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

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] *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)

Following the development in recent years of progressively accurate approximations to the exchange-correlation functional, the use of density functional theory (DFT) methods to examine increasingly large and complex systems has grown, in particular for solids and other condensed matter systems. However the cost of these calculations is high, often requiring the use of specialist HPC facilities. As such, for the purpose of large scale high-throughput screening of material properties, a hierarchy of simplified DFT methods has been proposed that allows the comparatively rapid calculation of the electronic structure of large systems, and we have recently extended this scheme to the solid state (sol-3c). Here, we analyze the applicability and scaling of the new sol-3c DFT methods to molecules and crystals composed of light-elements, such as small proteins and model DNA-helices. Furthermore, the calculation of the electronic electronic structure of large to very large porous systems, such as metal-organic frameworks and inorganic nanoparticles, is discussed. The new composite methods have been implemented in the CRYSTAL17 code, which efficiently implements hybrid functionals and enables routine application of the new methods to large scale calculations of such materials with excellent performance even on small-scale computing resources.

PACS numbers:

I. INTRODUCTION

Density functional theory, within the Kohn-Sham formalism, is a computationally efficient alternative to wave function methods to compute ground-state properties of atoms, molecules, and solids. Nowadays, it is the most widely used methodology for electronic structure calculations in quantum chemistry, solid-state physics, and materials science. The success of DFT lies in the simplicity of the theoretical framework[1,2], the favourable scaling of the computational cost, which is similar to Hartree-Fock[3,4], and the accuracy that can be achieved after three decades of research finding increasingly good approximations to the elusive exchange-correlation functional[5]. Modern density functionals improve the short- to medium-range correlation by the combination of exact constraints with various degrees of parametrization.[6–8] A substantial development in the last decade is the incorporation of dispersion forces in the DFT framework,[9–11] which extends the applicability of DFT to non-covalently bound molecular complexes and solids.

Furthermore, substantial progress has been made to efficiently implement DFT methodologies in a variety of program codes[12–16] and testing their reproduceability in solid state applications.[17] The trade-off between cost and accuracy so allowed has been the key that has opened up the applicability of DFT to various fields from physics to chemistry, from biology to materials science, and from catalysis to mineralogy. With the advent of increasingly powerful CPUs and a huge increase in the availability and exploitation of parallelism the amount of computational resources from commodity PC to supercomputers has reached the peta-flop era[18]. This has allowed DFT to be applied to large scale simulations thus pushing the limit from a few hundreds to many thousands of atoms.[18,19]

Among different families of exchange-correlation (XC) functionals, hybrid functionals have emerged in the last decades as the method of choice for molecular calculations, while solid-state applications are still dominated by generalized gradient approximated (GGA) or meta-GGA functionals.[20] However, a certain amount of Hartree-Fock exchange reduces the self interaction error and thus rather generally improves the electronic structure description of finite gap materials.[21–24] Indeed, it has been empirically shown that hybrid functionals improve over GGAs for many important properties like band gaps, themochemistry, kinetics, and noncovalent interactions.[5,25,26]

Recently a hierarchy of cost-effective composite electronic structure methods has been proposed.[27] This hierarchy provides a comprehensive description of the available compromises between cost and accuracy for calculations on (mainly) molecular systems of increasing size. They are based on the pure Hartree-Fock (HF) method or HF/DFT hybrid functionals with the target of yielding excellent structures and reasonable energetic properties. Recently, its applicability has been extended to solid-state applications by (i) employing exchange-correlation functionals developed for solids (i.e. PBEsol[28] and HSEsol[29]); (ii) reducing the amount of HF

exchange in DFT hybrid methods for a better description of electronic properties[30–32] (e.g. 20-25%) and (iii) applying a simple recipe to make molecular basis sets suitable for solid-state calculations. The resulting methods have been dubbed as PBEsol0-3c and HSEsol-3c.

Related composite approaches have been used in computing crystal energy landscapes of active pharmaceutical ingredients.[33,34]

As hardware costs have plummeted over the last few years, it is not uncommon to have computing systems containing up to around 80 cores, with an appropriate amount of memory, at a cost of about 10k - 15k euros. The ready availability of such computational systems raises the question what types of calculations as described within the above hierarchy can be performed on such hardware, and what size of system can be addressed. Of course to address this question properly, cost effective methods and cost effective hardware must be also combined with an efficient implementation that permits calculations of large systems even on the limited number of cores provided by desktops, small clusters or even small-scale cloud computing resources. To that purpose, the methods described above have been implemented in the CRYSTAL17[12] code. CRYSTAL is characterised by the use of an atom-centred basis sets of Gaussian-type orbitals to expand Bloch functions[35] and by an extensive exploitation of the symmetry[36]. When compared to plane-wave codes, the computational cost due to the inclusion of exact exchange in the functional is small, and as such calculations employing hybrid functionals are competitive with those using pure XC functionals.
CRYSTAL can be run in serial and in parallel[37]. For the latter, two versions are available, namely Pcrystal and MPPcrystal. The first one is based on a replicated-data treatment of matrices and vectors while the second one distributes data among the processors. Details of the implementation and scaling are reported elsewhere[38–41] and we refer to those works for the discussion of the algorithms employed in the implementation. In this work, we adopted the MPP version of the code that has been shown to efficiently scale up to 30000 cores or more. However here, as stated above, we concentrate not on what can be done on the highest end; rather we intend to try and describe what can be done with the more modest but more readily available hardware that all researchers can access; the truth is that in many cases access to very high end computing systems is either difficult, or impractical, or both.

The point is simple: anyone can now access small-scale computing capabilities in their laboratories, without literally having state-of-the-art HPC machines at their fingertips. We therefore will concentrate on examining what systems may be solved in the 40-80 core regime, with 3.2 Gbytes of memory available per core, and will show that solid state calculations with hybrid DFT composite methods can be run for physical systems with more than 1000 atoms on computational systems with a relatively small number of cores at a moderate cost in terms of CPU time and memory.

II. COMPUTATIONAL METHODS

A. Methodologies

In this section we briefly recall the main ingredients of the composite hybrid methods which are examined in this work. The total energy provided can be written as[42]

$$E_{\rm tot}^{\rm sol\text{-}3c} = E_{\rm tot}^{\rm DFT/basis} + E_{\rm disp}^{\rm D3} + E_{\rm BSSE}^{\rm gCP} \qquad (1)$$

The different contributions are discussed in detail below. $E_{\rm tot}^{\rm DFT/basis}$ denotes the total energy evaluated with the hybrid XC functional. In the present work, the theoretical methods are the global hybrid PBEsol0 functional[28] and the range-separated hybrid HSEsol functional[29]. The two methods were chosen because they were specifically devised for solids and they use 25% of exact exchange in the hybrid exchange functional, which has been shown to be a good compromise.[43] The basis is the revised def2-mSVP for PBEsol0-3c and HSEsol-3c.[44] The DFT/basis total energy is supplemented with a damped atom-atom two-body dispersion energy as defined in the D3 approach:

$$E_{\rm disp}^{\rm D3(BJ)} = -\frac{1}{2}\sum_{AB}\sum_{n=6,8} s_n \frac{C_n^{AB}}{R_{AB}^n} f_{\rm damp}^{(n)}(R_{AB}) \qquad (2)$$

Here, C_n^{AB} denotes the *nth*-order dispersion coefficient (orders = 6, 8) for each atom pair AB, R^{AB} is their internuclear distances and s_n are the order-dependent scaling factors. The rational Becke-Johnson damping function[45] has become the default in combination with D3:

$$f_{\rm damp}^{(n)}(R_{AB}) = \frac{R_{AB}^n}{R_{AB}^n + (a_1 R_0^{AB} + a_2)^n} \qquad (3)$$

The damping function incorporates radii for atomic pairs $R_0^{AB} = \sqrt{C_8^{AB}/C_6^{AB}}$ and functional-specific parameters a_1 and a_2 that have been refitted in the present work for the different composite methods. In addition, the Axilrod–Teller–Muto[46,47] (ATM) three-body dipole-dipole-dipole term is also included.

The removal of the basis set superposition error (BSSE) due to the use of small basis sets is accomplished through the geometrical counterpoise correction:

$$E_{\rm BSSE}^{\rm gCP} = \frac{\sigma}{2}\sum_{AB} V_A^{\rm gCP}(R_{AB})\, f_{\rm damp}^{\rm gCP}(R_{AB}) \qquad (4)$$

The difference in atomic energy between a large (nearly complete) basis set and the target basis set for each free atom is used as a measure to generate the repulsive potential $V_A^{\rm gCP}$ with fitting parameters α, β, η. All functional and basis set specific parameters have been optimized for each revised composite method as discussed in Ref. 48.

B. Computational details

The calculations have been performed with a development version of CRYSTAL17[12,37] in its massively parallel version, using an all-electron split valence double-ζ basis set recently revised to deal with solid state calculations on inorganic systems.[44,48] For the numerical integration of the exchange-correlation term a (75,974) pruned grid was adopted, corresponding to the XLGRID keyword as used by the CRYSTAL code. The threshold for convergence of the SCF energy was set to 10^{-7} Ha. The Fock matrix in the reciprocal space was diagonalized at the central point of the first Brillouin zone (Γ point). Due to the large unit cells and non-conducting density of states, electronic structure calculations are already converged at the Γ-point. The tolerances for one- and two-electron integrals calculation were set to 10^{-6}, 10^{-6} for the calculation of Coulomb integrals and to 10^{-6}, 10^{-6} and 10^{-12} for the exchange integrals.

If not otherwise specified, calculations have been carried out on a small cluster composed of 4 nodes, each equipped with 20 core Intel® Xeon® E5-2630 v4 processors and 3.2 GB of RAM per core, and connected through a Gigabit Ethernet network. Inter-process communication is via the OpenMPI implementation of MPI.

C. Applicability of the presented methods

The advantage of the presented composite methods is that they permit running calculations with state-of-the-art hybrid functionals from small-scale to large-scale computing resources, even though here we focus on the former. This is non-trivial to achieve by other theoretical approaches in solid state calculations (e.g. plane-waves/pseudopotentials) for which applications are usually limited to semilocal functionals. However, several numerical techniques have emerged that try to address these computational bottlenecks.[49–52] Here, the trade-off between accuracy and cost has been made possible by a well balanced mixing of a small-to-medium sized double-zeta quality basis set and semiclassical corrections to cope with dispersion energy and to correct for the basis set superposition error, thus making them competitive with more costly triple-zeta quality calculations.

Numerical applicability to large-scale simulations are related to memory usage, speed-up and inter-process communication, and we see the following advantages and disadvantages:

(+) Low memory usage because of a smaller size of the matrices and vectors that are usually kept in memory but can be distributed among the processors.

(+) Efficient scaling for the calculation of one- and two-electron integrals as well as the diagonalization and exploitation of all space group symmetries.

(+) Use of efficient distributed memory parallel libraries and avoidance of I/O throughout the calculation resulting in good scaling of the total calculation

(-) High computational cost of individual integral evaluation compared to plane wave basis.

(-) Necessity to orthogonalize the basis set overlap, which might lead to linear dependencies in dense systems.

(-) Less clear route towards complete basis set results.

In terms of material classes, we have extended the original composite methods substantially and can now seamlessly describe molecules and molecular crystals, biological materials, oxides, nanomaterials, and layered materials. However, there are a range of systems and materials, where the presented approaches have to be used carefully. Whenever strong correlation occurs, the high amount of Fock exchange will be problematic and pure semi-local functionals are known be more robust. The presented methods might also fail for materials with vanishing band gap (e.g. metals), where unscreened Fock exchange is not appropriate. Here, a semi-local functional is preferred and its evaluation in a plane-wave expansion will most likely be more efficient due to the delocalized nature of the valence band. Even the London dispersion correction is not well founded for metallic systems as it is based on a partitioning of the response function to local (atomic) contributions.

III. RESULTS AND DISCUSSION

A. Benchmarks and scaling

We first analyze the efficiency of the sol-3c composite methods for several benchmark systems with increasing size from 100 atoms to 3000 atoms, For this, we consider the wall-clock time for an SCF and gradient (both atomic and cell component) calculation on 40 and 80 cores using MPPcrystal. The selected systems are listed in Table III A and show a variety of periodicities ranging from 1D through 2D to 3D, and also molecules and clusters (0D). We focused our attention on:

(i) Biologically relevant systems such as the small protein of Crambin and its crystalline anhydrous structure, the triple helix of collagen and 1D models of DNA made of an infinite-chain of Adenine-Thymine (AT) and Cytosine-Guanine (CG) homopolynucleotides.

(ii) The external and inner surfaces of amorphous silica through 2D and 3D models, respectively, with adsorbed drug molecules.

(iii) A supercell 3D-model of the mineral forsterite along with a corresponding nanoparticle (0D-model).

TABLE I: HSEsol-3c wall-clock time (sec.) required for SCF and gradients calculation performed on 40 and 80 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 basis functions.

system	PBC	N_{op}	N_{at}	N_{AO}	S_{40}	S_{80}	T_{40} (N_{cyc})	T_{80} (N_{cyc})	G_{40}	G_{80} [a]
Crambin[53]	0D	1	642	5553	370	234	5801 (16)	4044 (16)	974	469
Forsterite[54]	0D	1	812	13224	1871	1260	29195 (14)	19424 (14)	8004	3752
Collagen[55]	1D	7	245	2128	24	29	401 (12)	426 (12)	74	42
poly(C)-poly(G)	1D	11	715	7359	157	219	2155 (14)	3749 (15)	176	92
poly(A)-poly(T)	1D	11	726	7381	153	249	2359 (15)	3816 (15)	179	96
silica45[56]	2D	1	111	1457	64	37	683 (10)	464 (10)	310	161
ASP@silica45[56]	2D	1	132	1668	74	73	907 (12)	677 (12)	349	181
IBU@silica45[56]	2D	1	144	1718	71	47	884 (11)	659 (11)	379	196
NIT@silica45[57]	2D	1	143	1794	77	56	998 (12)	747 (12)	403	208
CTZ@silica45	2D	1	237	2693	139	88	1982 (13)	1362 (13)	683	352
TPM[58]	3D	24	856	6184	31	-	487 (15)	-	324	-
MCM-41[59]	3D	1	579	7785	878	449	7027 (12)	5740 (12)	3706	1885
IBU@MCM-41[59]	3D	1	612	8046	561	439	7361 (12)	5863 (12)	3905	1983
7IBU@MCM-41[59]	3D	1	810	9612	794	605	10232 (12)	8143 (12)	5473	2753
Crambin[53]	3D	2	1536	12702	913	777	12530 (13)	11346 (13)	6353	3256
Forsterite[b]	3D	2	1008	16416	2272	1573	28149 (11)	19934 (11)	15693	7970
MIL-100 (Al)[60]	3D	16	2788	37128	-	2965	-	59846 (16)	-	1763

[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 gradient time, respectively, computed on N CPU cores.

[b] 4x3x3 supercell is used.

(iv) Several Metal-Organic Frameworks (MOFs) with increasing number of atoms including the so-called *giant*-MOF MIL-100.

Notably, this set covers systems that span molecular crystals and inorganic solids, dense phases and microporous materials, thus showing the wide applicability of the composite methods, along with the code, to model solid-state properties. Most of the benchmark systems have been studied previously, however, the use of 512-1024 cores or more on Tier-0 HPC centers was mandatory (PRACE project)[61–63].

The wall-clock times for a full SCF procedure (single SCF cycle and total SCF with the number of SCF cycles) and the calculation of both atomic and cell gradients with HSEsol-3c are reported in Table III A. For sake of conciseness, in the following we mostly refer to calculations with the HSEsol-3c method, while results for the global hybrid PBEsol-3c are gathered in the supplementary information.

Timings in Table III A show that all examined systems up to 1500 atoms (or 16000 atomic orbitals) are affordable on 40 cores while for the largest one (i.e. MIL-100) the calculation is possible only on 80 cores because of the memory requirements due to the size of the matrices and screening tables used by the code. Actually, the distributed-data version of the code allows for an efficient partitioning of the matrices over the cores so that usually it is enough to increase the number of cores to bypass memory limits, as is seen here by the success of the code on 80 cores, but its failure on 40. This largest system also gives some indication of the maximum size that may be studied on the computational resources that we are considering here, and it can be seen we could reach almost 2800 atoms (32000 AOs) by using 80 cores. Further improvements in MPPcrystal are currently being implemented, in particular the use of OpenMP, which should push this limit further.

Results show a satisfactory scaling of the wall-clock time when passing from 40 to 80 cores for most of the examined systems. While, on average, the cost of a single SCF cycle and the total SCF process decreases adequately with an efficiency of 55-70%, the speed-up for nuclear and cell gradients is overall excellent (efficiency $> 90\%$). This is due to the very good scalability of the one- and two-electron integrals. In contrast, the computational cost of the SCF cycle depends on both integral evaluation and diagonalization of the Kohn-Sham (KS) matrix with the performance of latter being critically influenced by communication among processors. Here we are using Gigabit Ethernet for the inter-node communication. This is hardly the most efficient fabric, and the relatively good scaling despite this is gratifying. For a

few systems the wall-clock time for the SCF process does not scale at all. They all have a rather high space group symmetry, which CRYSTAL effectively exploits, thus reducing the time to evaluate the integrals to a negligible time when compared to the diagonalisation. The resulting KS matrices are relatively small and we thus cannot benefit from an increasing number of cores. Here, the replicated-data version of the code (Pcrystal), which exploits this symmetry in reciprocal space and does not attempt to distribute the small matrices, would be more efficient than the distributed-data one (MPPcrystal).

For the two largest systems, i.e. Crambin (1536 atoms, 40 CPU cores) and MIL-100 (2788 atoms, 80 CPU cores) the computational time for SCF and gradient calculations are 5 h and 17 h, respectively. This implies that full geometry relaxations are feasible within weeks.

1. *Metal Organic Frameworks*

To show the weak and strong scaling of the code when using sol-3c composite methods, we refer to the class of hybrid organic-inorganic microporous materials known as Metal-organic frameworks[64,65]. They are composed of an inorganic cluster (or a metal) and an organic linker that self-assemble in a three dimensional, microporous, crystalline structure. They have a wide range of potential applications including gas storage[66] and separation[67], drug delivery[68], (photo)catalysis[69], sensing[70] and optoelectronics[71]. The modular structure, which permits one to vary and combine the two different secondary building blocks, imparts to MOFs an extraordinary chemical versatility such that a multitude of frameworks with different chemical composition, topology and size have been synthesized so far[72]. Therefore, they are ideal to illustrate the method scaling with system size.

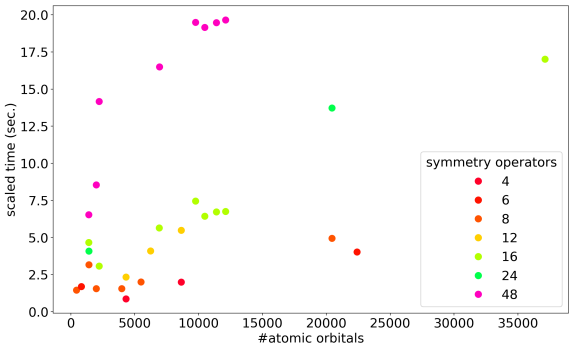


FIG. 1: HSEsol-3c single SCF scaled time for the MOF dataset. Data are plotted against the number of atomic orbitals. The plot includes results for both high symmetry (cubic) and the corresponding symmetry-lowered MOFs, coloring indicates the number of symmetry operators.

Single-point energy and gradient calculations have been performed with sol-3c composite methods on HFsol-3c[48] optimized geometries for a data set containing 21 MOFs with unit cells of increasing size and different symmetry (i.e. space groups). The full list of examined MOFs is reported as supporting information along with timings for the different calculation steps. Figure 1 shows the scaled time for a single SCF cycle (i.e. the wall-clock time of an SCF cycle multiplied by the number of symmetry operations and normalized to the number atoms) as plotted against the number of basis functions. Scaled computation times scale linearly with the number of atomic orbitals for systems with the same, or similar, number of symmetry operators. The slope of the data decreases when the symmetry reduces. A closer inspection of Figure 1 shows that three regimes can be highlighted for highly symmetric MOFs (cubic space groups (SG)), medium-symmetry frameworks (tetragonal and hexagonal SG) and low-symmetry systems (orthorhombic, monoclinic and triclinic SG). This means that the relative cost of the calculation decreases when the system becomes larger and the symmetry reduces.
This is not unexpected because MPPcrystal turns out to be more efficient for low-symmetry and large-size systems, and indeed it was designed *a priori* for this particular use case. For high-symmetry and small-size systems the replicated data version of the code is probably more efficient due to a better exploitation of the symmetry in the diagonalization of the Kohn-Sham matrix.
Overall, cost-effective sol-3c methods show a good weak scaling behaviour with the size of the system.

2. *MPPcrystal scaling for MIL-100 (Al)*

To demonstrate how sol-3c composite methods can be efficiently used even on a larger number of cores, we consider the MOF MIL-100(Al), which contains a 2788-atom primitive unit cell with a tetragonal space group (16 symmetry operators) and 37128 basis functions. Calculations were performed on an increasing number of processors up to 288 on ARCHER, whose compute nodes contain two 2.7 GHz, 12-core Intel® Xeon® E5-2697 v2 (Ivy Bridge) series processors, and 2.7 GByte per core. Each of the cores in these processors can support 2 hardware threads (Hyperthreads). Within the node, the two processors are connected by two Intel® QuickPath Interconnect (QPI) links.

Figure 2 documents the speedup of the code up to 288 cores relative to the timings on 48 cores for the HSEsol-3c composite method, but the cost of the calculation is rather similar for the PBEsol0 -3c one (see the supporting information). It can be clearly seen that for the different steps of a single SCF cycle (i.e. integrals evaluation and KS matrix diagonalization) and atomic gradient calculation the scaling is fairly good with an efficiency that is as high as 65-70% for integrals and atomic gradients calculation that are related to each other as previously discussed, and around 60% for the diagonalization and

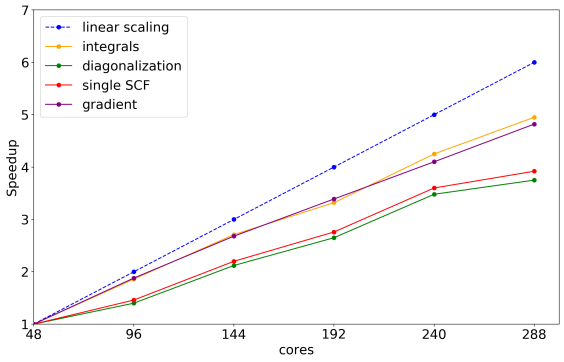


FIG. 2: HSEsol-3c wall-clock time speed-up (t_n/t_{48}, with n the number of cores) for integrals calculation, Kohn-Sham matrix diagonalization, single SCF and gradients calculation performed on increasing numbers of cores. Linear scaling is reported as dotted blue line.

in turn for the single SCF cycle whose efficiency mostly depends on the diagonalization step. On 288 cores, the wall-clock time for the whole SCF and atomic gradients with HSEsol-3c is just 1h15m and opens the possibility to extend calculations to geometry relaxation and other properties.

B. Applications

Given the satisfactory results for benchmarks and scaling as discussed in the previous section, we focus now on two applications. First, we extend calculations on DNA 1D models to show how structure, energetics, and electronic properties change between adenine-thymine (AT) and guanine-cytosine (GC) homo-polynucleotides. The second application encompasses the realm of drug-delivery and drug interaction with the external and internal surface of amorphous silica (2D and 3D periodic models) to compare computed interaction energies with data obtained in previous works.

1. *Biological Systems*

Complex biological systems are among the most important targets for large scale DFT simulations. Herein, we report the results obtained with sol-3c composite methods on poly(A)-poly(T) and poly(G)-poly(C) model systems containing both eleven basis pairs in the unit cell (i.e. a full helix). The A-form double-helix structure for the two homo-polynucleotides was completely relaxed with both HSEsol-3c and PBEsol0-3c and are shown in Figure 3(a) and 3(d). Relevant results of the present calculations are gathered in Table II.

The optimized lattice parameters computed for both systems and methods show that the CG infinite chain is slightly less compact than poly(A)-poly(T).[73] The lattice parameters for CG and AT infinite chains and, in particular the base stacking distances (i.e. 3.18 Åfor poly(A)-poly(T) and 3.33 Åfor poly(G)-poly(C), are consistent with average value of 3.2 from experimental data.[73]

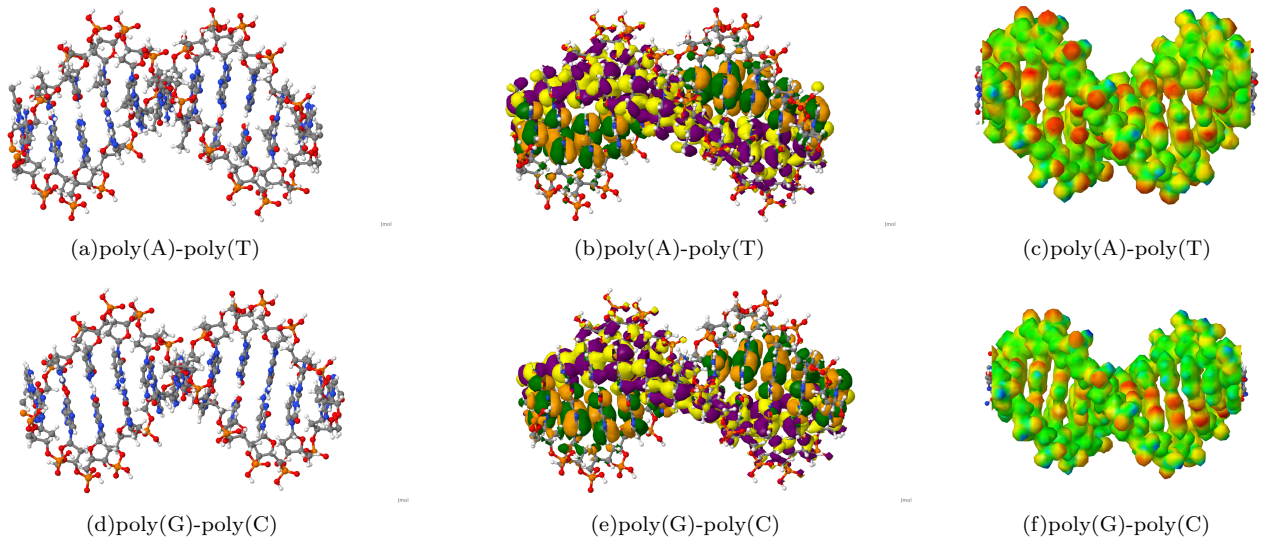
The energetics and electronic properties of the two double-strands have also been studied for the fully relaxed structures. We have calculated the interaction energy between the two strands in the frozen double-strand geometry (E_{int}) as well as the deformation energy upon relaxation (E_{def}), which describes the increase in intramolecular energy when the double-strand is formed. Results are also presented in Table II. As expected from Watson-Crick base pairing, the CG homo-polynucleotide shows a larger interaction energy than the AT one. The presence of an additional hydrogen bond between C and G bases makes the inter-strand interaction stronger and this likely leads to a slightly weaker base stacking that causes the longer lattice parameter with respect to the AT homo-polynucleotide. Notably, the energy difference between GC and AT double-strands agrees with the experimental evidence that CG sequences tends to assume an A-form DNA conformation while for AT sequences the B-form is favoured[74].

The electrostatic potential mapped on top of an electron density isosurface as shown in Figure 3(c) and 3(f) illustrates that negative regions (in red) are dominated by phosphate groups while small positive spots (in blue) are located around the hydrogen atoms that neutralize the whole structure. As expected the hydrophilic region is outside the double-strand while the electrostatic potential inside the minor and major groove is controlled by the nucleic bases for both systems.

Electronic and dielectric properties of DNA have become of relevance for DNA-based nanowires and nanodevices[75]. The band gap ranges from 3.04 eV for poly(G)-poly(C) system to 4.07 eV for poly(A)-poly(T) as computed with the HSEsol-3c level method while the PBEsol0-3c method gives slightly larger results, see Table II. For the GC infinite chain, the predicted band gap is larger and likely more reliable than the one computed with the PBE functional.[76] as expected for hybrid functionals[30–32]. Band structure and density-of-states for both methods can be found in the Supporting Informa-

TABLE II: Computed PBEsol0-3 and HSEsol-3c properties of the poly(A)-poly(T) and poly(G)-poly(C) double-strand.

	cell parameter [Å]	band gap [eV]	E_{int} [kcal/mol]	E_{def} [kcal/mol]
poly(A)-poly(T)				
PBEsol0-3c	34.98	4.79	-21.9	-7.5
HSEsol-3c	34.92	4.07	-22.3	-7.7
poly(G)-poly(C)				
PBEsol0-3c	36.63	3.73	-41.5	-11.9
HSEsol-3c	37.08	3.04	-41.9	-12.1



(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. 3: HSEsol-3c optimized structure of poly(A)-poly(T) and poly(G)-poly(C) DNA 1D models (a, d), highest occupied (green and orange) and lowest unoccupied (yellow and purple) crystal orbitals (b ,e), electrostatic potential mapped on an electron density isosurface (positive regions in blue and negative regions in red) (c, f). Color code for atoms: O in red, C in grey, H in white, N in blue, and P in orange .

tion. The topmost valence band for poly(G)-poly(C) is associated with the π-like highest occupied crystalline orbitals (HOCO) of the purine bases, while the bottom of the conduction bands (i.e. the lowest unoccupied crystalline orbitals (LUCO)) is mostly localized around the pyrimidine bases, as shown in Fig. 3(e) and 3(b).

2. *Amorphous silica surfaces*

Silica-based materials such as amorphous silica have relevant applications in many fields as solid phase in chromatography, support in catalysis and also as excipient and drug-delivery carrier in pharmaceutical formulations. Its usage is mainly related to peculiar surface properties in that surface Si-OH groups (silanols) act as adsorption centres via the formation of H-bonds with the functional groups of the adsorbed molecules. Recently, the adsorption of different drugs (i.e. aspirin, clotrimazole and ibuprofen) at the external surface of amorphous silica and the confinement in silica-based mesoporous materials (e.g. MCM-41) has been investigated. Realistic 2D and 3D models were devised in order to take into account the complexity of the disordered structure.

Here, we apply sol-3c methods to predict the interaction energy between several drugs and the amorphous silica substrate. The 2D model, dubbed silica45[77,78] corresponds to a fully hydroxylated silica surface while the 3D model is a long-range ordered model of the MCM-41 mesoporous silica[79]. We performed a single point SCF calculation with PBEsol0-3c and HSEsol-3c on optimized geometries taken from Ref.[56,59] and compared the results with previous works to assess the accuracy of composite methods. It is worth noting that Ugliengo and co-workers run calculations on Tier 0 HPC machines while here we used just 80 cores. Computed sol-3c interaction energies are reported in Table III.

TABLE III: Interaction energy of several drugs with silica45 and MCM-41, respectively. Single point PBEsol0-3c and HSEsol-3c calculations have been performed on optimized structures taken from Ref. 56,59. All values are in kJ/mol.

	[a]ASP@ silica45	[b]CTZ@ silica45	[c]IBU@ silica45	[c]IBU@ MCM-41	[c]7IBU@ MCM-41
PBEsol0-3c	-101.7	-129.8	-116.5	-121.6	-127.4
HSEsol-3c	-106.5	-130.2	-120.0	-126.0	-130.3
PBE-D[56]	-112.1	-	-118.0	-	-
B3LYP-D*[56,59]	-	[d]-78.4	-94.3	-99.5	-105.6

[a] Aspirin
[b] Clotrimazole
[c] Ibuprofen with one or seven seven molecules in the pore
[d] Binding of -134.1 kJ/mol reported earlier.[56]

Both PBEsol0-3c and HSEsol-3c provide results in agreement with PBE-D and B3LYP-D* data from Ref.[56,59]. Composite methods slightly overestimate the interaction energy between drug and substrate with respect to B3LYP-D* results. However, the relative bindings (e.g. for the three IBU adsorptions) are preserved and results are computed with substantially reduced computational overhead compared to the original studies.

IV. SUMMARY AND CONCLUSIONS

In this work we have shown how cost-effective quantum chemical methods tailored for solid-state calculations can be used to treat large-scale systems on small-scale computing resources.

We have addressed the following points:

1) *Do we need a HPC machine?*
We have selected systems for which calculations were previously run on supercomputing facilities and required hundreds, if not thousands, of cores. Using the sol-3c composite methods 40, cores may be enough for system up to 1500 atoms.

2) *Which is the largest system we can deal with when using 80 cores?*
We have shown that for a set of Metal-Organic frameworks of increasing size that we can run calculations for systems up to about 3000 atoms. A fairly good scaling has been observed and low-symmetry and large systems are efficiently treated using the MPP version of CRYSTAL.

3) *Do sol-3c hybrid methods scale with more than 80 cores?*
We have shown an excellent strong scaling up to 288 cores for the *giant* MOF MIL-100. In perspective, sol-3c methods do scale to larger HPC facilities, and these facilities can potentially be applied to investigate even larger systems.

Finally, we have demonstrated that the new composite methods can be applied to systems of scientific interest. The application to DNA-like 1D-model chains demonstrates how structure, energetics and electronic properties of AT and CG homo-polynucleotides can be computed at a quantum mechanical level. In a second application sol-3c methods have been employed to investigate drug-delivery with amorphous silica surface for both 2D (external surface) and 3D (internal surface) models. Results for interaction energies are in good agreement to previous theoretical data that were previously obtained on Tier-0 HPC machines.

In conclusion, the combination of hybrid DFT composite methods with their implementation in the CRYSTAL code allowed us to quickly and efficiently tackle large systems up to thousands of atoms, providing accurate results with a comparatively modest hardware requirement.

Conflicts of interest

There are no conflicts of interest to declare

Acknowledgements

This work used the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk). We thank Piero Ugliengo, Michele Cutini, and Lorenzo Zamirri for helpful discussions.

Electronic Supplementary Information

All tables and figures for miscellaneous systems, MOF dataset, MIL-100 scaling, DNA 1D models and amorphous silica as computed with PBEsol0-3c. Band structure and DOSs of DNA 1D models for HSEsol-3c.

[1] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.
[2] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.
[3] K. Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136(15):150901, 2012.
[4] R. J. Maurer, C. Freysoldt, A. M. Reilly, J. G. Brandenburg, O. T. Hofmann, T. Björkman, S. Lebègue, and A. Tkatchenko. Advances in density-functional calculations for materials modeling. *Annu. Rev. Mater. Res.*, 49:3.1–3.30, 2019.
[5] R. Peverati and D. G. Truhlar. Quest for a universal density functional: the accuracy of density functionals across a broad spectrum of databases in chemistry and physics. *Philos. Trans. R. Soc. A*, 372:20120476, 2014.
[6] J. Sun, A. Ruzsinszky, and J. P. Perdew. Strongly constrained and appropriately normed semilocal density functional. *Phys. Rev. Lett.*, 115:036402, 2015.
[7] N. Mardirossian and M. Head-Gordon. ωb97x-v: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy. *Phys. Chem. Chem. Phys.*, 16:9904–9924, 2014.
[8] Y. Wang, X. Jin, S. Y. Haoyu, D. G. Truhlar, and X. He. Revised m06-l functional for improved accuracy on chemical reaction barrier heights, noncovalent interactions, and solid-state physics. *Proc. Natl. Acad. Sci.*, 114:8487–8492, 2017.
[9] S. Grimme, A. Hansen, J. G. Brandenburg, and C.h Bannwarth. Dispersion-corrected mean-field electronic structure methods. *Chem. Rev.*, 116:5105–5154, 2016.
[10] J. Klimeš and A. Michaelides. Perspective: Advances and challenges in treating van der waals dispersion forces in density functional theory. *J. Chem. Phys.*, 137:120901, 2012.
[11] G. J. O. Beran. Modeling polymorphic molecular crystals with electronic structure theory. *Chem. Rev.*, 116:5567–5613, 2016.
[12] R. Dovesi, A. Erba, R. Orlando, C. M. Zicovich-Wilson, B. Civalleri, L. Maschio, M. Rérat, S. Casassa, J. Baima, S. Salustro, et al. Quantum-mechanical condensed matter

simulations with CRYSTAL. *WIREs Comput Mol Sci.*, 8:e1360, 2018.

[13] J. Hafner and G. Kresse. The vienna ab-initio simulation program vasp: An efficient and versatile tool for studying the structural, dynamic, and electronic properties of materials. In *Properties of Complex Inorganic Solids*, pages 69–82. Springer, 1997.

[14] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. I. J. Probert, K. Refson, and M. C. Payne. First principles methods using castep. *Z. Kristallogr. Cryst. Mater.*, 220:567–570, 2005.

[15] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, et al. Quantum espresso: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter*, 21(39):395502, 2009.

[16] J. Hutter, M. Iannuzzi, F. Schiffmann, and J. VandeVondele. cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 4:15–25, 2014.

[17] K. Lejaeghere, G. Bihlmayer, T. Björkman, P. Blaha, S. Blügel, V. Blum, D. Caliste, I. E. Castelli, S. J. Clark, A. Dal Corso, et al. Reproducibility in density functional theory calculations of solids. *Science*, 351:aad3000, 2016.

[18] L. E Ratcliff, S. Mohr, G. Huhs, T. Deutsch, M. Masella, and L. Genovese. Challenges in large scale quantum mechanical calculations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 7:e1290, 2017.

[19] A. Nakata, J. S. Baker, S. Y. Mujahed, J. TL Poulton, S. Arapan, J. Lin, Z. Raza, S. Yadav, L. Truflandier, T. Miyazaki, et al. Large scale and linear scaling dft with the conquest code. *J. Chem. Phys.*, 152(16):164112, 2020.

[20] K. Burke. Perspective on density functional theory. *J. Chem. Phys.*, 136:150901, 2012.

[21] J. P. Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B*, 23:5048–5079, 1981.

[22] A. J. Cohen, P. Mori-Sanchez, and W. Yang. Challenges for Density Functional Theory. *Chem. Rev.*, 112:289–320, 2012.

[23] D. Hait and M. Head-Gordon. Delocalization Errors in Density Functional Theory Are Essentially Quadratic in Fractional Occupation Number. *J. Phys. Chem. Lett.*, 9:6280–6288, 2018.

[24] P. Mori-Sanchez, A. J. Cohen, and W. Yang. *J. Chem. Phys.*, 125:201102, 2006.

[25] L. Goerigk, A. Hansen, C. Bauer, S. Ehrlich, A. Najibi, and S. Grimme. A look at the density functional theory zoo with the advanced gmtkn55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.*, 19:32184–32215, 2017.

[26] N. Mardirossian and M. Head-Gordon. Thirty years of density functional theory in computational chemistry: an overview and extensive assessment of 200 density functionals. *Mol. Phys.*, 115(19):2315–2372, 2017.

[27] E. Caldeweyher and J. G. Brandenburg. Simplified dft methods for consistent structures and energies of large systems. *J. Phys. Condens. Matter*, 30:213001, 2018.

[28] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, O. A. Vydrov, G. E. Scuseria, L. A. Constantin, X. Zhou, and K. Burke. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.*, 100:136406, 2008.

[29] L. Schimka, J. Harl, and G. Kresse. Improved hybrid functional for solids: The HSEsol functional. *J. Chem. Phys.*, 134:024116, 2011.

[30] P. Pernot, B. Civalleri, D. Presti, and A. Savin. Prediction uncertainty of density functional approximations for properties of crystals with cubic symmetry. *J. Phys. Chem. A*, 119:5288–5304, 2015.

[31] J. M. Crowley, J. Tahir-Kheli, and W. A. Goddard III. Resolution of the band gap prediction problem for materials design. *J. Phys. Chem. Lett.*, 7:1198–1203, 2016.

[32] A. J Garza and G. E. Scuseria. Predicting band gaps with hybrid density functionals. *J. Phys. Chem. Lett.*, 7:4165–4170, 2016.

[33] L. M. LeBlanc and E. R. Johnson. Crystal-energy landscapes of active pharmaceutical ingredients using composite approaches. *CrystEngComm*, 21:5995–6009, 2019.

[34] A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J Bygrave, R. Bylsma, J. E. Campbell, R. Car, et al. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Cryst. B*, 72:439–459, 2016.

[35] R. Dovesi, B. Civalleri, R. Orlando, C. Roetti, and V. R. Saunders. Ab initio quantum simulation in solid state chemistry. *Rev. Comp. Ch.*, 21:1, 2005.

[36] M. De La Pierre, R. Orlando, M. Ferrabone, C. M. Zicovich-Wilson, and R. Dovesi. Exploitation of symmetry in periodic self-consistent-field ab initio calculations: application to large three-dimensional compounds. *Sci. China Chem.*, 57:1418–1426, 2014.

[37] R. Dovesi, V. R. Saunders, C. Roetti, R. Orlando, C. M. Zicovich-Wilson, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'Arco, M. Llunell, M. Causà, Y. Noël, L. Maschio, A. Erba, M. Rerat, and S. Casassa. CRYSTAL17, Universitá di Torino, Torino. 2017.

[38] I. J. Bush, S. Tomić, B. G. Searle, G. Mallia, C. L. Bailey, B. Montanari, L. Bernasconi, J. M. Carr, and N. M. Harrison. Parallel implementation of the ab initio crystal program: electronic structure calculations for periodic systems. *Proc. R. Soc. A-Math. Phys. Eng. Sci.*, 467:2112–2126, 2011.

[39] R. Orlando, M. Delle Piane, I. J. Bush, P. Ugliengo, M. Ferrabone, and R. Dovesi. A new massively parallel version of crystal for large systems on high performance computing architectures. *J. Chem. Phys.*, 33:2276–2284, 2012.

[40] A. Erba, J. Baima, I. Bush, R. Orlando, and R. Dovesi. Large-scale condensed matter dft simulations: performance and capabilities of the crystal code. *J. Chem. Theory Comput.*, 13(10):5019–5027, 2017.

[41] R. Dovesi, F. Pascale, B. Civalleri, K. Doll, N. M. Harrison, I. J. Bush, P. D'arco, Y. Noël, M. Rérat, P. Carbonniere, et al. The crystal code, 1976–2020 and beyond, a long story. *J. Chem. Phys.*, 152:204111, 2020.

[42] E. Caldeweyher and J. G. Brandenburg. Simplified DFT methods for consistent structures and energies of large systems. *J. Phys.: Condens. Matter*, 30:213001, 2018.

[43] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The pbe0 model. *J. Chem. Phys.*, 110:6158–6170, 1999.

[44] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.*, 143:054107, 2015.

[45] E. R. Johnson and A. D. Becke. A post-hartree-fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.*, 124:174104, 2006.

[46] B. M. Axilrod and E. Teller. Interaction of the van der Waals type between three atoms. *J. Chem. Phys.*, 11:299–300, 1943.

[47] Y. Muto. Force between nonpolar molecules. *J. Phys. Math. Soc. Jpn*, 17:629–631, 1943.

[48] L. Doná, J. G. Brandenburg, and B. Civalleri. Extending and assessing composite electronic structure methods to the solid state. *J. Chem. Phys.*, 151:121101, 2019.

[49] J. Hutter M. Guidon and J. VandeVondele. Auxiliary density matrix methods for hartreefock exchange calculations. *J. Chem. Theory Comput.*, 6:2348–2364, 2010.

[50] L. Lin. Adaptively compressed exchange operator. *J. Chem. Theory Comput.*, 12:2242—-2249, 2016.

[51] M. Hutchinson and M. Widom. Vasp on a gpu: Application to exact-exchange calculations of the stability of elemental boron. *Comput. Phys. Commun.*, 183:1422—-1426, 2012.

[52] B. Santra X. Wu R. Car H.-Y. Ko, J. Jia and R. A. DiStasio Jr. Enabling large-scale condensed-phase hybrid density functional theory based ab initio molecular dynamics. 1. theory, algorithm, and performance. *J. Chem. Theory Comput.*, 16:3757–3785, 2020.

[53] M. Delle Piane, M. Corno, R. Orlando, R. Dovesi, and P. Ugliengo. Elucidating the fundamental forces in protein crystal formation: the case of crambin. *Chem. Sci.*, 7:1496–1507, 2016.

[54] L. Zamirri, M. Corno, A. Rimola, and P. Ugliengo. Forsterite surfaces as models of interstellar core dust grains: computational study of carbon monoxide adsorption. *ACS Earth Space Chem.*, 1:384–398, 2017.

[55] M. Cutini, S. Pantaleone, and P. Ugliengo. Elucidating the nature of interactions in collagen triple-helix wrapping. *J. Phys. Chem. Lett.*, 10:7644–7649, 2019.

[56] M. Delle Piane, M. Corno, and P. Ugliengo. Does dispersion dominate over h-bonds in drug–surface interactions? the case of silica-based materials as excipients and drug-delivery agents. *J. Chem. Theory Comput.*, 9:2404–2415, 2013.

[57] A. Gignone, M. Delle Piane, M. Corno, P. Ugliengo, and B. Onida. Simulation and experiment reveal a complex scenario for the adsorption of an antifungal drug in ordered mesoporous silica. *J. Phys. Chem. C*, 119:13068–13079, 2015.

[58] S. Rösel, H. Quanz, C. Logemann, J. Becker, E. Mossou, L. Canadillas-Delgado, E. Caldeweyher, S. Grimme, and P. R. Schreiner. London dispersion enables the shortest intermolecular hydrocarbon h··· h contact. *J. Am. Chem. Soc.*, 139:7428–7431, 2017.

[59] M. Delle Piane, M. Corno, A. Pedone, R. Dovesi, and P. Ugliengo. Large-scale b3lyp simulations of ibuprofen adsorbed in mcm-41 mesoporous silica as drug delivery system. *J. Phys. Chem. C*, 118:26737–26749, 2014.

[60] 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.

[61] B. Civalleri. Ab initio modelling of the adsorption in giant metal-organic frameworks: from small molecules to drugs, 2013. PRACE Project 201308168.

[62] M. D'Amore, B. Civalleri, I. J. Bush, E. Albanese, and M. Ferrabone. Elucidating the interaction of co2 in the giant metal–organic framework mil-100 through large-scale periodic ab initio modeling. *J. Phys. Chem. C*, 123:28677–28687, 2019.

[63] P. Ugliengo. Mesoporous silica for drug delivery: a quantum mechanical simulation, 2011. PRACE Project 2011050810.

[64] J. R. Long and O. M. Yaghi. The pervasive chemistry of metal–organic frameworks. *Chem. Soc. Rev.*, 38:1213–1214, 2009.

[65] S. R. Batten, N. R. Champness, X. Chen, J. Garcia-Martinez, S. Kitagawa, L. Öhrström, M. O'Keeffe, M. P. Suh, and J. Reedijk. Terminology of metal–organic frameworks and coordination polymers (iupac recommendations 2013). *Pure Appl. Chem.*, 85:1715–1724, 2013.

[66] S. Ma and H. Zhou. Gas storage in porous metal–organic frameworks for clean energy applications. *Chem. Comm.*, 46:44–53, 2010.

[67] H. Li, K. Wang, Y. Sun, C. T. Lollar, J. Li, and H. Zhou. Recent advances in gas storage and separation using metal–organic frameworks. 21:108–121, 2018.

[68] P. Horcajada, T. Chalati, C. Serre, B. Gillet, C. Sebrie, T. Baati, J. F. Eubank, D. Heurtaux, P. Clayette, C. Kreuz, et al. Porous metal–organic-framework nanoscale carriers as a potential platform for drug delivery and imaging. *Nat. Mater.*, 9:172–178, 2010.

[69] M. A. Nasalevich, F. Van der Veen, M .and Kapteijn, and J. Gascon. Metal–organic frameworks as heterogeneous photocatalysts: advantages and challenges. *CrystEngComm*, 16:4919–4926, 2014.

[70] Pawan Kumar, Akash Deep, and Ki-Hyun Kim. Metal organic frameworks for sensing applications. *Trends Anal. Chem.*, 73:39–53, 2015.

[71] V. Stavila, A. A. Talin, and M. D. Allendorf. Mof-based electronic and opto-electronic devices. *Chem. Soc. Rev.*, 43(16):5994–6010, 2014.

[72] M. Li, D. Li, M. O'Keeffe, and O. M. Yaghi. Topological analysis of metal–organic frameworks with polytopic linkers and/or multiple building units and the minimal transitivity principle. *Chem. Rev.*, 114:1343–1370, 2014.

[73] It is important to note that structural properties herein discussed can not be directly compared with the experimental results. A useful comparison would require the inclusion of solvation and temperature effects in the computational model.

[74] D. W. Ussery. Dna structure: A-, b-and z-dna helix families. *e LS*, 2001.

[75] P. J. De Pablo, F. Moreno-Herrero, J. Colchero, J. G. Herrero, P. Herrero, A. M. Baró, P. Ordejón, J. M. Soler, and E. Artacho. Absence of dc-conductivity in λ-dna. *Phys. Rev. Lett.*, 85:4992, 2000.

[76] E. Artacho, M. Machado, D. Sánchez-Portal, P. Ordejón, and J. Soler. Electrons in dry dna from density functional calculations. *Mol. Phys.*, 101:1587–1594, 2003.

[77] P. Ugliengo, M. Sodupe, F. Musso, I. J. Bush, R. Orlando, and R. Dovesi. Realistic models of hydroxylated amorphous silica surfaces and mcm-41 mesoporous material simulated by large-scale periodic b3lyp calculations. *Adv. Mater.*, 20:4579–4583, 2008.

[78] V. Bolis, C. Busco, and P. Ugliengo. Thermodynamic study of water adsorption in high-silica zeolites. *J. Phys. Chem. B*, 110:14849–14859, 2006.

[79] J. Scott Beck, J. C. Vartuli, W. J. Roth, M. E. Leonowicz, C. T. Kresge, K. D. Schmitt, C. T. W. Chu, D. H. Olson, E. W. Sheppard, S. B. McCullen, et al. A new family of mesoporous molecular sieves prepared with liquid crystal templates. *J. Am. Chem. Soc.*, 114:10834–10843, 1992.