Electronic Supplementary Material (ESI) for CrystEngComm.

Supporting Information

Computational Identification of Organic Porous Molecular Crystals

Jack D. Evans,[*a‡] David M. Huang,[a] Maciej Haranczyk,[bc] Aaron W. Thornton[ad], Christopher J. Sumby[a] and Christian J. Doonan[*a]

[a] *Centre for Advanced Nanomaterials, School of Physical Sciences, The University of Adelaide, Adelaide, Australia*
[b] *Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, USA*
[c] *IMDEA Materials Institute, C/Eric Kandel 2, 28906 Getafe, Madrid, Spain*
[d] *CSIRO Materials Science and Engineering, Clayton, Australia*

‡ Present address: Chimie ParisTech - IRCP, Equipe CTM, Paris, France
∗ jack.evans@chimie-paristech.fr
∗ christian.doonan@adelaide.edu.au

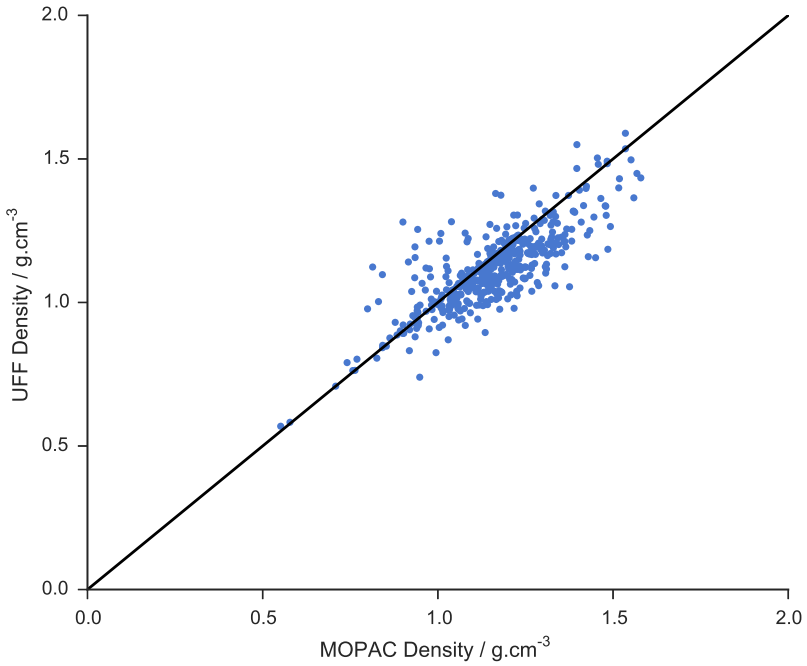

Figure SI- 1 Parity plot for density for UFF optimized structures and PM6-DH2x optimized structures.

Table SI- 1 Descriptors calculated by the RDKit program.

RDKit Descriptors
MolWt, exactMolWt, HeavyAtomMolWt, NumRadicalElectrons, NumValenceElectrons, HeavyAtomCount, NumHeteroatoms, NumRotatableBonds, RingCount
BalabanJ
BertzCT
Chi0v, Chi1v, Chi2v, Chi3v, Chi4v, ChiNv, Chi0n, Chi1n, Chi2n, Chi3n, Chi4n, ChiNn, HallKierAlpha , Kappa1, Kappa2, Kappa3
EState-VSA1 EState-VSA11 VSA-EState1 VSA-EState10
LabuteASA PEOE-VSA1 PEOE-VSA14 SMR-VSA1 SMR-VSA10 SlogP-VSA1 SlogP-VSA12
MolLogP, MolMR
TPSA

Table SI- 2 Molecular fragments calculated by the RDKit program.

fragment title	description
NumHAcceptors	Number of hydrogen bond acceptors
NumHDonors	Number of hydrogen bond donors
NHOHCount	Number of NHs and OHs
NOCount	Number of nitrogen and oxygen atoms
fr_ Al_ COO	Number of aliphatic carboxylic acids
fr_ Al_ OH_ noTert	Number of aliphatic hydroxyl groups excluding tert-OH
fr_ Al_ OH	Number of aliphatic hydroxyl groups
fr_ aldehyde	Number of aldehyde groups
fr_ alkyl_ carbamate	Number of alkyl carbamates
fr_ alkyl_ halide	Number of alkyl halide groups
fr_ allylic_ oxid	Number of allylic oxidation sites excluding steroid dienone
fr_ amide	Number of amide groups
fr_ amidine	Number of amidine groups
fr_ aniline	Number of aniline groups
fr_ Ar_ COO	Number of aromatic carboxylic acids
fr_ Ar_ N	Number of aromatic nitrogens
fr_ Ar_ NH	Number of aromatic amines
fr_ Ar_ OH	Number of aromatic hydroxyl groups
fr_ ArN	Number of N functional groups attached to aromatics
fr_ aryl_ methyl	Number of aryl methyl sites
fr_ azide	Number of azide groups
fr_ azo	Number of azo groups
fr_ barbitur	Number of barbiturate groups
fr_ benzene	Number of benzene rings
fr_ benzodiazepine	Number of benzodiazepines with no additional fused rings
fr_ bicyclic	Number of bicyclic rings
fr_ C_ O_ noCOO	Number of carbonyl O, excluding COOH
fr_ C_ O	Number of carbonyl groups
fr_ C_ S	Number of thiocarbonyl groups
fr_ COO	Number of carboxylic acids
fr_ COO2	Number of carboxylates
fr_ diazo	Number of diazo groups
fr_ dihydropyridine	Number of dihydropyridines
fr_ epoxide	Number of epoxide rings
fr_ ester	Number of esters
fr_ ether	Number of ether oxygens (including phenoxy)
fr_ furan	Number of furan rings

fragment title	description
fr_ guanido	Number of guanidine groups
fr_ halogen	Number of halogens
fr_ hdrzine	Number of hydrazine groups
fr_ hdrzone	Number of hydrazone groups
fr_ HOCCN	Number of C(OH)CCN-Ctert-alkyl or C(OH)CCNcyclic
fr_ imidazole	Number of imidazole rings
fr_ imide	Number of imide groups
fr_ Imine	Number of imine groups
fr_ isocyan	Number of isocyanates
fr_ isothiocyan	Number of isothiocyanates
fr_ ketone_ Topliss	Number of ketones excluding diaryl, a and b-unsat
fr_ ketone	Number of ketones
fr_ lactam	Number of beta lactams
fr_ lactone	Number of cyclic esters (lactones)
fr_ methoxy	Number of methoxy groups -OCH3
fr_ morpholine	Number of morpholine rings
fr_ N_ O	Number of hydroxylamine groups
fr_ Ndealkylation1	Number of XCCNR groups
fr_ Ndealkylation2	Number of tert-alicyclic amines (no heteroatoms, not quinine-like bridged N)
fr_ NH0	Number of tertiary amines
fr_ NH1	Number of secondary amines
fr_ NH2	Number of primary amines
fr_ Nhpyrrole	Number of H-pyrrole nitrogens
fr_ nitrile	Number of nitriles
fr_ nitro_ arom_ nonortho	Number of non-ortho nitro benzene ring substituents
fr_ nitro_ arom	Number of nitro benzene ring substituents
fr_ nitro	Number of nitro groups
fr_ nitroso	Number of nitroso groups, excluding NO2
fr_ oxazole	Number of oxazole rings
fr_ oxime	Number of oxime groups
fr_ para_ hydroxylation	Number of para-hydroxylation sites
fr_ phenol_ noOrthoHbond	Number of phenolic OH excluding ortho intramolecular hydrogen bond substituents
fr_ phenol	Number of phenols
fr_ phos_ acid	Number of phosphoric acid groups
fr_ phos_ ester	Number of phosphoric ester groups
fr_ piperdine	Number of piperdine rings
fr_ piperzine	Number of piperzine rings

fragment title	description
fr_ priamide	Number of primary amides
fr_ prisulfonamd	Number of primary sulfonamides
fr_ pyridine	Number of pyridine rings
fr_ quatN	Number of quarternary nitrogens
fr_ SH	Number of thiol groups
fr_ sulfide	Number of thioether
fr_ sulfonamd	Number of sulfonamides
fr_ sulfone	Number of sulfone groups
fr_ term_ acetylene	Number of terminal acetylenes
fr_ tetrazole	Number of tetrazole rings
fr_ thiazole	Number of thiazole rings
fr_ thiocyan	Number of thiocyanates
fr_ thiophene	Number of thiophene rings
fr_ unbrch_ alkane	Number of unbranched alkanes of at least 4 members (excludes halogenated alkanes)
fr_ urea	Number of urea groups
FractionCSP3	Fraction of C atoms that are SP3 hybridized

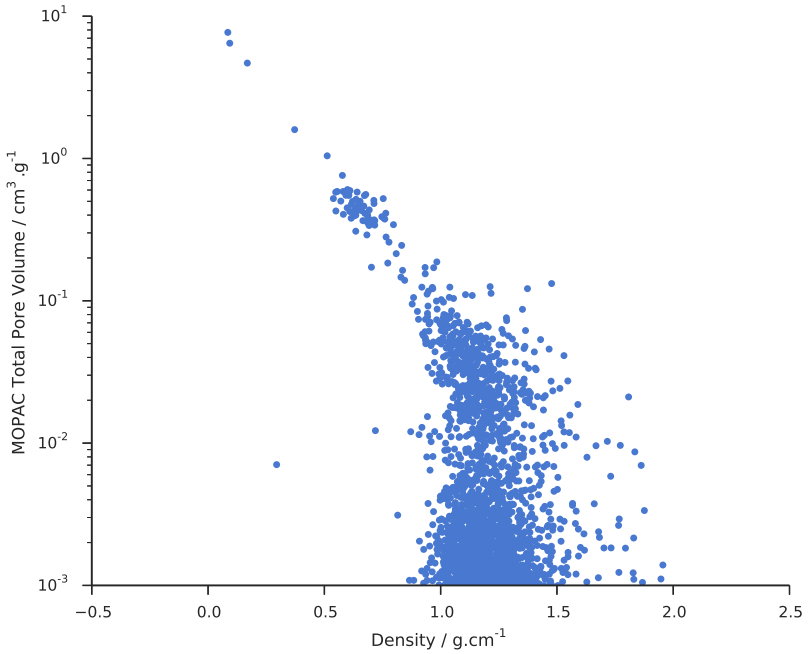

Figure SI- 2 Correlation of pore volume and crystal density demonstrated for the structures identified in the oPMC database.

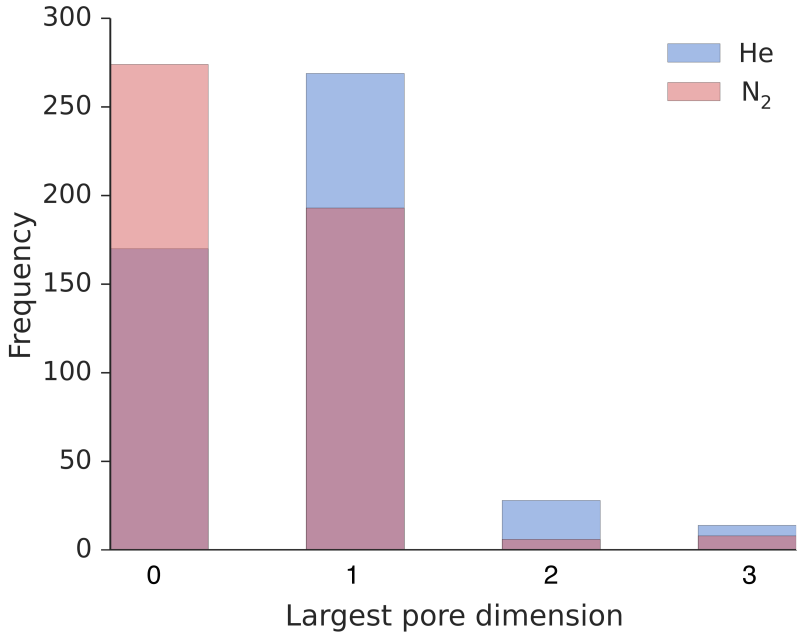

Figure SI- 3 Histogram of the dimensonality of largest pore identified for the oPMC database. Dimensionality was analyzed for probe sizes equivalent to the kinetic diameters of He (2.60 Å) and N_2 (3.64 Å).

Table SI- 3 List of identified structures with total surface areas (probe size equivalent to N_2) >1000 $m^2.g^{-1}$.

	total surface area / $m^2.g^{-1}$
FEQXAC	3855.37
GIPCAJ	2659.76
DEBXIT	2375.55
DIHGOR	1898.28
RERNEI	1843.72
MAVSIL	1515.66
TADWAY	457.37
CXPMSO01	1350.74
ZEXRIF02	1193.9
ABINOP	1188.37
VEWGOT	1178.08
ZEXRIF	1164.87
RIYQEW	1069.02
XOKTAS	1030.20
LOMDAS	1010.12
WUCJEJ	1009.07

Table SI- 4 List of identified structures with limiting pore diameters equivalent to the size of xenon (4.1 ± 0.1 Å).

	limiting pore diameter / Å
VOLCOP	4.19
HXANAL10	4.18
KEFJEK	4.18
DADLEA	4.18
HORMOQ	4.15
PILLOL	4.14
REBGUB	4.14
OVOPOF	4.14
LAXNEF	4.11
IFAREO	4.10
CEXGAP	4.10
TONYOL	4.09
BEMYAU	4.06
REYXAW	4.05
EVIBEQ	4.05
JURPOB	4.04
NEDPAN	4.04
QAMHAN	4.02
HEBLIJ	4.02
ZIKZOK	4.01
RAXTIT	4.01
QEKMUO	4.00

Table SI- 5 Average total, non-porous (NP) and porous(P) accuracy percentage for SVM classifiers applied to the labuteASA descriptor.

kernel function	**total / %**	**NP / %**	**P / %**
linear	69.7	74.6	64.8

A link to the database which includes a spreadsheet of the porosities calculated and the MOPAC optimized structures can be found below.

https://github.com/jackevansadl/oPMCdatabase