Journal of
Materials Chemistry A

PAPER

View Article Online
View Journal | View Issue

Check for updates

Cite this: *J. Mater. Chem. A*, 2019, 7, 9593

An extensive comparative analysis of two MOF databases: high-throughput screening of computation-ready MOFs for CH_4 and H_2 adsorption†

Cigdem Altintas,‡[a] Gokay Avci,‡[a] Hilal Daglar,‡[a] Ayda Nemati Vesali Azar,‡[a] Ilknur Erucar, ‡[b] Sadiye Velioglu *[a] and Seda Keskin *[a]

Received 4th February 2019
Accepted 7th March 2019

DOI: 10.1039/c9ta01378d

rsc.li/materials-a

Computation-ready metal–organic framework (MOF) databases (DBs) have tremendous value since they provide directly useable crystal structures for molecular simulations. The currently available two DBs, the CoRE DB (computation-ready, experimental MOF database) and CSDSS DB (Cambridge Structural Database non-disordered MOF subset) have been widely used in high-throughput molecular simulations. These DBs were constructed using different methods for collecting MOFs, removing bound and unbound solvents, treating charge balancing ions, missing hydrogens and disordered atoms of MOFs. As a result of these methodological differences, some MOFs were reported under the same name but with different structural features in the two DBs. In this work, we first identified 3490 common MOFs of CoRE and CSDSS DBs and then performed molecular simulations to compute their CH_4 and H_2 uptakes. We found that 387 MOFs result in different gas uptakes depending on from which DB their structures were taken and we identified them as 'problematic' MOFs. CH_4/H_2 mixture adsorption simulations showed that adsorbent performances of problematic MOFs, such as selectivity and regenerability, also significantly change depending on the DB used and lead to large variations in the ranking of materials and identification of the top MOFs. Possible reasons of different structure modifications made by the two DBs were investigated in detail for problematic MOFs. We described five main cases to categorize the problematic MOFs and discussed what types of different modifications were performed by the two DBs in terms of removal of unbound and bound solvents, treatment of missing hydrogen atoms, charge balancing ions *etc.* with several examples in each case. With this categorization, we aimed to direct researchers to computation-ready MOFs that are the most consistent with their experimentally reported structures. We also provided the new computation-ready structures for 54 MOFs for which the correct structures were missing in both DBs. This extensive comparative analysis of the two DBs will clearly show how and why the DBs differently modified the same MOFs and guide the users to choose either of the computation-ready MOFs from the two DBs depending on their purpose of molecular simulations.

1. Introduction

Metal–organic frameworks (MOFs) have been considered as highly promising porous materials for a large variety of applications. MOFs are crystalline materials with very large surface areas, high porosities, various pore sizes and shapes, reasonable thermal and mechanical stabilities. The advent of MOFs was triggered in 1999 when a 'building block approach'

---

[a] *Department of Chemical and Biological Engineering, Koc University, Rumelifeneri Yolu, Sariyer, Istanbul, 34450, Turkey. E-mail: skeskin@ku.edu.tr; svelioglu@ku.edu.tr; Tel: +90 212 338 1362*

[b] *Department of Natural and Mathematical Sciences, Faculty of Engineering, Ozyegin University, Cekmekoy, Istanbul, 34794, Turkey*

† Electronic supplementary information (ESI) available: Excel file listing and categorizing all problematic MOFs in five cases, the case descriptions, the refcodes of MOFs categorized in each case, common names of MOFs if available, chemical formulae of MOFs, types of the solvent molecules and CBIs present in MOFs, the links corresponding to experimental synthesis papers of MOFs, information about the stability of MOF that we derived from the literature if available and our comments/remarks about the structures. Comparison of calculated PLDs, LCDs and pore volumes of 3490 common MOFs in CoRE and CSDSS DBs; comparison of CH_4 and H_2 uptakes for commercially available MOFs; comparison of simulations with the experiments for CH_4 adsorption isotherm of a MOF. The RAR file contains the computation-ready structures for 54 MOFs. See DOI: 10.1039/c9ta01378d

‡ These authors equally contributed into this work.

was introduced to synthesize large families of isoreticular MOFs (IRMOFs) having the same topology with different pore sizes.[1,2] This approach provided the opportunity to synthesize a very large number of materials with different chemical compositions and functionalities.[3] Rational design of pore structures and controllable synthesis of MOFs led to a large diversity of materials with different geometries and chemical functionalities[4,5] and made MOFs strong alternatives to traditional porous materials such as zeolites in gas storage and separation. Several thousands of MOFs have been synthesized to date and this large materials space creates excellent opportunities to develop new gas storage and separation technologies based on MOFs. On the other hand, it is very challenging to identify the best materials for a target application using purely experimental techniques considering the continuous rapid increase in the number of synthesized MOFs. High-throughput computational screening approaches, especially molecular simulations, play a crucial role in the identification of the most promising materials among several thousands. Predicting the gas storage and gas separation potential of a MOF using molecular simulations is significantly faster than carrying out the corresponding experiments. Once the best MOFs are identified by molecular simulations, experimental efforts, time and resources can be directed to these materials for further examination. Grand Canonical Monte Carlo (GCMC) simulations have been successful in providing precise information about the gas adsorption potentials of MOFs and several excellent studies which reviewed simulations of MOFs are available in the literature.[6–10]

In order to perform molecular simulations, the crystallographic information files (CIFs) of MOFs containing the atomic positions are required. When a new MOF is synthesized, it is deposited in the Cambridge Structural Database (CSD),[3] the world's essential database of crystal structures containing over 973 630 entries, using a refcode (a six letter code to refer to MOFs). Early molecular simulation studies focused on a very limited number of MOFs, and simply searched for the CIFs of MOFs in the CSD to use them in molecular simulations. Due to the rapid increase in the number of synthesized MOFs, improved molecular simulation techniques and increased computational power, recent molecular simulations focused on screening very large numbers of MOFs to identify the top materials for a target application. Studying large numbers of materials also provides opportunities such as establishing structure–performance relations which is not straightforward when small numbers of materials are considered. The requirement of a comprehensive and accurate MOF library to be used in molecular simulations has become crucial due to the remarkable rate of MOF depositions in the CSD. However, MOFs deposited in the CSD are not specifically labeled as 'MOFs' and what is more challenging is that the database mostly contains several thousands of other types of crystal structures that are not MOFs. Another difficulty is that CIFs deposited in the CSD generally have several issues which we will discuss below to be handled before they become ready to be used in molecular simulations. Many structures contain solvent molecules that must be removed from the pores before performing molecular simulations to make the pores available for the adsorbate molecules. Other features such as the presence of disordered atoms, missing hydrogen atoms, partially occupied and/or overlapping atoms may be crystallographically meaningful, but these are problematical in simulations of structures. Removal of the solvent molecules and correction of the disorder can be done manually for a small number of materials. However, it is not straightforward to manually correct these problems one-by-one due to the large number of MOFs having multiple problems. Therefore, initial attempts on establishing a MOF database was simply focused on discarding the 'too difficult to fix' structures. For example, Goldsmith *et al.*[11] used data mining and automated structure analysis tools to identify ~40 000 MOFs from the CSD and excluded ~16 000 MOFs because of their problematic structures such as missing hydrogens and disordered atoms. Structures having interpenetrated frameworks and charge balancing ions (CBIs) were not considered in their database since these features also require special handling of the structures before molecular simulations.

Lack of publicly available, computation-ready MOF structures was a major impediment to high-throughput molecular simulations of MOFs until 2014. Snurr and co-workers[12] provided a comprehensive set of 3-D MOF structures that are derived directly from the experimental data and are immediately suitable for molecular simulations. They named this database the 'computation-ready, experimental (CoRE) MOF database' and we will refer to this database (DB) as CoRE DB throughout our manuscript. They started with >60 000 MOFs collected from the CSD, continued with >20 000 3-D MOFs and employed algorithms to retain CBIs and to remove solvent molecules bound to the unsaturated metal centers. They excluded highly disordered and difficult-to-correct materials and ended up with 5109 MOFs. Finally, they publicly provided the full atomic coordinates of 4764 MOFs that have been significantly modified with respect to the original structures obtained from the CSD to make them suitable for molecular simulations in addition to providing the refcodes of 345 unmodified MOFs. The CoRE DB has been very useful to investigate the gas storage and gas separation performances of MOFs. For example, Jiang's group[13] performed GCMC simulations to screen the CoRE DB for adsorption-based CO_2 separation from flue gas (CO_2/N_2) and natural gas (CO_2/CH_4) and identified the best 30 CoRE MOFs for these separations. The same group later used the CoRE DB to identify the best MOF membranes for separation of the $CO_2/N_2/CH_4$ mixture combining GCMC and molecular dynamics (MD) simulations.[14] Snurr's group[15] examined 2054 MOFs from the CoRE DB and identified the top 15 MOFs based on the ratio of Henry's law constants between CO_2 and H_2O. They then performed GCMC simulations of these top materials for CO_2/H_2O and $CO_2/H_2O/N_2$ separations. Not only gas separation but also gas storage performances of MOFs were examined using the CoRE DB. For example, MOFs that simultaneously achieve high gravimetric and volumetric H_2 densities were computationally identified by screening the CoRE DB.[16] Smit's group[17] used a materials genome approach to screen >650 000 nanoporous materials including CoRE DB for CH_4 storage and described structure–performance relations. The same group[18] recently reviewed the methods developed to build porous material DBs including MOFs and highlighted the insight gained from large-scale computational screening studies using these DBs in CO_2 sequestration, and CH_4 and H_2 storage.

In 2017, Jimenez's group[19] discussed that the CoRE DB only includes 3-D MOF structures and it is not integrated within the CSD to account for the additions of new MOF structures. They also stated that a small number of non-MOF structures are present in the CoRE DB whereas some MOFs are missing. Based on these arguments, they reported a CSD MOF subset which contains 69 666 synthesized MOFs including 1-D, 2-D, and 3-D MOFs and MOF-like structures with all kinds of pore sizes as well as nonporous structures. They named this database the 'CSD MOF subset' and we will refer to that as the CSDSS DB throughout our manuscript. The unbound solvents were removed from the structures using a CSD Python script which was made publicly available in their work in addition to the removal of the coordinated solvents for some specific types of MOFs. In contrast to the CoRE DB, handling of missing hydrogen atoms and retaining CBIs were not directly performed in the CSDSS DB. As a result, Jimenez's group ended up with 54 808 non-disordered, computation-ready MOFs. The CSDSS DB was reported as the most complete collection of MOFs maintained and updated by the CSD allowing direct searches and automatic updates with subsequent addition of new MOFs. This DB was also widely used in molecular simulation studies to investigate gas separation and gas storage performances of MOFs. Keskin's group screened the CSDSS DB for adsorption-based separation of CH_4/H_2, CO_2/CH_4, and CO_2/N_2 mixtures using GCMC simulations and identified the top MOF adsorbents for each separation process.[20,21] They also combined GCMC and MD simulations to screen the CSDSS DB for membrane-based separation of CH_4/H_2, CO_2/N_2, CO_2/H_2, and CO_2/CH_4 mixtures and showed that a large number of MOFs outperform traditional membranes, zeolites and polymers.[22–25] Snurr's group[26] very recently introduced a data-driven approach to accelerate materials screening and applied this approach to screen the CSDSS DB for H_2 storage in MOFs.

The currently available two DBs, CoRE and CSDSS, have tremendous value since they provide the computation-ready crystal structures of MOFs that are directly useable in high-throughput molecular simulations. Each DB was generated based on different methods to extract MOFs from the CSD and employed different algorithms to remove bound and unbound solvent molecules, to treat CBIs, missing hydrogens and disordered atoms of frameworks. As a result of these differences, the same MOF may have been reported in the two DBs with different structural features. The results of molecular simulations for gas storage and separation performances of these MOFs might also differ depending on the DB from where the MOF structure is taken. Considering the growing research on predicting gas adsorption and separation performances of MOFs using high-throughput molecular simulations, comparing the two computation-ready DBs is crucial to understand how (and why) the two DBs modified the same MOF and reported different structures. Motivated by this, we provided an extensive, comparative analysis of CoRE and CSDSS DBs in this study. We first identified the common MOFs in both DBs and then performed two separate molecular simulations for each common MOF, one using the MOF reported in the CoRE DB and the other using the CSDSS DB, to calculate single-component CH_4 and H_2 adsorption. After comparing the simulation results, we identified the 'problematic MOFs' resulting in significantly different gas uptakes depending on from which DB their structures were taken. Detailed analysis of problematic MOFs showed that these structures have different CIFs in CoRE and CSDSS DBs due to different algorithms and assumptions used in the establishment of the DBs. We compared these problematic computation-ready MOFs with their corresponding experimental structures which are used as reference, 'correct' structures to elucidate different modifications performed by each DB. We encountered five main cases which can be used to categorize the problematic MOFs and discussed what types of different modifications were performed by the two DBs in terms of removal of unbound and bound solvents, treatment of missing hydrogen atoms, charge balancing ions, *etc.* with several examples in each case. With this examination, we aimed to direct researchers to computation-ready MOFs which are the most consistent with their corresponding experimental structures. Our analysis showed that several MOFs which were modified by CoRE and CSDSS DBs to become computation-ready structures significantly deviated from their experimentally reported structures. For these MOFs, we provided new computation ready structures after manually editing them based on their experimental synthesis papers. We also compared the results of molecular simulations of problematic MOFs using the two different computation-ready DBs with the available experimental data for H_2 and CH_4 uptakes. Finally, we examined how differences in structures of MOFs reported in CoRE and CSDSS DBs would affect the predicted CH_4/H_2 separation performances of materials considering the fact that CH_4/H_2 separation is known as one of the most difficult refinery off-gas separation processes. Four adsorbent evaluation metrics, selectivity, working capacity, adsorbent performance score and regenerability, were computed for MOFs for separation of CH_4/H_2 mixtures. Our results showed that adsorbent performances of the problematic MOFs extensively differ depending on the DB used and lead to large variations in the rankings of materials and identification of the top performing MOFs. We showed some extreme examples where the top materials identified by molecular simulations using one of the DBs may not be even flagged as promising by molecular simulations using the other MOF DB. Overall, comparison of the two DBs that we provided in this work will be very useful to guide researchers to computation-ready MOFs and to reveal the sensitivity of the identification of the top MOFs to the DB.

2. Computational details

We provide a brief representation of how CoRE and CSDSS DBs were constructed in Fig. 1. More details about the construction of these DBs can be found in the literature.[12,19] There are many differences in the modification of MOFs by the two DBs as we will discuss in the next section in detail. For now, we

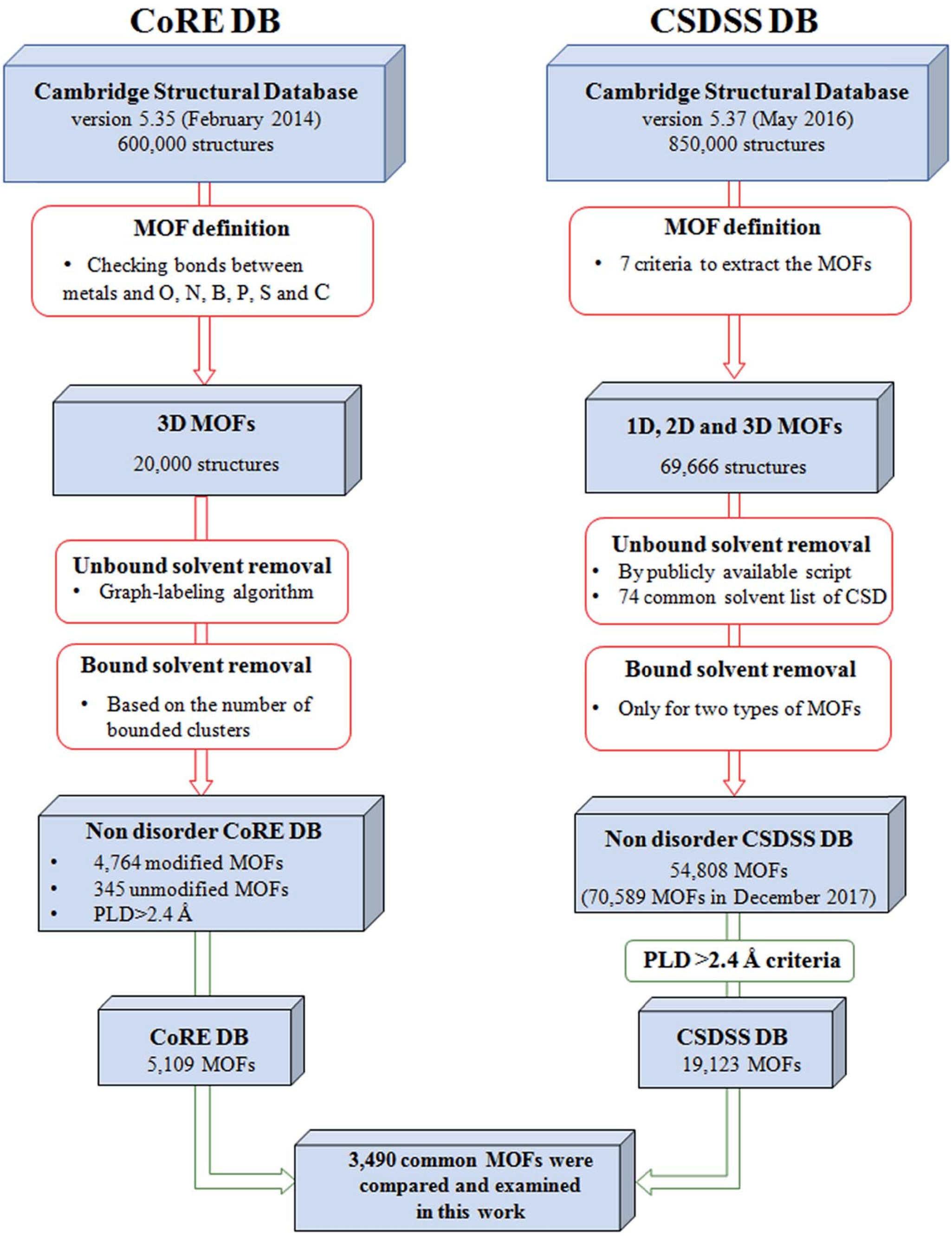

Fig. 1 Construction of CoRE and CSDSS DBs. Details of modifications shown by red boxes are available in the published papers of CoRE and CSDSS DBs. Green arrows represent how 3490 common MOFs of the two DBs were identified in this work.

summarized the procedures of extracting MOFs from the CSD: CoRE DB collected MOFs from version 5.35 of the CSD (February 2014) considering structures with more than one bond between metals and elements, O, N, B, P, S, and C, and structures forming any kinds of bonds from these six elements to N, P, S and C atoms. They only focused on 3-D MOF structures. The CSDSS DB used the version 5.37 of the CSD (May 2016) and searched for several keywords (such as -catena) to extract all 1-D, 2-D and 3-D MOFs. As a result, CoRE DB reported 5109 solvent-free MOFs with pore limiting diameters (PLDs) > 2.4 Å. CSDSS DB reported 70 589 MOFs without a pore size limitation in the updated version (December 2017) and we used their CSD Python script to obtain solvent-free structures.

We computed PLDs of 70 589 MOFs in the CSDSS DB and 5109 MOFs in the CoRE DB using Zeo++ (version 0.3)[27] software employing the covalent radii from the CSD for all framework atoms. The probe radius was used as 0 Å and the trial number was set to 50 000 for computing geometric pore volume. We found that there are 19 123 MOFs that have PLDs > 2.4 Å in the CSDSS DB. We then compared 5109 MOFs of the CoRE DB and 19 123 MOFs of the CSDSS DB and concluded that there are 3490 common MOFs having the same refcodes in both DBs.

PLDs, the largest cavity diameters (LCDs), and pore volumes of 3490 common MOFs present in both DBs were compared in Fig. S1† and the results showed that there are significant differences in the calculated pore sizes and pore volumes of many MOFs, indicating that many structures were differently modified by the two DBs.

GCMC simulations were performed for the common 3490 MOFs and their single-component CH_4 and H_2 uptakes were computed under practical conditions, 1 bar and 298 K. For each MOF, two GCMC simulations were performed, one using the CIF given in the CoRE DB, the other using the CIF given in the CSDSS DB. MOF structures in the CoRE DB were provided in P1 symmetry and we converted each MOF in the CSDSS DB to P1 symmetry using Materials Studio 2017 R2 (ref. 28) to simplify symmetry operations. GCMC simulations were performed as implemented in the RASPA simulation code.[29] Single-site spherical Lennard-Jones (LJ) 12-6 potential was used to model H_2 (ref. 30) and CH_4 (ref. 31) molecules. Potential parameters of MOF atoms were taken from the Universal Force Field (UFF).[32] These potentials and force fields were selected based on the results of our previous simulation studies in which we showed the very good agreement between our simulations using single-site LJ potentials and experiments for CH_4 and H_2 adsorption isotherms of various MOFs.[20,33,34] For example, we validated the accuracy of our CH_4 simulations by comparing with >200 experimental adsorption data points in a large number of MOFs at various pressures and temperatures,[33] and also presented the good agreement between simulated H_2 uptake and experimentally reported data of many types of MOFs.[34] We recently showed the very good agreement between experimentally measured and simulated adsorption isotherms of CH_4 and H_2 in some prototypical MOFs, IRMOF-1, CuBTC, UiO-66 and ZIF-8.[20] Furthermore, these single-site LJ potentials were shown to well reproduce the thermodynamics properties of bulk gases.[31] Previous studies also showed that charge–quadrupole interactions between MOFs and adsorbed H_2 molecules can be negligible.[35,36] The Lorentz–Berthelot mixing rules were employed and the cut-off distance for truncation of the intermolecular interactions was set to 13 Å. The simulation cell lengths were increased to at least 26 Å along each dimension and periodic boundary conditions were applied in all simulations. MOFs were assumed to be rigid in their reported crystallographic structures during simulations to save computational time. For each MOF, simulations were carried out for 15 000 cycles with the first 5000 cycles for initialization and the last 10 000 cycles for taking the ensemble averages. For MC, a cycle is max (20, *n*) move attempts with *n* being the number of adsorbed molecules. In one cycle, each molecule experienced on average one MC move attempt either accepted or rejected.[29] Finally, the Peng–Robinson equation of state was used to convert the pressure to the corresponding fugacity.

We also performed GCMC simulations to compute adsorption of equimolar CH_4/H_2 mixtures in MOFs, using the CIFs from CoRE and CSDSS DBs. Simulations were performed at two different pressures, 1 and 10 bar, at 298 K. The results of adsorption simulations of CH_4/H_2 mixtures were used to estimate four different adsorbent evaluation metrics. These are adsorption selectivity ($S_{CH_4/H_2} = N_{CH_4}/N_{H_2}$), working capacity ($\Delta N_{CH_4} = N_{ads,CH_4} - N_{des,CH_4}$), adsorbent performance score ($APS = S_{CH_4/H_2} \times \Delta N_{CH_4}$) and percent regenerability ($R\% = \Delta N_{CH_4}/N_{ads,CH_4} \times 100$) where N represents the gas uptake calculated from the GCMC simulations in the unit of mol gas per kg MOF, subscripts ads and des represent adsorption and desorption, respectively. Adsorption and desorption pressures were set as 10 and 1 bar, respectively, in computing ΔN_{CH_4}, APS and $R\%$. These performance metrics were computed for MOFs from both CoRE and CSDSS DBs, and the results were compared to evaluate the effect of the DB used on the ranking and identification of the top materials.

We compared the CIFs of computation-ready MOFs reported in the two DBs with the experimental ones which were used as reference structures. If we found that a MOF was modified both by CoRE and CSDSS DBs in a way to become very different than its experimentally reported structure, we manually corrected its structure following the experimental synthesis paper. Editing of MOF structures was carried out using Materials Studio 2017 R2.[28] If the MOF had missing charge balancing ions (CBIs), they were added into the framework according to the chemical formula of the structure by the fixed loading task using the configurational bias algorithm available in the sorption module. GCMC simulations were carried out for 2.5×10^6 cycles where the first 5×10^5 cycles were for initialization to find the probable conformation and position of ions within the framework. The structure with the lowest energy configuration was found and used in the gas adsorption simulations. For manually edited CBIs, the energy minimization with a convergence tolerance quality of fine (energy tolerance value of 10^{-4} kcal mol^{-1}) and geometry optimization using Smart algorithm with maximum iteration number of 1000 were performed using the Forcite module. The missing or disordered CBIs were modeled using the UFF and QEq method.[37] The disordered CBIs, solvents and guest molecules were handled by the Forcite module using a 500 ps MD run in the NVE ensemble to define their probable conformation and positions within the framework. The missing hydrogen atoms of MOFs were automatically added as implemented in Materials Studio and optimized using the Forcite module. New computation-ready CIFs of 54 MOFs that we prepared are provided as a separate RAR file in the ESI.†

3. Results and discussions

3.1 Identifying MOFs that have been differently reported in DBs

The main motivation of our work in comparing two DBs is to identify the MOFs that were reported in both DBs under the same refcodes but resulting in significantly different simulated gas uptakes. If these MOFs can be identified, their reported structures in CoRE and CSDSS DBs can be compared to understand how differently these structures were modified by the two DBs. We compared the results of molecular simulations for CH_4 and H_2 uptakes of 3490 common MOFs in the two DBs at 1 bar and 298 K as shown in Fig. 2(a) and (b), respectively. Each point in these figures corresponds to a single MOF reported with the same refcode in both DBs. A large number of

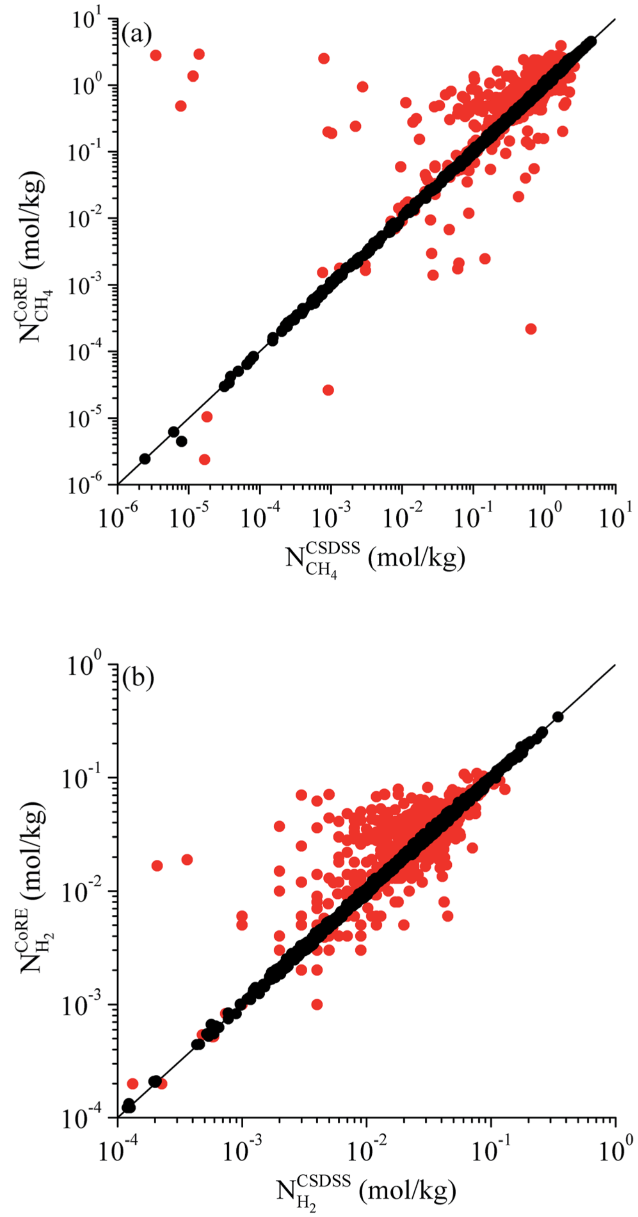

Fig. 2 Comparison of molecular simulations for (a) CH_4 and (b) H_2 uptakes of 3490 common MOFs taken from the CoRE DB and CSDSS DB at 1 bar and 298 K. Black (red) symbols represent non-problematic (problematic) MOFs for which Ratio$_{CH_4}$ and Ratio$_{H_2}$ were computed to be between 0.90 and 1.10 (<0.90 and >1.10).

MOFs are located on the diagonal line indicating that the CH_4 and H_2 uptakes of these MOFs were predicted to be exactly the same regardless of from which DB the MOF was taken. These are the computation-ready MOFs for which using either the CoRE DB or CSDSS DB will not lead to different gas uptakes under the conditions we considered. On the other hand, CH_4 and H_2 uptakes were predicted to be significantly different for many MOFs depending on from which DB the MOF was taken. For example, Fig. 2(a) shows that simulations predicted a negligible CH_4 uptake of 1.15×10^{-5} mol kg^{-1} for a MOF when its structure was taken from the CSDSS DB but the CH_4 uptake of the same MOF was computed to be moderately high, 1.36 mol kg^{-1}, when the same MOF structure was taken from the CoRE DB. A similar type of large discrepancy was observed for H_2 uptakes in Fig. 2(b) where molecular simulations using the CoRE DB generally predicted higher gas uptakes than the ones using the CSDSS DB.

In order to quantitatively identify the problematic MOFs that give different gas uptakes depending on the DB used, we defined the ratio of the simulated gas uptake of MOFs taken from the CoRE DB to the simulated gas uptake of MOFs taken from the CSDSS DB as follows: Ratio$_{CH_4}$ = $N_{CH_4}{}^{\text{CoRE DB}}/N_{CH_4}{}^{\text{CSDSS DB}}$ and Ratio$_{H_2}$ = $N_{H_2}{}^{\text{CoRE DB}}/N_{H_2}{}^{\text{CSDSS DB}}$. Ratio$_{CH_4}$ was calculated to be between 3.3×10^{-4} and 8.1×10^{5} whereas Ratio$_{H_2}$ was computed to be between 0.13 and 23.7, indicating that deviations are more pronounced for CH_4. If the simulated gas uptake of a MOF using the structure reported in the CoRE DB is the same as the one using the structure given in the CSDSS DB, then the ratios become unity. Therefore, MOFs for which ratios were computed to be between 0.90 and 1.10 were accepted as 'non-problematic', giving almost the same gas uptake regardless of the DB. However, we identified 387 MOFs among 3490 common MOFs for which both ratios were computed to be <0.90 and >1.10. These 387 MOFs were considered as 'problematic' and they are shown by red symbols in Fig. 2. We examined these problematic MOFs in detail as explained below. There is a possibility that some non-problematic MOFs for which gas uptakes were predicted to be the same, as shown by the black points in Fig. 2, may have been differently reported in the two DBs. We simply assumed that MOFs are non-problematic if their simulated gas uptakes were found to be the same regardless of the DB used throughout our manuscript.

3.2 Examining the structural differences in MOFs

Problematic MOFs were further examined to understand why and how these structures were reported differently in the two DBs. We compared CIFs of 387 MOFs reported in CoRE and CSDSS DBs and found that these MOFs have different structures in the two DBs due to different techniques used in handling of solvent molecules, CBIs, and missing atoms. In these comparisons, experimentally reported structures in the CSD and corresponding synthesis studies were used as references to compare the DBs. Structural information given in the experimental synthesis papers were specifically considered to understand what type of modification was done on MOFs by the computation-ready DBs. We described 5 main cases to explain why the structures were reported differently in the two DBs. These cases are related to missing hydrogen atoms (Case-1), removal of unbound solvents (Case-2), removal of bound solvents (Case-3), retaining CBIs (Case-4), and missing an essential part of MOFs such as linkers, atoms *etc.* (Case-5). We provided a very detailed excel file in the ESI† including the descriptions of the five cases, the refcodes of MOFs categorized in each case, common names of MOFs if available, chemical formulae of MOFs, types of the solvent molecules and CBIs present in MOFs, the links corresponding to experimental synthesis papers of MOFs, information about the stability of

Table 1 Number of MOFs categorized in each case

	Number of MOFs
Case-1	34
Case-2	29
Case-3	116
Case-4	54
Case-5	115
Case-1 + Case-3	1
Case-1 + Case-5	1
Case-2 + Case-3	2
Case-2 + Case-5	2
Case-3 + Case-4	2
Case-3 + Case-5	7

MOFs that we derived from the literature if available and our related comments about the structures. The excel file also includes some important properties that we computed in this work such as simulated CH_4 and H_2 uptakes, computed PLDs and LCDs of MOFs for the two DBs. The number of problematic MOFs categorized in each case is given in Table 1. We note that some problematic MOFs were categorized in more than one case. For example, a MOF may have missing hydrogen atoms and CBIs, so it is categorized both into Case-1 and Case-4. The information we provided in the excel file for each case clearly shows how CoRE and CSDSS DBs differently modified the same MOF. Considering these data, users can choose either of the computation-ready MOF structures depending on their purpose of molecular simulation. We discussed each case by providing interesting examples as follows:

Case-1: missing hydrogen atoms. A large number of MOF structures has been reported in the CSD with missing hydrogen atoms. During the preparation of the CoRE DB, 63 MOFs were manually edited to add their missing hydrogen atoms and geometrically optimized using Materials Studio. A special treatment for missing hydrogen atoms was not reported by the CSDSS DB. There are 34 problematic MOFs in this case which were reported in the CSDSS without required hydrogen atoms while their hydrogens were added by the CoRE DB. The Ratio_{CH_4} and Ratio_{H_2} vary from 0.02 to 8.73 and from 0.40 to 5.13, respectively, for these MOFs. We found that 2 MOFs (OCUNAC and EHALOP) have additional problems in their structures and they were also categorized in other cases (Cases-3 and 5). The remaining 32 MOFs, for which the only structural problem was missing hydrogen atoms, were compared and our analysis showed that addition of missing hydrogens either limits the gas uptake by narrowing the pore size of the MOF if hydrogen atoms are pointed towards the pore center or enhances gas uptakes since additional hydrogens act as new adsorption sites for gas molecules. NHBZZN10 is a good example showing that existence of hydrogen atoms may determine the pore diameter and accordingly the gas uptake capacity of a MOF. The lowest CH_4 and H_2 uptake ratios, 0.02 and 0.40, respectively, belong to this MOF. The PLD and LCD (2.94 and 3.41 Å) of NHBZZN10 taken from the CoRE DB, which has the required hydrogen atoms, were narrower than the pore sizes of NHBZZN10 taken from the CSDSS DB (3.48 and 3.74 Å) due to the presence of hydrogen atoms in the former. As a result, molecular simulations predicted almost no CH_4 uptake for the MOF reported in the CoRE DB and high CH_4 uptake for the MOF reported in the CSDSS DB, leading to a very small Ratio_{CH_4}. This result indicates that the absence of hydrogen atoms in MOFs reported in the CSDSS DB may cause overprediction of gas uptake capacities of some MOFs. Among the MOFs categorized in this case, there is a MOF (OFERUN02) with a common name, ZIF-8.[38] This MOF was reported with hydrogens in the CoRE DB and without hydrogens in the CSDSS DB. The LCDs of OFERUN02 structures taken from both DBs were large enough (>11 Å) to accommodate gas molecules. Therefore, the presence of hydrogen atoms in the CoRE DB structure provided additional adsorption sites and enhanced gas uptakes resulting in ratios of 3.14 and 1.24 for CH_4 and H_2, respectively. Finally, it is important to note that we identified some MOFs, for which hydrogen atoms belonging to the framework were complete but there were still missing hydrogens in the bound solvents and/or functional groups such as H_2O, OH or NH_2. Detailed comments related to MOFs having this problem were added in Case-1 of the excel file provided in the ESI.†

Case-2: removal of unbound solvents. MOFs are generally synthesized by solvothermal methods[39] and most of them are reported with uncoordinated solvent molecules in the CSD. Before gas adsorption measurements, MOFs are activated to remove the residual solvents to make pores available for gas molecules. Removal of the unbound solvents before molecular simulations is important to imitate the experimental activation process of MOFs. Both DBs aimed to remove unbound solvents from MOFs. The CoRE DB used an efficient graph-labeling algorithm to identify and remove the unbound solvents, whereas the CSDSS DB introduced a Python script which recognizes the free solvents based on a library consisting of 74 common solvents. We categorized 29 problematic MOFs in this case for which the unbound solvents were not completely removed by either of the DBs. While the CoRE DB generally removed all unbounded solvents, the CSDSS DB did not remove all of them because most of the solvents reported in the CSD have disorder or missing hydrogens and the Python script of the CSDSS DB could not recognize the solvents having these features. The Ratio_{CH_4} and Ratio_{H_2}vary from 0.11 to 1.20×10^5 and from 0.13 to 7.84, respectively, for these MOFs. The highest CH_4 ratio belongs to RAXCEA. Since the CoRE DB removed both dimethylformamide (DMF) and water, CH_4 can enter into the pores of this MOF whereas the CSDSS DB only removed water resulting in a negligible CH_4 uptake. Depending on the location of the removed solvent molecules, simulated gas uptakes of MOFs may differ. For example, QEGGUE has dichlorobenzene molecules located inside the pores which were removed by the CoRE DB leading to a PLD of 6.65 and LCD of 7.79 Å, whereas the CSDSS DB kept them leading to a PLD of 6.02 and LCD of 6.76 Å. Consequently, simulations using the CoRE DB predicted higher gas uptakes compared to those using the CSDSS DB and led to high CH_4 and H_2 ratios of 3.09 and 2.06, respectively. Another example is HOXJIO for which the CoRE DB removed the bulky nitrobenzene solvent. The PLD of this MOF did not change since the solvent was not inside the pores but located between the ligands. As a result, both ratios were computed to be only slightly higher than unity, 1.39 and 1.48. An important point is that solvent molecules may be necessary to stabilize some MOFs. The algorithms of DBs cleaned the unbound solvent molecules without considering whether the structural integrity of the framework would be retained or not. We referred to the experimental thermogravimetric analysis data (TGA) of MOFs in the synthesis studies to examine if the absence of the unbound solvent molecules causes a problem and added our related comments in Case-2 of the excel file provided in the ESI.† Examining the synthesis papers of MOFs is always suggested for detailed stability information, at least for the best materials identified from high-throughput screening studies that use computation-ready DBs.

Case-3: removal of bound solvents. The CoRE DB removed the coordinated solvent molecules such as water, DMF, methanol, and pyridine for all types of MOFs using a trial cut on all bonds between metal centers and oxygen atoms. The CSDSS DB removed the coordinated solvents for two types of MOFs, the ones containing copper–copper paddlewheels (*e.g.* CuBTC) and MOF-74-type metal clusters. 116 MOFs, for which at least one coordinated solvent molecule was improperly removed by the DBs, were categorized in this case. Similar to the previous case, we referred to the synthesis papers to understand whether the MOF retains its stability after removal of the coordinated solvent. If the information about the history of structure decomposition was not provided in the experimental studies, we noted this in Case-3 of the excel file provided in the ESI† and left it open for discussion.

Water and DMF were found as the most common coordinated solvents present in 86 and 20 of MOFs, respectively. The Ratio_{CH_4} and Ratio_{H_2} vary from 3.3×10^{-4} to 6.3×10^{4} and from 0.20 to 9.98, respectively, for the MOFs categorized in this case. While the highest ratios of this case were generally attributed to the removal of coordinated water molecules by the CoRE DB, the lowest uptake ratios can be explained by the removal of bulky solvents such as DMF, dioxane, pyridine by the CSDSS DB. For example, the highest Ratio_{CH_4} was reported for LUTDAG for which the CoRE DB removed coordinated water molecules connected to Cu and as a result, CH_4 was adsorbed into the pores. On the other hand, simulations using the CSDSS DB gave almost no CH_4 uptake due to the presence of water molecules in the pores. The highest Ratio_{H_2} was found for LENKIA for which the CoRE DB removed the bulky DMF solvent leading to a large PLD of 10.73 Å. The CSDSS DB could not recognize this coordinated solvent due to its disorder and did not remove it from the pores leading to a narrow PLD of 3.47 Å and very low H_2 uptake in the MOF. We checked the TGA data provided in the experimental work of this MOF[40] and concluded that bound DMF molecules can be removed, as the CoRE DB did, without causing structural instability. An interesting example is QOPCII for which coordinated water molecules were kept by the CoRE DB but removed by the CSDSS DB. Gas uptakes and pore sizes of this MOF were reasonably close to each other in both DBs as can be seen in Case-3 of the excel file. Experimental synthesis paper reported that elimination of coordinated water molecules might cause structural instability of this MOF.[41] This analysis indicates that removal of the coordinated solvents may not significantly affect the gas adsorption properties of some MOFs, but structural stabilities of materials should be always checked from the corresponding experimental studies prior to any application of interest.

Case-4: retaining CBIs. CBIs are required to keep the electric neutrality of some MOFs. The CoRE DB searched for '+' and '−' symbols in the chemical formula of MOFs to identify materials with CBIs and retained them to make the overall framework charge neutral when necessary. During this analysis, we visually realized that half of the MOFs have disordered CBIs that should be removed from the DB. We categorized 54 problematic MOFs in this case. We assume that the CoRE DB deleted ions of MOFs in the solvent removal process because either the chemical formulae of MOFs did not have '+' and '−' symbols or CBIs were disordered. The CSDSS DB did not report a procedure to deal with CBIs. Our analysis showed that the presence of CBIs has an important effect on the simulated gas uptakes of MOFs. The Ratio_{H_2} (Ratio_{CH_4}) was calculated as 0.39–16.75 (0.15–217.51) for the MOFs in this case. If the PLD of a MOF is very close to the kinetic diameters of gas molecules, even a small increase in PLD due to the removal of CBIs leads to an increase in gas uptakes. For example, XINWUO was found to have one of the highest Ratio_{CH_4} values. Its chemical formula, ($n(C_5H_7N_3O_2Zn)$, H_2O) does not have any + or − symbols to mark CBIs therefore, the CoRE DB could not identify acetate ions, CH_3COO^-, and removed them considering as bound solvents. We examined its experimental synthesis paper[42] and found that the bound acetate ions should be retained in the structure to neutralize the MOF. The PLD (LCD) of XINWUO in the CoRE DB was calculated as 4.29 (6.29) Å, whereas the structure taken from the CSDSS DB resulted in a narrower PLD (LCD) of 3.10 (3.58) Å with a limited CH_4 uptake compared to the structure taken from the CoRE DB. FABFOF ($n(C_{10}H_2O_8Zn_2)^-$, $2n(H_4N)^+$, and $6n(H_2O)$) is another interesting example with its very low uptake ratios. Hydrogen atoms of counter ions (NH_4^+) were missing in the CSDSS DB whereas they were complete in the CoRE DB. The PLD and LCD of this MOF were calculated as 2.89 and 4.12 Å for the CoRE DB and 4.08 and 5.65 Å for the CSDSS DB, respectively. Both gases could be adsorbed in the structure taken from the CSDSS DB whereas adsorption of CH_4 was found to be very low in the CoRE DB due to the narrow pore size of the structure. If both the PLD and LCD of the MOF are large enough to allow adsorption of gases, removal of the CBIs might decrease the number of interaction sites and decrease the gas uptakes. For example, the PLD and LCD of VEHXEN were computed to be large (>6 Å) for structures taken from both DBs. Its chemical formula ($5n(C_6H_8N)$, $10n(C_{36}H_{24}FeN_6)$, $2n(C_{18}H_6Cl_{13}O_{30}Pr_4S_3Sb_{12})$, and $80n(H_2O)$) does not include any symbol related to CBIs. The unbound cationic parts, $(\text{Fe}(1,10\text{-phen})_3)^{2+}$ and $(2\text{-MepyH})^+$, were removed by the CoRE DB whereas they were not modified by the CSDSS DB. Consequently, simulations using the CSDSS DB resulted in higher CH_4 uptake than the ones using the CoRE DB since CBIs acted as additional adsorption sites for gases.

Case-5: missing parts in MOFs (ligand, metal, guest, *etc.*). This case can be considered as problems occurred due to the misidentification and removal of an essential part of the MOF during the solvent removal process of DBs. The missing parts can be (i) a main part of the framework such as a ligand, a metal core, side groups or atoms, and (ii) guest molecules or adsorbed gases within the framework. 115 MOFs were categorized in this case and we checked the corresponding experimental studies for each of them to determine whether these parts should be removed or kept. If we were able to clearly identify the type of the missing part, we noted this as the ligand, metal *etc.* in Case-5 of the excel file. The Ratio_{H_2} (Ratio_{CH_4}) was calculated as 0.34–9.83 (0.19–339) for the MOFs in this case. The CoRE DB assumed that if the distance between two atoms is less than the sum of their covalent radii plus a skin distance (0.3 Å), a bond exists between them. This skin distance was considered slightly smaller than the recommended value by the CSD (0.4 Å) to avoid terminal atoms connected to metals making new bonds with other atoms. However, the metal connected to the solvents was accidentally removed in some MOFs during the solvent removal process. We found that the most commonly removed part of MOFs is oxygen atoms double bonded to metals, halogens and hydroxyl groups. For example, fluorine side groups and oxygen atoms of OXUPUT were removed by the CoRE DB. Simulations using the CoRE DB predicted higher gas uptakes than the CSDSS DB since the PLD (LCD) of OXUPUT was calculated as 8.40 (9.06) Å in the CoRE DB and 5.53 (6.49) Å in the CSDSS DB. Similarly, removal of hydroxyl groups of VIRPOC by the CoRE DB caused higher pore sizes and gas uptakes compared to the CSDSS DB. We also noticed that pyridine ligands of several MOFs were accidentally removed by the CSDSS DB. For example, removal of the pyridine ligand of DEPXIG by the CSDSS DB resulted in lower CH_4 uptake compared to the CoRE DB since the pyridine linker was acting as an adsorption site. In a series of MOFs, (REXCOO, REXDAB, REXDIJ, REXFAD, and REXFOR in Case-5 of the ESI†) clathrate guest molecules were deleted by the CoRE DB and these MOFs were found to have low Ratio_{CH_4} indicating that guest molecules may be favorable adsorption sites for CH_4. Adsorbed gases in MOFs already occupied the favorable adsorption sites and removal of these molecules may create free adsorption sites and dramatically increase simulated gas uptakes. For example, adsorbed D_2 molecules in CAXVOO were removed by the CoRE DB resulting in ratios of 16.60 and 4.86 for CH_4 and H_2, respectively. All these examples underline the importance of checking the structural integrity of MOFs especially after the solvent removal process.

Additional observations. The CoRE DB reported 4764 MOFs that have been modified with respect to the original structures obtained from the CSD to make them suitable for molecular simulations. The refcodes of unmodified 345 MOFs were also provided in the CoRE DB study and these MOFs were referred to as 'solvent free, ready to be used in molecular simulations'. We found that 9 of these MOFs (ARUYES, FERCAH, KIYMIQ, NEVPUA, NEVQAH, QUQGAL, SOWYIM, WAJHOG and XAGCEO01) are problematic because the CSDSS DB removed their bound solvents which led to different structures than the ones reported in the CoRE DB. These MOFs were categorized in Cases-3 and 5 of the excel file. We also note that MOFs that have been widely studied in the literature were found to be non-problematic and their computation-ready structures have been reported to be exactly the same in CoRE and CSDSS DBs. IRMOF-1, CuBTC and ZIF-8 are three commercially available MOFs and their experimental structures have been deposited in the CSD using various refcodes. We examined SAHYIK and MIBQAR for IRMOF-1, DOTSOV and FIQCEN for CuBTC, and OFERUN and TUDKEJ for ZIF-8. Molecular simulations were performed for these MOFs using CIFs both from CoRE and CSDSS DBs and the results in Fig. S2† showed that their simulated CH_4 and H_2 adsorption isotherms were the same regardless of which DB was used.

3.3 Editing problematic MOF structures

We so far considered two scenarios as shown in Fig. 3: (i) a problematic MOF is categorized in one case, mismodified by one of the DBs, and the user is directed to the other DB. (ii) A problematic MOF is categorized in two cases, mismodified by one of the DBs in both cases, and the user is directed to the other DB. We categorized 333 out of 387 problematic MOFs in 5 main cases as described above and aimed to direct the simulators to the 'correct' structure of problematic MOFs in either of the DBs. There are two more complicated scenarios for the remaining MOFs as shown in Fig. 3: (iii) a problematic MOF is categorized in one case but mismodified by both DBs. (iv) A problematic MOF is categorized in two different cases but the 'correct' structure solving the problems of both cases is missing. We found 54 problematic MOFs in (iii) and (iv) for which correct computation-ready structures are absent. These MOFs were

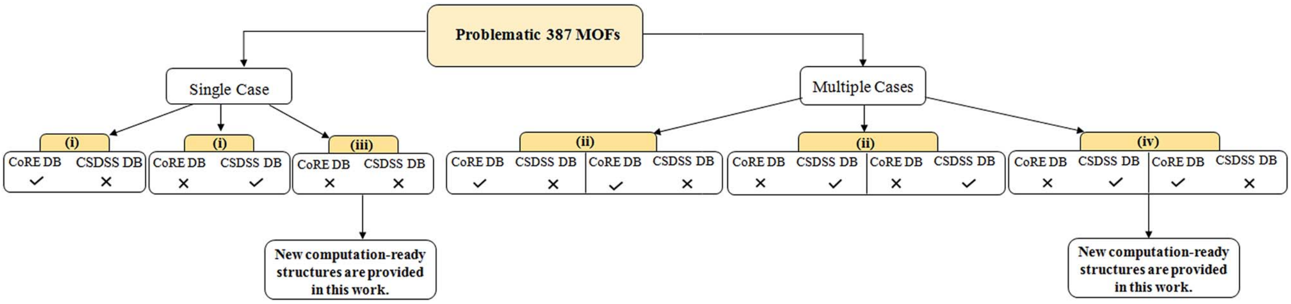

Fig. 3 Categorization of 387 problematic MOFs. ✓ (✗) symbol indicates the 'Computation-ready MOF structure is (not) consistent with the experimentally reported MOF structure.'

edited as explained in Section 2 and their new computation-ready CIFs are provided as a separate RAR file in the ESI.† Table S1† in the excel file lists the refcodes of 54 MOFs that were edited in this work based on their original structures as reported in their synthesis papers. Information about the problems of structures in computation-ready DBs, H_2 and CH_4 uptakes, PLDs and LCDs of MOFs computed using structures from CoRE and CSDSS DBs, and simulated gas uptakes and pore sizes of MOFs computed using the corrected structure we provided are all given in Table S1.† The most observable problem of these 54 MOFs was either the total or partial absence of CBIs. If the CBIs were completely removed by DBs, we used the original CIF provided in the CSD. If (i) CBIs were partially removed by DBs, (ii) CBIs were not completely reported in the CSD or (iii) CBIs were reported with disorder, then we manually edited these MOFs using Materials Studio to include the CBIs in the framework. The second most observed problem was the missing of some parts of MOFs such as ligands. For these, CIFs given in the CSD were used and the solvents, if any, were removed manually to ensure structural integrity.

We compared simulated H_2 and CH_4 uptakes of the corrected structures with the results of simulations using structures reported in the DBs in Fig. 4. For each MOF, there are two data points aligning horizontally on the *y* axis with the same color, representing the simulation results using the structures from CoRE (circle) and CSDSS DBs (squares). The large deviations between the gas uptakes of the corrected structures and MOFs in the CSDSS DB can be explained by the accidental removal of bulky side groups. For example, the CSDSS DB removed both the pyridine molecules and bound DMF in DEJCEB, whereas the CoRE DB only removed DMF. As we discussed before, pyridine molecules were acting as linkers in some MOFs and their removal by the CSDSS DB resulted in a much lower CH_4 uptake (1.2 mol kg^{-1}) than the one predicted by the CoRE DB (2 mol kg^{-1}). The experimental synthesis study of this MOF[43] indicated that both DMF and pyridine molecules should be present in the structure and once we corrected the structure keeping both in the MOF, CH_4 molecules could not adsorb into the framework as tabulated in Table S1.† Other interesting examples are SONVEX and SONVIB which have disordered, bulky CBIs. These ions were removed by the CoRE DB, whereas the CSDSS DB kept the disordered CBIs. We provided the corrected structures with non-disordered CBIs for these MOFs and obtained lower CH_4 and (similar) H_2 uptakes compared to the CoRE DB (CSDSS DB) as given in Table S1.† We note that some MOFs have been deposited in the CSD with CBIs but the ions have missing hydrogen atoms. When we corrected these MOFs by adding missing hydrogens of CBIs, we observed small deviations in the simulated gas uptakes of the corrected MOFs and the ones reported in the CSDSS DB since the latter used the CBIs as reported in the CSD. At that point, it is useful to discuss some issues related to the experimental CIFs deposited in the CSD. Some MOFs have been deposited without their solvent molecules although their presence may be important for structural integrity and some have been deposited with disordered solvent molecules. Another issue that is difficult to track is that the space groups of some MOFs reported in the CSD may be defective, as discussed in the CoRE DB study,[12] which may cause a problem during the packing of the structure using simulation tools. As we noted before, MOFs deposited in the CSD were considered as complete and accurate to compare the two computation-ready DBs.

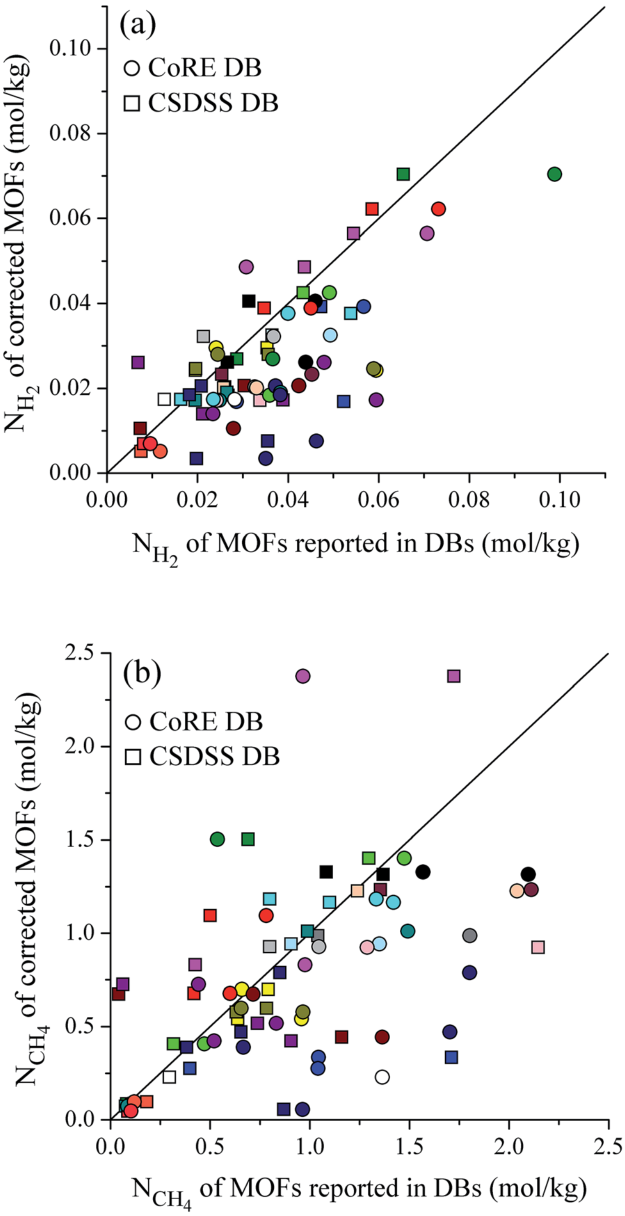

Fig. 4 Comparisons of simulated (a) H_2 and (b) CH_4 uptakes of problematic MOF structures corrected in this work and the ones taken from the DBs. For each MOF, there are two data points aligning horizontally on the *y* axis with the same color, representing the simulations using the structures from CoRE and CSDSS DBs. Detailed information about the refcodes of MOFs, simulated gas uptakes using the CoRE DB and the CSDSS DB and the ones obtained using the corrected structures are given in Table S1.†

3.4 Comparing simulations with experiments

We compared the results of molecular simulations of problematic MOFs with the available experimental data for H_2 and CH_4 uptakes in this section. Two GCMC simulations were

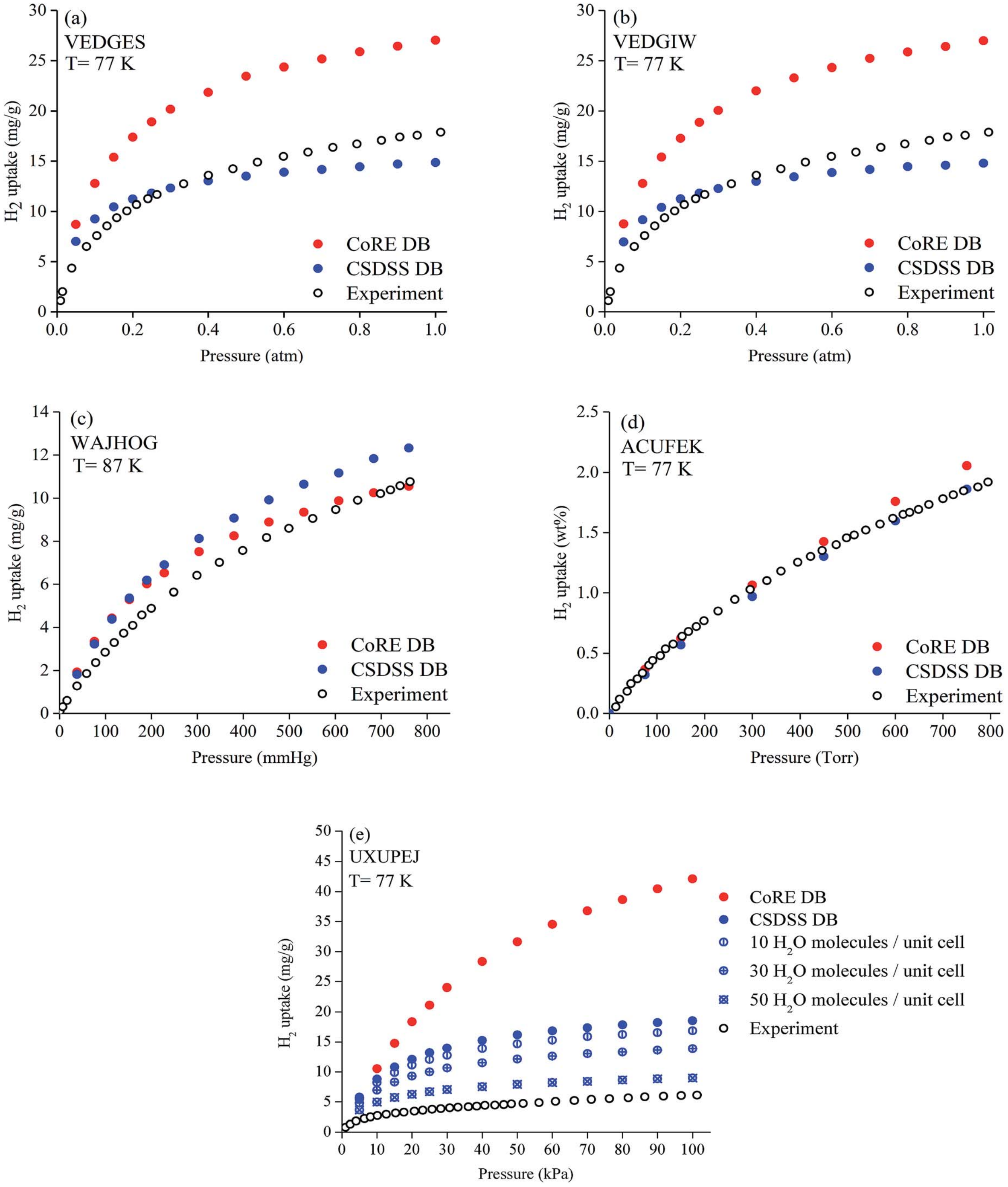

Fig. 5 Comparison of simulations with the experiments for H_2 adsorption isotherms (excess) of MOFs. Experimental data for adsorption isotherms of (a) VEDGES,[44] (b) VEDGIW,[44] (c) WAJHOG,[45] (d) ACUFEK,[46] and (e) UXUPEJ[47] (for which the isomer structure was used) were taken from the literature.

performed for each MOF, one using the structure from the CoRE DB and the other using the structure from the CSDSS DB. Fig. 5 compares H_2 adsorption isotherms in 5 different problematic MOFs, VEDGES, VEDGIW, WAJHOG, ACUFEK and UXUPEJ, for which we found experimental gas adsorption isotherms. We note that these 5 problematic MOFs were not among the 54 corrected MOFs discussed above for scenarios (iii) and (iv). VEDGES and VEDGIW were synthesized as two enantiomorphous MOFs with two chiral 1-phenylethanol guest molecules.[44] These guest molecules were removed by the CoRE DB, whereas they were kept by the CSDSS DB and MOFs were categorized in Case-5. Fig. 5(a) and (b) show that the simulated H_2 adsorption isotherms using MOFs from the CSDSS DB agreed well with the experimental data. Removal of guest molecules by the CoRE DB led to larger pore volumes and higher simulated gas uptakes. Fig. 5(c) compares experiments and simulations for WAJHOG, which was categorized in Case-3. Simulations using the CoRE DB were in a better agreement with the experiments. The overprediction of H_2 uptake by the CSDSS DB can be explained by the removal of bound water molecules. We inferred that bound water molecules can be removed only at high temperatures based on TGA analysis provided in the synthesis paper without changing the cell parameters of the MOF.[45] Fig. 5(d) compares simulations and experiments for ACUFEK (also known as PCN-6) which was categorized in Case-3. The CoRE DB removed bound water molecules whereas the CSDSS DB kept them. The synthesis study suggested that guest solvent molecules and axial aqua ligands can be removed by solvent exchange and thermal activation,[46] and simulations were found to be in slightly better agreement with the structure having the bound solvents. We finally investigated H_2 uptake in UXUPEJ and the results in Fig. 5(e) indicate that simulations using structures from CoRE and CSDSS DBs were not in good agreement with the experiments. The CoRE DB removed (CF_3SO^{3-}) anions, whereas the CSDSS DB kept them as discussed in Case-4. The experimental synthesis study of this MOF indicates that Cu sites of the framework should be surrounded by these anions to construct its octahedral coordination geometry.[47] Although the CSDSS DB kept these anions, simulations overpredicted experimental H_2 uptake. In order to understand if this overprediction may be due to the residual solvent molecules within the pores of UXUPEJ, we manually added unbound H_2O molecules into the framework and repeated simulations using this edited structure. Simulation of the H_2O-loaded MOF was found to have a better agreement with experiments indicating that experimental measurements may have been carried out using the MOF having the solvent in the framework. Finally, we showed a similar type of comparison for the CH_4 uptake of YEZKIZ. Fig. S3† indicates that simulations using the structure from the CSDSS DB were in relatively better agreement with the experiments since the CoRE DB removed the double-bonded oxygen atoms of the anionic part, whereas the CSDSS DB kept them as discussed in Case-4. Before closing this section, we would like to highlight that different research groups working on exactly the same MOF may even report different gas uptake results due to several reasons such as the purity and activation of the MOF used in measurements, and differences between gravimetric and volumetric measurement methods.

3.5 Predicting mixture adsorption in problematic MOFs

As we summarized in the literature review, most of the molecular simulation studies used computation-ready MOF DBs, either the CoRE or CSDSS DB, to study gas separation performance of MOF adsorbents and provided a list of the top MOFs for the gas separation of interest. We aimed to examine how predicted adsorbent performances of MOFs would change depending on the DB used in molecular simulations. We computed adsorption of equimolar CH_4/H_2 mixtures in 387 problematic MOFs at 1 bar, 298 K and the results are shown in Fig. 6(a). Similar to the single-component gas uptake data given in Fig. 2, simulations using MOFs from the CoRE DB generally resulted in higher CH_4 and H_2 uptakes than the ones using the CSDSS DB. One interesting feature of Fig. 6(a) is that there are several MOFs for which the CoRE DB predicts large CH_4 uptakes whereas the CSDSS DB estimated very low, even negligible CH_4 uptakes. This means the predicted CH_4/H_2 selectivities of MOFs may significantly change depending on the DB used. We compared equimolar mixture selectivities (S_{CH_4/H_2}) computed at 1 bar in Fig. 6(b) and showed that there are many problematic MOFs for which S_{CH_4/H_2} varies by several orders of magnitudes depending on the DB.

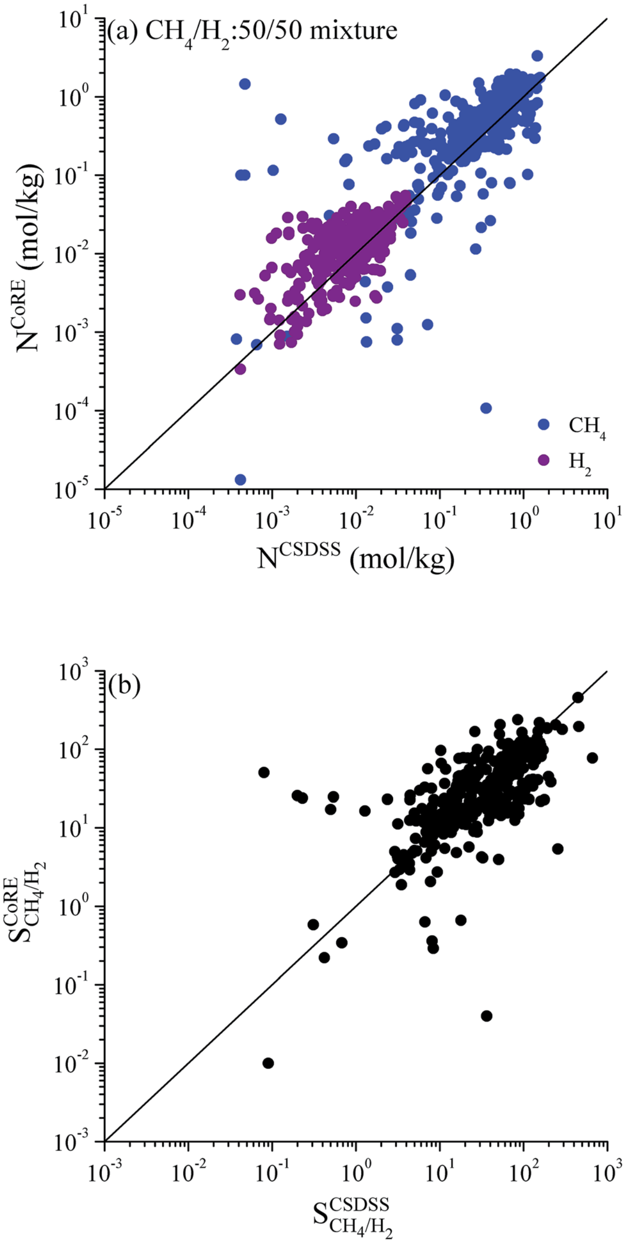

Fig. 6 Comparison of (a) simulated CH_4/H_2 mixture uptakes, (b) CH_4/H_2 selectivities of 387 problematic MOFs obtained from molecular simulations using the two different DBs at 1 bar, 298 K.

In order to evaluate the sensitivity of adsorbent performance evaluation metrics on the DB, we compared S_{CH_4/H_2}, ΔN_{CH_4}, APS and $R\%$ obtained from molecular simulations using two

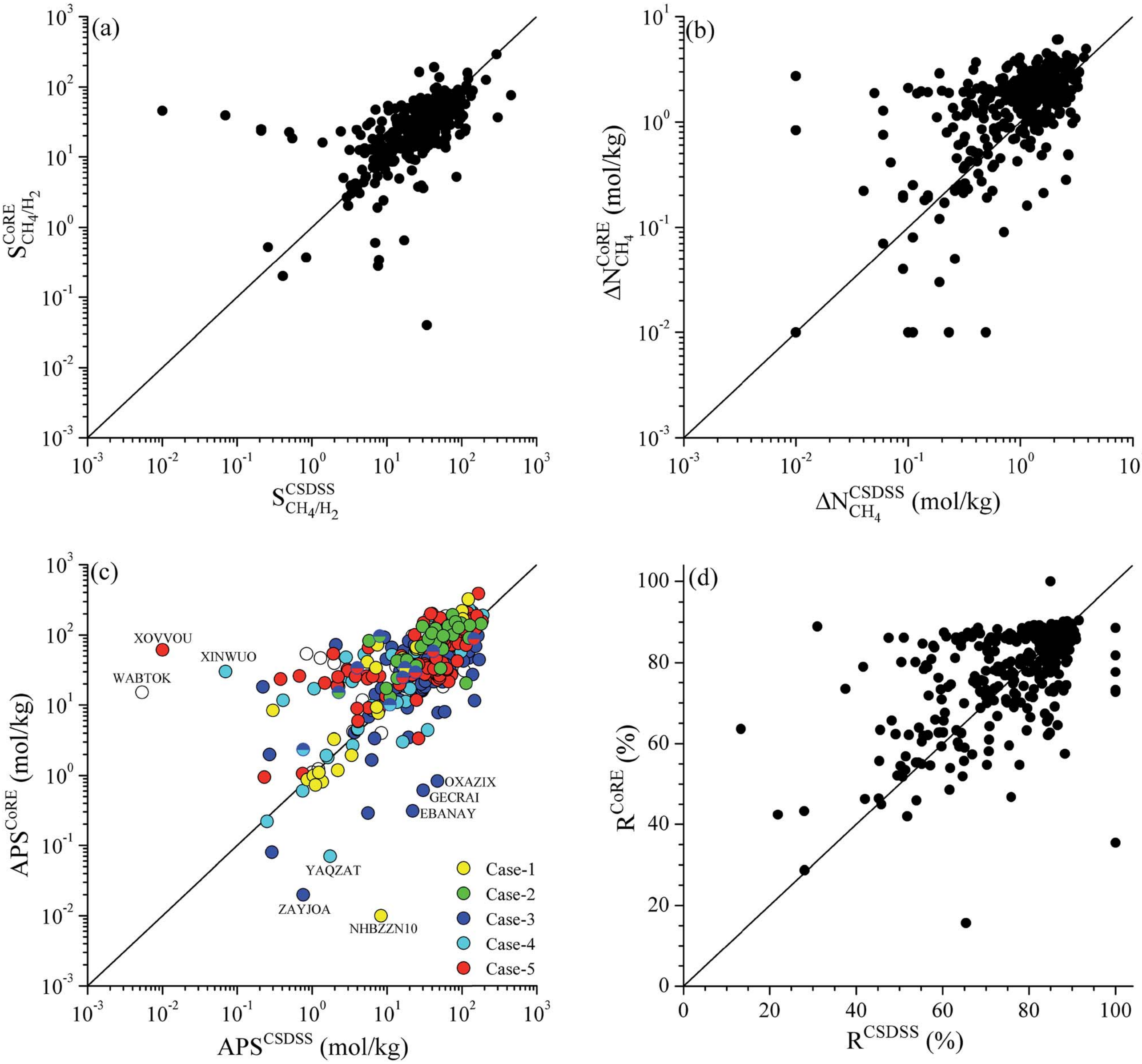

Fig. 7 Comparison of (a) CH_4/H_2 selectivities computed at 10 bar, (b) working capacities, (c) APSs, (d) *R*% of 387 problematic MOFs obtained from molecular simulations using the two different DBs. MOFs in (c) were color-coded based on the cases they were categorized in where uncolored data points represent the MOFs corrected in this study and double-colored data points represent the MOFs categorized in two cases.

different DBs in Fig. 7(a)–(d). These simulations were performed under practical operating conditions of CH_4/H_2 separation, at an adsorption (desorption) pressure of 10 (1) bar, 298 K. There are very large differences between S_{CH_4/H_2} and ΔN_{CH_4} computed using CoRE and CSDSS DBs for many MOFs as shown in Fig. 7(a) and (b). The multiplication of S_{CH_4/H_2} and ΔN_{CH_4}, APS, is a metric generally used to rank MOF adsorbents. Fig. 7(c) shows the large varieties in APSs computed using the two DBs. We also color-coded the problematic MOFs depending on the case they were categorized in and the results showed that MOFs that were treated differently by the two DBs during solvent removal (Case-3) and MOFs having missing parts after DB modification (Case-5) have the largest discrepancies in their APSs. *R*% is an important criterion to screen MOF adsorbents since high *R*% is desired for adsorption-based gas separation applications. Fig. 7(d) shows that many MOFs identified to have *R*% > 85% based on the simulations using the CoRE DB were predicted to have very low *R*%, <50% based on the simulations using the CSDSS DB and *vice versa*.

We previously suggested to identify the most promising MOF adsorbents by selecting the MOFs having *R*% > 85% and ranking them based on their APSs.[20] We followed this procedure to rank the 387 problematic MOFs and identified the top 10 materials for each DB in Table 2. The performance metrics computed for each DB are given in Table 2 to discuss the differences. For example, the top MOF from the CoRE DB was identified as UXUPEJ for separation of CH_4/H_2 mixtures with an *R*% of 86% and APS of 177 mol kg^{-1}. If the structure of this MOF

Table 2 Comparison of the ranking of the top 10 MOFs among 387 problematic MOFs

Top MOFs from CoRE DB	APS^{CoRE} (mol kg^{-1})	R^{CoRE} (%)	APS^{CSDSS} (mol kg^{-1})	R^{CSDSS} (%)	Top MOFs from CSDSS DB	APS^{CSDSS} (mol kg^{-1})	R^{CSDSS} (%)	APS^{CoRE} (mol kg^{-1})	R^{CoRE} (%)
UXUPEJ	176.95	86.23	129.26	71.15	KOCWEF	72.50	86.14	88.54	82.04
UXUPIN	172.40	86.48	124.88	72.73	DEPXIG	67.26	85.77	118.97	73.72
SAKRAZ	83.22	86.09	118.78	47.53	RIBDAJ	58.46	85.47	73.37	86.00
NURMUJ	81.40	85.29	19.45	84.50	TIRLIQ	58.37	85.98	97.78	69.10
YUCZOM	80.97	85.66	60.34	76.70	ATOWOW	55.20	86.85	62.55	88.21
FIFNUE	80.02	85.69	57.82	75.41	ATOWIQ	54.27	86.65	65.29	87.47
FIFPAM01	79.11	86.20	55.59	77.05	DUVNIS01	47.89	86.90	93.56	77.11
FIFNUE01	75.08	85.52	54.22	76.64	ESEQUO	44.02	86.00	70.15	78.79
RIBDAJ	73.37	86.00	58.46	85.47	KISPOS	42.34	86.44	22.32	84.70
SAKRED	71.06	86.16	113.51	51.00	ICAMEG	38.02	88.30	38.07	82.63

was taken from the CSDSS DB for the molecular simulations, its *R*% and APS would be calculated as 71% and 129 mol kg^{-1}, respectively. This means due to its *R*% (<85%) this MOF would not be considered as promising if the CSDSS DB was used. Similarly, the top MOF from the CSDSS DB was identified as KOCWEF (where its bound solvent was deleted) with an *R*% of 86% and APS of 72.5 mol kg^{-1}. If the structure of this MOF was taken from the CoRE DB for the molecular simulations, its *R*% would be calculated as 82% and the MOF would not be identified as promising. Table 2 shows that simulations using the CSDSS DB generally predict lower *R*% and APSs and the top materials identified by using the CoRE DB may not be even flagged as promising if the CSDSS DB is used. It is also important to note that there is no common MOF in the top 10 lists of the two DBs indicating that different structural modifications of DBs strongly affect ranking and identification of the best materials. Detailed explanation about how these top 10 MOFs were differently modified by the two DBs can be seen in the 'Overall' sheet of the excel file given in the ESI.†

3.6 DBs used for identification of the top materials in previous studies

We finally examined the effect of the DB used in previous high-throughput molecular simulations of MOFs on the identification of the best materials. In our previous work,[20] we screened the CSDSS DB and identified the top 20 MOF adsorbents based on their CH_4/H_2 selectivities. Nine of these MOFs (CAYSIE, CAYSOK, CAYYEG, DABWUA, FEHCOM, KEWZOD, OXUPUT, PACZUQ and VOCXUH) were found to be common with CoRE DBs. 7 of the 9 MOFs were non-problematic and we obtained the same CH_4/H_2 mixture selectivities independent of which DB was used in molecular simulations. We found that 2 of the top MOFs (OXUPUT and PACZUQ) were differently reported in DBs. OXUPUT was already discussed in the previous section, so we examined PACZUQ[48] here. It was reported with a coordinated water molecule which has a missing hydrogen atom in the CoRE DB whereas the CSDSS DB removed this coordinated water. Molecular simulations using the CSDSS DB resulted in a CH_4/H_2 selectivity of 28 at 10 bar for an equimolar mixture and simulations using the CoRE DB resulted in a selectivity of 14 under the same conditions. The originally reported structure of PACZUQ in the CSD has both bound and unbound water molecules, where bound water molecules have only one hydrogen atom. The coordinated water molecules can be removed at high temperatures without structural integrity problems according to the TGA data.[48] Our analysis shows that simulations using activated PACZUQ as reported in the CSDSS DB result in twice the selectivity of non-activated PACZUQ reported in the CoRE DB. In other words, the same MOF was reported as it is and as activated in the two DBs and simulations using different structures led to different selectivity predictions for the same MOF. Snurr's group[26] very recently investigated 54 776 MOFs using the CSDSS DB for H_2 storage and identified the top 25 MOFs. One of the MOFs, VAZTOG, was in our problematic MOFs in this work. It was categorized in Case-5 because guest H_2 molecules were present in the CSDSS DB, whereas they were removed by the CoRE DB. We performed molecular simulations using both structures and found similar simulated H_2 uptakes (10–12 g L^{-1}) for MOFs taken from CoRE and CSDSS DBs at 2 bar, 77 K.

Jiang's group[13] screened the CoRE DB and identified the top 30 MOF adsorbents for CO_2/CH_4 and CO_2/N_2 separations. We found that 17 of these MOFs are common with 3490 MOFs that we considered in this work and 2 of them (SEVWEV and VEHRIL) were categorized as problematic MOFs. SEVWEV was reported with missing hydrogens in the CSD and CSDSS DB whereas the CoRE DB added these missing hydrogen atoms. VEHRIL has CBIs, $[SiMo_{12}O_{40}]^{4-}$, which were removed by the CoRE DB, whereas the CSDSS DB kept them. Our molecular simulations showed that the CH_4 and H_2 uptakes of VEHRIL taken from the CoRE DB are significantly higher than the results obtained for the structure taken from the CSDSS DB. Since CBIs may have more pronounced effects for the electrostatic interactions between the MOF atoms and CO_2 molecules, molecular simulations using VEHRIL with CBIs may lead to different CO_2 selectivities. In this work, we did not study CO_2 adsorption in MOFs, we only considered the CH_4 and H_2 uptakes of MOFs and our future work will focus on examining the effects of different structure modifications made by the two DBs on the predicted CO_2 uptakes and selectivities of MOFs.

4. Conclusions

We provided an extensive, comparative analysis of the two widely used, computation-ready MOF DBs in this work. Both

DBs have tremendous value since they provide crystal structures of MOFs to be used in molecular simulations. CoRE and CSDSS DBs were generated based on different methods to extract the MOFs from the CSD and used different techniques to treat missing atoms, to remove bound/unbound solvent molecules, and to retain CBIs. As a result of these differences in methods, some MOFs were reported in the two DBs under the same refcode with different structural features which may lead to differences in simulated gas uptakes and separation performances of MOFs depending on the DB used. We identified 3490 common MOFs present in both DBs and performed two molecular simulations for each MOF, one using the structure from the CoRE DB, one using the structure from the CSDSS DB. We then compared simulated CH_4 and H_2 uptakes and identified 387 problematic MOFs giving different gas uptakes depending on the DB. These 387 MOFs were differently reported in CoRE and CSDSS DBs and we provided a detailed analysis of these MOFs by comparing computation-ready structures with the original and experimental structures in the CSD with the aim of directing the simulators to the computation-ready structure that is the most consistent with the experiments. Possible reasons of different structure modifications made by the two DBs were investigated in detail and categorized in five main cases. Here, we note that this categorization and comments provided for each problematic MOF in the excel file provided in the ESI† were limited to our own observations and knowledge. Although we attempted to identify all problematic MOFs to the best of our knowledge, it is possible that there may be additional problems in some MOFs other than the ones we specified. We also edited the MOFs which we found to be modified by both DBs in a way to become significantly different than their experimentally reported structures and provided their new computation-ready structures. One important outcome of our work is that adsorbent performances of the problematic MOFs significantly differ depending on the DB used and lead to large variations in rankings and identification of the top performing materials for CH_4/H_2 separations. Our comparison of the two computation-ready DBs was only based on CH_4 and H_2 uptakes in MOFs and it can be extended to other gas molecules in future studies. Finally, we note that it is not completely possible to fully automate establishment of a MOF DB because treatment of some MOFs requires personal, hands-on manipulation and detailed chemistry knowledge. As we discussed above, there were some complex examples where even the experimental data was not clear enough to determine whether the solvent should remain inside the MOF or not for the structural integrity. Even if a MOF was reported to be stable and identified as the top promising material for a specific gas separation application, it may decompose during practical applications such as exposure to air and/or impurities in gas mixtures. Therefore, experimental information about the stability of MOFs under practical operating conditions is very valuable to establish the most recent and accurate MOF DB. We anticipate that there will be continuous updates in computation-ready DBs and contributions from users will be extremely useful in assisting the high-throughput molecular simulations of MOFs to identify the best materials for target applications.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

S. K. acknowledges ERC-2017-Starting Grant. This research has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC-2017-Starting Grant, grant agreement no. 756489-COSMOS).

References

1 H. Li, M. Eddaoudi, M. O'Keeffe and O. M. Yaghi, *Nature*, 1999, **402**, 276–279.
2 M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O'Keeffe and O. M. Yaghi, *Science*, 2002, **295**, 469–472.
3 F. H. Allen, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 380–388.
4 O. K. Farha and J. T. Hupp, *Acc. Chem. Res.*, 2010, **43**, 1166–1175.
5 N. Stock and S. Biswas, *Chem. Rev.*, 2011, **112**, 933–969.
6 J. W. Jiang, R. Babarao and Z. Q. Hu, *Chem. Soc. Rev.*, 2011, **40**, 3599–3612.
7 S. Keskin, J. Liu, R. B. Rankin, J. K. Johnson and D. S. Sholl, *Ind. Eng. Chem. Res.*, 2009, **48**, 2355–2371.
8 J. W. Jiang, *Mol. Simul.*, 2014, **40**, 516–536.
9 J. W. Jiang, *Curr. Opin. Chem. Eng.*, 2012, **1**, 138–144.
10 J. W. Jiang, in *Coordination Polymers and Metal Organic Frameworks: Properties, Types and Applications*, ed. O. L. Ortiz and L. D. Ramirez, Nova Science Publishers 2012, ch. 7, pp. 225–247.
11 J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella and D. J. Siegel, *Chem. Mater.*, 2013, **25**, 3373–3382.
12 Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl and R. Q. Snurr, *Chem. Mater.*, 2014, **26**, 6185–6192.
13 Z. W. Qiao, K. Zhang and J. W. Jiang, *J. Mater. Chem. A*, 2016, **4**, 2105–2114.
14 Z. W. Qiao, Q. S. Xu and J. W. Jiang, *J. Membr. Sci.*, 2018, **551**, 47–54.
15 S. Li, Y. G. Chung and R. Q. Snurr, *Langmuir*, 2016, **32**, 10368–10376.
16 A. Ahmed, Y. Y. Liu, J. Purewal, L. D. Tran, A. G. Wong-Foy, M. Veenstra, A. J. Matzger and D. J. Siegel, *Energy Environ. Sci.*, 2017, **10**, 2459–2471.
17 C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr and B. Smit, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.
18 P. G. Boyd, Y. J. Lee and B. Smit, *Nat. Rev. Mater.*, 2017, **2**, 1–15.

19 P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chem. Mater.*, 2017, **29**, 2618–2625.

20 C. Altintas, I. Erucar and S. Keskin, *ACS Appl. Mater. Interfaces*, 2018, **10**, 3668–3679.

21 C. Altintas, G. Avci, H. Daglar, A. N. V. Azar, S. Velioglu, I. Erucar and S. Keskin, *ACS Appl. Mater. Interfaces*, 2018, **10**, 17257–17268.

22 C. Altintas, G. Avci, H. Daglar, E. Gulcay, I. Erucar and S. Keskin, *J. Mater. Chem. A*, 2018, **6**, 5836–5847.

23 H. Daglar and S. Keskin, *J. Phys. Chem. C*, 2018, **122**, 17347–17357.

24 G. Avci, S. Velioglu and S. Keskin, *ACS Appl. Mater. Interfaces*, 2018, **10**, 33693–33706.

25 C. Altintas and S. Keskin, *ACS Sustainable Chem. Eng.*, 2018, **7**, 2739–2750.

26 B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri and R. Q. Snurr, *Mol. Syst. Des. Eng.*, 2019, **4**, 162–174.

27 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.

28 Dassault Systèmes BIOVIA, *Materials Studio 2017 R2*, Dassault Systèmes, San Diego, 2017.

29 D. Dubbeldam, S. Calero, D. E. Ellis and R. Q. Snurr, *Mol. Simul.*, 2016, **42**, 81–101.

30 V. Buch, *J. Chem. Phys.*, 1994, **100**, 7610–7629.

31 M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.

32 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.

33 K. B. Sezginel, A. Uzun and S. Keskin, *Chem. Eng. Sci.*, 2015, **124**, 125–134.

34 Y. Basdogan and S. Keskin, *CrystEngComm*, 2015, **17**, 261–275.

35 M. Fischer, F. Hoffmann and M. Fröba, *ChemPhysChem*, 2009, **10**, 2647–2657.

36 J. Liu, R. B. Rankin and J. Karl Johnson, *Mol. Simul.*, 2009, **35**, 60–69.

37 A. K. Rappe and W. A. Goddard, *J. Phys. Chem.*, 1991, **95**, 3358–3363.

38 J. C. Saint Remi, T. Remy, V. Van Hunskerken, S. van de Perre, T. Duerinck, M. Maes, D. De Vos, E. Gobechiya, C. E. A. Kirschhock, G. V. Baron and J. F. M. Denayer, *ChemSusChem*, 2011, **4**, 1074–1077.

39 S. T. Meek, J. A. Greathouse and M. D. Allendorf, *Adv. Mater.*, 2011, **23**, 249–267.

40 R. Sanz, F. Martinez, G. Orcajo, L. Wojtas and D. Briones, *Dalton Trans.*, 2013, **42**, 2392–2398.

41 Y. Y. Zhu, J. Li, Z. G. Sun, J. Zhang, Y. Zhao, X. Lu, L. Liu and N. Zhang, *Z. Anorg. Allg. Chem.*, 2009, **635**, 171–174.

42 A. X. Zhu, Y. Liu and W. X. Zhang, *Inorg. Chem. Commun.*, 2013, **30**, 88–91.

43 J. F. Eubank, R. D. Walsh, P. Poddar, H. Srikanth, R. W. Larsen and M. Eddaoudi, *Cryst. Growth Des.*, 2006, **6**, 1453–1457.

44 H.-X. Zhang, F. Wang, Y.-X. Tan, Y. Kang and J. Zhang, *J. Mater. Chem.*, 2012, **22**, 16288–16292.

45 L.-H. Xie, J.-B. Lin, X.-M. Liu, Y. Wang, W.-X. Zhang, J.-P. Zhang and X.-M. Chen, *Inorg. Chem.*, 2010, **49**, 1158–1165.

46 D. Sun, S. Ma, Y. Ke, D. J. Collins and H.-C. Zhou, *J. Am. Chem. Soc.*, 2006, **128**, 3896–3897.

47 A. Kondo, H. Kajiro, H. Noguchi, L. Carlucci, D. M. Proserpio, G. Ciani, K. Kato, M. Takata, H. Seki, M. Sakamoto, Y. Hattori, F. Okino, K. Maeda, T. Ohba, K. Kaneko and H. Kanoh, *J. Am. Chem. Soc.*, 2011, **133**, 10512–10522.

48 Y. Xie, F. Y. Bai, Y. H. Xing, Z. Wang, H. Y. Zhao and Z. Shi, *Z. Anorg. Allg. Chem.*, 2010, **636**, 1585–1590.