# Supplementary information

## SmartCIF: A Context-aware Multi-agent System for Refining MOF Topologies in CIFs

Qixiang Zhang[a,b,†], Chen Zhang[a,b,†], and Liwei Wang[a,b,*]

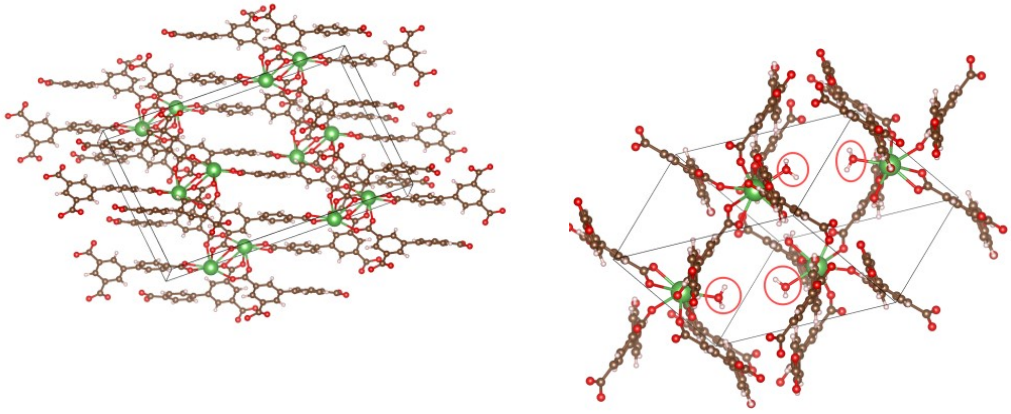
a. Institute of Refrigeration and Cryogenics, Key Laboratory of Power Machinery and Engineering of MOE, Shanghai Jiao Tong University, Shanghai, 200240, China.

b. Shanghai Non-Carbon Energy Conversion and Utilization Institute, Shanghai, China.

† Q Zhang and C Zhang contributed equally to this work.

* Corresponding authors: lwwang@sjtu.edu.cn (Liwei Wang).

The chemical full name of APAZOI is catena-[(μ-biphenyl-3,4',5-tricarboxylato)-diaqua-lanthanum unknown solvate].

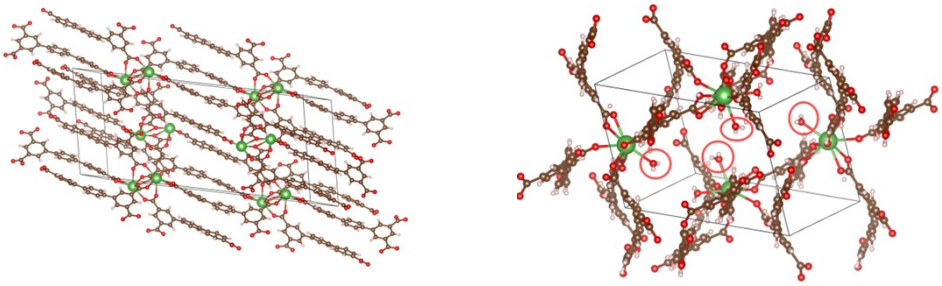


(a) (b)

Difference: The two water molecules coordinated to the La atom1

Figure S1 The structure difference of APAZOI between（a）Context-free and (b) SmartCIF

The chemical full name of APAYUN is catena-[(μ-1,1':4',1"-terphenyl-3,4",5-tricarboxylato)-diaqua-lanthanum unknown solvate]

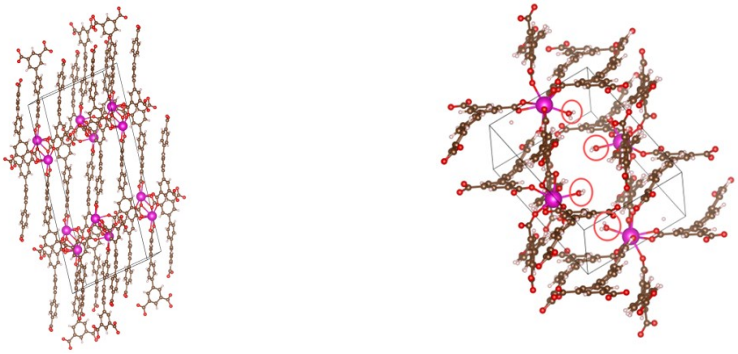


(a) (b)

Difference: The two water molecules coordinated to the La atom[1]

Figure S2 The structure difference of APAYUN between（a）Context-free and (b) SmartCIF

The chemical full name of APAYIB is catena-[(μ-1,1':4',1"-terphenyl-3,4",5-tricarboxylato)-diaqua-europium unknown solvate]

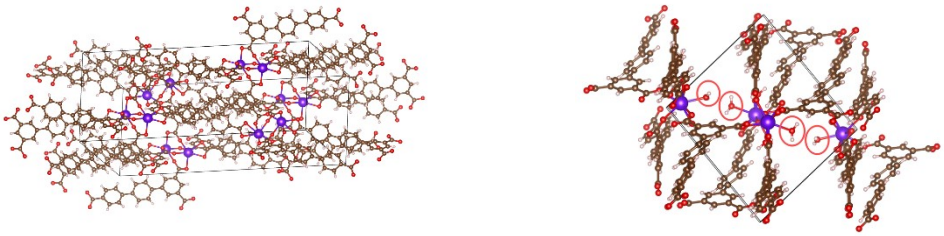


(a) (b)

Difference: The two water molecules coordinated to the Eu atom[1]

Figure S3 The structure difference of APAYIB between（a）Context-free and (b) SmartCIF

The chemical full name of APAYOH is catena-[(μ-1,1':4',1"-terphenyl-3,4",5-tricarboxylato)-diaqua-terbium unknown solvate]

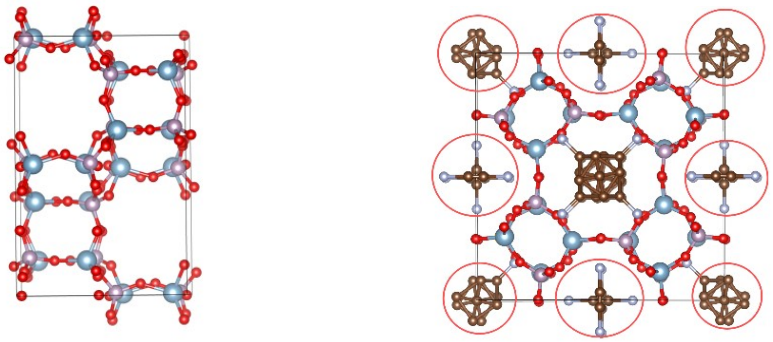


(a) (b)

Difference: The water molecules coordinated to the Tb atom[1]

Figure S4 The structure difference of APAYOH between（a）Context-free and (b) SmartCIF

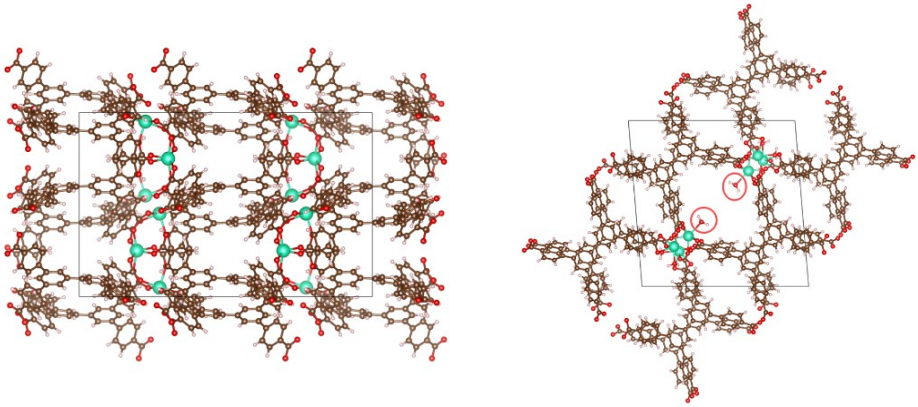
The chemical full name of BEFLUV is catena-(methylammnoim ethane-1,2-diaminium (μ4-phosphato)-aluminium). The organic cations are essential for charge balance, neutralizing the anionic inorganic phosphate-aluminate framework. They act as structural templates, directing the assembly of the specific one-dimensional polymeric chain topology. Their hydrogen-bonding interactions with the inorganic chains provide critical inter-chain stability for the overall crystal structure.[2]



(a) (b)

Figure S5 The structure difference of BEFLUV between（a）Context-free and (b) SmartCIF

The chemical full name of ATEZIK is catena-[bis(μ-biphenyl-3,3',5,5'-tetra(phenyl-4-carboxylato))-triaqua-tri-lutetium(iii) hydroxide N,N-dimethylacetamide solvate]

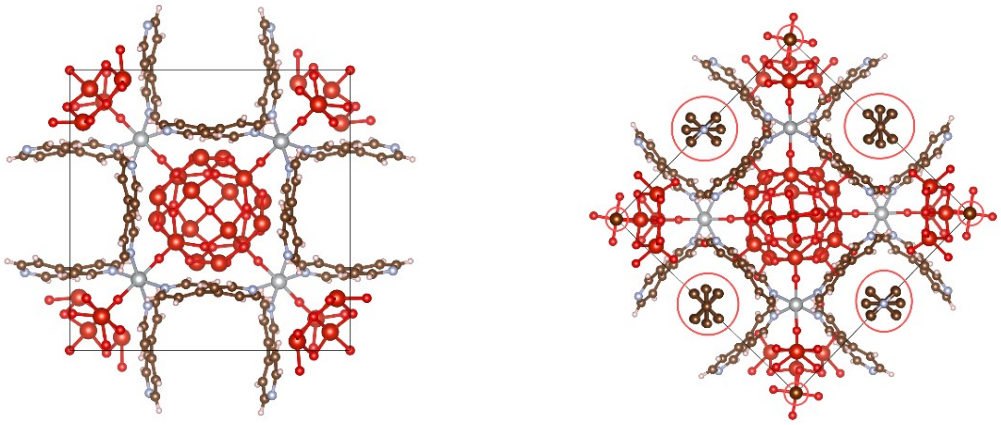


(a) (b)

Difference: The three water molecules coordinated to the Lu atom

Figure S6 The structure difference of ATEZIK between（a）Context-free and (b) SmartCIF

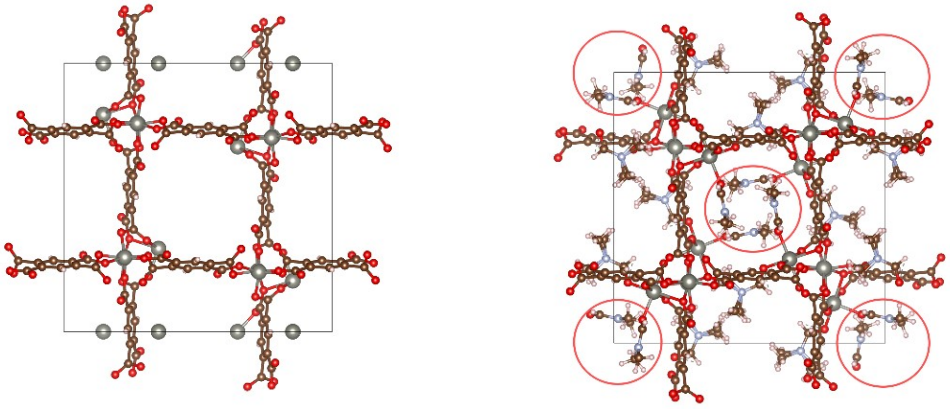
The chemical full name of ADOCEC is catena-(icosakis(μ3-oxo)-tetrakis(μ2-4,4'-bipyridine)-hexakis(μ2-oxo)-dodecaoxo-di-nickel-hepta-vanadium(iv)-nona-vanadium(v) bromide 4,4'-bipyridine hydrate) The bromide anions and bridging 4,4'-bipyridine ligands are integral structural components and are not removable. The bromides are essential for maintaining charge balance, while the bipyridine linkers are critical for sustaining the one-dimensional polymeric chain. Removal of either component would fundamentally compromise the material's structural integrity, likely causing it to collapse[3].



(a) (b)

Figure S7 The structure difference of ADOCEC between（a）Context-free and (b) SmartCIF

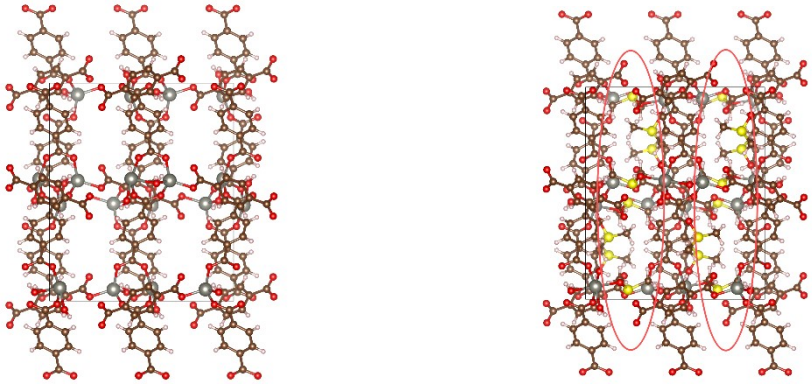
The chemical full name of ATAFIK is catena-((μ7-1,2,4,5-Benzenetetracarboxylato)-bis(N,N-dimethylformamide)-di-zinc(ii)). According to the original paper, the coordinated DMF molecules are integral components of the zinc coordination sphere, directly bonded to the metal center as essential ligands. Their removal would create unsaturation and geometric distortion at the metal site, destabilizing the local structure. Since these DMF molecules also occupy and support the channels within the 3D network, deleting them would risk the collapse of the entire framework[4].



(a) (b)

Figure S8 The structure difference of ATAFIK between（a）Context-free and (b) SmartCIF

The chemical full name of ABUWOJ is catena-(bis(μ4-1,4-Benzenedicarboxylate)-(μ2-1,4-benzenedicarboxylate)-bis(μ3-hydroxo)-tetrakis(dimethyl sulfoxide)-tetrazinc dihydrate)

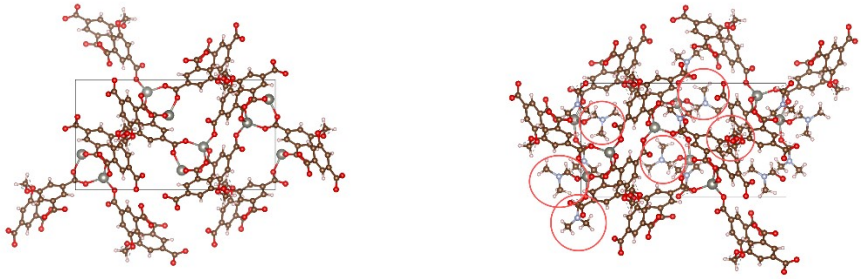


(a) (b)

Difference: The two DMSO coordinated to the Lu atom

Figure S9 The structure difference of ABUWOJ between（a）Context-free and (b) SmartCIF

The chemical full name of AZIXUD is catena-[[bis(dimethylformamide)[μ7-5,5'-(methylenedioxy)diisophthalato]dizinc] dimethylformamide solvate]. Although there are some free DMFs in the original files, the coordinated DMF ligands are integral components of the zinc coordination sphere, and their direct removal would create unsaturation and geometric distortion at the metal centers, destabilizing the local structure. Furthermore, these DMF molecules occupy and support the channels within the three-dimensional network, meaning their deletion would risk the collapse of the entire porous framework and irreversibly alter the material's properties[5].



(a) (b)

Figure S10 The structure difference of AZIXUD between（a）Context-free and (b) SmartCIF

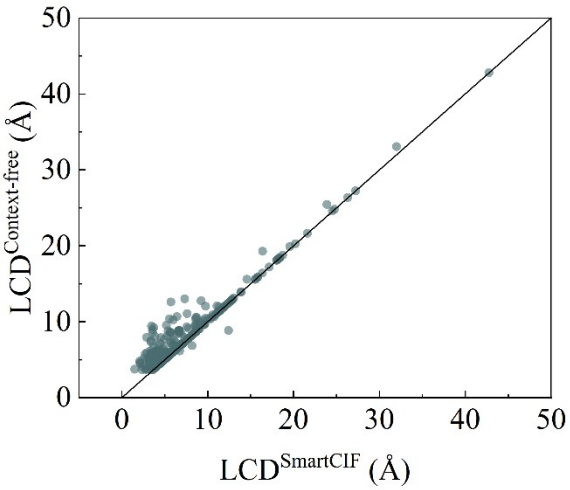



Figure S11 LCD of the cifs processed by SmartCIF and context-free method.

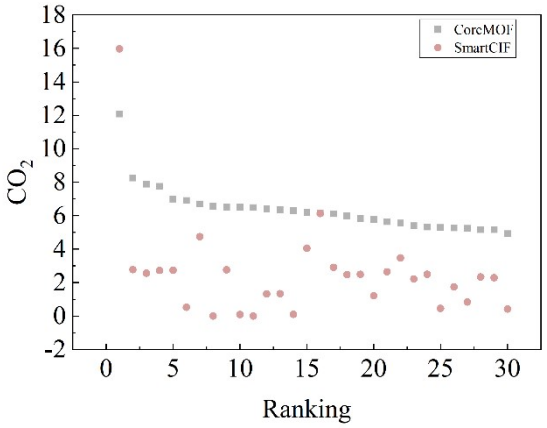



Figure S12 The CO2 adsorption of the top 10% of materials with the greatest difference in CO2 adsorption capacity

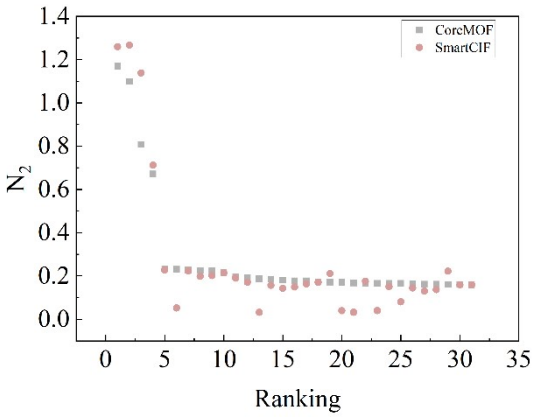



Figure S13 The $N_2$ adsorption of the top 10% of materials with the greatest difference in $CO_2$ adsorption capacity

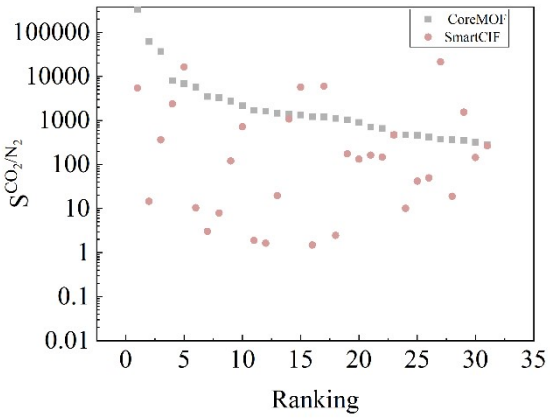



Figure S14 The selectivity of the top 10% of materials with the greatest difference in $CO_2$ adsorption capacity

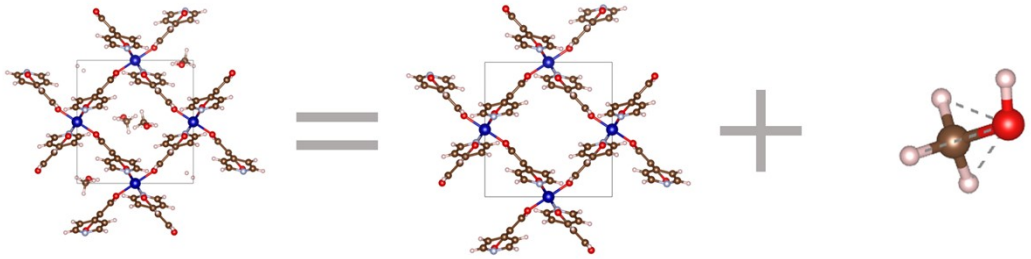


Figure S15 Connectivity-based clustering result for ADIQEL at skin = 0.15. Under this skin parameter, the structure is partitioned into one framework-connected cluster with composition C32Co2H24N4O8 and four coordinated methanol clusters.

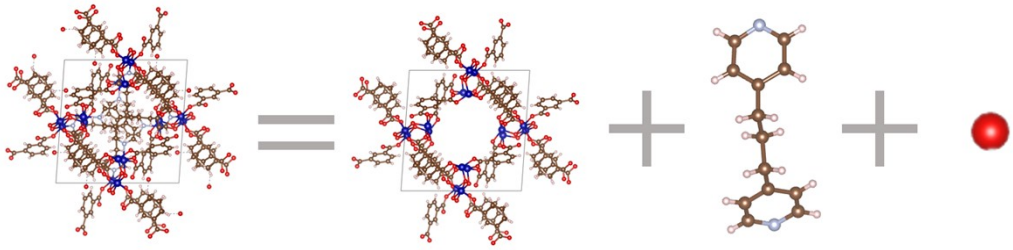


Figure S16 Connectivity-based clustering result for ANIDIL at skin = 0.08. The structure is partitioned into one main framework cluster (C86Co14H46O48), four organic linker clusters (C13H14N2), and four water-molecule clusters.

Table S1 Detailed results of the BET calculation

| MOF | BET_Paper | CSD_BET | Topology_only | SmartCIF_BET |
|---|---|---|---|---|
| ADIQEL | 568 | 0.00 | 640.06 | 871.90 |
| AHEBAS | 200.56 | 368.59 | 696.23 | 363.42 |
| AJOVEB | 170.4 | 136.09 | 296.63 | 165.26 |
| AKOXIJ | 945 | 925.25 | 818.25 | 980.25 |
| AMILUE | 820 | 1144.75 | 1353.81 | 1308.61 |
| ANAWIX | 125 | 0.00 | 714.23 | 577.66 |
| ANIDIL | 47.2 | 268.18 | 304.92 | 297.72 |
| AROFAP | 231 | 908.60 | 1305.70 | 1020.16 |
| ASAJOU | 382 | 0.00 | 671.97 | 549.55 |
| ASOSUY | 1145.9 | 1323.08 | 1745.05 | 1411.70 |
| ASUJIJ | 493 | 107.87 | 550.96 | 229.93 |
| ATAYEB | 1580 | 0.00 | 2147.10 | 2170.60 |
| ATIXUX | 285.4 | 146.35 | 590.11 | 146.35 |
| ATUZEW | 1919.2 | 1285.87 | 908.61 | 1572.82 |
| AVIMIC | 216 | 0.00 | 266.33 | 266.33 |
| BAKYIY | 207 | 268.44 | 1253.91 | 262.93 |
| BAXSIE | 450 | 0.00 | 585.60 | 1952.05 |
| BEXPAX | 733 | 983.72 | 2456.24 | 854.54 |
| BUSQIQ | 308 | 0.00 | 541.13 | 541.13 |
| CAHQOS | 271 | 0.00 | 467.74 | 868.39 |
| CAJQEL | 2896 | 2495.67 | 2707.10 | 2405.84 |
| CANYIB | 1386 | 1170.34 | 1303.66 | 1361.40 |
| CAWVIH | 729.1 | 0.00 | 786.45 | 971.24 |
| CAXVII | 1360 | 1442.04 | 1364.20 | 1330.49 |
| CAXVUU | 1360 | 805.00 | 1360.29 | 941.82 |
| CAXWAB | 1360 | 856.84 | 1364.95 | 995.84 |
| CAXWEF | 1360 | 572.35 | 1351.11 | 1033.56 |
| CAXWIJ | 1360 | 465.07 | 1364.64 | 1007.19 |
| CAZFOA | 405.7 | 0.00 | 502.03 | 399.50 |
| CENPUI | 950.8 | 623.75 | 1098.48 | 2043.65 |
| CIFMEL | 17.4 | 113.11 | 2772.06 | 2625.34 |
| CIFMIP | 340.6 | 925.78 | 1260.02 | 953.62 |
| CIVTEH | 268 | 0.00 | 900.07 | 1514.19 |
| COJHIT | 2670 | 98.22 | 2344.59 | 2323.29 |
| COQHUN | 711 | 776.73 | 1141.73 | 911.45 |
| COVJAA | 1 | 141.59 | 420.60 | 90.49 |
| CUNWEO | 1005 | 0.00 | 913.10 | 1692.94 |
| CUQRIR | 722 | 0.00 | 1151.40 | 1259.68 |
| CUXFIL | 350 | 84.43 | 784.93 | 382.56 |
| CUXFUX | 306.3 | 0.00 | 898.01 | 892.43 |
| DAJHUS | 484 | 0.00 | 524.81 | 543.65 |

| | | | | |
|---|---|---|---|---|
| DARSEV | 601 | 503.69 | 1614.54 | 2347.16 |
| DEJJIN | 1200 | 878.11 | 1310.04 | 850.91 |
| DIRPIF | 1563 | 1169.26 | 1606.11 | 1278.70 |
| DIWNAA | 390 | 987.22 | 2786.71 | 980.91 |
| DOKDEN | 1715 | 2081.34 | 2486.81 | 2070.43 |
| DUTKAG | 83 | 1764.23 | 2178.05 | 1695.13 |
| ECASOS | 69 | 0.00 | 233.52 | 224.11 |
| ECOLEP | 202 | 3782.21 | 5482.40 | 3832.45 |
| EHODOV | 354.1 | 359.50 | 1155.51 | 256.03 |
| EKEKIQ | 839.04 | 825.42 | 811.36 | 777.82 |
| EKIGIQ | 25.8 | 0.00 | 414.84 | 326.69 |
| EKIGOW | 86.2 | 0.00 | 260.63 | 295.28 |
| EPEWAY | 146.7 | 0.00 | 629.80 | 623.71 |
| EQERAU | 810 | 1147.37 | 1716.98 | 987.06 |
| EQOXIT | 140 | 918.73 | 930.46 | 869.86 |
| ERIWAF | 2134 | 0.00 | 2169.94 | 2146.92 |
| ESIFET | 2.7 | 738.14 | 872.57 | 1111.10 |
| EWUCOP | 810 | 34.37 | 1567.46 | 857.80 |
| EZUCIM | 870 | 20.60 | 1236.59 | 628.20 |
| FAPXIG | 1892 | 0.00 | 1776.10 | 1150.24 |
| FAPXOM | 1892 | 0.00 | 1633.44 | 1337.27 |
| FAQLUH | 1892 | 0.00 | 1222.84 | 1331.49 |
| FATLUJ | 808.5 | 2.82 | 856.92 | 757.74 |
| FECKAA | 396 | 577.28 | 1163.14 | 648.70 |

The linear range for BET analysis was determined by manual selection based on the BET-transformed isotherm. We first defined the relative pressure as $x = P/P_0$using a unified saturation pressure $P_0 = 97150$ Pa, and converted the adsorption data to the linear BET form $y = P/[n(P_0 - P)]$, where $n$is the $N_2$ uptake at 77 K. To ensure consistency across samples, we focused on the low-pressure regime $0.01P_0$_$0.30P_0$ and identified, by inspection, a visually linear segment comprising at least five consecutive points within this interval. The selected fitting window was chosen to maximize linearity while avoiding obvious deviations associated with incipient plateauing or low-pressure fluctuations, and the validity of the chosen interval was reported a posteriori through the corresponding linear regression $R^2$(from fitting $y = ax + b$over the selected points). This manual procedure is illustrated with three representative cases, including a smoothly increasing isotherm, an early plateau requiring exclusion of post-plateau points, and a fluctuating low-pressure region where a later consecutive segment yields a more reliable linear fit.

Example A: smooth monotonic uptake (ASAJOU)

The low-pressure isotherm increases smoothly with minimal fluctuations.The optimal BET window spans $x = [0.04P_0\text{-}0.08P_0]$ using 5 consecutive points, yielding the highest $R^2$among all candidate windows.

| Pressure_Pa | 3886.0 | 4857.5 | 5829.0 | 6800.5 | 7772.0 |
|---|---|---|---|---|---|
| Adsorption_n | 110.9528921016 | 119.5034938620 | 121.3959485249 | 124.3251967046 | 126.1121210281 |
| x=P/P0 | 0.040000000 | 0.050000000 | 0.060000000 | 0.070000000 | 0.080000000 |
| y=x/(n*(1-x)) | 0.0003755348 | 0.0004404187 | 0.0005257983 | 0.0006054188 | 0.0006895176 |
| Selected (TRUE/FALSE) | false | true | true | true | true |
| | 8743.5 | 9715.0 | 14572.5 | 19430.0 | 29145.0 |
| | 127.4056489469 | 128.2710458077 | 128.6075890313 | 129.4348843914 | 131.6173671967 |
| | 0.090000000 | 0.100000000 | 0.150000000 | 0.200000000 | 0.300000000 |
| | 0.0007762693 | 0.0008662213 | 0.0013721631 | 0.0019314731 | 0.0032561921 |
| | true | false | false | false | false |

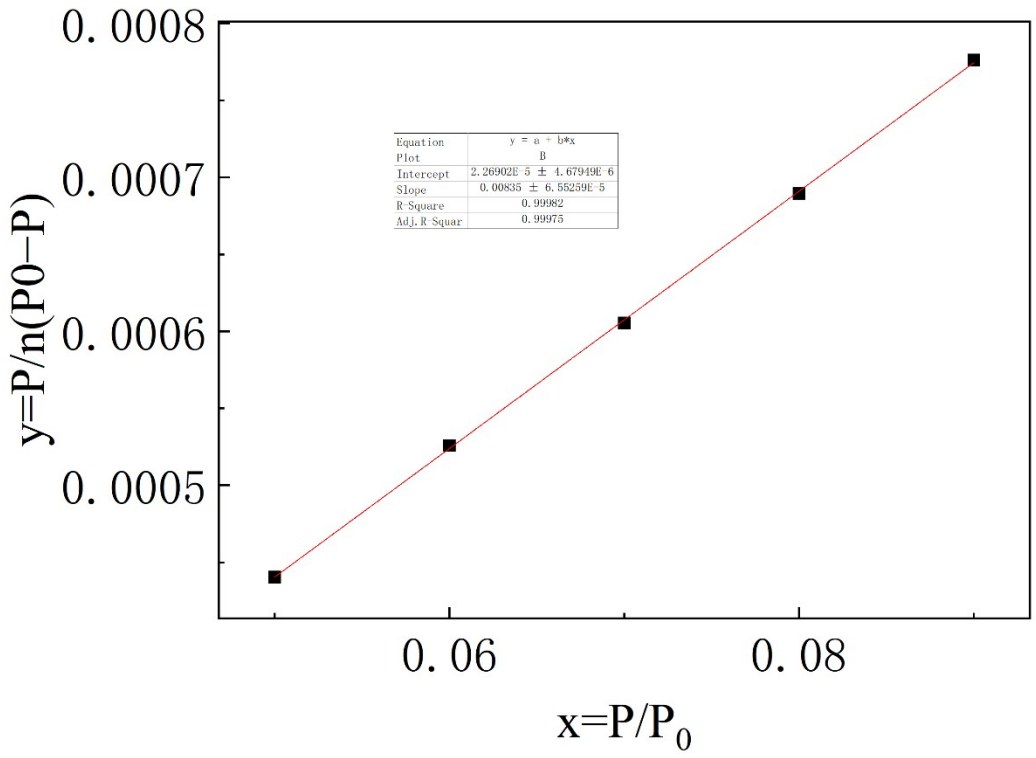



Figure S17 Representative determination of the BET linear range(ASAJOU).

Example B: early plateau (AROFAP)

The uptake increases rapidly at low pressure and reaches an early plateau. Including plateau points causes deviations from linearity in the BET transform; therefore, the selected BET window is restricted to the pre-plateau region $x = [0.04P_0–0.08P_0]$, which maximizes $R^2$.

| Pressure_Pa | 3886.0 | 4857.5 | 5829.0 | 6800.5 | 7772.0 |
|---|---|---|---|---|---|
| Adsorption_n | 236.3511607365 | 251.0050877300 | 252.0610618281 | 253.9126970179 | 255.4056489906 |
| x=P/P0 | 0.040000000 | 0.050000000 | 0.060000000 | 0.070000000 | 0.080000000 |
| y=x/(n*(1-x)) | 0.0001762914 | 0.0002096833 | 0.0002532314 | 0.0002964358 | 0.0003404644 |
| Selected (TRUE/FALSE) | false | true | true | true | true |
| | 8743.5 | 9715.0 | 14572.5 | 19430.0 | 29145.0 |
| | 255.4056489906 | 255.4056489906 | 255.4056489906 | 255.4064297065 | 255.4056489906 |
| | 0.090000000 | 0.100000000 | 0.150000000 | 0.200000000 | 0.300000000 |
| | 0.0003872314 | 0.0004350378 | 0.0006909424 | 0.0009788321 | 0.0016780029 |
| | true | false | false | false | false |

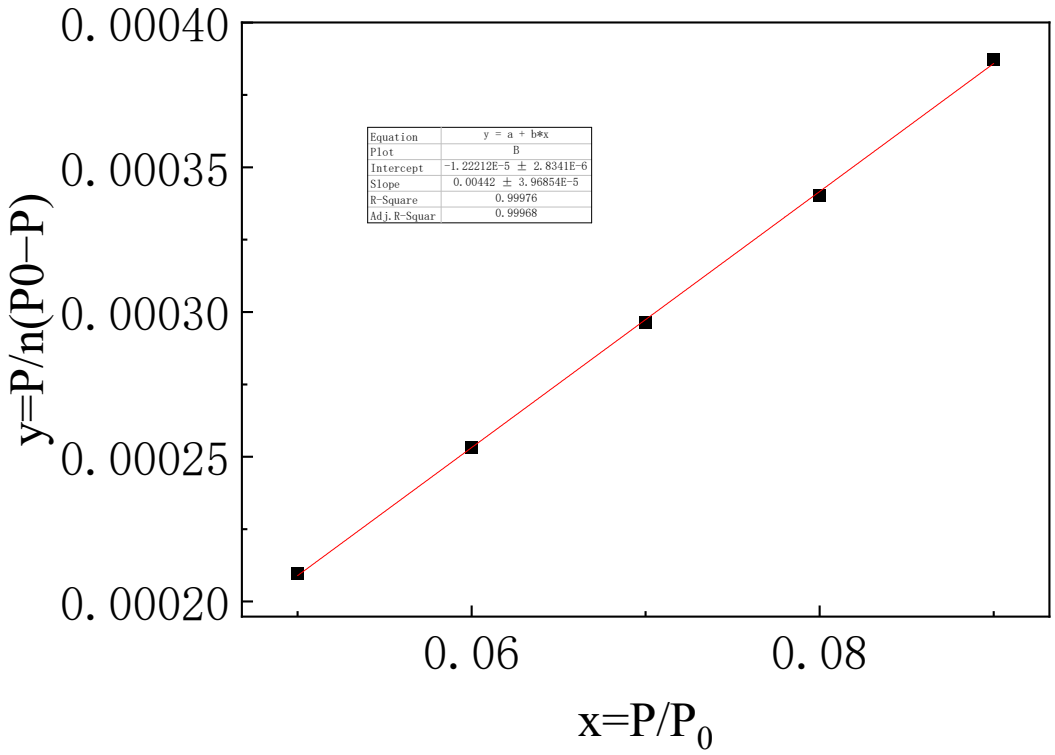



Figure S18 Representative determination of the BET linear range(AROFAP)

.Example C: low-pressure fluctuations (CUXFIL)

The low-pressure uptake shows noticeable point-to-point fluctuations and a slight non-monotonic tail. The window-selection scan excludes unstable points and identifies a later consecutive window $x = [0.06P_0–0.10P_0]$ with improved linearity and maximal $R^2$.

| Pressure_Pa | 3886.0 | 4857.5 | 5829.0 | 6800.5 | 7772.0 |
|---|---|---|---|---|---|
| Adsorption_n | 164.2113004687 | 170.1721620866 | 170.3140997574 | 171.7076222926 | 173.1747164322 |
| x=P/P0 | 0.04000000 00 | 0.05000000 00 | 0.06000000 00 | 0.07000000 00 | 0.08000000 00 |
| y=x/(n*(1-x)) | 0.0002537381 | 0.0003092843 | 0.0003747769 | 0.0004383545 | 0.0005021317 |
| Selected (TRUE/FALSE) | false | false | true | true | true |
| | 8743.5 | 9715.0 | 14572.5 | 19430.0 | 29145.0 |
| | 176.4476117205 | 178.7359662460 | 179.1740990317 | 180.4898855321 | 179.1138883083 |
| | 0.09000000 00 | 0.10000000 00 | 0.15000000 00 | 0.20000000 00 | 0.30000000 00 |
| | 0.0005605125 | 0.0006216494 | 0.0009849113 | 0.0013851192 | 0.0023927314 |
| | true | true | false | false | false |

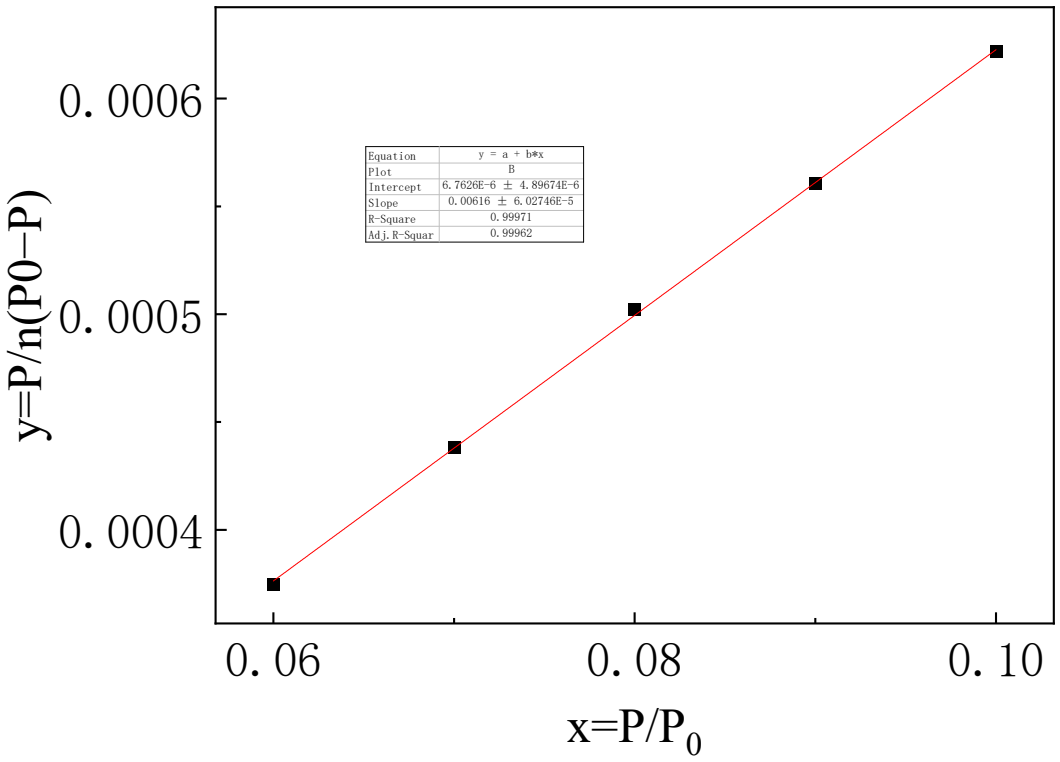



Figure S19 Representative determination of the BET linear range(CUXFIL)

# Reference:


1. H. Wang, F. Cheng, C. Zou, Q. Li, Y. Hua, J. Duan and W. Jin, *CrystEngComm*, 2016, **18**, 5639–5646.
2. P. Feng, X. Bu and G. D. Stucky, *Nature*, 1997, **388**, 735–741.
3. C.-D. Zhang, S.-X. Liu, B. Gao, C.-Y. Sun, L.-H. Xie, M. Yu and J. Peng, *Polyhedron*, 2007, **26**, 1514–1522.
4. S.-Y. Yang, L.-S. Long, R.-B. Huang, L.-S. Zheng and S. W. Ng, *Applied Organometallic Chemistry*, 2004, **18**, 91–92.
5. C.-Q. Li, W.-G. Qiu and H. He, *Acta Crystallographica Section E*, 2011, **67**, m1406.
6. S. Su, Z. Guo, G. Li, R. Deng, S. Song, C. Qin, C. Pan, H. Guo, F. Cao, S. Wang and H. Zhang, *Dalton Transactions*, 2010, **39**, 9123–9130.