

## SUPPLEMENTARY INFORMATION for:

Early stages of phase selection in MOF formation observed in molecular Monte Carlo simulations.

Stephen A. Wells<sup>1</sup>, Naomi F. Cessford<sup>2</sup>, Tina Düren<sup>1</sup>

<sup>1</sup>Centre for Advanced Separations Engineering, Department of Chemical Engineering, University of Bath, Bath, UK; <sup>2</sup>Institute for Materials and Processes, School of Engineering, The University of Edinburgh, Edinburgh, UK.

## CONTENTS:

S.1: EFFECT OF  $P_{\text{LINK}}$  PARAMETER ON CLUSTER FORMATION.

S.2: DETAILS OF EXPLICIT-SOLVENT MOLECULAR MONTE CARLO SIMULATIONS AND PRODUCTION OF POTENTIALS OF MEAN FORCE.

S.3: POTENTIALS OF MEAN FORCE AS USED IN MONTE CARLO SIMULATIONS

S.4: CONSTRUCTION AND TESTING OF  $vO$ - $vO$  POTENTIAL.

S.5: STERIC RADII

S.6: ON CLUSTER IDENTIFICATION FOR PHASES A AND F

S.7: COMPARISON OF CONTACT CLUSTER MONTE CARLO TO ENERGETIC CLUSTERING SCHEMES

S.8: STATISTICS FOR MONTE CARLO SIMULATION RUNS

S.9: EFFECT OF  $P(\text{COLLECT})$  VALUE

## REFERENCES FOR SUPPLEMENTARY MATERIAL

---

<sup>1</sup> Email: s.a.wells@bath.ac.uk

## S.1: EFFECT OF $P_{\text{LINK}}$ PARAMETER ON CLUSTER FORMATION.

Since the collective move scheme is constructed to obey detailed balance, it is necessary that the formation of new contacts as a result of a move be penalised in a way exactly equivalent to the breaking of contacts when the moving cluster is formed. This implies a certain symmetry in the behaviour as a function of the link probability parameter  $P_{\text{LINK}}$ . At low values of  $P_{\text{LINK}}$  clusters are unlikely to be moved collectively, but newly formed contacts are likely to be accepted; whereas at high values, collective movement of clusters is favoured but the formation of new contacts is penalised. On the basis of this symmetry we expect  $P_{\text{LINK}}=0.5$  to be a sensible parameter choice.

To examine this behaviour, we have carried out multiple test simulations; a representative case is reported below. In this case, we used the metal-rich C1 composition and carried out  $1 \times 10^8$  attempted Monte Carlo moves in each simulation, using the collective move scheme with  $P_{\text{COLLECTIVE}} = 0.5$  and  $P_{\text{LINK}}$  values of 0.1, 0.3, 0.5, 0.7 and 0.9, as well as a simulation with only single-object moves and no collective moves. The properties of F-like and A-like clusters formed by the end of the simulation are tabulated below.

At this metal-rich C1 composition, the formation of A-like clusters is disfavoured regardless of the move scheme, with the largest such clusters being of size 4 or 5 in all cases. The formation of F-like clusters is favoured by collective moves and is most favoured by  $P_{\text{LINK}}$  values of 0.3 and 0.5, with a value of 0.1 being comparably effective. It appears that higher values of  $P_{\text{LINK}}$  are disadvantageous. This is understandable as the simulation begins, by construction, in an initial configuration without clusters; the

effect of the higher  $P_{\text{LINK}}$  value in disfavouring the formation of new contacts thus leads to poorer performance. On this basis we use  $P_{\text{LINK}}=0.5$  as our default parameter choice.

<b>Table S.1.1: Cluster-forming performance of simulations with different link probabilities</b>			
$P_{\text{LINK}}$	<b>Largest A-like</b>	<b>Largest F-like</b>	<b>Objects in F-like clusters</b>
No collective	5	6	25
0.1	4	8	66
0.3	5	10	75
0.5	5	9	71
0.7	4	6	30
0.9	5	7	51

## S.2: DETAILS OF EXPLICIT-SOLVENT MOLECULAR MONTE CARLO SIMULATIONS AND PRODUCTION OF POTENTIALS OF MEAN FORCE.

The effective energy of interaction between the various sites in our implicit-solvent model is represented by a set of potentials of mean force (PMFs) inferred from all-atom empirical-potential Monte Carlo simulations. The MC simulations, carried out using a modified version of the MUSIC code<sup>1</sup>, made use of the TIP3P water model<sup>2</sup> and appropriate force field parameters for the divalent cobalt ion<sup>3</sup> and for the succinate ion, using the OPLS-UA forcefield<sup>4, 5</sup> supplemented by the OPLS-AA force field where necessary<sup>6-8</sup>. The simulation temperature was 348 K, corresponding to the experimental synthesis of phases A and F, and the simulation box was cubic with a 25 Å edge length. The succinate ion,  $\text{COO}^- - \text{CH}_2 - \text{CH}_2 - \text{COO}^-$ , was treated as a rigid body and the  $\text{CH}_2$  group was treated as a united atom. Electrostatic interactions were handled in real space using the Wolf method<sup>9, 10</sup> and all interactions were cut off at a distance of 12 Å. This work is discussed at length in the doctoral thesis of NFC<sup>11</sup>.

<b>Geometry</b>	<b>Bond length (Å) / Angle (°) / Dihedral angle (°)</b>	<b>Reference</b>
CH <sub>2</sub> -CH <sub>2</sub>	1.53	5
C-CH <sub>2</sub>	1.522	8
C-O	1.231	4
C-CH <sub>2</sub> -CH <sub>2</sub>	112.0	8
CH <sub>2</sub> -C-O	120.5	8
O-C-O	131.0	4
C-CH <sub>2</sub> -CH <sub>2</sub> -C	-180.0	8
CH <sub>2</sub> -CH <sub>2</sub> -C-O	180.1	7
O-H (water)	0.9572	2
H-O-H	104.52	

<b>Table S.2.2: LJ AND CHARGE PARAMETERS FOR INTERACTION SITES IN MC SIMULATIONS</b>				
<b>Site</b>	<b><math>\sigma(\text{\AA})</math></b>	<b><math>\epsilon/k_B</math> (K)</b>	<b><math>q(\text{eu})</math></b>	<b>Reference</b>
O (carboxylate)	2.960	105.676	-0.800	4
C	3.750	52.838	0.700	4
CH <sub>2</sub>	3.905	59.380	-0.100	4,8
Co	2.186	14.392	2.000	3
O (water)	3.151	76.526	-0.834	2
H	0.0	0.0	0.417	

For every pair of interacting atom types from the set (Co, C, CH, O) a PMF was obtained using umbrella sampling and the Weighted Histogram Analysis Method (WHAM)<sup>12-14</sup>. For the umbrella sampling simulations, interatomic separation distances were set at distances from 1.0-2.0 Å up to 12.0 Å, with window dimensions of 0.1 Å, and constrained by a harmonic confinement potential with force constant of 25 kcal mol<sup>-1</sup> Å<sup>-2</sup> in general, or higher if necessary to ensure sufficient confinement. Each window was sampled for 3×10<sup>8</sup> MC simulation cycles. The self-consistent calculation of the PMF using WHAM proceeded until all points remained converged to within 0.001 kcal mol<sup>-1</sup> over 100 iterations. The final PMFs were shifted to coincide with the macroscopic Coulomb potential at the maximum separation examined, in line with the practice of Madhusoodanan and Tembe<sup>15</sup>.

### S.3: POTENTIALS OF MEAN FORCE AS USED IN MONTE CARLO SIMULATIONS

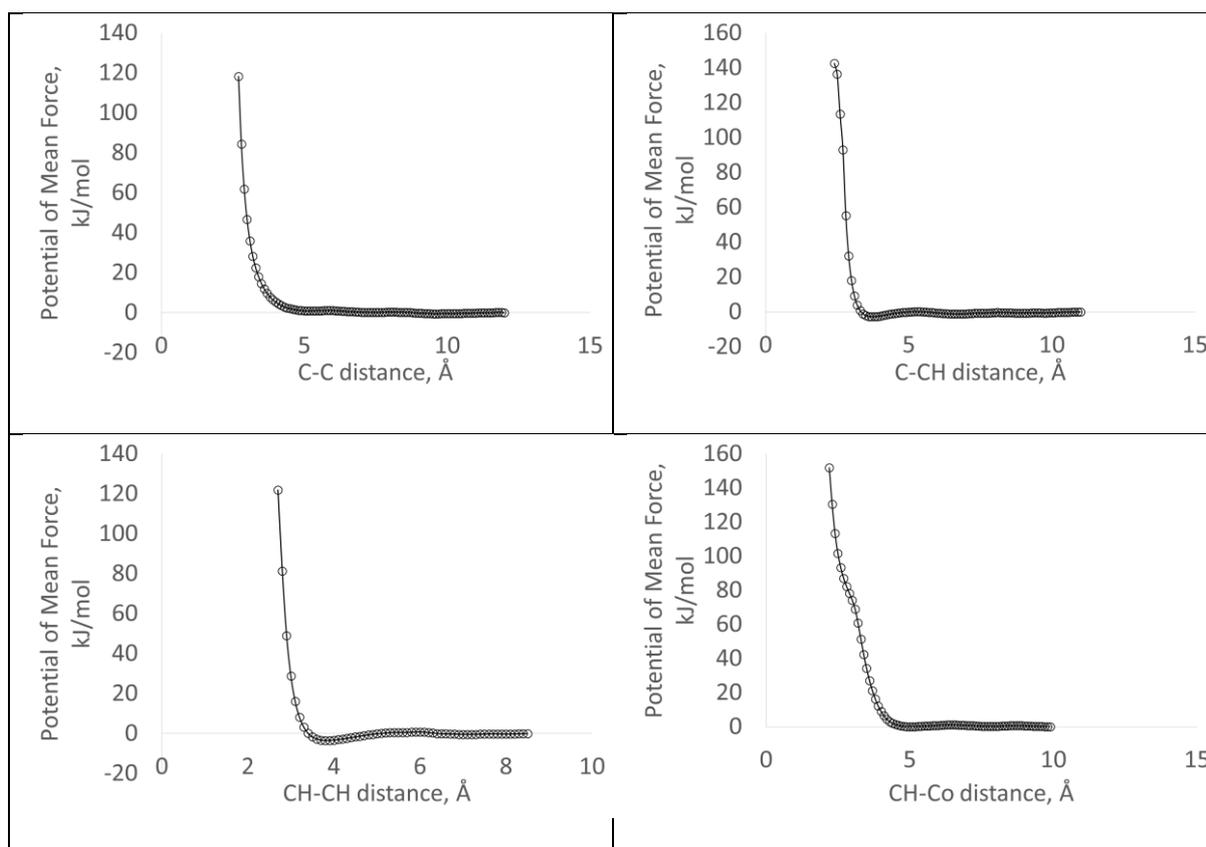
The PMFs used for intersite interactions in our Monte Carlo simulations are provided as a separate text file (in the form in which they are provided as input to our Monte Carlo code). Interactions are headed by the keyword POT followed by the two species involved in the interaction, and are terminated by the keyword END. All potentials are specified as distances in Ångstroms followed by an interaction energy in  $\text{kJ mol}^{-1}$ . Most potentials are specified at spacings of  $0.1 \text{ \AA}$ . Within the code, the energy of interaction at a given distance is calculated using a cubic spline interpolation between the two nearest defined distance points.

The majority of the PMFs defined here are as derived from the MC simulations, with the exception of a very small shift to ensure that the longest-range value becomes zero to avoid a discontinuity in the energy. The interaction of the cobalt ion with the carboxylate oxygen atom is re-interpreted in terms of the “virtual site” system described in the main text. The octahedron of virtual sites surrounding each cobalt ion is constructed with a Co-vO distance of  $1.9 \text{ \AA}$  corresponding to the minimum of the Co-O PMF.

Since in these simulations oxygen sites may be either “real” (carboxyl group oxygens in the ligand) or “virtual”, oxygen potentials have a numerical rider, readable by our code, specifying the sites to which they apply. Potentials labelled as “POT X o 1” describe interactions of species X with carboxyl oxygens. The potential labelled “POT o o 11” describes the interaction of two carboxyl oxygens in different ligands. The potential labelled “POT o o 12” describes the attractive interaction of a carboxyl oxygen

with a virtual site (O-vO), when the ligand coordinates a metal; there is therefore no potential defined between the cobalt itself and the carboxyl oxygen. The potential labelled “POT o o 22” describes the attractive interaction of two virtual sites in the formation of a metal-hydroxyl-metal bridge, and is constructed as a multiple of the “POT o o 12” attractive potential.

The potentials are graphed below, to illustrate that most are relatively featureless with the exception of the O-vO case. The C-Co potential is not graphed, as for this interaction a simple hard sphere exclusion potential was used, with a nominal value of 10,000 kJ/mol for  $r < 2.968 \text{ \AA}$  and a value of 0 at  $r > 2.968 \text{ \AA}$ . The construction of the vO-vO potential is discussed in the following section.



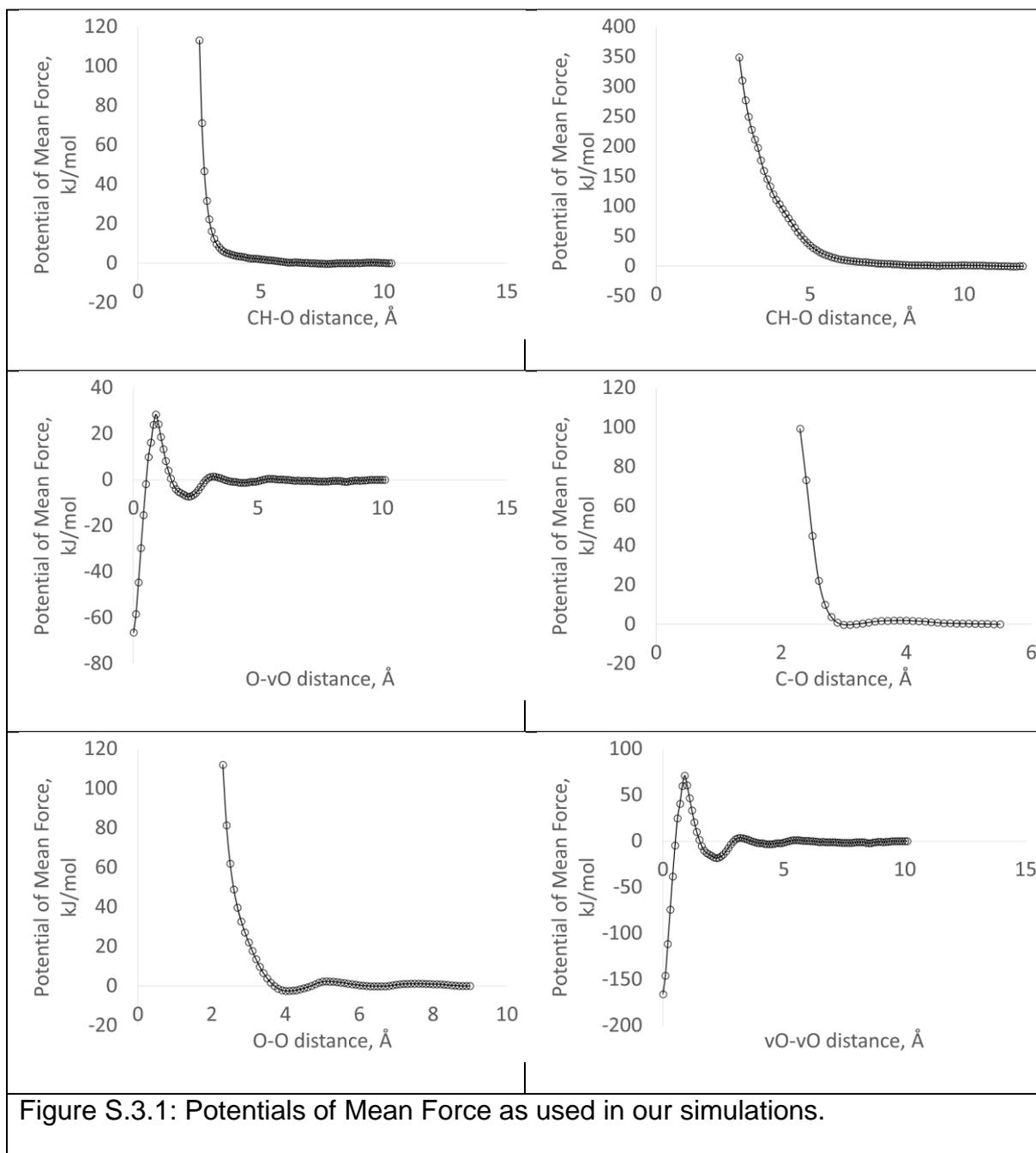
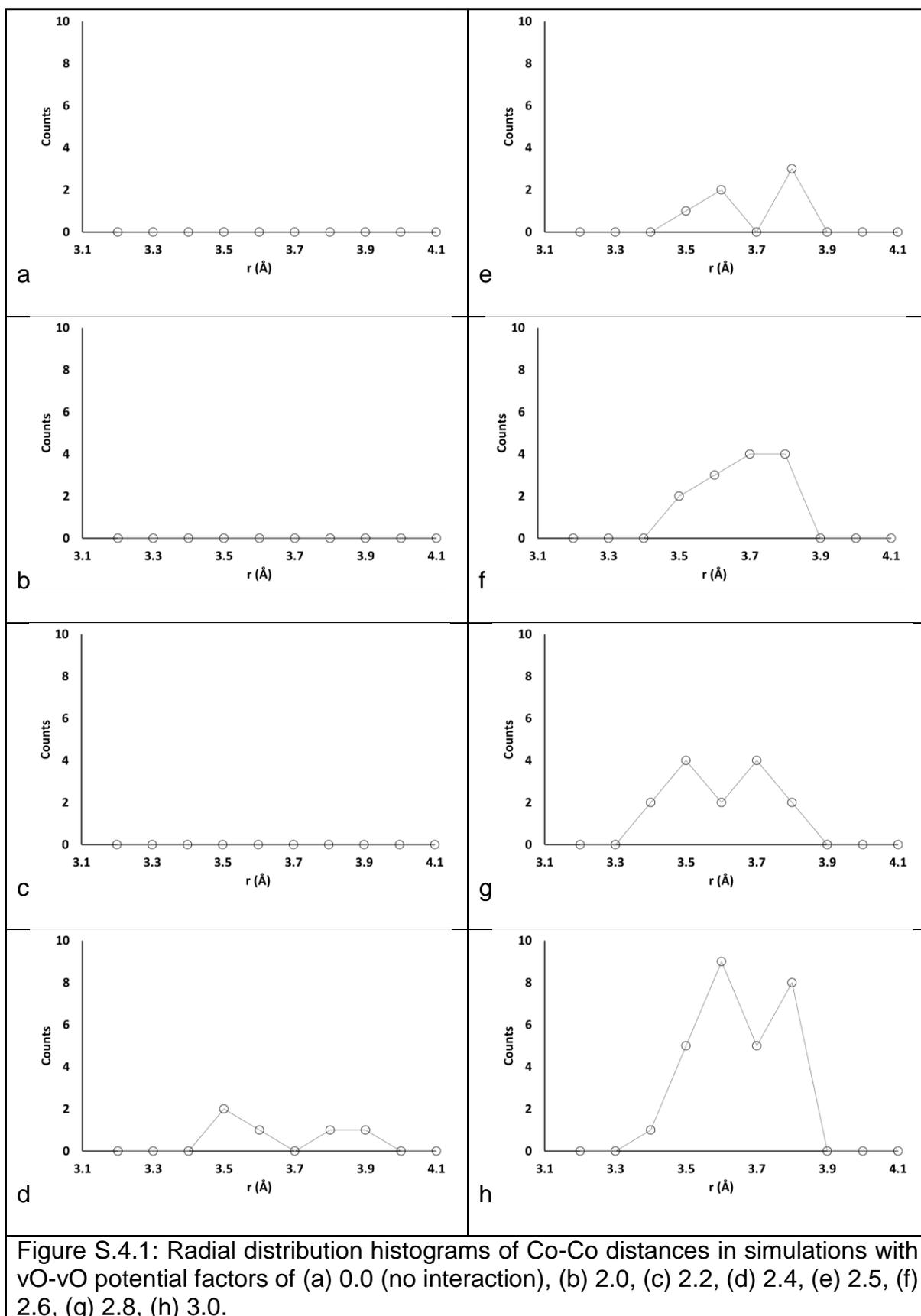


Figure S.3.1: Potentials of Mean Force as used in our simulations.

#### S.4: CONSTRUCTION AND TESTING OF vO-vO POTENTIAL.

Our model also requires an attractive vO-vO interaction, favouring a zero vO-vO distance, in order to generate metal-metal associations as seen in the denser phases. This vO-vO interaction is absent from the set of calculated PMFs and so we employ a scaling of the vO-O interaction, as follows. A single vO-vO interaction represents interactions between two  $\text{Co}^{2+}$  ions and the bridging oxygen species; and we assume that the bridging species is a hydroxyl with charge  $q_{\text{OH}} = -1$ , whereas the charge of a carboxylate oxygen in the MC simulations is  $q_{\text{O}} = -0.8$ . This gives a scaling factor of  $2 \cdot (1/0.8) = 2.5$  under the assumption that the Coulomb interaction is the dominant factor in this interaction.

To confirm the validity of this constructed potential, we have carried out Monte Carlo simulations using a variety of scaling factors. The graphs below were constructed as follows. We generated vO-vO potentials by multiplying the vO-O potential by factors of: 0.0 (that is, no interaction), 2.0, 2.2, 2.4, 2.5, 2.6, 2.8, 3.0. With each potential set we carried out  $2 \times 10^6$  attempted Monte Carlo moves on the metal-rich C1 composition, using the collective move system with both  $P_{\text{LINK}}$  and  $P_{\text{COLLECTIVE}}$  set to 0.5. We exported radial distribution histograms of the Co-Co distances at the end of each simulation, using histogram bins of width  $0.1 \text{ \AA}$ , and examined the range between 3.1 and  $4.1 \text{ \AA}$  to detect the occurrence of linked cobalt octahedra. It is clear that the calculated scaling factor of 2.5 is sufficient to allow the formation of linked Co octahedra, whereas no such linkages are observed for factors less than 2.4. Linkages between cobalt octahedra can be favoured to an arbitrary degree by increasing the scaling factor. For our main study we used the calculated scaling factor of 2.5.



## S.5: STERIC RADII

Hard-sphere radii are assigned to the atomic species in the simulation, to enable rapid rejection of intolerable steric clashes without the need to evaluate interaction energies. Virtual oxygen (vO) sites are not assigned a radius and are permitted to overlap freely with other species. The radii assigned to the C, CH and carbonyl O species are chosen to be slightly less than half the Lennard-Jones  $\sigma$  values discussed in section S.3 above. This ensures that the hard sphere radii used for summary clash rejection are slightly smaller than the radii corresponding to the  $\sigma$  values, and small overlaps (steric contacts) are permitted. The Co ion is assigned a very small radius to ensure that when a carboxyl O overlaps a virtual vO site, no steric clash between Co and O is registered.

Table S.5.1: Hard-sphere radii for species in the simulation, for rapid rejection of steric clashes without energy evaluation.		
Species	Radius (Å)	Notes
C	1.8	Carbonyl carbon
CH	1.8	United atom representing CH <sub>2</sub> group
Co	0.5	Cobalt ion
O	1.35	Carbonyl oxygen

## S.6: ON CLUSTER IDENTIFICATION FOR PHASES A AND F

The identification of A-like clusters is based on the recognition of linear chains of alternating metal (M) and ligand (L) objects, schematically -M-L-M-L-. This is achieved by a double loop over all units in the simulation, during which labels are applied as follows. In a first loop, the label “Primary” is attached to any metal (M) unit with exactly two ligand (L) overlapping neighbours and no M neighbour, and likewise to any L unit with exactly two M neighbours and no L neighbours. In the second loop, a label “Secondary” is applied to some of the neighbours of Primary objects, according to the following rules. An L object whose only overlapping neighbour is an M Primary object, or an M object whose only overlapping neighbour is an L Primary object, is recognised as a chain termination (dangling end) and gets the Secondary label. There may also be units lying in the middle of a chain which did not receive the Primary label due to some defect in the bonding; for example an M unit might overlap with two L units forming a chain, but have additional overlaps with other M or L units. Therefore, an M unit which has exactly two L Primary neighbours, or an L unit which has exactly two M Primary neighbours, receives the Secondary label and is recognised as part of a linear chain. Each chain of overlapping Primary and Secondary objects constitutes an A-like cluster.

The identification of “F-like” clusters likewise proceeds by a labelling process. In this case the characteristic structural feature to be recognised are groups of multiple M objects bridged by L objects. Identification proceeds through three rounds of labelling and one of filtering, as follows. In the first round, the Primary label is assigned to M objects with at least two overlapping neighbours of which at least one is another M

object. In the second round, the Secondary label is attached to any M or L object which is the neighbour of a Primary object. In the third round, the Tertiary label is applied to any M object which is the neighbour of a Secondary object. Objects with the Primary, Secondary or Tertiary label are then grouped into clusters. In a filtering round, only those clusters which contain at least one group of three connected M objects (-M-M-M-) are retained; a strict criterion intended to avoid false-positive identification of F-like clusters.

Due to the labelling process, an A-like or F-like cluster may be identified as a part of a larger cluster, some of whose members do not fit the definition; for example, an A-like linear chain with a single additional object coordinating one of its metal units. It is for this reason that in the paper, it is possible for the number of A- or F-like objects of a given size to exceed the census number of all objects of that size.

A final point to note is that this identification system results in the clusters L-M-L and M-L-M being identified as A-like. However, these three-object clusters can also be identified in the structure of phase F. We therefore do not count such clusters of size three as A-like, and when assessing how many objects are members of A-like clusters, we consider only clusters of size 4 or more.

## S.7: COMPARISON OF CONTACT CLUSTER MONTE CARLO TO ENERGETIC CLUSTERING SCHEMES

It is informative to briefly contrast the properties of CCMC to those of energetic clustering schemes<sup>16, 17</sup> in which the link formation probability for two objects  $x$  and  $y$  is given by a Boltzmann-like exponential term:  $1 - P_{\text{LINK},xy} = \exp(-\Delta E_{xy}/kT)$ , where  $\Delta E_{xy}$  is the change in energy when one object undergoes a proposed move and the other does not. We note that exponentiation is a relatively expensive mathematical operation; timing tests on our code show that the exponentiation of an energy term takes roughly three times as many processor cycles as a call to our pseudo-random number generator. In the energetic methods, energies of interaction between objects lying within the cluster are found and exponentiated individually during cluster construction, as are the energies of interaction between objects in the cluster and other objects in the initial and final states. Energies of interaction must be considered during cluster construction for all objects lying within the range of the interaction potential of the moving cluster, which in these simulations is on the order of 10 Å. Moves which create a highly favourable interaction between the moving cluster and an object in the trial state are heavily penalised by the exponential term controlling the link probability, to maintain detailed balance, and so are likely to be rejected. CCMC considers only objects in close contact when constructing the cluster, and thus needs to consider a smaller number of candidate objects; only a single Boltzmann exponential factor is evaluated, when applying the Metropolis criterion to the total energy change between the initial and final states; and the flat link probability makes it easy for favourable interactions to be both formed and broken during the process of proposing the cluster move. The price of these simplifications, as noted, is that CCMC does not approximate physical dynamics.

## S.8: STATISTICS FOR MONTE CARLO RUNS

We report illustrative statistics (acceptance rates and causes of rejection) for simulation runs with  $P_{\text{COLLECT}} = 0.5$ . For the ligand-poor composition C1 over the course of  $200 \times 10^6$  attempted Monte Carlo moves,  $30.5 \times 10^6$  were accepted (an acceptance rate of  $\sim 15\%$ ). Of the  $169.5 \times 10^6$  rejected moves,  $16.9 \times 10^6$  were rejected for steric clashes and  $6.3 \times 10^6$  for the link probability correction, while the great majority ( $146.3 \times 10^6$ ) were rejected on energetic grounds. In the case of the equivalent simulation run at the ligand-rich composition C2 with  $P_{\text{COLLECT}} = 0.5$ ,  $16.7 \times 10^6$  out of  $200 \times 10^6$  were accepted (an acceptance rate of  $\sim 8\%$ ).  $37.6 \times 10^6$  moves were rejected for steric clashes and  $5.4 \times 10^6$  for the link probability correction, while  $140.2 \times 10^6$  were rejected on energetic grounds. The greater rejection rate overall is thus accounted for by steric clashes in this more densely populated system (for C2, 480 ligands and 400 metal units are present in the simulation box, for C1, 400 metal units but only 80 ligands, corresponding to the experimental synthesis conditions for phase A and F, respectively.). For both systems, the link probability correction is not a dominant source of rejected moves.

### S.9: EFFECT OF P(COLLECT) VALUE

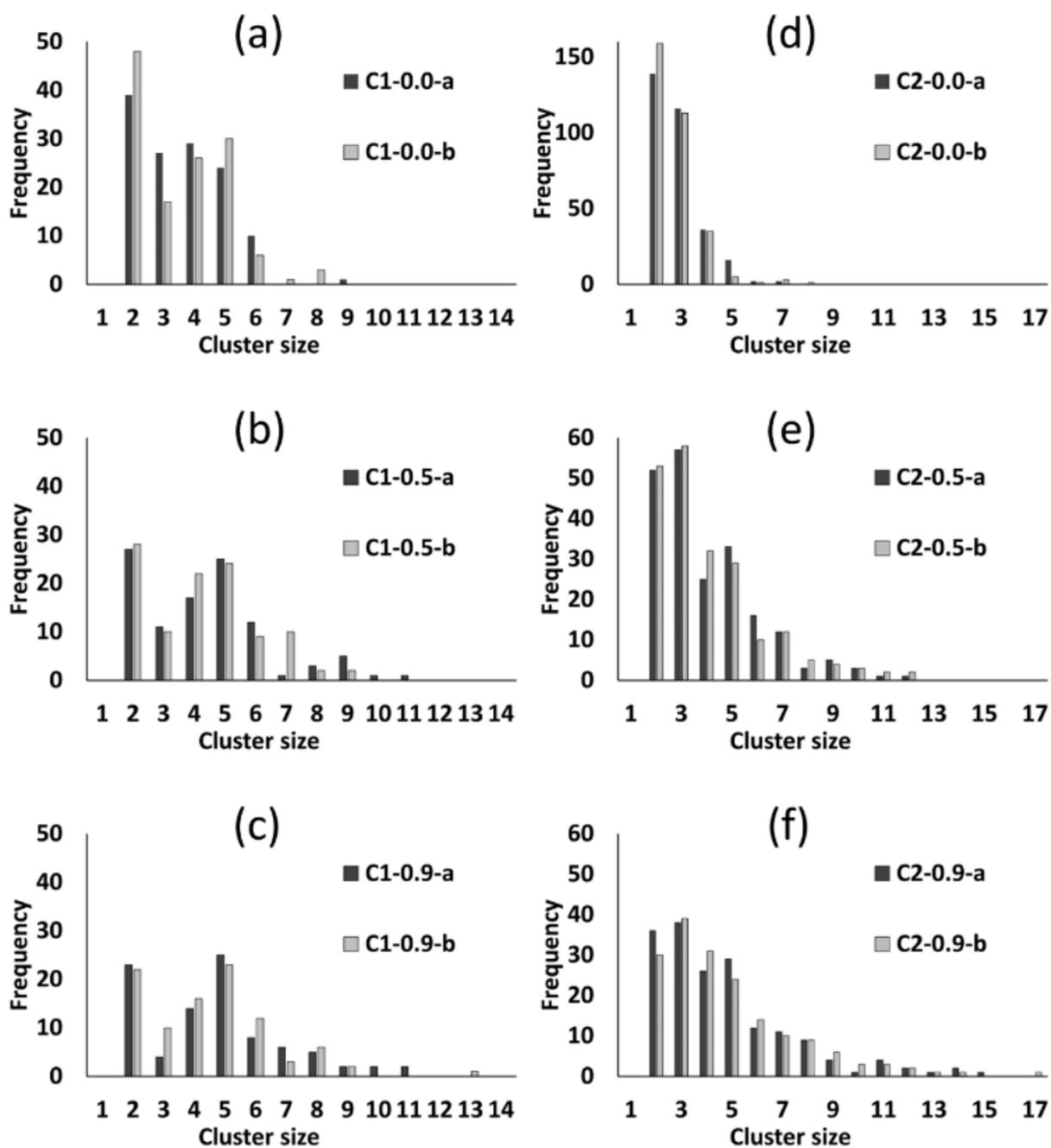


Figure S.9.1: Cluster census data from simulations at compositions C1 (left) and C2 (right) using PCOLLECT values of 0.0 (a,d), 0.5 (b,e) and 0.9 (c,f). Within each panel, data series -a and -b represent two independent Monte Carlo runs with identical parameters but different random seeds.

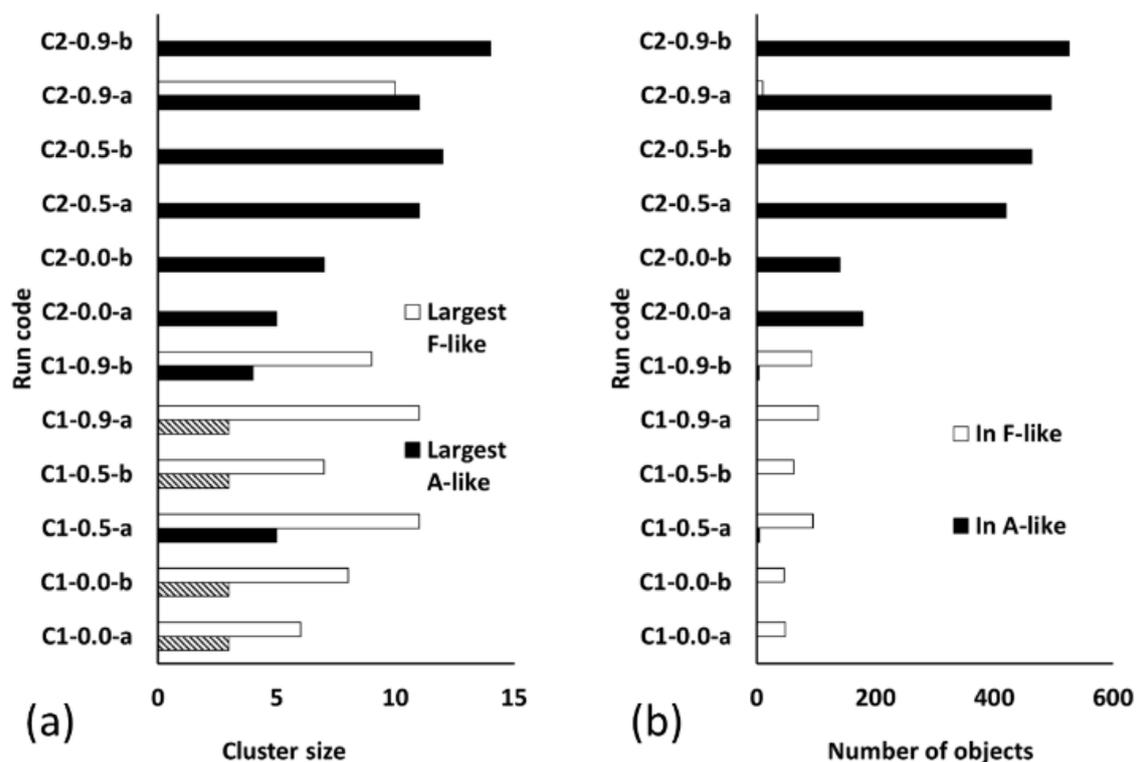


Figure S.9.2: Charts showing (a) the size of the largest A-like and F-like clusters, and (b) the number of molecular objects contained in A-like and F-like clusters. Simulations are coded by compositions (C1, ligand-poor, and C2, ligand-rich) and value of  $P_{\text{COLLECT}}$  (0.0, 0.5, 0.9), with runs –a and –b using different random seeds. Cases where the largest “A-like” cluster is of size 3 are hatched, as in this case the phase identification is ambiguous.

## REFERENCES FOR SUPPLEMENTARY MATERIAL

Details of Dr. Cessford's thesis (reference 11 below) can be found at:

<https://www.era.lib.ed.ac.uk/handle/1842/17610> (accessed 08 February 2019).

1. A. Gupta, S. Chempath, M. J. Sanborn, L. A. Clark and R. Q. Snurr, *Mol Simulat* **29** (1), 29-46 (2003).
2. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J Chem Phys* **79** (2), 926-935 (1983).
3. C. S. Babu and C. Lim, *J Phys Chem A* **110** (2), 691-699 (2006).
4. W. L. Jorgensen and J. Gao, *J Phys Chem-Us* **90** (10), 2174-2182 (1986).
5. W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J Am Chem Soc* **106** (22), 6638-6646 (1984).
6. D. J. Price, J. D. Roberts and W. L. Jorgensen, *J Am Chem Soc* **120** (37), 9672-9679 (1998).
7. W. L. Jorgensen, D. S. Maxwell and J. TiradoRives, *J Am Chem Soc* **118** (45), 11225-11236 (1996).
8. W. L. Jorgensen and C. J. Swenson, *J Am Chem Soc* **107** (3), 569-578 (1985).
9. D. Wolf, P. Keblinski, S. R. Phillpot and J. Eggebrecht, *J Chem Phys* **110** (17), 8254-8282 (1999).
10. C. J. Fennell and J. D. Gezelter, *J Chem Phys* **124** (23) (2006).
11. N. F. Cessford, Edinburgh, 2014.
12. S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg, *J Comput Chem* **13** (8), 1011-1021 (1992).
13. H. Wang, J. Stalnaker, H. Chevreau and J. P. Lewis, *Chem Phys Lett* **457** (1-3), 26-30 (2008).
14. J. S. Hub, B. L. de Groot and D. van der Spoel, *J Chem Theory Comput* **6** (12), 3713-3720 (2010).
15. M. Madhusoodanan and B. L. Tembe, *J Phys Chem-Us* **98** (28), 7090-7094 (1994).
16. A. Bhattacharyay and A. Troisi, *Chem Phys Lett* **458** (1-3), 210-213 (2008).
17. S. Whitelam and P. L. Geissler, *J Chem Phys* **127** (15) (2007).