Electronic Supplementary Material (ESI) for Faraday Discussions.
This journal is © The Royal Society of Chemistry 2020

Supporting Information:

Cost-effective composite methods for large scale solid-state calculations

L. Donà,[1] J. G. Brandenburg,[2] I. J. Bush,[3,4] and B. Civalleri*[1]

[1]*Dipartimento di Chimica, Università di Torino and NIS (Nanostructured Interfaces and Surfaces) Centre, Via P. Giuria 5, 10125 Torino, Italy.*

[2] *Chief Digital Organization, Merck KGaA, Frankfurter Str. 250, 64293 Darmstadt, Germany*

[3]*Oxford e-Research Centre, Department of Engineering Science, University of Oxford, 7 Keble Road, Oxford OX1 3QG, U.K.*

[4]*Scientific Computing Department, Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot OX11 0QX, United Kingdom*

(Dated: June 23, 2020)

PACS numbers:

Contents

S1. Metal Organic Frameworks 2

A. MOF dataset 2

B. MPPcrystal scaling for MIL-100 (Al) 11

S2. Biological systems 11

S3. Amorphous silica surfaces 12

S4. Miscellaneous systems 14

References 16

S1. METAL ORGANIC FRAMEWORKS

A. MOF dataset

MOF dataset contains 21 Metal Organic Frameworks selected by us for different molecular architecture in terms of number of atoms, metal cluster, organic linker and topology. In Tables S1, S2 the computational time for single and total SCF, as well for gradient calculation is reported at PBEsol0-3c and HSEsol-3c level of theory, respectively.

TABLE S1: PBEsol0-3c wall-clock time (sec.) required for SCF and gradients calculation performed on 80 CPU cores Intel® Xeon® E5-2630 v4 for the MOF dataset.

system	N_{at}	N_{op}	N_{AO}	S	T (N_{cyc})	A	G[a]
MIL-53 (Al)[1]	38	8	446	22	150 (24)	6	11
CPO-27 (Zn)[2]	54	6	828	8	180 (15)	12	16
MOF-5 (Zn)[3]	106	48	1406	16	247 (12)	21	13
UiO-66 (Zr)[4]	114	24	1418	17	312 (13)	24	36
IRMOF-10 (Zn)[3]	166	48	1994	29	407 (13)	31	22
HKUST-1 (Zn)[5]	156	48	2220	30	372 (11)	34	21
IRMOF-9 (Zn)[3]	332	8	3988	59	965 (13)	74	71
FJI-H6 (Zr)[6]	396	12	4322	74	1115 (14)	80	65
MIL-127 (Al)[7]	424	8	5496	109	1455 (12)	121	115
NU-1000 (Zr)[8]	558	12	6246	206	2408 (12)	140	140
PCN-60 (Zn)[9]	504	48	6696	201	2818 (13)	217	63
MOF-399 (Zn)[10]	636	48	6924	226	2992 (13)	230	89
meso-MOF-2 (Cd)[11]	528	48	7104	236	3051 (13)	235	66
DUT-122 (Zr)[12]	696	12	8648	286	4478 (14)	320	187
NU-100 (Zn)[13]	792	48	9768	324	5120 (14)	366	106
NU-109 (Zn)[14]	840	48	10488	354	6023 (15)	402	116
NU-108 (Zn)[15]	984	48	11400	387	5958 (13)	458	146
NU-110 (Zn)[14]	1032	48	12120	345	6026 (15)	402	117
DUT-49 (Zn)[16]	1728	24	20448	965	16266 (14)	1162	375
MOF-210 (Zn)[17]	1854	6	22410	1217	20123 (14)	1437	1167
MIL-100 (Al)[18]	2788	16	37128	2963	103165 (25)	4127	1677

[a] N_{at} is the total number of atoms, N_{op} is the number of symmetry operators,N_{AO} is the total number of atomic orbitals, N_{cyc} is the total number of SCF cycles, S is the single SCF time, T is the total SCF time, A is the average SCF time, G is the gradient time.

TABLE S2: HSEsol-3c wall-clock time (sec.) required for SCF and gradients calculation performed on 80 CPU cores Intel® Xeon® E5-2630 v4 for the MOF dataset.

system	N_{at}	N_{op}	N_{AO}	S	T (N_{cyc})	A	G[a]
MIL-53 (Al)	38	8	446	7	62 (10)	6	13
CPO-27 (Zn)	54	6	828	15	139 (11)	13	22
MOF-5 (Zn)	106	48	1406	14	321 (12)	27	13
UiO-66 (Zr)	114	24	1418	19	313 (12)	26	48
IRMOF-10 (Zn)	166	48	1994	30	403 (13)	31	23
HKUST-1 (Zn)	156	48	2220	46	388 (11)	35	22
IRMOF-9 IP (Zn)	332	8	3988	64	956 (13)	74	79
FJI-H6 (Zr)	396	12	4322	77	1120 (13)	86	98
MIL-127 (Al)	424	8	5496	106	1519 (12)	127	1230
NU-1000 (Zr)	558	12	6246	190	2447 (12)	204	168
PCN-60 (Zn)	504	48	6696	207	3052 (14)	218	70
MOF-399 (Zn)	636	48	6924	218	3084 (13)	237	86
meso-MOF-2 (Cd)	528	48	7104	209	3191 (14)	228	3230
DUT-122 (Zr)	696	12	8648	318	4576 (14)	327	226
NU-100 (Zn)	792	48	9768	322	5410 (15)	361	109
NU-109 (Zn)	840	48	10488	345	6026 (15)	402	117
NU-108 (Zn)	984	48	11400	399	6352 (14)	454	150
NU-110 (Zn)	1032	48	12120	422	6843 (14)	489	139
DUT-49 (Zn)	1728	24	20448	988	16296 (12)	1356	389
MOF-210 (Zn)	1854	6	22410	1243	20346 (14)	1453	1214
MIL-100 (Al)	2788	16	37128	2965	59846 (16)	3740	1763

[a] N_{at} is the total number of atoms, N_{op} is the number of symmetry operators, N_{AO} is the total number of atomic orbitals, N_{cyc} is the total number of SCF cycles, S is the single SCF time, T is the total SCF time, A is the average SCF time, G is the gradient time.

The scaled time reported in Figures S1, S2, S3, S4, S5, S6, S7, S8, S9, S10, S11, and S12 has been calculated for each system as follow: single SCF, total SCF and gradient time have been multiplied by the number of symmetry operators and divided by the number of atoms and atomic orbitals (AO), respectively Data presented in Figure are gathered according to the number of symmetry operators of the system. All calculations have been performed on 80 cores Intel® Xeon® E5-2630 v4.

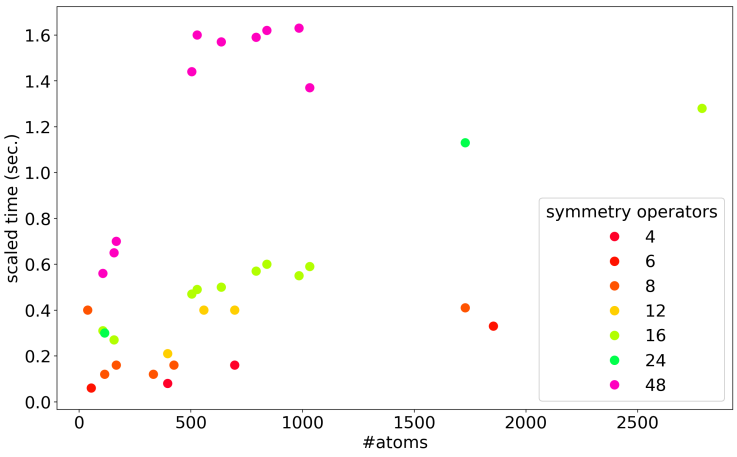


FIG. S1: Single SCF scaled time computed with PBEsol0-3c method for the MOF data set and plotted against the number of atoms. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

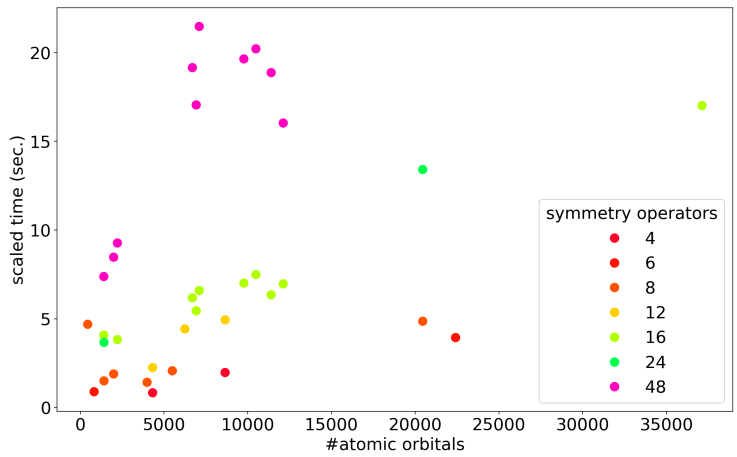


FIG. S2: Single SCF scaled time computed with PBEsol0-3c method for the MOF data set and plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

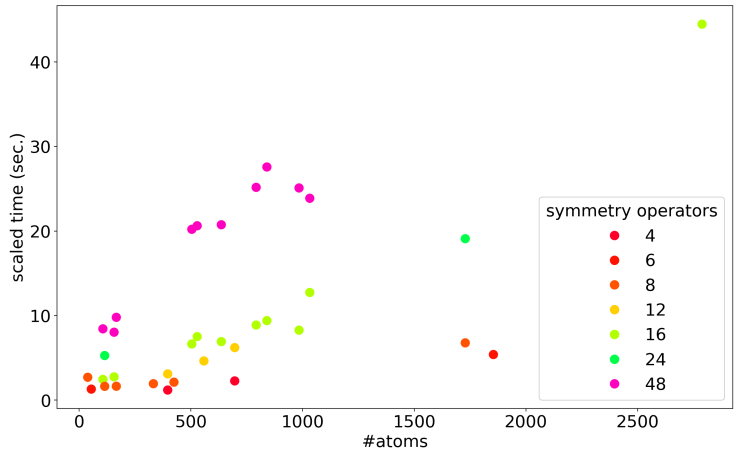


FIG. S3: Total SCF scaled time computed with PBEsol0-3c method for the MOF data set and plotted against the number of atoms. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

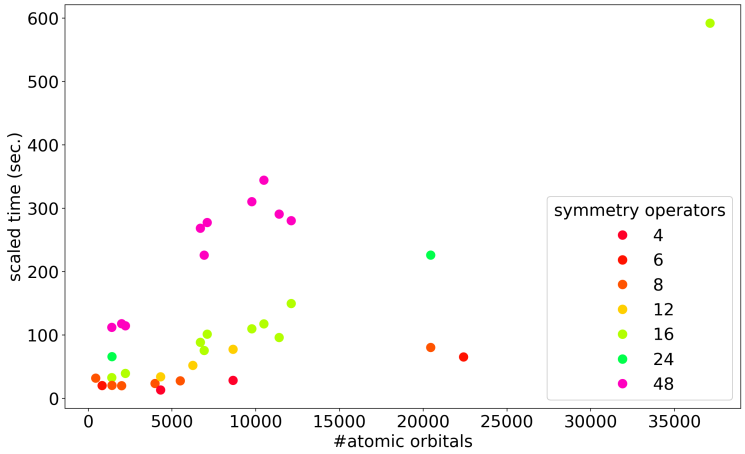


FIG. S4: Total SCF scaled time computed with PBEsol0-3c method for the MOF data set and plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

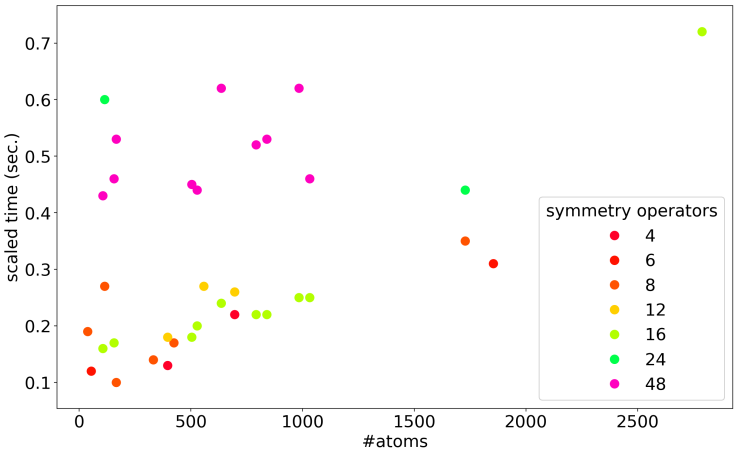


FIG. S5: Gradient scaled time computed with PBEsol0-3c method for the MOF data set and plotted against the number of atoms. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

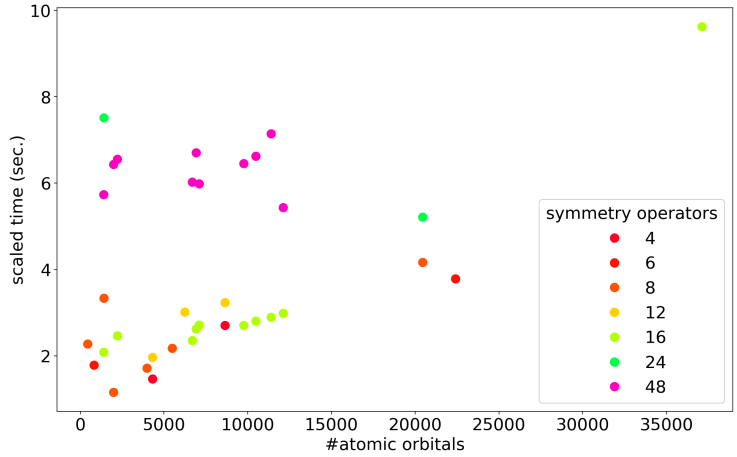


FIG. S6: Gradient scaled time computed with PBEsol0-3c method for the MOF data set and plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

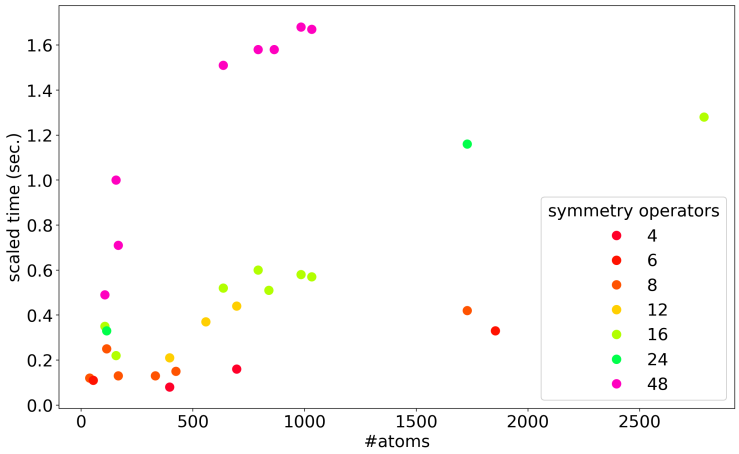


FIG. S7: Single SCF scaled time computed with HSEsol-3c method for the MOF data set and plotted against the number of atoms. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

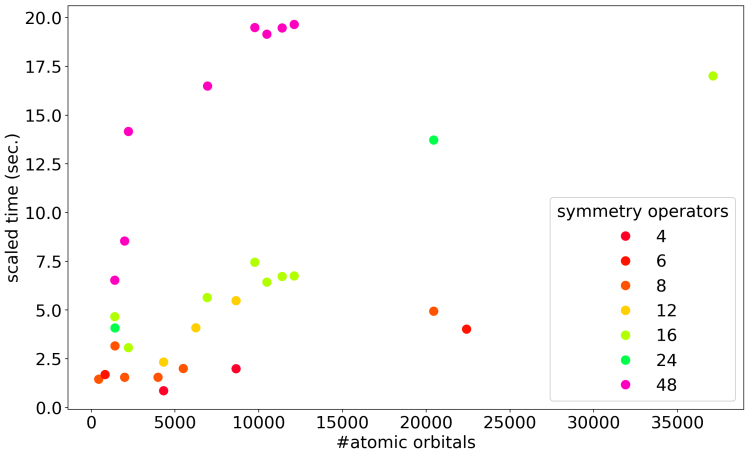


FIG. S8: Single SCF scaled time computed with HSEsol-3c method for the MOF data set and plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

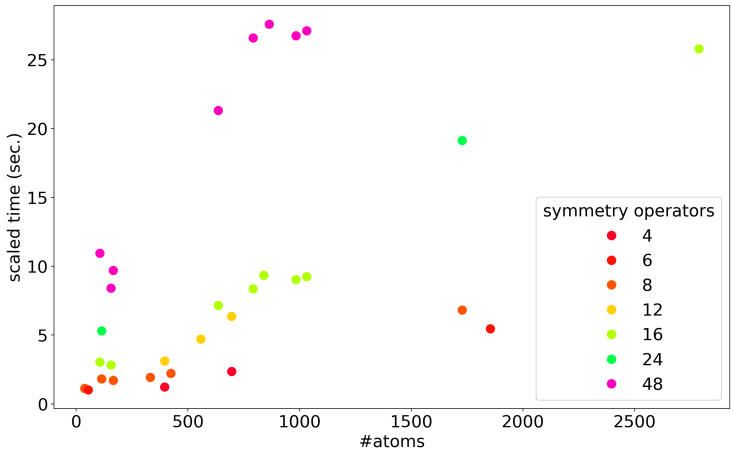


FIG. S9: Total SCF scaled time computed with HSEsol-3c method for the MOF data set and plotted against the number of atoms. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

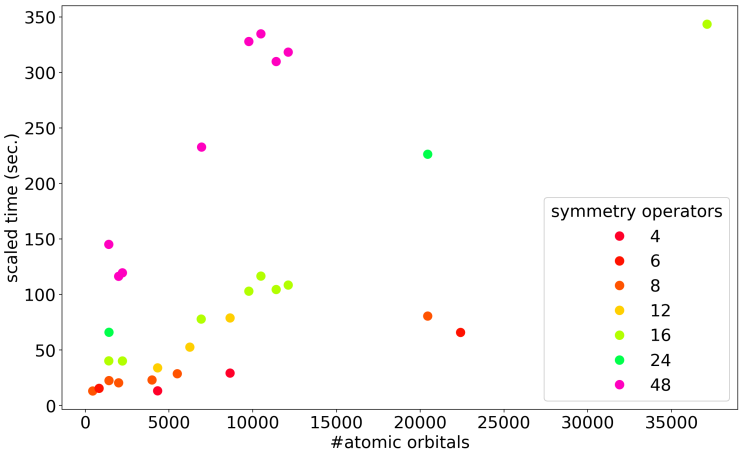


FIG. S10: Total SCF scaled time computed with HSEsol-3c method for the MOF data set and plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

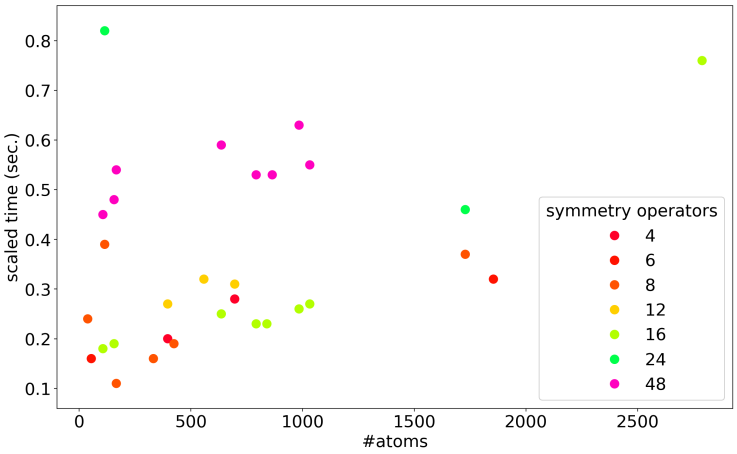


FIG. S11: Gradient scaled time computed with HSEsol-3c method for the MOF data set and plotted against the number of atoms. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

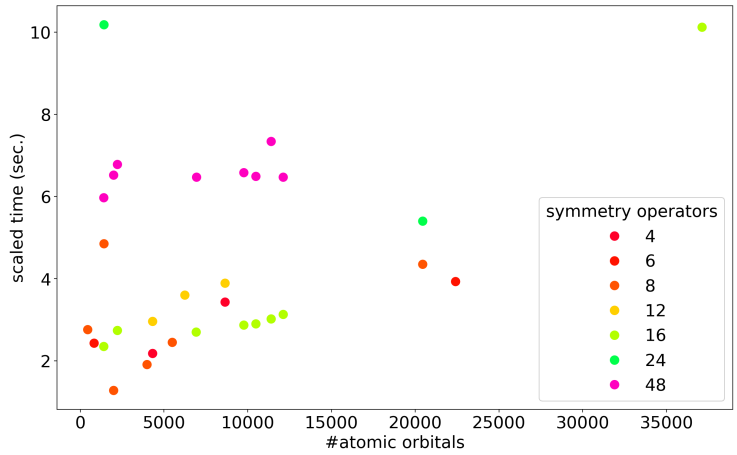


FIG. S12: Gradient scaled time computed with HSEsol-3c method for the MOF data set and plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs.

B. MPPcrystal scaling for MIL-100 (Al)

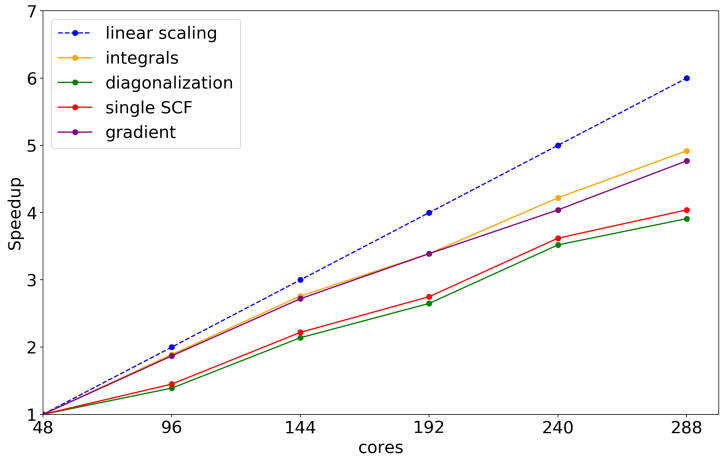
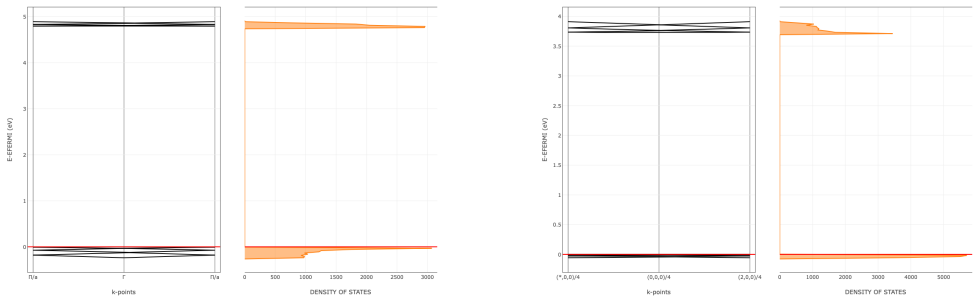


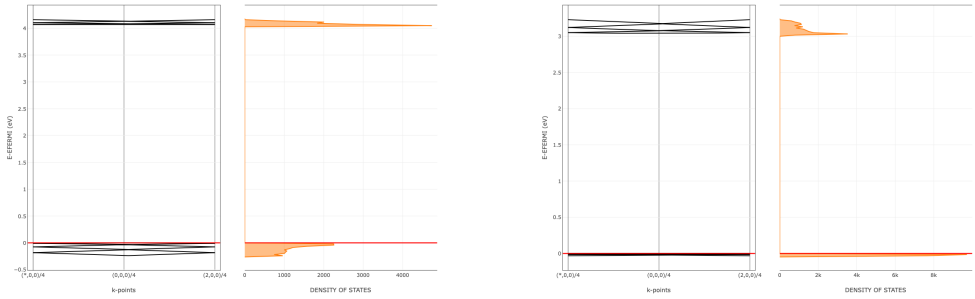
FIG. S13: PBEsol0-3c wall-clock time speed-up (t_n/t_{48}, with n the number of cores) for integrals calculation, Hamiltonian matrix diagonalization, single SCF and gradients calculation performed on increasing numbers of cores. Linear scaling is reported as dotted blue line.

S2. BIOLOGICAL SYSTEMS



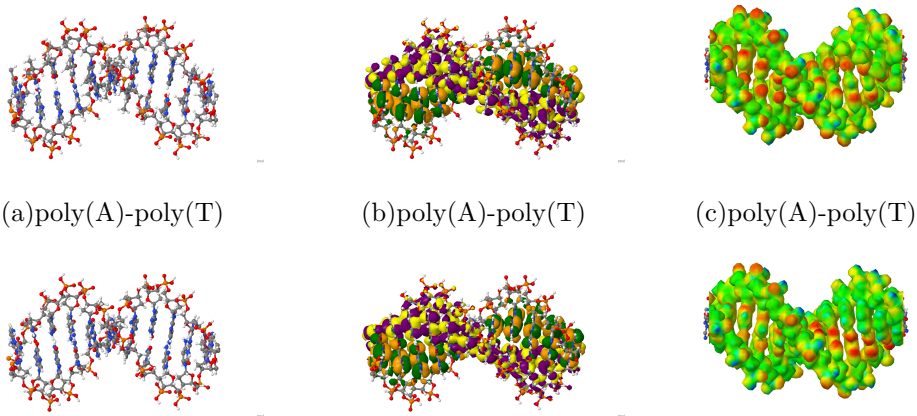
(a)poly(A)-poly(T) (b)poly(G)-poly(C)

FIG. S14: Unified band structure and density of states plot computed with CRYSPLOT[19] for poly(A)-poly(T) (a) and poly(G)-poly(C) (b) double strand calculated at PBEsol0-3c level of theory.



(a)poly(A)-poly(T) (b)poly(G)-poly(C)

FIG. S15: Unified band structure and density of states plot computed with CRYSPLOT[19] for poly(A)-poly(T) (a) and poly(G)-poly(C) (b) double strand calculated at HSEsol-3c level of theory.



(a)poly(A)-poly(T) (b)poly(A)-poly(T) (c)poly(A)-poly(T)

(d)poly(G)-poly(C) (e)poly(G)-poly(C) (f)poly(G)-poly(C)

FIG. S16: PBEsol0-3c optimized structure (a, d), highest occupied (green and orange) and lowest unoccupied (yellow and purple) crystal orbitals (b, e), electrostatic potential mapped on the electron density (c, f) for poly(A)-poly(T) and poly(G)-poly(C) DNA models. Color code for atoms: O in red, C in grey, H in white, N in blue, and P in orange

S3. AMORPHOUS SILICA SURFACES

TABLE S3: PBEsol0-3c wall-clock time (sec.) required for SCF and gradients calculation on 80 CPU cores Intel® Xeon® E5-2630 v4 for different drug@silica systems.

system	N_{op}	N_{at}	N_{AO}	S	T (N_{cyc})	G[a]
ASP@silica45	1	132	1668	50	588 (12)	138
CTZ@silica45	1	237	2693	81	1251 (13)	268
IBU@silica45	1	144	1718	53	593 (12)	143
NIT@silica45	1	143	1794	45	727 (13)	152
IBU@MCM-41	1	612	8046	419	5783 (12)	1706
7IBU@MCM-41	1	810	9612	581	8260 (13)	2402

[a]N_{at} is the total number of atoms, N_{op} is the number of symmetry operators, N_{AO} is the total number of atomic orbitals, N_{cyc} is the total number of SCF cycles, S is the single SCF time, T is the total SCF time, A is the average SCF time, G is the gradient time.

TABLE S4: HSEsol-3c wall-clock time (sec.) required for SCF and gradients calculation on 80 CPU cores Intel® Xeon® E5-2630 v4 for different drug@silica systems.

system	N_{op}	N_{at}	N_{AO}	S	T (N_{cyc})	G[a]
ASP@silica45	1	132	1668	60	695 (12)	193
CTZ@silica45	1	237	2693	95	1457 (13)	379
IBU@silica45	1	144	1718	63	705 (11)	210
NIT@silica45	1	143	1794	59	803 (12)	225
IBU@MCM-41	1	612	8046	439	6040 (12)	1983
7IBU@MCM-41	1	810	9612	605	8143 (12)	2753

[a]N_{at} is the total number of atoms, N_{op} is the number of symmetry operators, N_{AO} is the total number of atomic orbitals, N_{cyc} is the total number of SCF cycles, S is the single SCF time, T is the total SCF time, A is the average SCF time, G is the gradient time.

S4. MISCELLANEOUS SYSTEMS

TABLE S5: PBEsol0-3c wall-clock time (sec.) required for SCF and gradients calculation performed on 40 and 80 CPU cores Intel® Xeon® E5-2630 v4 for several systems. Data are gathered with respect to the periodicity of the system and sorted by increasing number of atomic functions.

system	PBC	N_{op}	N_{at}	N_{AO}	S_{40}	S_{80}	T_{40} (N_{cyc})	T_{80} (N_{cyc})	G_{40}	$G^a{}_{80}$
Crambin	0D	1	642	5553	336	209	5135 (16)	3691 (16)	680	331
Forsterite	0D	1	812	13224	1452	1074	21244 (13)	15796 (13)	4641	2319
Collagen	1D	7	245	2128	45	39	370 (12)	419 (12)	56	32
poly(C)-poly(G)	1D	11	715	7359	140	245	2069 (14)	3615 (14)	140	74
poly(A)-poly(T)	1D	11	726	7381	147	254	1925 (13)	3310 (13)	141	76
silica45	2D	1	111	1457	42	43	589 (11)	458 (11)	208	108
ASP@silica45	2D	1	132	1668	53	43	678 (12)	537 (12)	239	124
IBU@silica45	2D	1	144	1718	60	45	851 (12)	632 (12)	262	135
NIT@silica45	2D	1	143	1794	61	48	855 (13)	692 (13)	276	143
CTZ@silica45	2D	1	237	2693	105	76	1625 (13)	1173 (13)	493	258
TPM	3D	24	856	6184	105	194	1422 (12)	4438 (22)	275	159
MCM-41	3D	1	579	7785	492	412	6236 (12)	5434 (12)	3133	1586
IBU@MCM-41	3D	1	612	8046	496	419	6544 (12)	5783 (12)	3293	1706
7IBU@MCM-41	3D	1	810	9612	685	581	9827 (13)	8260 (13)	4735	2402
Crambin	3D	2	1536	12702	836	771	13072 (14)	11869 (14)	5950	3007
Forsterite[b]	3D	2	1008	16416	1884	1396	21528 (10)	16069 (10)	12143	6061
MIL-100 (Al)	3D	16	2788	37128	-	2963	-	103165 (25)	-	1677

[a] N_{op} is the number of symmetry operators, N_{at} is the total number of atoms, N_{AO} is the total number of atomic orbitals, N_{cyc} is the total number of SCF cycles, S_N, T_N, and G_N are the single SCF, total SCF, and graient time, respectively, computed on N CPU cores. [b] 4x3x3 supercell is used.

[1] Y. Liu, J. Her, A. Dailly, A. J. Ramirez-Cuesta, D. Neumann, and C. M. Brown. Reversible structural transition in mil-53 with large temperature hysteresis. *J. Am. Chem. Soc.*, 130:11813–11818, 2008.

[2] N. L Rosi, J. Kim, M. Eddaoudi, B. Chen, M. O'Keeffe, and O. M. Yaghi. Rod packings and metal- organic frameworks constructed from rod-shaped secondary building units. *J. Am. Chem. Soc.*, 127:1504–1518, 2005.

[3] M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe, and O. M. Yaghi. Systematic design of pore size and functionality in isoreticular mofs and their application in methane storage. *Science*, 295:469–472, 2002.

[4] J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga, and K. P. Lillerud. A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *J. Am. Chem. Soc.*, 130:13850–13851, 2008.

[5] H. Furukawa, K. E. Cordova, M. O'Keeffe, and O. M. Yaghi. The chemistry and applications of metal-organic frameworks. *Science*, 341:1230444, 2013.

[6] J. Zheng, M. Wu, F. Jiang, W. Su, and M. Hong. Stable porphyrin zr and hf metal–organic frameworks featuring 2.5 nm cages: high surface areas, scsc transformations and catalyses. *Chem. Sci.*, 6:3466–3470, 2015.

[7] Y. Belmabkhout, R. S. Pillai, D. Alezi, O. Shekhah, P. M. Bhatt, Z Chen, K. Adil, S. Vaesen, G. De Weireld, M. Pang, et al. Metal–organic frameworks to satisfy gas upgrading demands: fine-tuning the soc-mof platform for the operative removal of h 2 s. *J. Mater. Chem. A*, 5:3293–3303, 2017.

[8] P. Deria, J. E Mondloch, E. Tylianakis, P. Ghosh, W. Bury, R. Q Snurr, J. T. Hupp, and O. K. Farha. Perfluoroalkane functionalization of nu-1000 via solvent-assisted ligand incorporation: synthesis and co2 adsorption studies. *J. Am. Chem. Soc.*, 135:16801–16804, 2013.

[9] D. Zhao, D. Yuan, D. Sun, and H. Zhou. Stabilization of metal- organic frameworks with high surface areas by the incorporation of mesocavities with microwindows. *J. Am. Chem. Soc.*, 131:9186–9188, 2009.

[10] H. Furukawa, Y. B. Go, N. Ko, Y. K. Park, F. J. Uribe-Romo, J. Kim, M. O'Keeffe, and O. M. Yaghi. Isoreticular expansion of metal–organic frameworks with triangular and square building

units and the lowest calculated density for porous crystals. *Inorg. Chem.*, 50:9147–9152, 2011.

[11] D. Yuan, D. Zhao, D. J. Timmons, and H. Zhou. A stepwise transition from microporosity to mesoporosity in metal–organic frameworks by thermal treatment. *Chem. Sci.*, 2:103–106, 2011.

[12] F. Drache, V. Bon, I. Senkovska, M. Adam, A. Eychmüller, and S. Kaskel. Vapochromic luminescence of a zirconium-based metal–organic framework for sensing applications. *Eur. J. Inorg. Chem.*, 2016:4483–4489, 2016.

[13] O. K. Farha, A. Yazaydın, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen, R. Q. Snurr, and J. T. Hupp. De novo synthesis of a metal–organic framework material featuring ultrahigh surface area and gas storage capacities. *Nat. Chem.*, 2:944–948, 2010.

[14] O. K. Farha, I. Eryazici, N. C. Jeong, B. G. Hauser, C. E. Wilmer, A. A. Sarjeant, R. Q. Snurr, S. T. Nguyen, A. Ö. Yazaydın, and J. T. Hupp. Metal–organic framework materials with ultrahigh surface areas: is the sky the limit? *J. Am. Chem. Soc.*, 134:15016–15021, 2012.

[15] I. Eryazici, O. K. Farha, B. G. Hauser, A. Ö. Yazaydın, A. A. Sarjeant, S. T. Nguyen, and J. T. Hupp. Two large-pore metal–organic frameworks derived from a single polytopic strut. *Cryst. Growth Des.*, 12:1075–1080, 2012.

[16] U. Stoeck, S. Krause, V. Bon, I. Senkovska, and S. Kaskel. A highly porous metal–organic framework, constructed from a cuboctahedral super-molecular building block, with exceptionally high methane uptake. *Chem. Comm.*, 48:10841–10843, 2012.

[17] H. Furukawa, N. Ko, Y. B. Go, N. Aratani, S. B. Choi, E. Choi, A. Yazaydin, R. Q. Snurr, M. O'Keeffe, J. Kim, et al. Ultrahigh porosity in metal-organic frameworks. *Science*, 329:424–428, 2010.

[18] G. Férey, C. Serre, C. Mellot-Draznieks, F. Millange, S. Surblé, J. Dutour, and I. Margiolaki. A hybrid solid with giant pores prepared by a combination of targeted chemistry, simulation, and powder diffraction. *Angew. Chem. Int. Ed.*, 43:6296–6301, 2004.

[19] G. Beata, G. Perego, and B. Civalleri. Crysplot: A new tool to visualize physical and chemical properties of molecules, polymers, surfaces, and crystalline solids. *J. Comp. Chem.*, 40:2329–2338, 2019.