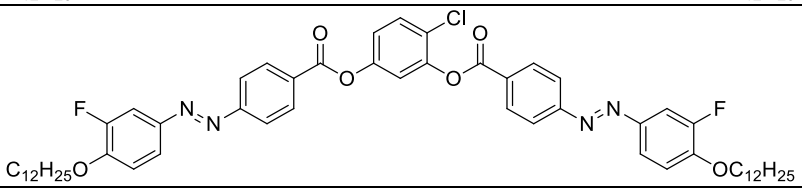
Table S1. Molecular structure and liquid crystal behavior of the modelled compounds

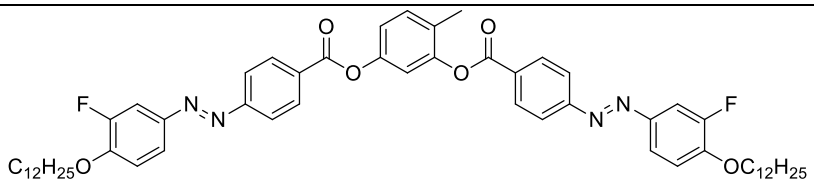
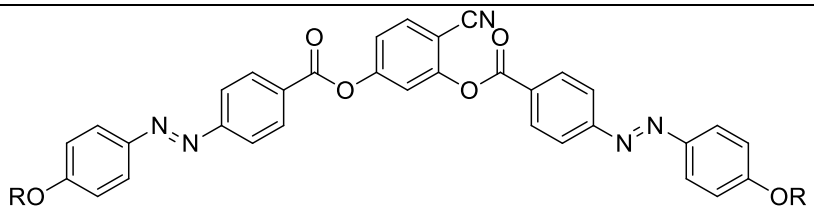
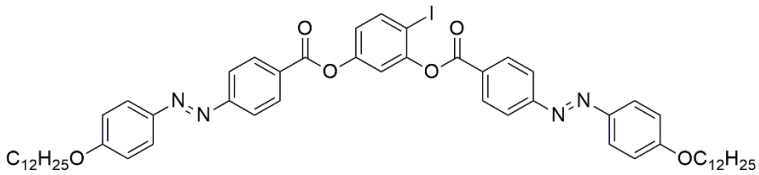
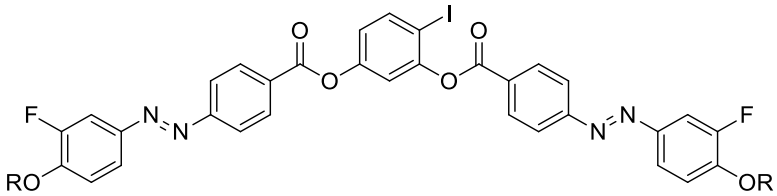
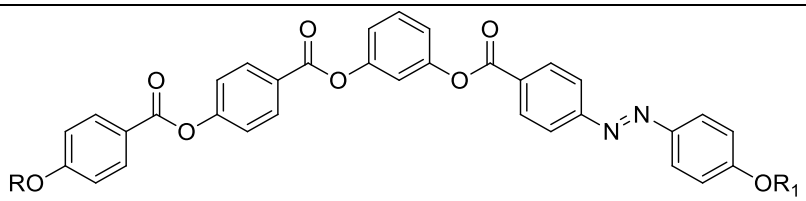
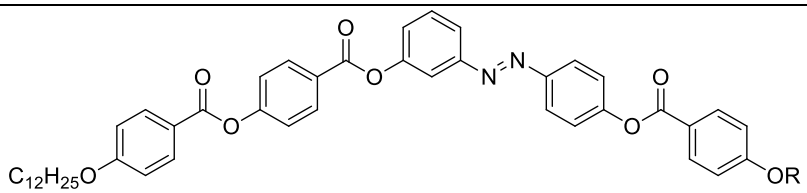
Comp.	Molecular structure	LC/NLC	Ref.
			
1	R = C_6H_{13}	LC	[1]
2	R = C_8H_{17}	LC	[1]
3	R = $C_{10}H_{21}$	LC	[1]
4	R = $C_{12}H_{25}$	LC	[1]
			
5	R = C_4H_9	LC	[2]
6	R = C_5H_{11}	LC	[2]
7*	R = C_6H_{13}	LC	[2]
8*	R = C_7H_{15}	LC	[2]
9	R = C_8H_{17}	LC	[2]
10	R = C_9H_{19}	LC	[2]
11	R = $C_{10}H_{21}$	LC	[2]
12	R = $C_{11}H_{23}$	LC	[2]
13	R = $C_{12}H_{25}$	LC	[2]
14*	R = $C_{14}H_{29}$	LC	[2]
15	R = $C_{16}H_{33}$	LC	[2]
			
16	R = C_4H_9	LC	[2]
17*	R = C_5H_{11}	LC	[2]
18	R = C_6H_{13}	LC	[2]
19	R = C_7H_{15}	LC	[2]
20	R = C_8H_{17}	LC	[2]
21	R = C_9H_{19}	LC	[2]
22	R = $C_{10}H_{21}$	LC	[2]
23	R = $C_{11}H_{23}$	LC	[2]
24	R = $C_{12}H_{25}$	LC	[2]

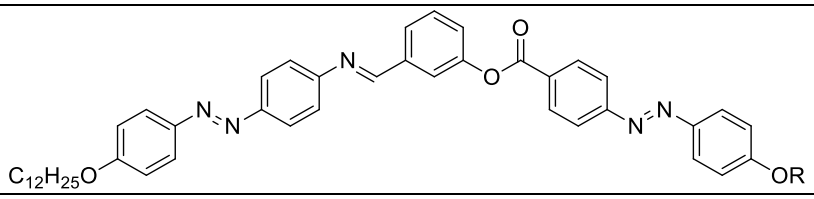
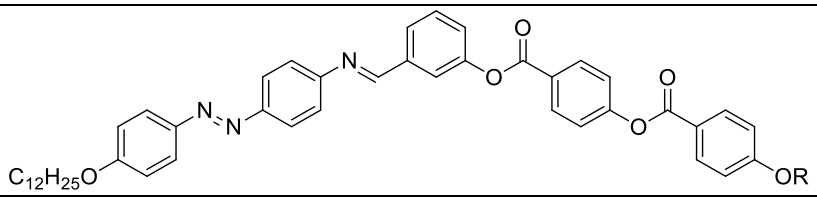
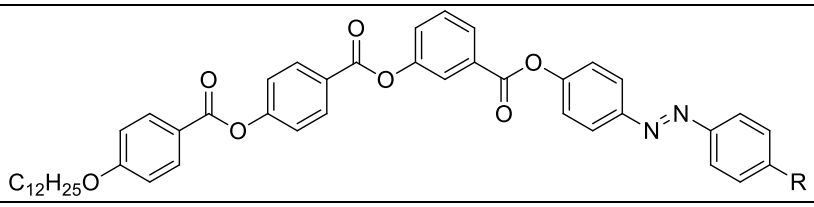
			
25	R = C_8H_{17}	NLC	[2]
26	R = C_9H_{19}	LC	[2]
27	R = $C_{10}H_{21}$	LC	[2]
28	R = $C_{11}H_{23}$	LC	[2]
29	R = $C_{12}H_{25}$	LC	[2]
			
30	R = C_5H_{11}	NLC	[2]
31	R = C_6H_{13}	LC	[2]
32	R = C_7H_{15}	LC	[2]
33*	R = C_8H_{17}	LC	[2]
34	R = C_9H_{19}	LC	[2]
35	R = $C_{10}H_{21}$	LC	[2]
36	R = $C_{11}H_{23}$	LC	[2]
37	R = $C_{12}H_{25}$	LC	[2]
38	R = $C_{14}H_{29}$	LC	[2]
39	R = $C_{16}H_{33}$	LC	[2]
			
40	R = C_4H_9	LC	[2]
41	R = C_6H_{13}	LC	[2]
42	R = C_8H_{17}	LC	[2]
43*	R = $C_{10}H_{21}$	LC	[2]
44	R = $C_{12}H_{25}$	LC	[2]
			
45	R = $C_{12}H_{25}$	NLC	[2]
46	R = $C_{14}H_{29}$	LC	[2]
47	R = $C_{16}H_{33}$	LC	[2]
48	R = $C_{18}H_{37}$	LC	[2]

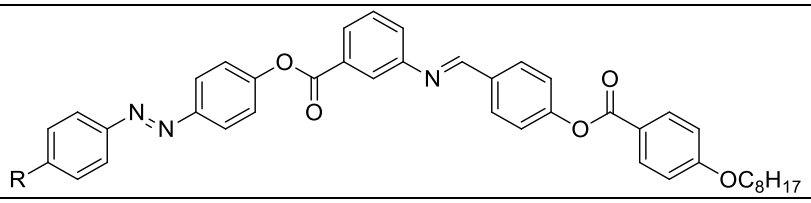
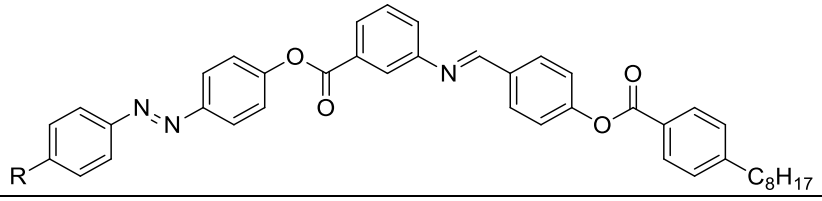
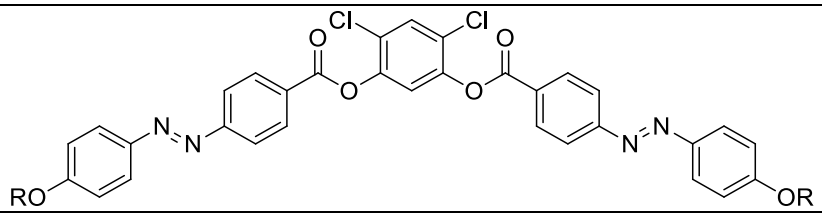
49		NLC	2
50		NLC	2
51		NLC	2
52*		NLC	2
53		NLC	2
			
54	R = C_6H_{13}	LC	3
55	R = C_8H_{17}	LC	3
56	R = $C_{10}H_{21}$	LC	3
57	R = $C_{12}H_{25}$	LC	3
			
58*	R = C_6H_{13}	LC	4
59*	R = C_8H_{17}	LC	4
60	R = C_9H_{19}	LC	4
61	R = $C_{10}H_{21}$	LC	4
62	R = $C_{12}H_{25}$	LC	4
63	R = $C_{14}H_{29}$	LC	4
64	R = $C_{16}H_{33}$	LC	4
65	R = $C_{18}H_{37}$	LC	4
66	R = $C_{20}H_{41}$	LC	4

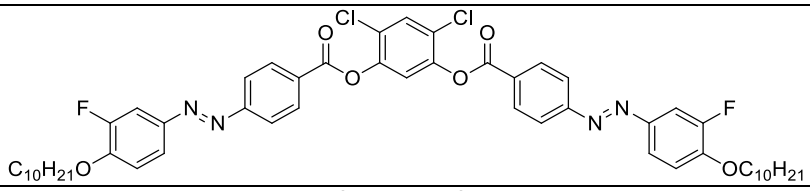
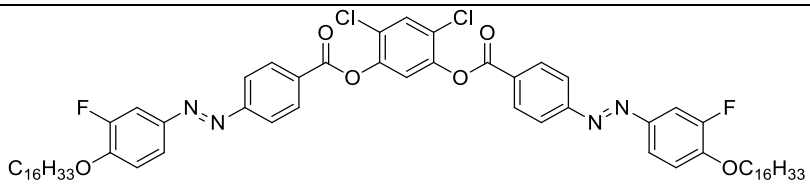
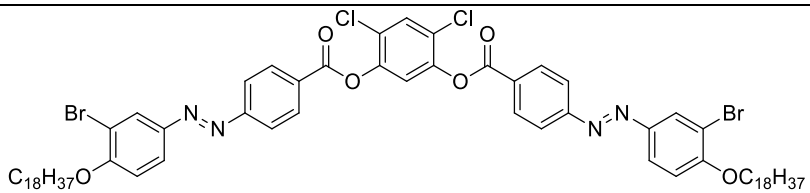
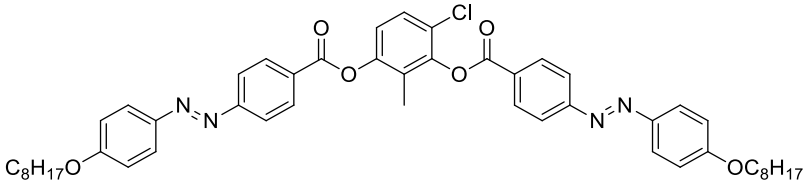
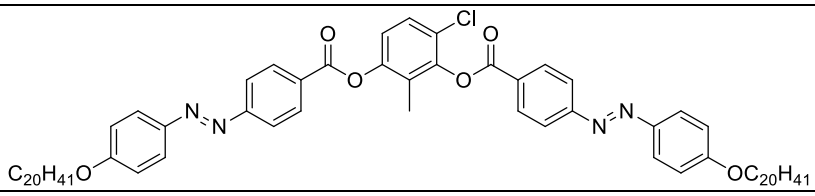
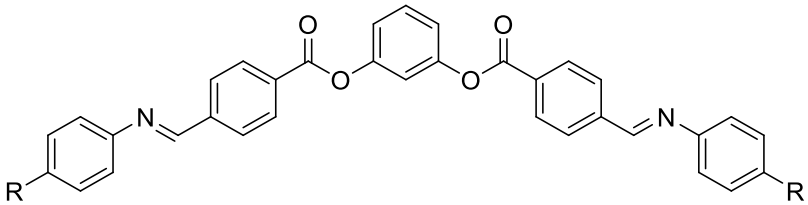
67	R = $C_{22}H_{45}$	NLC	4
			
68	R = C_6H_{13}	LC	4
69	R = $C_{14}H_{29}$	LC	4
70	R = $C_{20}H_{41}$	LC	4
71		NLC	4
72		LC	4
			
73	X = H, Y = H	LC	5
74*	X = H, Y = Cl	LC	5
75	X = H, Y = COOH	NLC	5
76	X = F, Y = H	LC	5
77	X = F, Y = Cl	LC	5
78	X = F, Y = COOH	NLC	5
			
79	R = C_8H_{17}	LC	6
80	R = $C_{10}H_{21}$	LC	6
81	R = $C_{12}H_{25}$	LC	6
82*	R = $C_{14}H_{29}$	LC	6
83	R = $C_{16}H_{33}$	LC	6
			
84	R = C_8H_{17}	LC	6
85*	R = $C_{10}H_{21}$	LC	6

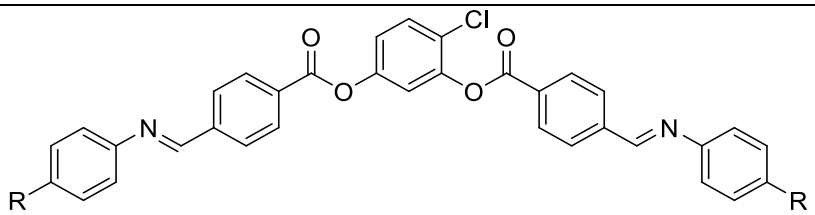
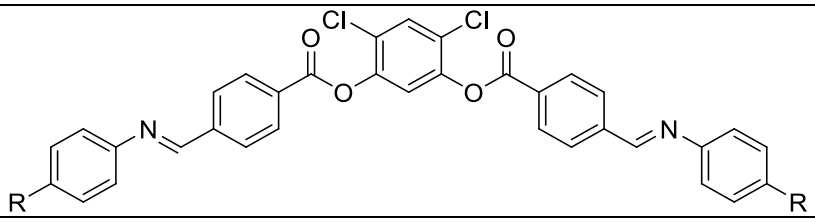
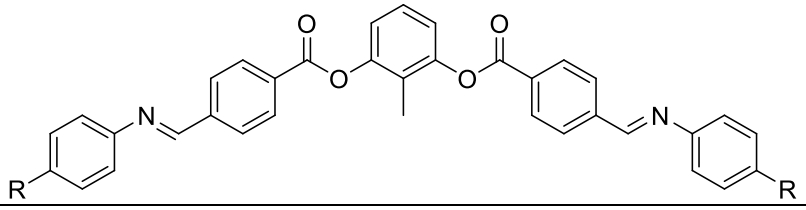
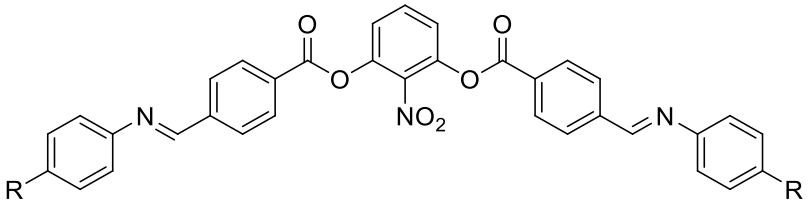
86	R = $C_{12}H_{25}$	LC	6
87	R = $C_{14}H_{29}$	LC	6
88	R = $C_{16}H_{33}$	LC	6
			
89	R = C_8H_{17}	NLC	6
90	R = $C_{10}H_{21}$	NLC	6
91	R = $C_{12}H_{25}$	NLC	6
92*	R = $C_{14}H_{29}$	NLC	6
93	R = $C_{16}H_{33}$	LC	6
			
94	R = C_4H_9	NLC	7
95*	R = C_6H_{13}	LC	7
96	R = C_8H_{17}	LC	7
97	R = $C_{10}H_{21}$	LC	7
98	R = $C_{12}H_{25}$	LC	7
99	R = $C_{14}H_{29}$	LC	7
100*	R = $C_{16}H_{33}$	LC	7
101	R = $C_{18}H_{37}$	NLC	7
102		LC	7
103		NLC	7
104		LC	7
105		LC	7

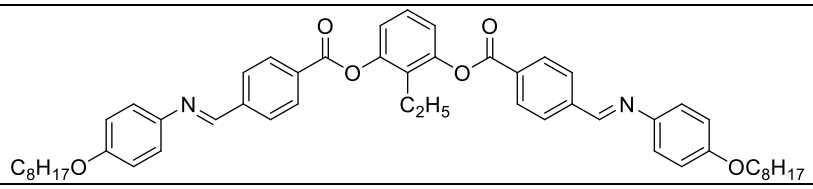
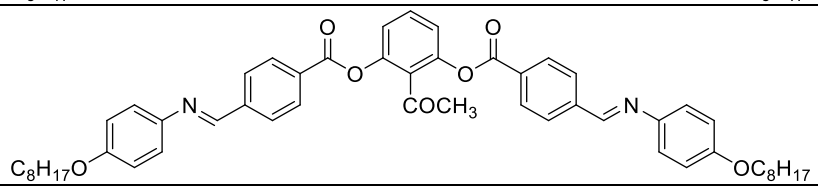
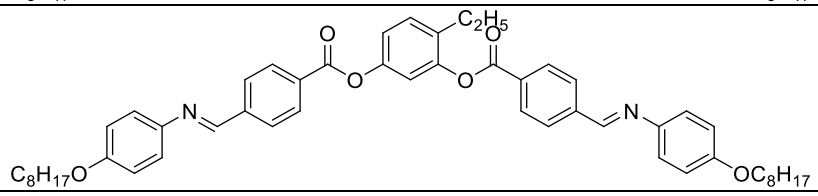
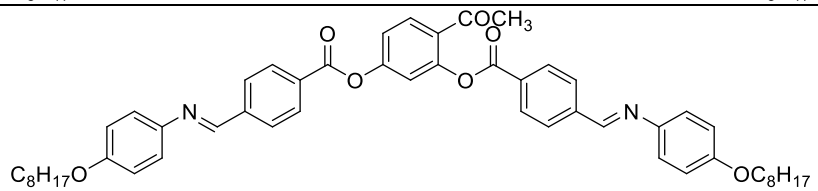
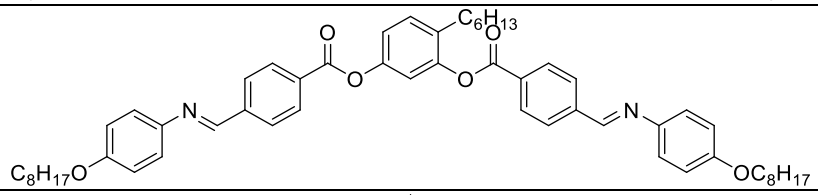
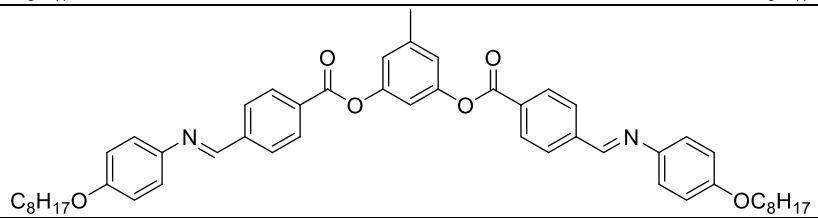
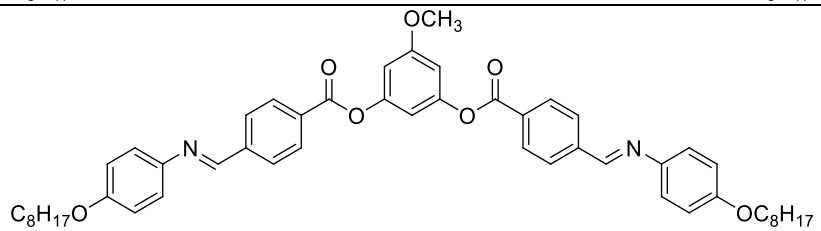
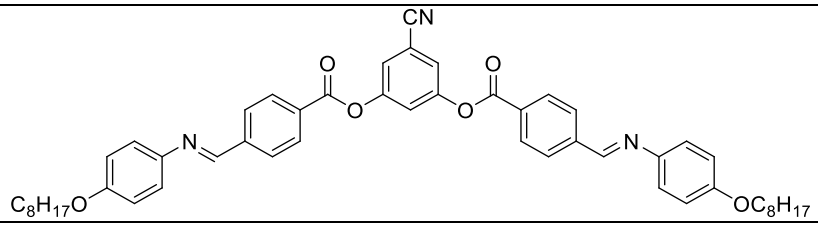
106		LC	7
			
107	R = C_8H_{17}	LC	8
108	R = $C_{10}H_{21}$	LC	8
109	R = $C_{12}H_{25}$	LC	8
110	R = $C_{14}H_{29}$	LC	8
111*	R = $C_{16}H_{33}$	LC	8
112	R = $C_{18}H_{37}$	LC	8
113		NLC	9
			
114	R = C_8H_{17}	LC	9
115	R = $C_{10}H_{21}$	LC	9
116	R = $C_{12}H_{25}$	LC	9
117	R = $C_{14}H_{29}$	LC	9
			
118	R = $C_{12}H_{25}$; R_1 = $C_{14}H_{29}$	NLC	10
119	R = $C_{12}H_{25}$; R_1 = $C_{16}H_{33}$	LC	10
120	R = $C_{16}H_{33}$; R_1 = $C_{12}H_{25}$	LC	10
121	R = $C_{16}H_{33}$; R_1 = $C_{16}H_{33}$	LC	10
			
122	R = C_4H_9	NLC	10
123*	R = C_5H_{11}	LC	10
124	R = C_6H_{13}	LC	10

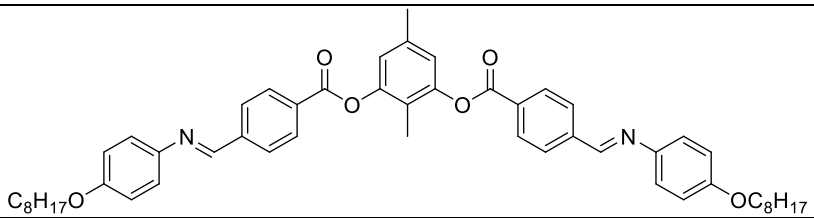
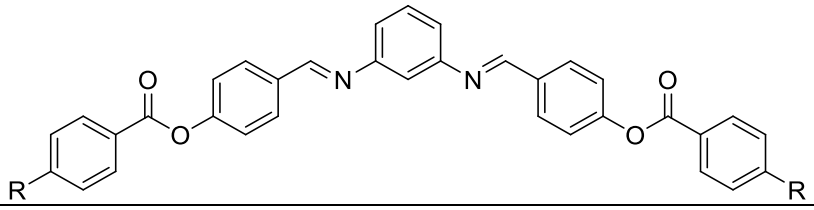
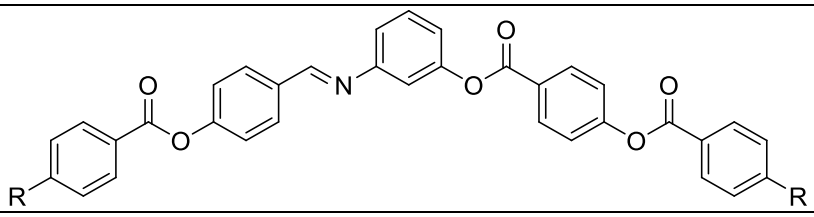
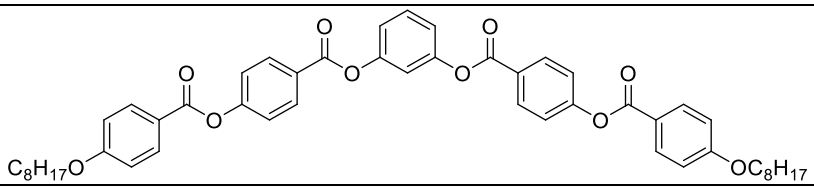
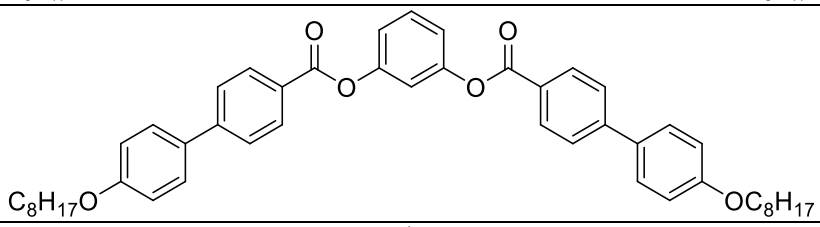
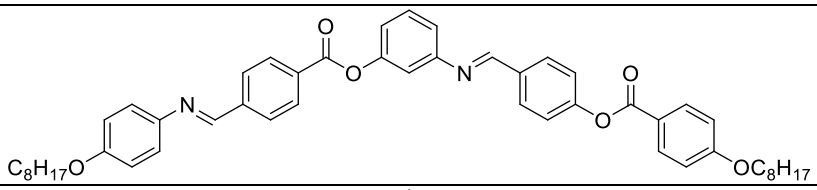
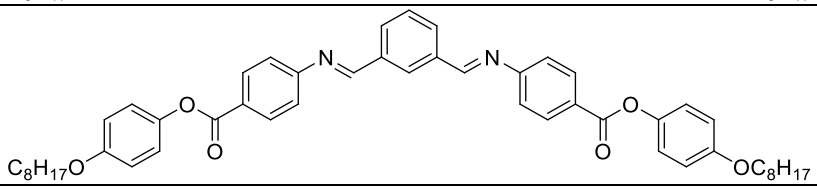
125	R = C_7H_{15}	LC	[10]
126	R = C_8H_{17}	LC	[10]
127	R = C_9H_{19}	LC	[10]
128*	R = $C_{10}H_{21}$	NLC	[10]
129	R = $C_{11}H_{23}$	LC	[10]
130	R = $C_{12}H_{25}$	LC	[10]
131	R = $C_{14}H_{29}$	LC	[10]
132	R = $C_{16}H_{33}$	LC	[10]
			
133	R = C_4H_9	LC	[11]
134	R = C_5H_{11}	LC	[11]
135	R = C_6H_{13}	LC	[11]
136	R = C_7H_{15}	LC	[11]
137	R = C_8H_{17}	LC	[11]
138*	R = C_9H_{19}	LC	[11]
139*	R = $C_{10}H_{21}$	LC	[11]
140*	R = $C_{11}H_{23}$	LC	[11]
141	R = $C_{12}H_{25}$	LC	[11]
142	R = $C_{14}H_{29}$	LC	[11]
143	R = $C_{16}H_{33}$	LC	[11]
			
144	R = C_4H_9	LC	[11]
145*	R = C_5H_{11}	LC	[11]
146	R = C_6H_{13}	LC	[11]
147	R = C_7H_{15}	LC	[11]
148	R = C_8H_{17}	LC	[11]
149	R = C_9H_{19}	LC	[11]
150	R = $C_{10}H_{21}$	LC	[11]
151	R = $C_{11}H_{23}$	LC	[11]
152	R = $C_{12}H_{25}$	LC	[11]
153	R = $C_{14}H_{29}$	LC	[11]
154	R = $C_{16}H_{33}$	LC	[11]
			
155	R = OC_4H_9	LC	[12]

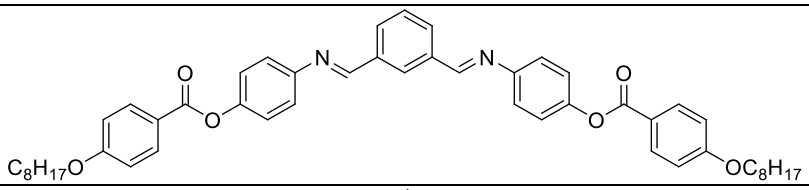
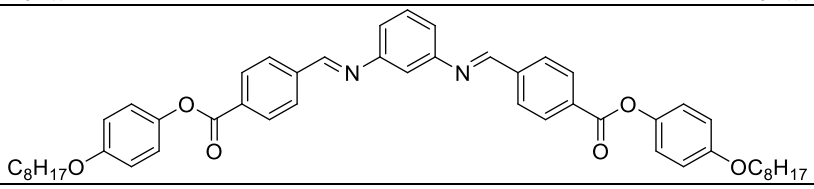
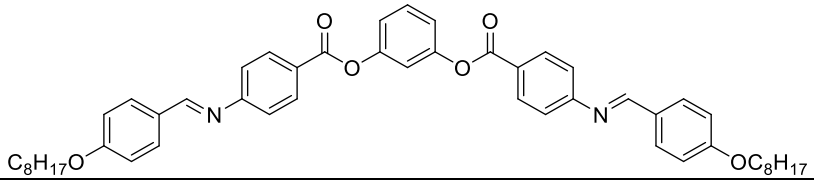
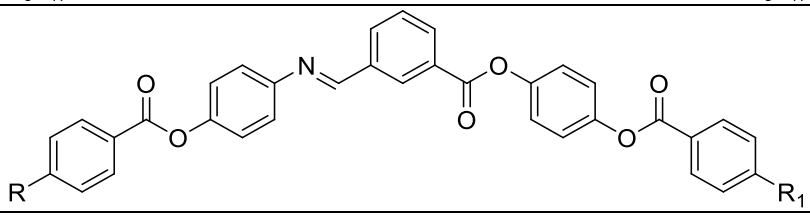
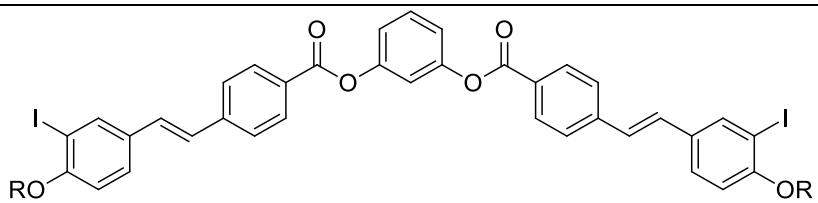
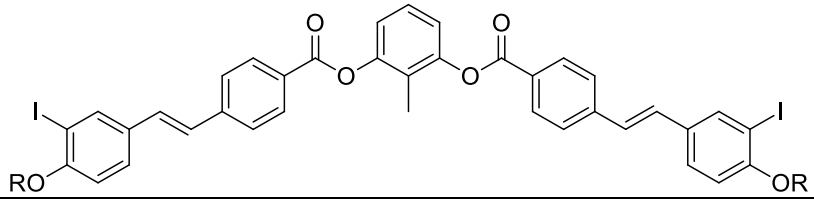
156	R = OC_6H_{13}	LC	[12]
157	R = OC_8H_{17}	LC	[12]
158	R = $OC_{10}H_{21}$	LC	[12]
159	R = $OC_{12}H_{25}$	LC	[12]
160	R = C_4H_9	LC	[12]
161	R = C_6H_{13}	LC	[12]
162	R = C_8H_{17}	LC	[12]
163	R = $C_{10}H_{21}$	LC	[12]
164	R = $C_{12}H_{25}$	LC	[12]
			
165	R = C_4H_9	LC	[13]
166*	R = C_6H_{13}	LC	[13]
167	R = C_8H_{17}	LC	[13]
168	R = $C_{10}H_{21}$	LC	[13]
169	R = $C_{12}H_{25}$	LC	[13]
170	R = OC_4H_9	LC	[13]
171	R = OC_6H_{13}	LC	[13]
172	R = OC_8H_{17}	LC	[13]
173	R = $OC_{10}H_{21}$	LC	[13]
174	R = $OC_{12}H_{25}$	LC	[13]
			
175	R = C_6H_{13}	NLC	[14]
176	R = C_8H_{17}	LC	[14]
177	R = $C_{10}H_{21}$	LC	[14]
178	R = $C_{12}H_{25}$	LC	[14]
			
179	R = C_8H_{17}	LC	[15]
180	R = $C_{10}H_{21}$	LC	[15]
181	R = $C_{12}H_{25}$	LC	[15]
182*	R = $C_{14}H_{29}$	LC	[15]
183	R = $C_{16}H_{33}$	LC	[15]

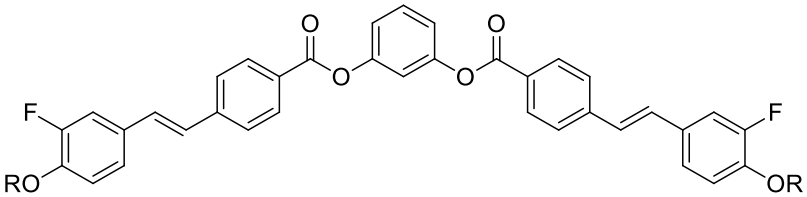
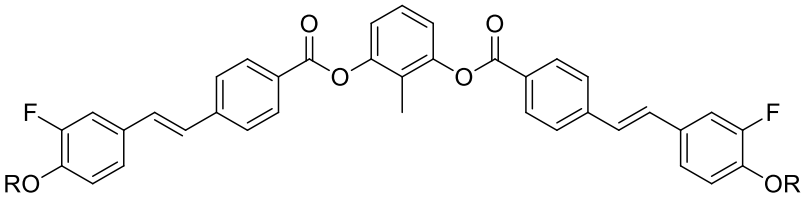
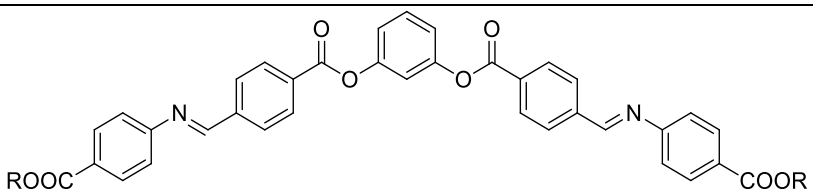
184*		NLC	15
185		NLC	15
186		NLC	15
187		NLC	15
188		NLC	15
			
189	R = OC_5H_{11}	LC	16
190	R = OC_6H_{13}	LC	16
191	R = OC_7H_{15}	LC	16
192	R = OC_8H_{17}	LC	16
193	R = $OC_{10}H_{21}$	LC	16
194*	R = $OC_{12}H_{25}$	LC	16
195	R = $OC_{16}H_{33}$	LC	16
196	R = C_5H_{11}	LC	16
197	R = C_6H_{13}	LC	16
198	R = C_8H_{17}	LC	16
199	R = $C_{10}H_{21}$	LC	16
200	R = $C_{12}H_{25}$	LC	16
201	R = $C_{14}H_{29}$	LC	16

			
202	R = OC_6H_{13}	NLC	[16]
203	R = OC_7H_{15}	LC	[16]
204	R = OC_8H_{17}	LC	[16]
205*	R = OC_9H_{19}	LC	[16]
206	R = $OC_{10}H_{21}$	LC	[16]
207	R = $OC_{12}H_{25}$	LC	[16]
208*	R = C_8H_{17}	LC	[16]
209	R = $C_{10}H_{21}$	LC	[16]
210	R = $C_{12}H_{25}$	LC	[16]
211	R = $C_{14}H_{29}$	LC	[16]
			
212*	R = OC_6H_{13}	LC	[16]
213	R = OC_7H_{15}	LC	[16]
214	R = OC_8H_{17}	LC	[16]
215	R = OC_9H_{19}	LC	[16]
216	R = $OC_{10}H_{21}$	LC	[16]
217	R = $OC_{12}H_{25}$	LC	[16]
218	R = $C_{14}H_{29}$	LC	[16]
			
219	R = OC_6H_{13}	NLC	[16]
220	R = OC_7H_{15}	LC	[16]
221	R = OC_8H_{17}	LC	[16]
222	R = OC_9H_{19}	LC	[16]
223	R = $OC_{10}H_{21}$	LC	[16]
224*	R = $OC_{12}H_{25}$	LC	[16]
			
225*	R = OC_6H_{13}	LC	[16]
226	R = OC_7H_{15}	LC	[16]

227	R = OC_8H_{17}	LC	16
228*	R = OC_9H_{19}	LC	16
229	R = $OC_{12}H_{25}$	LC	16
230*		NLC	16
231		NLC	16
232		NLC	16
233		NLC	16
234		NLC	16
235		NLC	16
236		NLC	16
237		NLC	16

238		NLC	16
			
239	R = OC_6H_{13}	LC	16
240*	R = OC_7H_{15}	LC	16
241	R = OC_8H_{17}	LC	16
242*	R = OC_9H_{19}	LC	16
243*	R = $OC_{10}H_{21}$	LC	16
			
244	R = OC_6H_{13}	LC	16
245	R = OC_7H_{15}	LC	16
246	R = OC_8H_{17}	LC	16
247	R = OC_9H_{19}	LC	16
248		LC	16
249		NLC	16
250		NLC	16
251		LC	16

252		LC	16
253		LC	16
254		NLC	16
			
255	R = $OC_{12}H_{25}$; R_1 = OC_8H_{17}	LC	16
256	R = $OC_{12}H_{25}$; R_1 = OC_9H_{19}	LC	16
257	R = $OC_{12}H_{25}$; R_1 = $OC_{10}H_{21}$	LC	16
258	R = $OC_{12}H_{25}$; R_1 = $OC_{11}H_{23}$	LC	16
259	R = $OC_{12}H_{25}$; R_1 = $OC_{12}H_{25}$	LC	16
260	R = $OC_{12}H_{25}$; R_1 = $OC_{13}H_{27}$	LC	16
261	R = $OC_{12}H_{25}$; R_1 = $OC_{14}H_{29}$	LC	16
262	R = $OC_{10}H_{21}$; R_1 = OC_8H_{17}	LC	16
263	R = $OC_{10}H_{21}$; R_1 = OC_9H_{19}	LC	16
264	R = $OC_{10}H_{21}$; R_1 = $OC_{10}H_{21}$	LC	16
			
265	R = OC_8H_{17}	NLC	17
266	R = $C_{10}H_{21}$	NLC	17
267*	R = $C_{11}H_{23}$	NLC	17
268	R = $C_{12}H_{25}$	LC	17
269	R = $C_{13}H_{27}$	LC	17
			
270	R = OC_8H_{17}	NLC	17
271	R = $C_{10}H_{21}$	NLC	17

272	R = $C_{11}H_{23}$	NLC	[17]
273*	R = $C_{12}H_{25}$	NLC	[17]
			
274	R = OC_8H_{17}	LC	[17]
275	R = $C_{10}H_{21}$	LC	[17]
276	R = $C_{11}H_{23}$	LC	[17]
277	R = $C_{12}H_{25}$	LC	[17]
278	R = $C_{13}H_{27}$	LC	[17]
			
279	R = OC_8H_{17}	LC	[17]
280	R = $C_{10}H_{21}$	LC	[17]
281	R = $C_{11}H_{23}$	LC	[17]
282	R = $C_{12}H_{25}$	LC	[17]
283*	R = $C_{13}H_{27}$	LC	[17]
			
284	R = C_6H_{13}	NLC	[18]
285*	R = C_7H_{15}	LC	[18]
286*	R = C_8H_{17}	LC	[18]
287	R = C_9H_{19}	LC	[18]
288*	R = $C_{10}H_{21}$	LC	[18]
289	R = $C_{11}H_{23}$	LC	[18]
290	R = $C_{12}H_{25}$	LC	[18]
291	R = $C_{13}H_{27}$	LC	[18]
292*	R = $C_{14}H_{29}$	LC	[18]
293*	R = $C_{16}H_{33}$	LC	[18]
294	R = $C_{18}H_{37}$	LC	[18]

* Compounds used in prediction set

Molecular descriptors: Basic terms

The terminology used in the field of molecular descriptors is well described by Balaban and Ivanciuc[19]:

"Chemical formulas may be viewed as graphs that is as non-empty sets *V* and *E* of elements. Elements *V* are called vertices and symbolize atoms. Elements *E* are called edges; they are binary relations between elements *V*, i.e. unordered pairs, and they symbolize covalent bonds between atoms. Unless otherwise stated, hydrogen atoms will be ignored as organic chemists usually do when they write a benzene ring as a hexagon; such a hydrogen-depleted graph includes only non-hydrogen atoms. Two vertices connected by an edge are called adjacent, and the graph can be uniquely described by its adjacent matrix **A**. This matrix has elements a_{ij} equal to 1 for adjacent vertices *i* and *j*, and zero otherwise. It was initially called the Hückel matrix. A path is a succession of non-repeating edges such that there is no break between edges. Walks can have repeating edges. Chemical graphs are connected graphs, as there is at least one path from any vertex to any other vertex of the graph. The (topological) distance d_{ij} between two vertices *i* and *j* is the number of edges along the shortest path between these vertices. The graph is also uniquely determined by its distance matrix **D**, whose elements are distances d_{ij}."

The LC equations of created 2D-MARS (Eq. (S1)) and 2&3D-MARS (Eq. (S2)) models:

$$2D\text{-}MARS\ (LC) = 1.13 - 9.71 \cdot B_1 + 1.84 \cdot 10^{-2} \cdot B_2 - 1.63 \cdot 10^{-3} \cdot B_3 \cdot B_4 - 8.43 \cdot 10^{-4} \cdot B_5 \cdot B_3 \cdot B_6 + 3.34 \cdot 10^{-3} \cdot B_7 \cdot B_3 \cdot B_6 - 2.18 \cdot 10^{-3} \cdot B_8 \cdot B_3 \cdot B_6 - 3.11 \cdot 10^{-2} \cdot B_9 \cdot B_3 \cdot B_6 + 3.21 \cdot B_9 \cdot B_{10} \cdot B_3 \cdot B_6 + 2.94 \cdot 10^{-1} \cdot B_9 \cdot B_{11} \cdot B_3 \cdot B_6 - 5.68 \cdot 10^{-2} \cdot B_{12} \cdot B_3 \cdot B_6 + 3.00 \cdot 10^{-4} \cdot B_8 \cdot B_{13} \cdot B_3 \cdot B_6 - 4.57 \cdot 10^{-4} \cdot B_8 \cdot B_{14} \cdot B_3 \cdot B_6 - 8.12 \cdot 10^{-1} \cdot B_9 \cdot B_{15} \cdot B_3 \cdot B_6 + 1.75 \cdot 10^{-3} \cdot B_{16} \cdot B_3 \cdot B_4 + 9.24 \cdot 10^{-3} \cdot B_{17} \cdot B_3 \cdot B_4 + 4.37 \cdot 10^{-4} \cdot B_{18} \cdot B_{12} \cdot B_3 \cdot B_6 + 1.26 \cdot 10^{-4} \cdot B_{19} \cdot B_{12} \cdot B_3 \cdot B_6 - 2.43 \cdot B_{20} \cdot B_{21} - 2.87 \cdot 10^{-1} \cdot B_{22} \cdot B_9 \cdot B_3 \cdot B_6 - 5.58 \cdot 10^{-5} \cdot B_{23} \cdot B_9 \cdot B_3 \cdot B_6 + 3.51 \cdot B_{24} \cdot B_3 \cdot B_6 + 1.16 \cdot 10^{1} \cdot B_{25} \cdot B_3 \cdot B_6 - 3.10 \cdot B_{25} \cdot B_{26} \cdot B_3 \cdot B_6 + 6.15 \cdot 10^{-2} \cdot B_9 \cdot B_{27} \cdot B_3 \cdot B_6 \quad \textbf{(S1)}$$

$$2\&3D\text{-}MARS\ (LC) = 9.02 \cdot 10^{-2} - 5.69 \cdot 10^{1} \cdot B_1 + 8.80 \cdot B_2 - 1.81 \cdot 10^{-3} \cdot B_3 \cdot B_4 - 7.83 \cdot 10^{-2} \cdot B_3 \cdot B_5 - 6.83 \cdot 10^{-1} \cdot B_3 \cdot B_6 - 3.16 \cdot 10^{-2} \cdot B_3 \cdot B_7 + 1.26 \cdot 10^{-1} \cdot B_3 \cdot B_8 + 8.33 \cdot 10^{-2} \cdot B_3 \cdot B_9 - 1.64 \cdot B_{10} - 6.00 \cdot B_{11} \cdot B_{12} - 7.08 \cdot 10^{-1} \cdot B_{11} \cdot B_{13} + 2.44 \cdot 10^{-1} \cdot B_{13} - 6.09 \cdot 10^{-2} \cdot B_2 \cdot B_{15} - 3.19 \cdot 10^{-3} \cdot B_2 \cdot B_{16} - 8.74 \cdot 10^{1} \cdot B_{17} \cdot B_{18} - 7.65 \cdot 10^{-2} \cdot B_{19} \cdot B_{18} - 1.03 \cdot B_{20} \cdot B_3 - 2.33 \cdot 10^{-5} \cdot B_{21} \cdot B_3 - 5.30 \cdot 10^{1} \cdot B_{11} \cdot B_{22} - 1.94 \cdot 10^{-2} \cdot B_{23} \cdot B_3 + 2.23 \cdot 10^{-2} \cdot B_{24} - 3.75 \cdot 10^{-1} \cdot B_{25} \cdot B_3 + 1.12 \cdot 10^{-1} \cdot B_{26} \cdot B_1 \quad \textbf{(S2)}$$

Table S2. List of descriptors used in the MARS models with short description

Group	Label	Short description
2D descriptors		
Path Count	MPC10	Molecular path count of order 10
Wiener Numbers	WPOL	Wiener polarity number
BCUT	BCUTp-1h	nlow highest polarizability weighted BCUTS
Chi Path	VP-4	Valence path, order 4
	VP-6	Valence path, order 6
	AVP-0	Average valence path, order 0
Chi Chain	SCH-7	Simple chain, order 7
Chi Path Cluster	VPC-4	Valence path cluster, order 4
	VC-5	Valence cluster, order 5
Detour Matrix	VE1_Dt	Coefficient sum of the last eigenvector from detour matrix
	VE2_Dt	Average coefficient sum of the last eigenvector from detour matrix
	VE3_Dt	Logarithmic coefficient sum of the last eigenvector from detour matrix
Molecular distance edge	MDEC-12	Molecular distance edge between all primary and secondary carbons
Barysz Matrix	VE3_Dzv	Logarithmic coefficient sum of the last eigenvector from Barysz matrix weighted by van der Waals volumes
	VE3_Dzs	Logarithmic coefficient sum of the last eigenvector from Barysz matrix weighted by I-state
	VE3_Dze	Logarithmic coefficient sum of the last eigenvector from Barysz matrix weighted by Sanderson electronegativities
	SpAbs_Dzp	Graph energy from Barysz matrix weighted by polarizabilities
Information Content	IC2	Information content index (neighbourhood symmetry of 2-order)
	MIC5	Modified information content index (neighbourhood symmetry of 5-order)
Molecular linear free energy relation	MLFER_S	Combined dipolarity/polarizability
	MLFER_L	Solute gas-hexadecane partition coefficient
	MLFER_BO	Overall or summation solute hydrogen bond basicity
Extended Topochemical Atom	ETA_EtaP_F	Functionality index EtaF relative to molecular size
	ETA_Eta_R	Composite index Eta for reference alkane
3D descriptors		
Petitjean Shape Index	geomDiameter	Geometrical diameter (maximum geometric eccentricity)
Charged partial surface area	WNSA-2	Partial negative surface area * total negative charge on the molecule * total molecular surface area / 1000
	RPCS	Relative positive charge surface area = most positive surface area * relative positive charge
WHIM	E3s	3rd component accessibility directional WHIM index weighted by relative I-state
	P2m	2nd component shape directional WHIM index weighted by relative mass

Table S3. Short description of descriptors selected using feature selection and genetic algorithms

Label	Short description
2D descriptors	
SM1_Dzi	Spectral moment of order 1 from Barysz matrix weighted by first ionization potential
SM1_Dzs	Spectral moment of order 1 from Barysz matrix weighted by I-state
SM1_Dzp	Spectral moment of order 1 from Barysz matrix weighted by polarizabilities
EE_Dzi	Estrada-like index from Barysz matrix weighted by first ionization potential (ln(1+x))
EE_Dzm	Estrada-like index from Barysz matrix weighted by mass (ln(1+x))
EE_Dzv	Estrada-like index from Barysz matrix weighted by van der Waals volumes (ln(1+x))
BCUTp-1h	nlow highest polarizability weighted BCUTS
BCUTc-1h	nlow highest partial charge weighted BCUTS
C1SP3	Singly bound carbon bound to one other carbon
C3SP2	Doubly bound carbon bound to three other carbons
SCH-6	Simple chain, order 6
SCH-7	Simple chain, order 7
VCH-6	Valence chain, order 6
VCH-7	Valence chain, order 7
VC-3	Valence cluster, order 3
VC-5	Valence cluster, order 5
VPC-4	Valence path cluster, order 4
VPC-5	Valence path cluster, order 5
VPC-6	Valence path cluster, order 6
piPC5	Conventional bond order ID number of order 5 (ln(1+x)
piPC7	Conventional bond order ID number of order 7 (ln(1+x)
TpiPC	Total conventional bond order (up to order 10) (ln(1+x))
MPC8	Molecular path count of order 8
ETA_Beta	A measure of electronic features of the molecule
ETA_dAlpha_B	A measure of count of hydrogen bond acceptor atoms and/or polar surface area
ETA_dPsi_A	A measure of hydrogen bonding propensity of the molecules
ETA_BetaP	A measure of electronic features of the molecule relative to molecular size
ETA_Beta_ns_d	A measure of lone electrons entering into resonance
MDEC-13	Molecular distance edge between all primary and tertiary carbons
MLFER_E	Excessive molar refraction
GGI6	Topological charge index of order 6
GGI9	Topological charge index of order 9
JGI9	Mean topological charge index of order 9
SIC3	Structural information content index (neighbourhood symmetry of 3-order)
Mare	Mean atomic Allred-Rochow electronegativities (scaled on carbon atom)
VE3_D	Logarithmic coefficient sum of the last eigenvector from topological distance matrix
VR2_Dzs	Normalized Randic-like eigenvector-based index from Barysz matrix weighted by I-state
3D descriptors	
Dp	D total accessibility index / weighted by relative polarizabilities
L1s	1st component size directional WHIM index weighted by relative I-state

The SKNN and CPNN optimization results. An example of bubble plot.

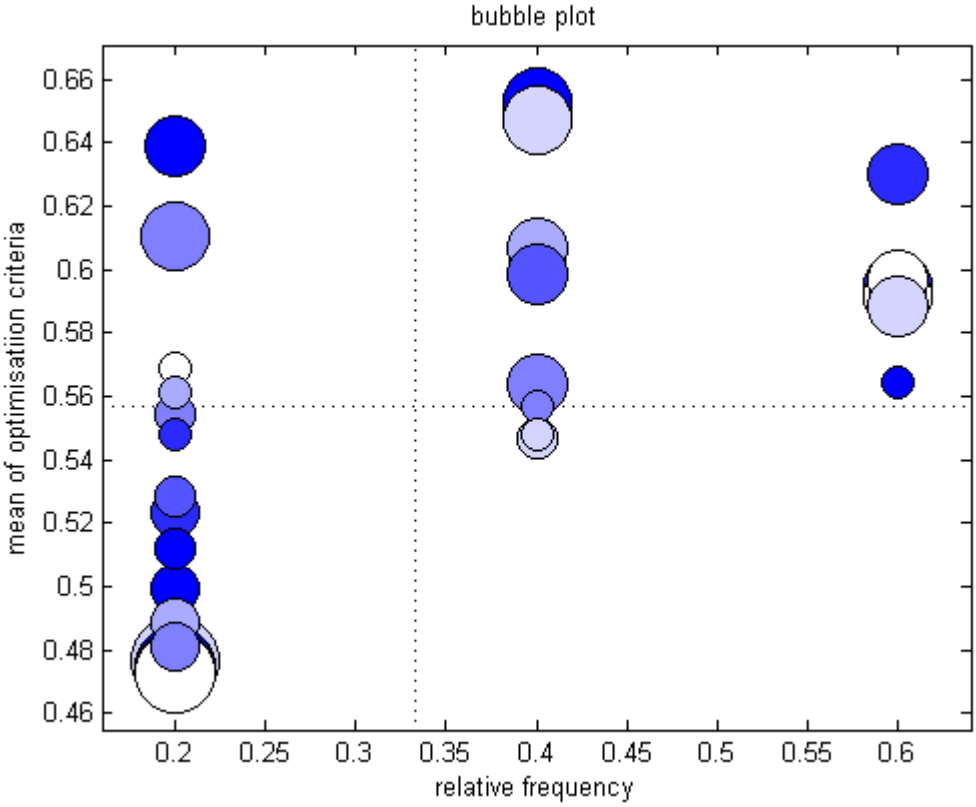


Figure S1. Plot of relative frequency of selection and mean value of optimization criteria used for the selection of the 2D-FLS-SKNN architecture and training parameters. The dimension of each bubble is proportional to the network size, the colour of the bubbles is proportional to the number of epochs; the darker colour means the higher number of epochs. The dotted line represents the overall means of frequencies and fitness function.

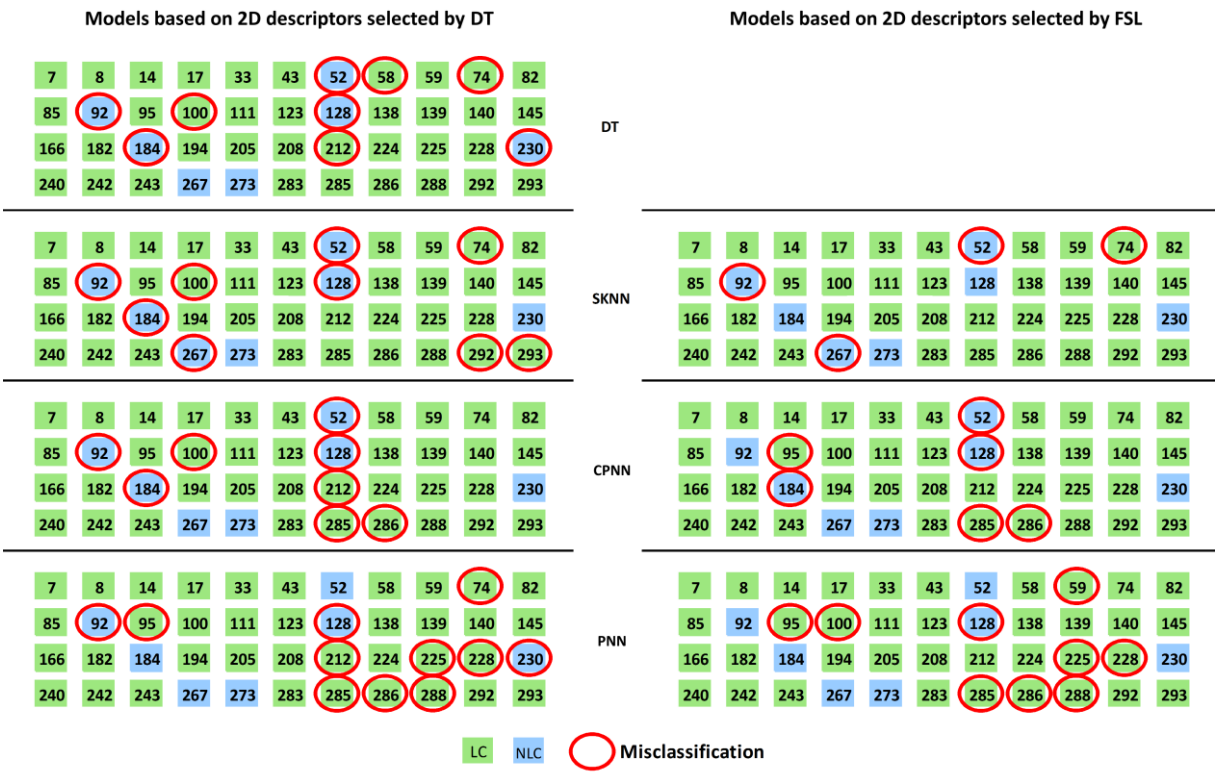


Figure S2. The prediction of LCs by DT model and ANN models, which are based on descriptors selected by DT and FSL

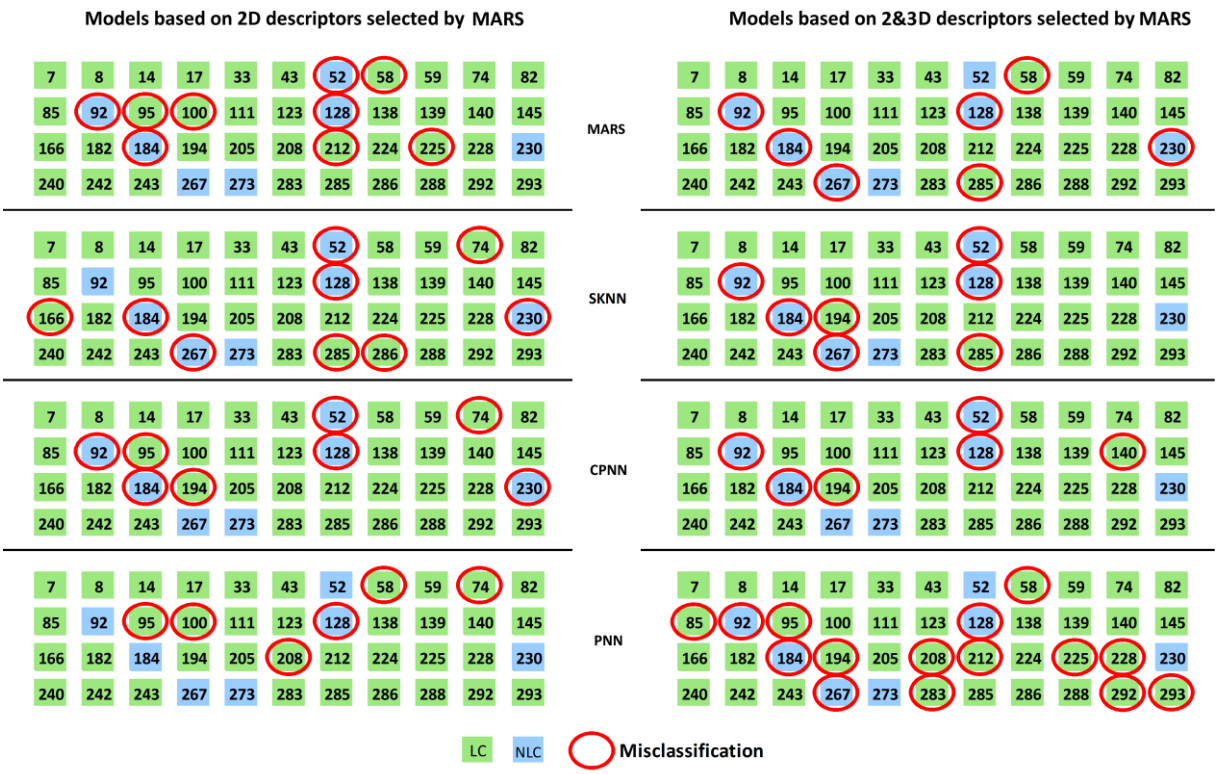


Figure S3. The prediction of LCs by MARS models and ANN models, which are based on descriptors selected by MARS

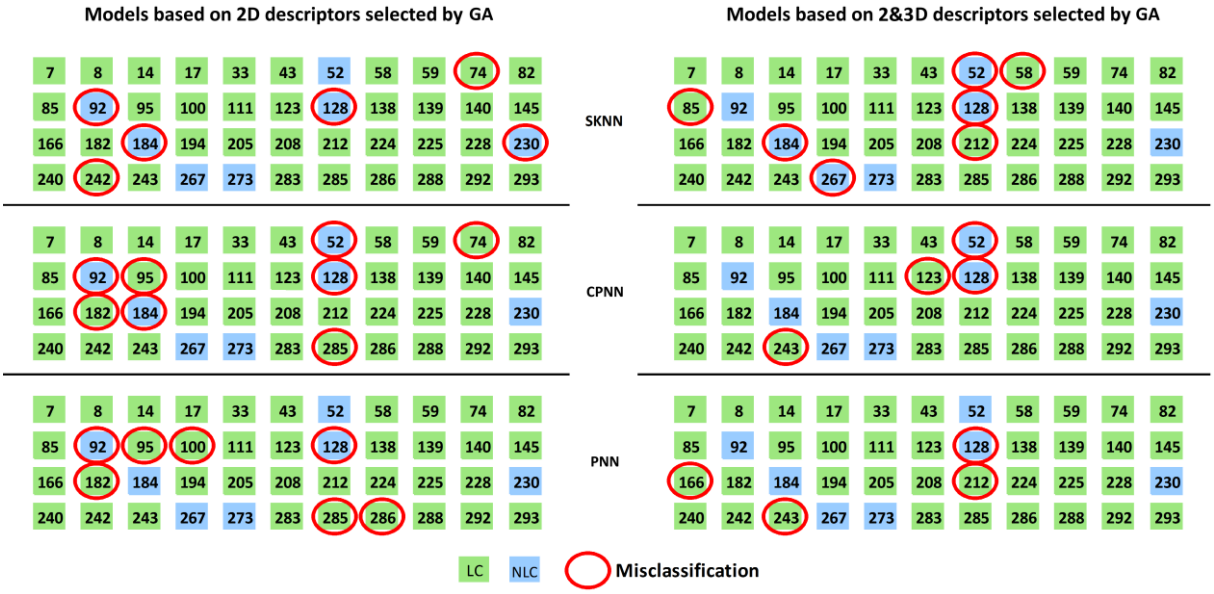


Figure S4. The prediction of LCs by ANN models based on descriptors selected by GAs

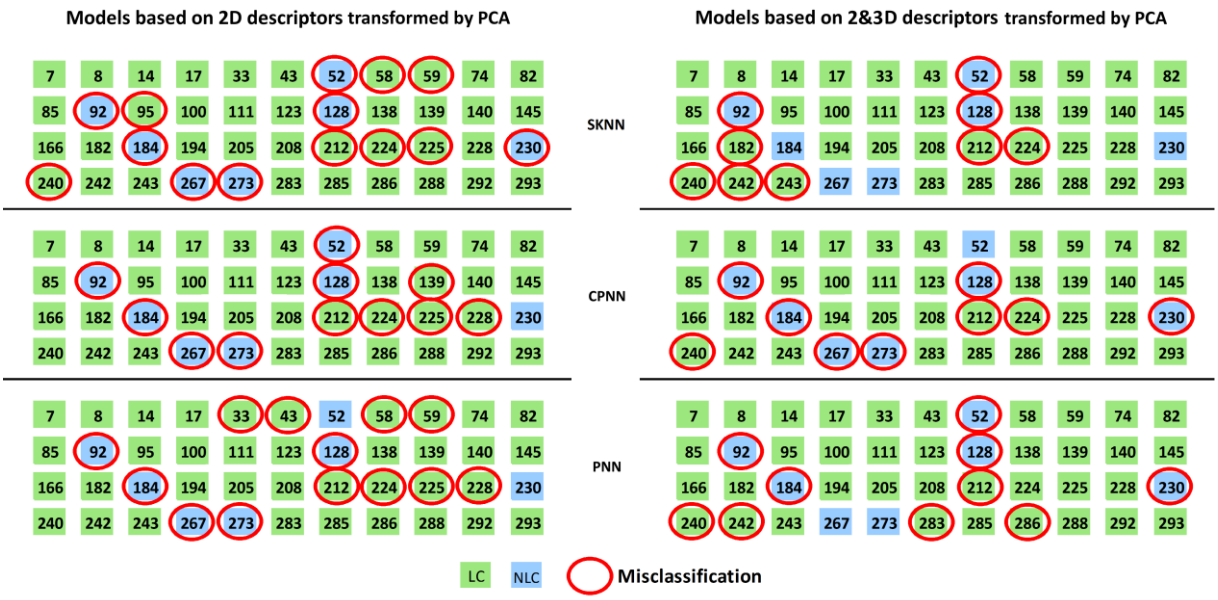


Figure S5. The prediction of LCs by ANN models based on descriptors selected by PCA

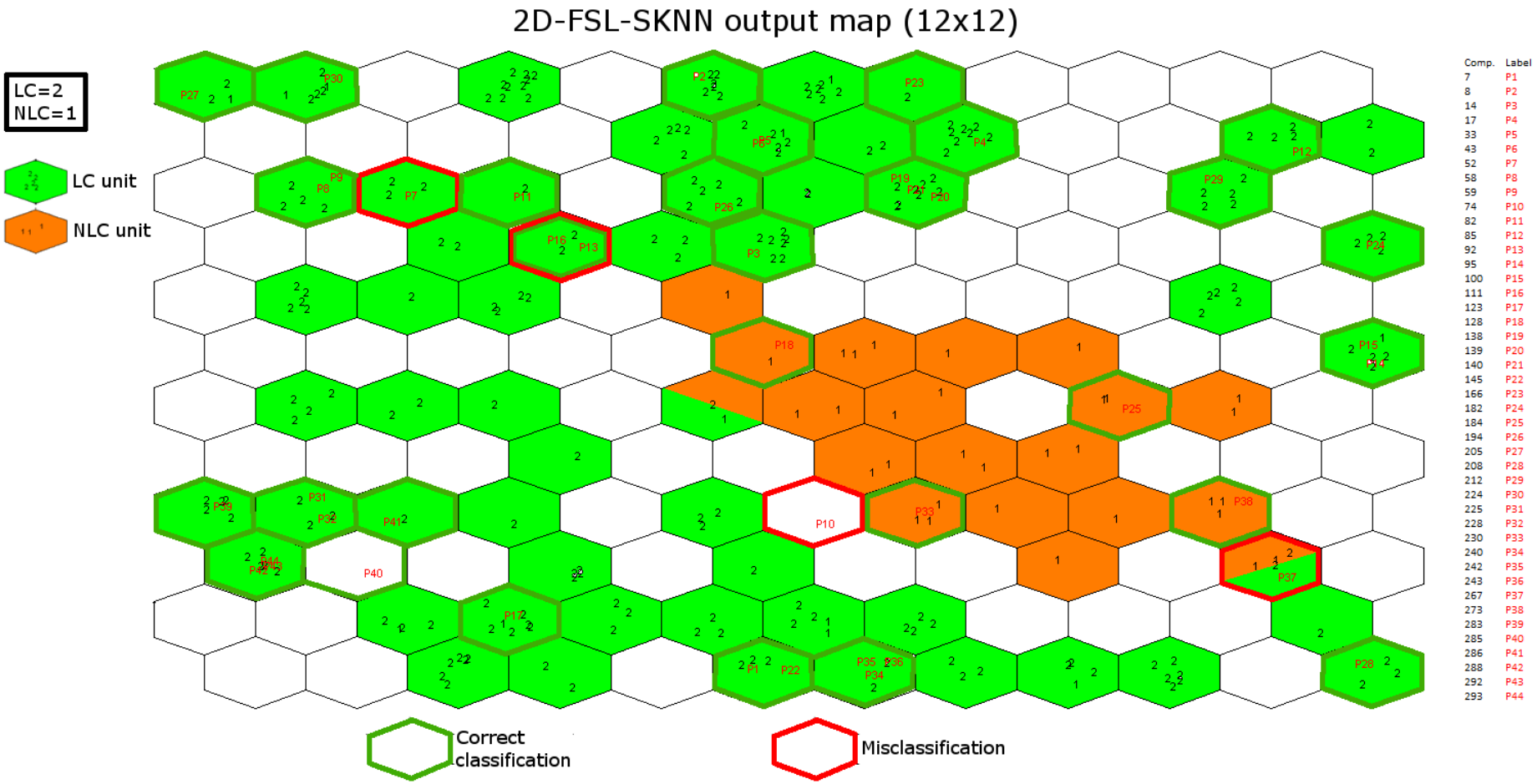


Figure S6. The 2D-FSL-SKNN output map (12x12) with the location of compounds used to test the model

Table S4. The 2&3D-GA-PNN output probabilities

Compound	Observed behavior	Probability of being LC (p_{LC}) (%)	Classification
7	LC	86	LC
8	LC	86	LC
14	LC	82	LC
17	LC	100	LC
33	LC	85	LC
43	LC	62	LC
52	NLC	10	NLC
58	LC	50	LC
59	LC	70	LC
74	LC	100	LC
82	LC	77	LC
85	LC	65	LC
92	NLC	18	NLC
95	LC	99	LC
100	LC	56	LC
111	LC	77	LC
123	LC	55	LC
128	**NLC**	**81**	**LC**
138	LC	99	LC
139	LC	99	LC
140	LC	99	LC
145	LC	83	LC
166	**LC**	**47**	**NLC**
182	LC	71	LC
184	NLC	27	NLC
194	LC	95	LC
205	LC	77	LC
208	LC	82	LC
212	**LC**	**19**	**NLC**
224	LC	51	LC
225	LC	62	LC
228	LC	76	LC
230	NLC	37	NLC
240	LC	68	LC
242	LC	57	LC
243	**LC**	**6**	**NLC**
267	NLC	20	NLC
273	NLC	5	NLC
283	LC	92	LC
285	LC	57	LC
286	LC	77	LC
288	LC	82	LC
292	LC	91	LC
293	LC	91	LC

References

(1) Srinivasan, M. V.; Kannan, P.; Roy, A. Photo and electrically switchable B7 mesophase exhibiting asymmetric bent-core liquid crystals. *New J. Chem.* **2013**, *37*, 1584–1590.

(2) Nagaveni, N. G.; Roy, A.; Prasad, V. Achiral bent-core azo compounds: effect of different types of linkage groups and their direction of linking on liquid crystalline properties. *J. Mater. Chem.* **2012**, *22*, 8948–8959.

(3) Vijay Srinivasan, M.; Kannan, P.; Roy, A. Investigations on photo and electrically switchable asymmetric bent-core liquid crystals. *J. Mater. Sci.* **2013**, *48*, 2433–2446.

(4) Alaasar, M.; Prehm, M.; Brautzsch, M.; Tschierske, C. 4-Methylresorcinol based bent-core liquid crystals with azobenzene wings – a new class of compounds with dark conglomerate phases. *J. Mater. Chem. C* **2014**, *2*, 5487–5501.

(5) Lutfor, M. R.; Hegde, G.; Kumar, S.; Tschierske, C.; Chigrinov, V. G. Synthesis and characterization of bent-shaped azobenzene monomers: Guest-host effects in liquid crystals with azo dyes for optical image storage devices. *Opt. Mater.* **2009**, *32*, 176–183.

(6) Alaasar, M.; Prehm, M.; Tschierske, C. Influence of halogen substituent on the mesomorphic properties of five-ring banana-shaped molecules with azobenzene wings. *Liq. Cryst.* **2013**, *40*, 656–668.

(7) Alaasar, M.; Prehm, M.; Brautzsch, M.; Tschierske, C. Dark conglomerate phases of azobenzene derived bent-core mesogens – relationships between the molecular structure and mirror symmetry breaking in soft matter. *Soft Matter* **2014**, *10*, 7285–7296.

(8) Alaasar, M.; Prehm, M.; May, K.; Eremin, A.; Tschierske, C. 4-Cyanoresorcinol-Based Bent-Core Mesogens with Azobenzene Wings: Emergence of Sterically Stabilized Polar Order in Liquid Crystalline Phases. *Adv. Funct. Mater.* **2014**, *24*, 1703–1717.

(9) Alaasar, M.; Prehm, M.; Tschierske, C. A new room temperature dark conglomerate mesophase formed by bent-core molecules combining 4-iodoresorcinol with azobenzene units. *Chem. Commun.* **2013**, *49*, 11062–11064.

(10) Nagaveni, N. G.; Raghuvanshi, P.; Roy, A.; Prasad, V. Azo-functionalised achiral bent-core liquid crystalline materials: effect of presence of –N=N– linkage at different locations in the molecular architecture. *Liq. Cryst.* **2013**, *40*, 1238–1254.

(11) Nagaveni, N. G.; Prasad, V.; Roy, A. Azo functionalised achiral bent-core liquid crystals: observation of photo-induced effects in B7 and B2 mesophases. *Liq. Cryst.* **2013**, *40*, 1405–1416.

(12) Prasad, V.; Kang, S.-W.; Qi, X.; Kumar, S. Photo-responsive and electrically switchable mesophases in a novel class of achiral bent-core azo compounds. *J. Mater. Chem.* **2004**, *14*, 1495–1502.

(13) Prasad, V.; Kang, S.-W.; Kumar, S. Novel examples of achiral bent-core azo compounds exhibiting B1 and anticlinic-antiferroelectric B2 mesophases. *J. Mater. Chem.* **2003**, *13*, 1259–1264.

(14) Prasad, V.; Jákli, A. Achiral bent-core azo compounds: observation of photoinduced effects in an antiferroelectric tilted smectic mesophase. *Liq. Cryst.* **2004**, *31*, 473–479.

(15) Alaasar, M.; Prehm, M.; Tschierske, C. New azobenzene containing bent-core liquid crystals based on disubstituted resorcinol. *Liq. Cryst.* **2014**, *41*, 126–136.

(16) Pelzl, G.; Diele, S.; Weissflog, W. Banana-shaped compounds - a new field of liquid crystals. *Adv. Mater.* **1999**, *11*, 707–724.

(17) Pyc, P.; Mieczkowski, J.; Pociecha, D.; Gorecka, E.; Donnio, B.; Guillon, D. Bent-core molecules with lateral halogen atoms forming tilted, synclinic and anticlinic, lamellar phases. *J. Mater. Chem.* **2004**, *14*, 2374–2379.

(18) Umadevi, S.; Jákli, A.; Sadashiva, B. K. Odd-even effects in bent-core compounds containing terminal n-alkyl carboxylate groups. *Soft Matter* **2006**, *2*, 875–885.

(19) Balaban, A. T.; Ivanciuc, O. Historical development of topological indices. In *Topological Indices and Related Descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers, Inc.: Netherlands, 1999; pp 21–57.