

Table 1: Small molecule building blocks. CG particle type, corresponding chemical building block, and examples of molecules in which such a block appears. The atoms the CG block is taken to represent are also shown in the 2D chemical structures, with T-beads, S-beads, and R-beads depicted in blue, green, and red, respectively.

<sup>†</sup> Indicates that the (group of) atom(s) is shared with neighboring beads.

\* Indicates molecules considered in this work; others are from DOI: 10.1038/s41592-021-01098-3.















type	chemical building block	examples	
		2D	name (mapping)
P2	$\text{CH}_3\text{CH}_2\text{COOH}$		propanoic acid (P2)
SP2	$\text{CH}_3\text{COOH}$		acetic acid (SP2)
P1	$(\text{CH}_3)_2\text{CHOH}$		isopropanol (P1)
TP1d	$-\text{CH}_2\text{OH}$ (bonded to $-\text{CF}_3$ )		2,2,2-trifluoroethanol (SX4e-TP1d)
SN6	$-\text{CH}^\dagger=\text{C}(-\text{OH})-\text{CH}^\dagger=$		*phenol ((TC5) <sub>2</sub> -SN6)
	$-\text{CH}_2^\dagger-\text{CH}(-\text{OH})-\text{CH}_2^\dagger-$		*cyclopentanol (SN6-SC3)
SN6d	$-\text{CH}^\dagger=\text{C}(-\text{NH}_2)-\text{CH}^\dagger=$		*aniline ((TC5) <sub>2</sub> -SN6d)
TN6d	$=\text{CH}^\dagger-\text{NH}-\text{CH}^\dagger=$		*pyrrole ((TC5) <sub>2</sub> -TN6d)
TN6	$=\text{C}(-\text{OH})-$		*p-cresol ((TC5) <sub>2</sub> -TC5-TN6)
N6a	$-(\text{CH}_2)_2-\text{CH}(=\text{O})$		heptanal (C1-N6a)
TN6a	$-\text{CH}=\text{N}-$		*pyridine ((TC5) <sub>2</sub> -TN6a)
SN5a	$-\text{O}-\text{CH}_2-\text{O}-$		*1,3-dioxolane (SN5a-TC3)
TN5a	$=\text{C}=\text{O}$		*benzoquinone (TN5a-(TC5) <sub>2</sub> -TN5a)
N4a	$-\text{C}(=\text{O})-\text{O}-\text{CH}_3$		*methyl-benzoate (N4a-(TC5) <sub>3</sub> )

Table 1: (continued)


































	$\text{CH}_3-\text{C}(=\text{O})-\text{O}-\text{CH}_3$		methyl-acetate (N4a)
SN4a	$-\text{CH}_2-\text{O}-\text{CH}_2-$		*tetrahydropyran (SC3-SN4a)
	$-\text{C}(=\text{O})-\text{CH}_3$		*acetophenone (SN4a-(TC5) <sub>3</sub> )
TN4a	$-\text{CH}_2^\dagger-\text{O}-\text{CH}_2^\dagger-$		*tetrahydrofuran (SC3-TN4a)
	$-\text{C}\equiv\text{N}$		*benzonitrile ((TC5) <sub>3</sub> -TN4a)
	$-\text{CH}(=\text{O})$		*benzaldehyde ((TC5) <sub>3</sub> -TN4a)
SN3a	$-\text{CH}_2-\text{O}-\text{CH}_2-$		*1,4-dioxane (SN3a-SN3a)
SN3a	$-\text{NO}_2$		*nitrobenzene ((TC5) <sub>3</sub> -SN3a)
SN2a	$=\text{C}(-\text{O}-\text{CH}_3)-$		*o-methylanisole (SN2a-TC4-(TC5) <sub>2</sub> )
TN2a	$=\text{CH}^\dagger-\text{O}-\text{CH}^\dagger=$		*furan ((TC5) <sub>2</sub> -TN2a)
	$-\text{O}-\text{CH}_3$		*methoxybenzene ((TC5) <sub>3</sub> -TN2a)
SN1	$-\text{N}(-\text{CH}_3)_2$		*N,N-dimethylaniline ((TC5) <sub>3</sub> -SN1)
TN1	$-\text{N}(-\text{CH}_3)-$		*1-methylimidazole (TC5-TN1-TN6a)
SC6	$-\text{CH}^\dagger=\text{C}(-\text{SH})-\text{CH}^\dagger=$		*thiophenol ((TC5) <sub>2</sub> -SC6)
	$-\text{CH}_2-\text{S}-\text{CH}_2-$		*tetrahydrothiophene (SC6-TC3)
TC6	$=\text{CH}^\dagger-\text{S}-\text{CH}^\dagger=$		*thiophene ((TC5) <sub>2</sub> -TC6)
TC5	$-\text{CH}=\text{CH}-$		*benzene ((TC5) <sub>3</sub> )
	$-\text{CH}=\text{CH}_2$		*styrene ((TC5) <sub>4</sub> )

Table 1: (continued)

TC5e	$(-)_2C=C(-)_2$		*naphthalene ((TC5) <sub>2</sub> -TC5e-(TC5) <sub>2</sub> )
C4	$CH_3-CH=CH-CH_3$		1-butene (C4)
SC4	$-CH^\dagger=C(-CH_3)-CH^\dagger=$		*toluene ((TC5) <sub>2</sub> -SC4)
TC4	$=C(-CH_3)-$		*o-methylanisole (SN2a-TC4-(TC5) <sub>2</sub> )
SX4e	$-CF_3$		*benzotrifluoride ((TC5) <sub>3</sub> -SX4e)
	$-CF_3$		2,2,2-trifluoroethanol (SX4e-TP1d)
SC3	$-(CH_2)_3-$		*cyclohexane (SC3-SC3)
	$-CH_2^\dagger-CH(-CH_3)-CH_2^\dagger-$		*methylcyclopentane (SC3-SC3)
TC3	$-CH_2-CH_2-$		*tetrahydrothiophene (SC6-TC3)
SX3	$-CH^\dagger=C(-Cl)-CH^\dagger=$		*chlorobenzene ((TC5) <sub>2</sub> -SX3)
SC2	$-(CH_2)_2-(CH_3)$		*3-propylthiophene (TC6-(TC5) <sub>2</sub> -SC2)
X2	$CHCl_3$		*trichloromethane (X2)
SX2	$-CH^\dagger=C(-Br)-CH^\dagger=$		*bromobenzene ((TC5) <sub>2</sub> -SX2)
C1	$CH_3-(CH_2)_2-CH_3$		butane (C1)
X1	$-CH^\dagger=C(-I)-CH^\dagger=$		*iodobenzene ((TC5) <sub>2</sub> -X1)