Electronic Supplementary Material (ESI) for Journal of Materials Chemistry A.

Rational design of hydrocarbon-based sulfonated copolymers for proton exchange membranes

Lunyang Liu[a,b], Wenduo Chen[a], Tingli Liu[a,c], Xiangxin Kong[a,c], Jifu Zheng[a] and Yunqi Li[a,c†]

[a] Key Laboratory of High-Performance Synthetic Rubber and its Composite Materials & Key Laboratory of Polymer Ecomaterials, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun 130022, PR China; [b] University of Chinese Academy of Sciences, Beijing 100049, PR China; [c] University of Science and Technology of China, Hefei 230026, China

[†] To whom correspondence should be addressed. E-mail: yunqi@ciac.ac.cn. Phone: +86 (0)431 85262535

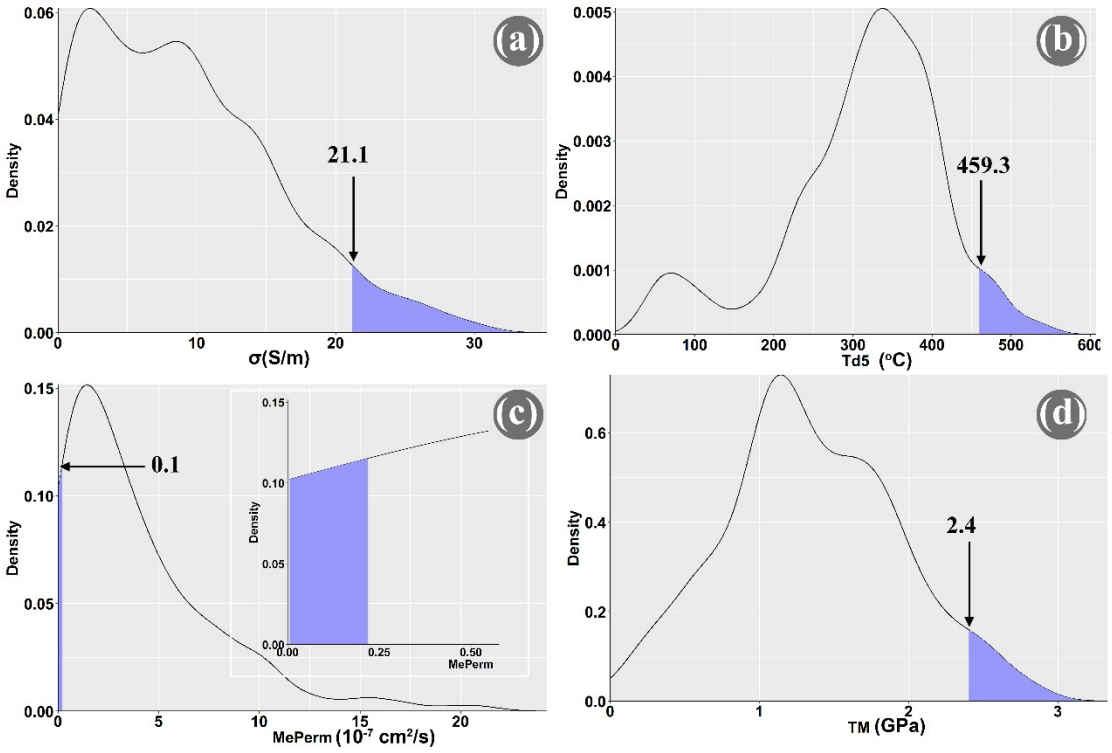

Fig. S1. Probability density function(PDF) for (a) proton conductivity(σ), (b) methanol permeability(MePerm), (c) the degradation temperature at 5% weight loss (Td5) and (d) tensile modulus(TM) of hydrocarbon-based sulfonated copolymer PEMs . Highlighted region with preferred significance (L_{sig}) of 0.05.

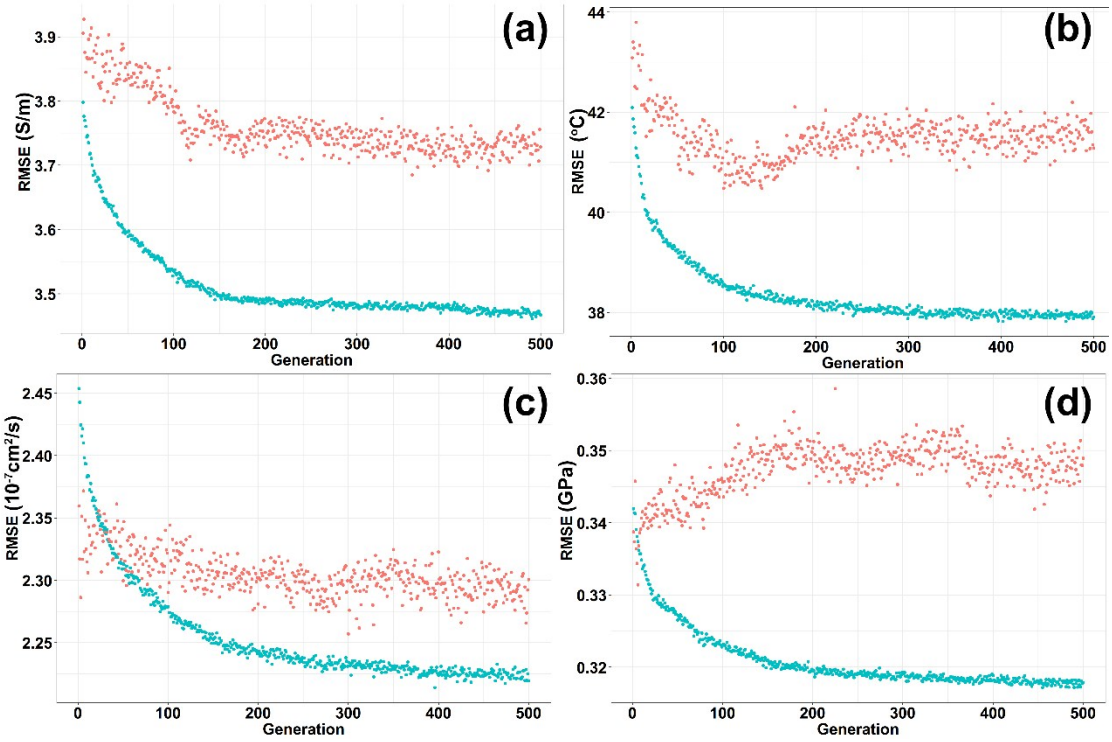

Fig. S2 Optimal combination of features to predict each of the (a) σ, (b) Td5, (c) MePerm and d (TM) four performance indexes. RMSE against the generation of GA was plotted in 5-fold cross validation for training and test. Red points refer to testing performance and blue points are training performance.

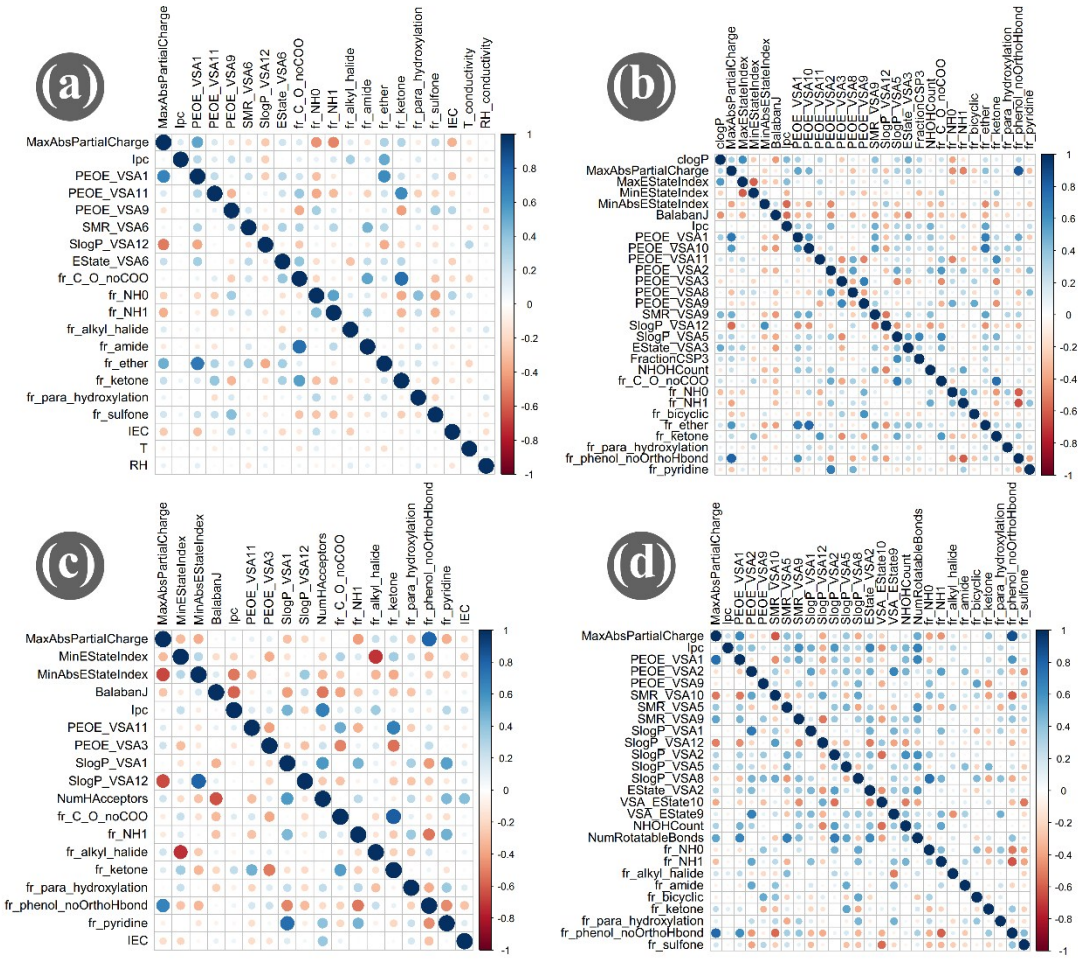

Fig. S3 Pearson ($CORR_p$, lower triangle) and Spearman ($CORR_{sp}$, upper triangle) correlation matrix for the selected features to predict σ (a),Td5 (b), MePerm (c), TM (d), respectively.

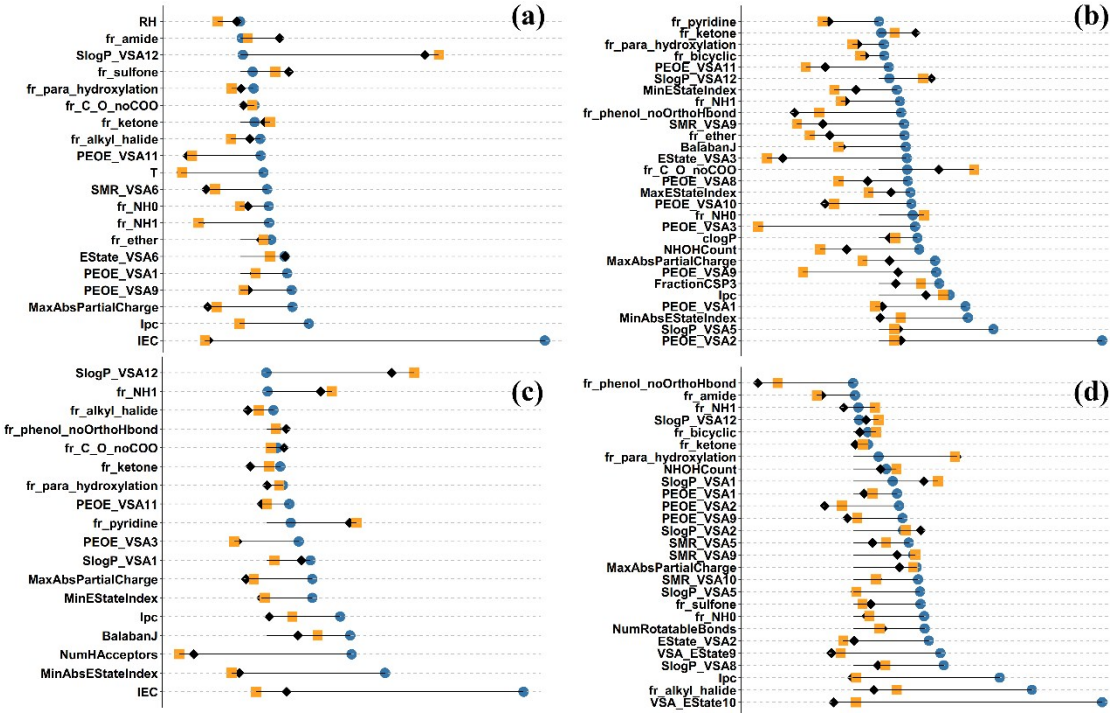

Fig. S4 Ranking of feature importance(blue circle) using RF models, Pearson R_p(black diamond) and Spearman R_{sp} (yellow square) correlation coefficients between features and performance indexes for σ(a), Td5(b), MePerm(c), TM(d).

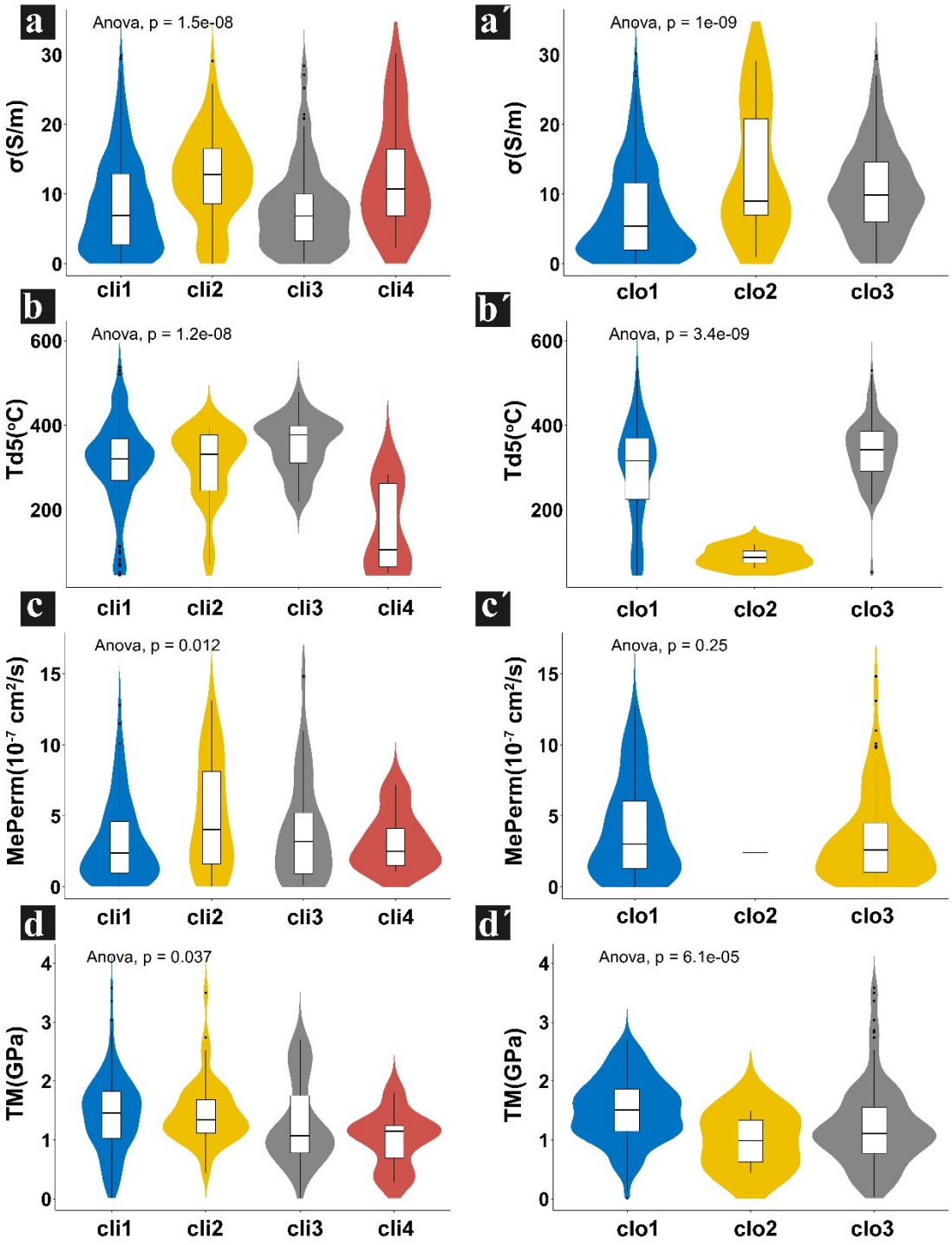

Fig. S5. The violin plots of the distributions of performance indexes contributed from the monomers in each cluster. ANOVA analysis of hydrophilic (a, b c, d) and hydrophobic (a´, b´, c´, d´) clusters. The box and violin plots show the distribution of membrane properties (σ, Td5, MePerm and TM) in each cluster.

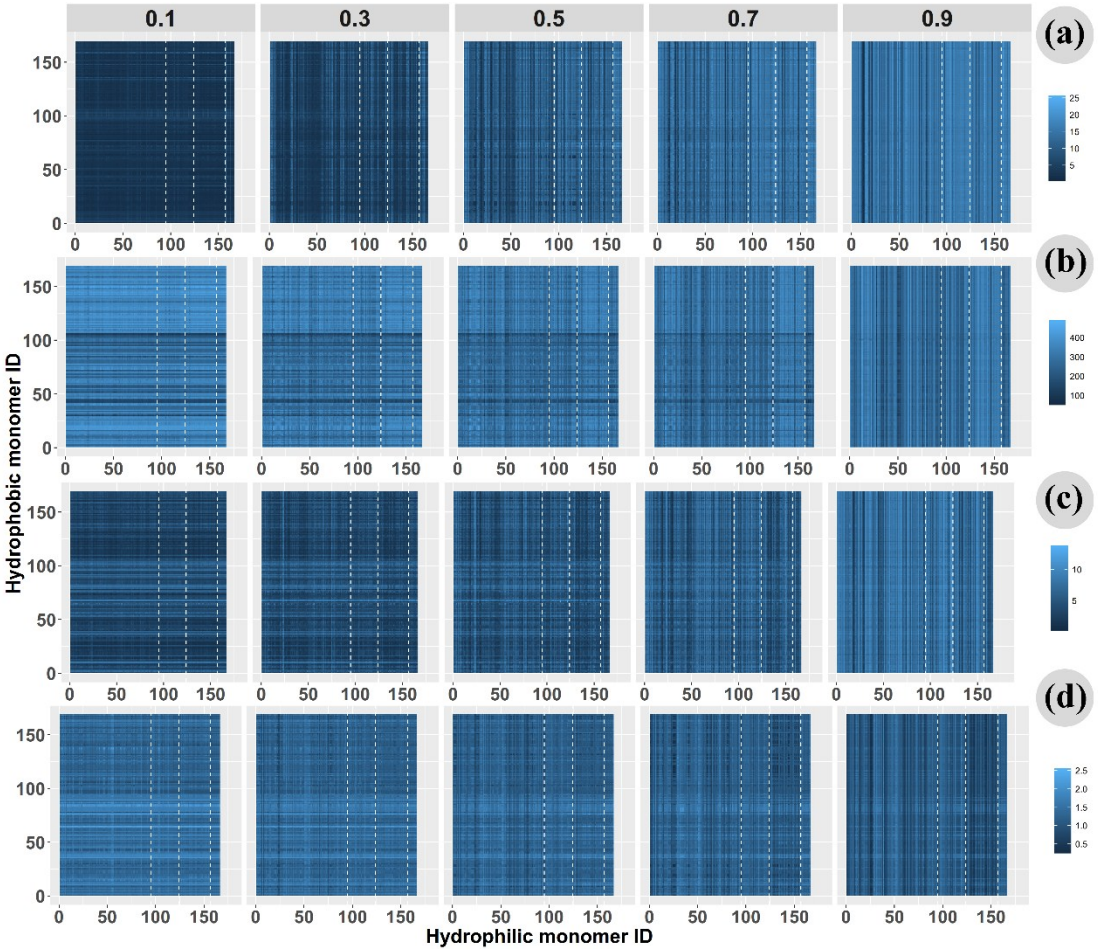

Fig. S6. Screening copolymers for σ (a), Td5(b), MePerm(c), TM(d) with RF model. The RF predicted performance indexes are mapped onto the color. The dashed vertical line separate hydrophilic monomer into four clusters, cli1 to cli4 from left to right using the same order as shown Fig. 3c. The x, y axis are monomer IDs which are presented in Table S1 and S2 in supporting information.

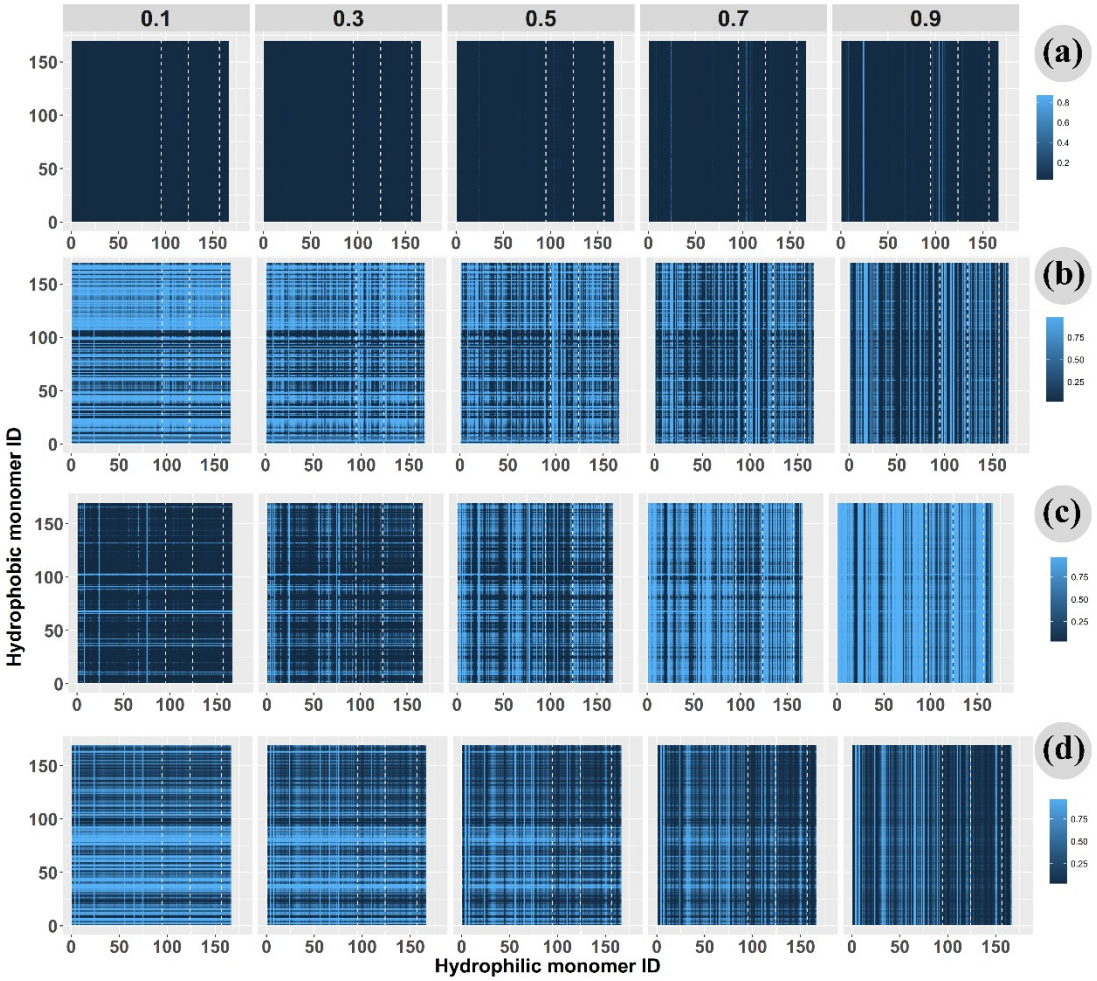

Fig. S7 Copolymer screening results of σ (a), Td5(b), MePerm(c), TM(d) with LR model. Each identical copolymer represented by a unique block and the predicted performance index are mapped onto the color of block. The dashed vertical line separate hydrophilic monomer into four clusters, the same order as shown Fig. 3c. The x, y axis are monomer IDs which are presented in Table S1 and S2.

Tables

Table S1 List of hydrophilic monomers.

ID	Name	ID	Name
1	S-PEEK-1	84	SEF
2	SPAEK-40	85	SPE4
3	SPAEKKS-1	86	SN30
4	SPAEK-X15Y4	87	SMP-PAEK-90
5	X4Y4	88	SPAES-1
6	SPEEKK-10	89	LSPA-0.49
7	sPAS-7/3	90	S2-PAES-40
8	SPEEKK-60	91	BM-1
9	PAES-2S-b60	92	SPAEK40-TRI
10	SPAE-9	93	SC-SPAE90
11	1a	94	PAEN80
12	SPI-B1	95	SPPBI-20
13	sPBI-60	96	SPI-4a
14	DQB50	97	SPI-4b
15	8a	98	SPI-8(50)
16	SPI-40	99	SPI-8-p
17	SPI-30	100	SPI-4c
18	II-b	101	SPI-4d
19	S-75	102	SPEKEBI-1
20	SPI 40	103	SPI-2.2
21	P-DTN70	104	SCT-SAPs-0.8

22	DAN-50	105	6F-SNPAEK-50
23	DHNH40	106	2-SPAES-50
24	DQN40	107	4-SPAES-25
25	SPBIBI-a(70)	108	SOPAES-0.3
26	SPBIBI-b(70)	109	M10N5-SO3H
27	SPPQ-80	110	SPAES-25
28	SPPQ(o)-80	111	SPFENO-20
29	SPI-1	112	D-SPAEENH-30
30	dsPPEPO-70	113	m-SPAEENH-50
31	SPI-N1	114	P-SPAEENH-50
32	SPAEK-20-COOH	115	SPI90-TFVBPA10
33	S-FPEK(HF90)	116	CSPI90-6FATFVP10
34	NB-90	117	sPBT70
35	IBQS-20	118	sPBT-PE52.5
36	S4-PAES-30	119	P1-1.0
37	SPE1	120	SPAE20
38	SPAEK-20-BI	121	If-60
39	sPEKkP-450	122	sPBT-6FA45
40	PAES-2S-a60	123	SPBT-PS50
41	3a-1	124	SPBT-6F70
42	3b-1	125	S-PES-40
43	S4PH-20-PS_1	126	B20V80-SDPA
44	S4PH-20-PS_2	127	SPES-C-70

45	SMPES-15	128	SPPBP20
46	SPAEEKK-B90	129	SPEES-10
47	4-PES-30	130	S-APES-20
48	semi-SPEK-12	131	SPP-1
49	SPI-1.88	132	SPAES-HQ-40
50	6F-PAEK-SP22	133	HQSH-20
51	S4-PAES-20	134	SPAEPO-55
52	SPES-25-TPD	135	sPATPO-60
53	S2-PAES-30	136	SPTPOF-60
54	SPES-1	137	tsPTPO-80
55	sPBI-NA50	138	tsPEPOF-100
56	sPBI-30	139	DsPEPOF-100
57	2F-SPAES-20	140	sPESPO-30
58	SPEEKK-1	141	SPE2
59	OBA-40	142	DHDPE40
60	PTEHSH-60	143	6F-BPA40
61	PTHQSH-60	144	BPAQS-20
62	PTAQSH-50	145	1
63	SPTA-50	146	SP35
64	PTFQSH-70	147	6FBPAQSH-20
65	PTATSH60	148	SPSSF30
66	SPTA1	149	6FBPAQSH-0-20
67	PAQS-30	150	SPAES-6FBPA-40
68	SPA-0-70	151	SPAES

69	SPA-70	152	SEB
70	HPPQSH-10PS	153	M-30
71	sPAEPO-60	154	C-SPAEK-40
72	sPEKP-350	155	BP-1
73	SPE3	156	SPAEKBO-10
74	HPPQSH-10	157	SPAEK-6F-20
75	1-tSPPco-PAEK-1.0	158	HFAS82
76	PSA-SPAE-80	159	SPEN-40
77	SPAE-1	160	B20P80
78	SPAEK-60	161	NPI(1)/SPP(40)
79	HS-PAEK-10	162	Ia-70
80	TS-0-PAEK-15	163	SPA-0.49
81	TS-PAEK-15	164	PFMPP5-r-PFSPP5
82	DHN40	165	3a
83	SPAES-PQ-20	166	PBOS2-r-PSBOS8

Table S2 List of hydrophobic monomers.

ID	**Name**	**ID**	**Name**
1	SPFENO-20	89	8a
2	sPAEPO-60	90	SPI-40
3	SPE3	91	SPI-N2
4	SPEN-40	92	SPI-4a
5	dsPPEPO-70	93	P1-1.0

6	SPAEK-20-COOH	94	SPAE20
7	OBA-40	95	PBOS2-r-PSBOS8
8	sPEKP-350	96	B20P80
9	M10N5-SO3H	97	SPP-2
10	3a	98	SPP-3
11	SPAEK40-TRI	99	SPI-10
12	BM-1	100	SPI-B5
13	SPAEEKK-H90	101	I-40
14	sPEKkP-450	102	FSPPI-50
15	SPEEKK-1	103	Ic-70
16	SPE1	104	If-60
17	2F-SPAES-20	105	LSPA-0.49
18	SPAEK-X15Y4	106	SPA-0.49
19	SPBT-PS50	107	SPTPOF-60
20	SCT-SAPs-0.8	108	DsPEPOF-100
21	NB-90	109	tsPTPO-80
22	SPAEEKK-B90	110	sPATPO-60
23	PTATSH60	111	sPESPO-30
24	PTAQSH-50	112	SPE2
25	PTFQSH-70	113	SPES-25-TPD
26	X4Y4	114	SPAEPO-55
27	BPAQS-20	115	4-PES-30
28	6FBPAQSH-20	116	SMPES-15
29	PTEHSH-60	117	SPES-1

30	PTHQSH-60	118	6F-BPA40
31	sPBI-SF70	119	S4PH-20-PS
32	SPEKEBI-1	120	SPSSF30
33	sPBI-30	121	PAES-2S-a60
34	SPPI-60	122	PSA-SPAE-80
35	sPBI-NAA70	123	S2-PAES-40
36	sPBI-NA50	124	SPAES-6FBPA-40
37	sPBI-60	125	2-SPAES-50
38	SPTA3	126	1
39	SPA-0-70	127	S-APES-20
40	SPA-70	128	SPAES-25
41	SPI-9	129	SOPAES-0.3
42	SPI90-TFVBPA10	130	SP35
43	CSPI90-6FATFVP10	131	SPEES-10
44	Id-70	132	SPAES-HQ-40
45	Ia-70	133	HQSH-20
46	Ib-70	134	SPBT-6F70
47	DHNH40	135	SPBT-6F55
48	SPI-N1	136	sPBT70
49	DAN-50	137	sPBT-PE52.5
50	DThioN70	138	SPPBP20
51	DPyN70	139	SPAEN-1
52	M-DTN70	140	SEB

53	P-DTN70	141	SPAE-1
54	DQN40	142	m-SPAEENH-50
55	SPI-N3	143	PAEN80
56	NPI(1)/SPP(40)	144	SN30
57	SPI-30	145	PFMPP5-r-PFSPP5
58	S-75	146	SPP-1
59	DQB50	147	SPP-60
60	1a	148	B20V80-SDPA
61	IBQS-20	149	3a-1
62	SPI-B2	150	SPAE-9
63	SPBIBI-b(70)	151	6F-PAEK-SP22
64	SPI-B1	152	TS-PAEK-15
65	SPI-B3	153	6F-SNPAEK-50
66	HPPQSH-10PS	154	BP-1
67	SPPQ-80	155	SPI-1.88
68	SPPQ(o)-80	156	SMP-PAEK-90
69	SPBIBI-a(70)	157	SPAEK-60
70	1-tSPPco-PAEK-1.0	158	SPEEKK-10
71	HPPQSH-10	159	M-30
72	SPPBI-20	160	semi-SPEK-12
73	PAQS-30	161	C-SPAEK-40
74	SPAEK-20-BI	162	SPTA1
75	SPAEK-40	163	SPTA-50

76	SPAEKKS-1	164	sPAS-7/3
77	SPEEK-60	165	S2-PAES-30
78	S-PEEK-1	166	DHN40
79	SPAEKBO-10_1	167	PAES85
80	SPAEKBO-10_2	168	SPAES-PQ-20
81	S-FPEK(HF90)	169	SEF
82	SPI-2.2	170	SPAES-1
83	TPPO-40	171	SES
84	II-50	172	SPAES
85	SPI-8-m	173	S-PES-40
86	SPI-8(50)	174	SPE4
87	SPI-8-p	175	HFAS82
88	SPI-1		

Table S3. Statistical summary of performance indexes of hydrocarbon-based sulfonated copolymer PEMs. Threshold is the value for binary classification, and value at L_{sig} is the cutoff value of a performance index where a solution is preferred at L_{sig} significance level.

Properties	**Records**	**Min.**	**Mean**	**Max.**	**Threshold**	**Value at L_{sig}**
σ(S/m)	589	0.004	8.9	29.8	9.4	21.1
MePerm(10^{-7}cm^2/s)	211	0.005	4.0	21.1	3.8	0.1
WU(%)	607	1.8	47.3	670.0	30.8	120.2
SR(%)	427	7.1	17.7	240.0	18.6	50.0
Tg(°C)	148	107.0	229.2	324.0	240.6	294.4
Td5(°C)	385	44.7	312.8	537.0	328.4	459.3

TS(MPa)	395	5.9	49.8	123.0	52.3	85.4
TM(GPa)	312	0.01	1.3	4.5	1.6	2.4
EAB(%)	396	1.3	27.5	247.1	28.9	101.6

Table S4. Features for modeling performance index of copolymers. Bold features with a, b, c, d superscript were used to construct predictive models for σ, Td5, MePerm and TM, respectively.

ID	Label	Description
1	FraLic	Fraction of hydrophilic monomers in copolymer
2	**IEC**[a,c]	Ion exchange capacity of copolymer
3	**T**[a]	Temperature for measurements
4	**RH**[a]	Relative humidity
5	**clogP**[b]	Calculated LogP value
6	**MaxAbsPartialCharge**[a,b,c,d]	Theoretic charge value for fully dissociated or protonated monomer
7	**MinEStateIndex**[b,c]	EState indices for the molecule, encode electronic and topological information
8	**MaxEStateIndex**[b]	EState indices for the molecule, encode electronic and topological information
9	**MinAbsEStateIndex**[b,c]	EState indices for the molecule
10	**BalabanJ**[b,c]	Balaban's connectivity topological index
11	**Ipc**[a,b,c,d]	the information content of the coefficients of the characteristic polynomial of the adjacency matrix of a hydrogen-suppressed graph of a molecule.
12-20	**PEOE_VSA1**[b,d],**2**[b,d],**3**[b,c], **6**[a],**7**,**8**[b],**9**[a,b,d],**10**[b],**11**[a,b,c]	MOE-type descriptors using partial charges and surface area contributions

21-24	**SMR_VSA5**[d],**6**[a],**9**[b,d],**10**[d]	MOE-type descriptors using molecular refractivity contributions and surface area contributions
25-29	**SlogP_VSA1**[c,d],**2**[d],**5**[b,d],**8**[d],**12**[a,b,c,d]	MOE-type descriptors using LogP contributions and surface area contributions
30-37	**EState_VSA2**[d],**3**[b],**4**,**5**,**6**[a],**7**,**8**,**9**	MOE-type descriptors using EState indices and surface area contributions
38-39	**VSA_Estate9**[d],**10**[d]	MOE-type descriptors using EState indices and surface area contributions
40	**FractionCSP3**[b]	The fraction of C atoms with SP3 hybrid orbital, i.e., C in sigma bonds
41	**NHOHCount**[b,d]	Number of NH or OH groups
42	**NumHAcceptors**[c]	Number of Hydrogen Bond Acceptors
43	**NumRotatableBonds**[d]	Number of Rotatable Bonds, i.e., sigma bonds
44	**fr_C_O_noCOO**[a,b,c]	Number of oxygen atoms in carbonyl group excluding COOH
45	**fr_NH0**[a,b,d]	Number of Tertiary amines
46	**fr_NH1**[a,b,c,d]	Number of Secondary amines
47	**fr_alkyl_halide**[a,c,d]	Number of alkyl halides
48	**fr_amide**[a,d]	Number of amides
49	**fr_bicyclic**[b,d]	Number of Bicyclic
50	**fr_ether**[a,b]	Number of ether oxygens (including phenoxy)
51	**fr_ketone**[a,b,c,d]	Number of ketones
52	**fr_para_hydroxylation**[a,b,c,d]	Number of para-hydroxylation sites
53	**fr_phenol_noOrthoHbond**[b,c,d]	Number of phenolic OH excluding ortho intramolecular Hbond substituents
54	**fr_pyridine**[b,c]	Number of pyridine rings
55	**fr_sulfone**[a,d]	Number of sulfonic groups