5 Supplementary Material

5.1 Defect Setup

Defects were set up by removing the layers from the underlying (111) Au surface as discussed in the main body of the paper. The characteristic Au-Au distance of 2.87Å has been maintained in all cases. Similarly, a 30^o angle with respect to the normal to the plane for each SAM chain has been maintained. As shown in Fig.12 for the *flat surface* case, S-atoms on the 16-MHDA chains will typically sit at specific surface points in-between 3 Au-atoms, separated by a characteristic distance of 4.98Å. This characteristic S positioning has been maintained for all the defects we generated. Setting up the configuration for *Vacancy type B* has required a minor shift of 1.34Å of the S-atoms as indicated in the figure. This was necessary as otherwise the depth of the defect in combination with the orientation of the chains causes overlap between chains in the starting configuration. In Fig.12 we have highlighted the original placement in blue as well as the movement of the S-atoms in red.

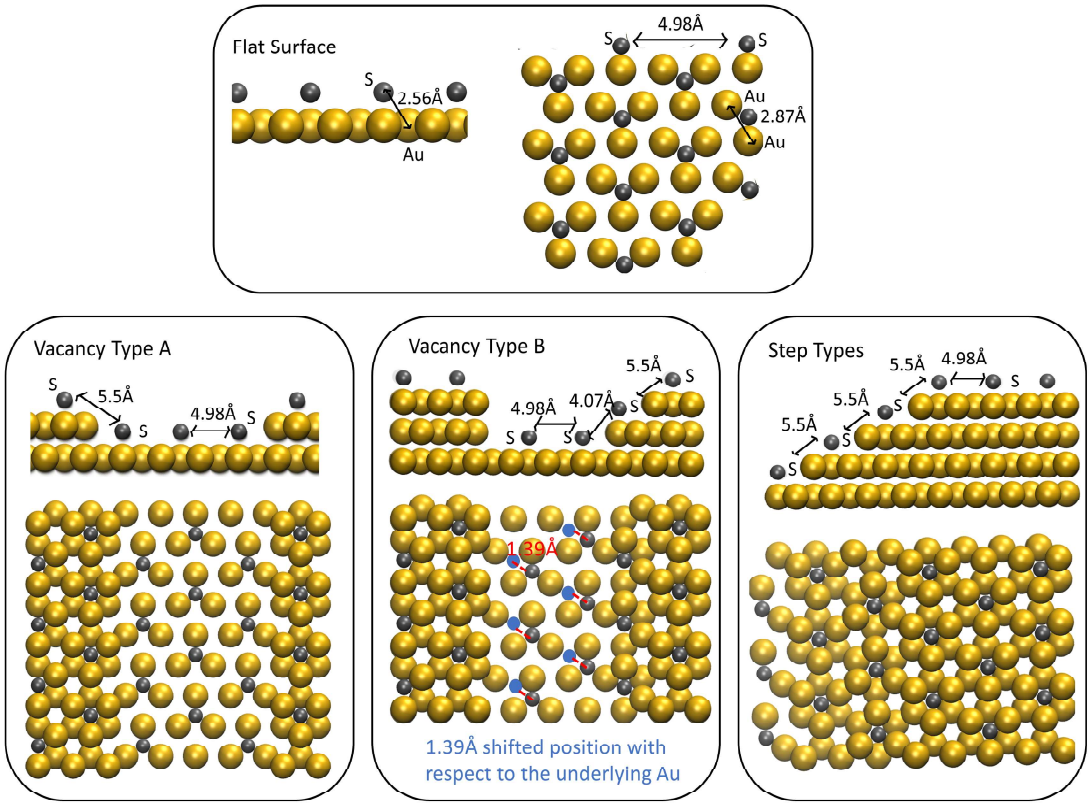

Fig. 12 Schematic of defect configuration setup. Au-atoms are depicted in yellow, while S-atoms have been shown in grey for clarity. Two different aspects have been provided for each surface topology - an *xz*-plane (left) and an *xy*-plane (right) views.

5.2 Full List of Forcefield Parameters

Tables 1 and 2 in this section give the non-bonded and bonded parameters used to set up the forcefield interactions. A reference to the atom labels can be found in Fig.13. The forcefield setup follows that of Raiteri et al.[29], where all calcium and carbonate intermolecular terms are tapered to 0 over the range of 6-9Å.

16-MHDA was parametrised using the standard `antechamber` procedure which generates atomic charges for the AMBER forcefield, while charges for $CaCO_3$ and H_2O are those of the Raiteri et al.[29] forcefield. Whenever the Lennard-Jones potential has been used, we use the standard 12/6 format, given as:

$$E = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{1}$$

Table 2 Non-bonded interactions.

Lennard-Jones	ε [eV]	σ [Å]
C_3-C_3	0.00474820	3.39961
C_3-C	0.00420798	3.39974
C_3-O_H	0.00658129	3.23316
C_3-O	0.00657459	3.17990
C_3-H_C	0.00180006	3.02437
C-C	0.00373303	3.39958
C-O_H	0.00583914	3.23297
C-O	0.00583048	3.17984
C-H_C	0.00159251	3.02491
O_H-O_H	0.00912738	3.06654
O_H-O	0.00912302	3.01315
O_H-H_C	0.00249707	2.85771
O-O	0.00911094	2.95997
O-H_C	0.00248998	2.80488
H_C-H_C	0.00067729	2.65083
O_W-O_W	0.00674	3.16549
C_3-O_W	0.00565711	3.28255
C-O_W	0.00501604	3.28253
O_H-O_W	0.00784338	3.11602
O-O_W	0.00783631	3.06273
H_C-O_W	0.00213658	2.90816
C_3-Au	0.00283275	3.1673
C-Au	0.00251174	3.1673
O_H-Au	0.0039275	3.00077
O-Au	0.00392396	2.94749
H_C-Au	0.00106987	2.79291
Ca-O_W	0.00095	3.35
C_3-C_C	0.00420798	3.39974
C-C_C	0.00373303	3.39958
O_H-C_C	0.00583914	3.23297
O-C_C	0.00583048	3.17984
H_C-C_C	0.00159251	3.02491
C_3-O_C	0.00967158	3.07231
C-O_C	0.00857559	3.07229
O_H-O_C	0.0134093	2.90577
O-O_C	0.0133972	2.85249
H_C-O_C	0.00365275	2.69792

Buckingham	A [eV]	ρ [Å]	C [eVÅ^6]
Ca-O_C	3161.63	0.27151	0.0
Ca-C_C	120000000	0.12	0.0
O_C-O_C	63840.20	0.19891	27.899
O_C-O_W	12534.46	0.2020	12.090
O_C-H_W	396.30	0.2170	0.0
Ca-O	1820.0	0.27151	0.0
Ca-O_H	1999.8	0.27151	0.0
Ca-C	120000000	0.12	0.0
Ca-C_3	120000000	0.12	0.0

Table 3 Bonded interactions.

Harmonic	k_2 [eV Å^{-2}]	r_0 [Å]
C_3-C_3	13.042	1.5375
C_3-C	13.567	1.5250
C_3-H_C	14.330	1.0969
C-O_H	17.342	1.3513
C-O	27.642	1.2183
O_H-H_O	16.098	0.9730
C_3-S	13.042	1.8450
O_W-H_W	22.965	1.012
C_C-O_C	17.950	1.313

Angle-bending	k_θ [eV rad^{-2}]	θ_0 [deg]
C_3-C_3-C_3	2.726	111.51
C_3-C_3-H_C	2.006	109.80
C_3-C_3-C	2.743	111.04
C_3-C-O_H	2.964	112.73
C_3-C-O	2.921	123.2
C-C_3-H_C	2.032	108.77
C-O_H-H_O	2.162	106.55
O_H-C-O	3.289	122.1
H_C-C_3-H_C	1.707	107.58
H_W-O_W-H_W	1.645	113.24
O_C-C_C-O_C	6.000	120.0

Dihedrals	k_ϕ [eV]	n	δ [deg]
C_3-C_3-C_3-C_3	0.0078	3	0
C_3-C_3-C_3-C_3	0.0087	1	180
C_3-C_3-C_3-C_3	0.0108	2	180
C_3-C_3-C_3-H_C	0.0069	3	0
C-C_3-C_3-H_C	0.0067	3	0
C_3-C-O_H-H_O	0.0997	2	180
O-C-C_3-H_C	0.0347	1	0
O-C-C_3-H_C	0.0035	3	180
O-C-O_H-H_O	0.0824	1	0
H_C-C_3-C_3-H_C	0.0069	3	0

Out of plane	k_2 [eV Å^{-2}]	k_4 [eV Å^{-4}]
C_C-O_C-O_C-O_C	20.796	360

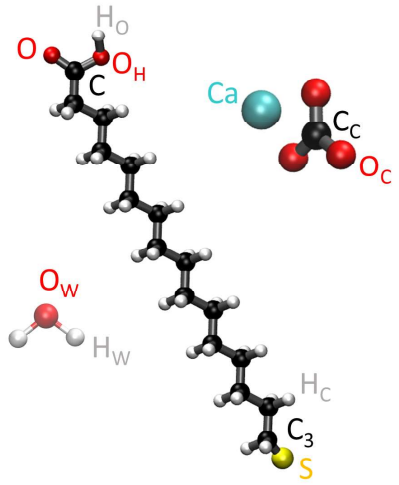

Fig. 13 Structure of all system components, where atom labels have been introduced to ease reading of the forcefield tables in this section.

5.3 Interquartile Range Analysis

Interquartile range (IQR) analysis was performed when calculating the mean cluster diffusion coefficients for each case. This procedure allows us to report a representative cluster diffusion by removing any outliers. In practice, outliers correspond to ions which have detached from the cluster and have much larger diffusion coefficient for the duration of the simulation. The IQR is identified using the following procedure:

1. Identify the median in the data set and create two subgroups of data points containing the lower half values and the upper half values

2. Find the median of each subgroup, which are referred to as Q1, the 25^{th} percentile, and Q3, the 75^{th} percentile.

3. The IQR is the difference of Q3−Q1

Outliers were flagged where they fall outside the boundary of Q1-1.5×IQR and Q3+1.5×IQR.

The table below reports all diffusion coefficients, where those highlighted in red have been flagged up as outliers following the aforementioned procedure. For reference, all IQR analysis plots have also been provided.

Table 4 Diffusion data on individual ions in each run, where the outliers discarded from the calculation of the ion cluster diffusion have been highlighted in red. Data has been provided in units of $10^{-11}\mathrm{m}^2\mathrm{s}^{-1}$

flat surface			vacancy type A			vacancy type B			step type A			step type B		
set 1	set 2	set 3	set 1	set 2	set 3	set 1	set 2	set 3	set 1	set 2	set 3	set 1	set 2	set 3
0.49	0.08	0.37	1.49	0.13	0.15	0.004	0.038	0.138	0.086	0.543	0.017	0.946	0.469	0.033
2.56	0.22	0.36	0.16	0.09	0.21	0.008	0.163	0.011	0.283	0.003	0.002	0.369	0.865	2.028
0.11	0.19	0.12	0.022	0.22	0.23	0.040	0.351	0.007	0.004	0.003	0.164	0.424	0.346	0.262
0.0042	0.19	0.027	1.00	0.12	0.21	0.011	0.146	0.008	0.441	0.033	0.010	1.632	0.661	0.179
0.031	2.27	0.029	0.11	0.26	0.27	0.113	0.045	0.003	0.002	0.227	0.310	0.128	0.024	0.061
0.012	0.086	0.17	0.011	0.12	0.069	0.088	0.143	0.038	0.544	0.087	0.091	0.158	0.075	0.049
0.076	0.19	0.028	4.45	0.024	0.031	0.015	0.014	0.007	0.020	0.034	0.252	0.176	0.052	0.016
0.53	0.45	0.079	0.14	0.59	0.049	0.064	0.008	0.006	0.248	0.289	0.010	2.161	0.061	0.149
6.78	0.027	0.13	0.69	0.075	0.087	0.012	0.005	0.007	0.041	1.257	0.005	1.380	0.188	0.043
0.32	0.089	3.1	0.14	14.9	1.55	0.009	0.281	0.005	0.003	0.244	0.043	0.310	0.091	0.795
0.49	0.0027	0.0036	0.046	0.041	0.005	0.143	0.018	0.007	0.209	0.044	0.002	0.111	0.052	0.000
0.007	0.22	0.036	0.029	19.6	0.007	0.006	0.108	0.041	0.134	0.159	0.009	0.268	0.200	0.013
0.029	0.075	0.029	0.006	0.07	0.096	0.007	0.067	0.009	0.016	0.892	0.076	1.935	0.264	0.130
0.21	0.0017	0.19	1.05	0.098	0.012	0.331	0.010	0.021	0.059	0.015	0.012	1.121	0.011	0.031
3.15	0.13	0.014	0.064	0.055	0.111	0.007	0.011	0.034	0.015	0.026	0.018	0.021	0.024	0.835
0.41	0.096	0.0047	0.48	0.65	0.032	0.039	0.306	0.004	0.135	0.202	0.002	0.243	0.258	0.069
7.2	1.88	0.037	0.03	0.011	0.020	0.006	0.009	0.007	0.013	0.005	0.008	1.524	0.630	0.016
0.0033	0.12	0.068	0.11	0.016	0.153	0.005	0.019	0.005	0.011	0.011	0.063	0.242	0.169	0.089
0.051	0.12	0.13	6.27	0.13	1.99	0.006	0.010	0.009	0.003	0.110	0.007	0.030	0.627	0.187
0.024	0.022	1.44	0.012	0.017	0.013	0.008	0.003	0.021	0.003	0.004	0.028	0.009	0.635	1.698

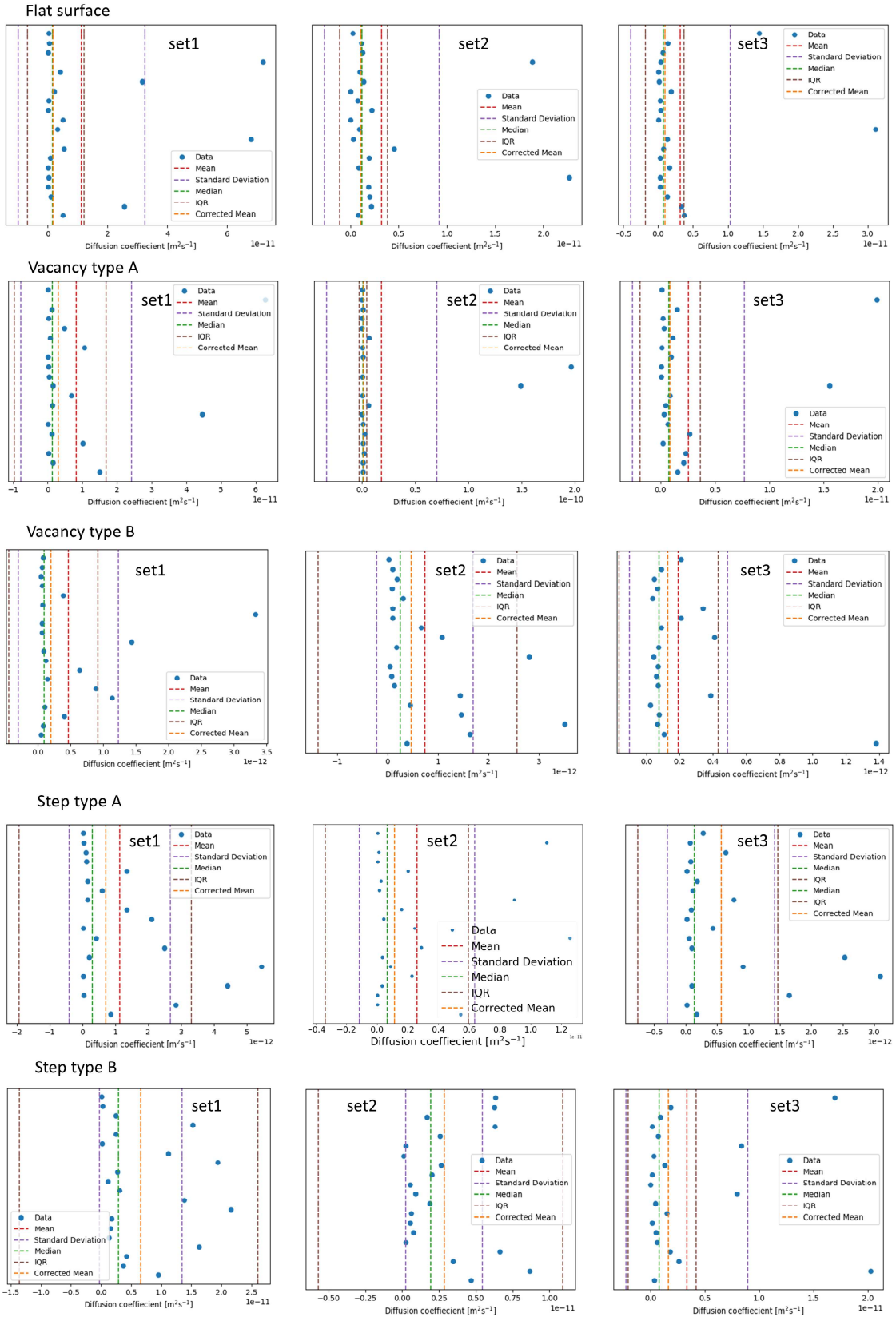

Fig. 14 Figures of IQR analysis for all surface topologies and each simulation set - data points (blue scatter), mean(red), standard deviation(purple), median(green), corrected mean (orange), IQR (brown).

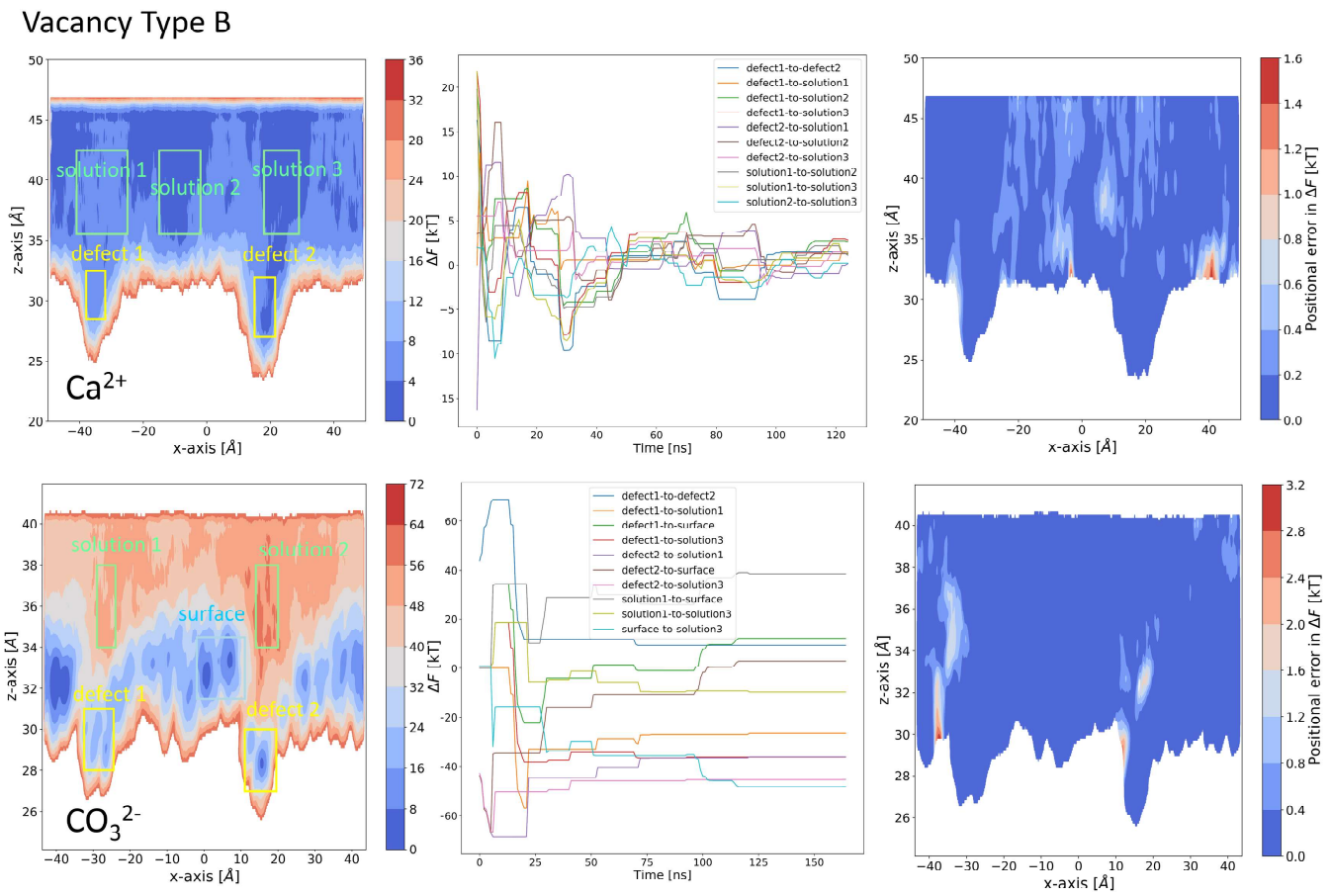

Fig. 15 From left to right: Free energy profile recovered from PLUMED; Convergence analysis in time based on the highlighted sections in the free energy profiles; Positional standard error in the free energy over the last 20ns of simulation time.

5.4 WTmetaD Convergence Analysis

This section discusses the convergence analysis for the free energy profiles of a Ca^{2+} and CO_3^{2-} ion at a SAM surface which has topological defects of the type *vacancy type B*. Fig.15 reports each free energy profile in units of k_BT, where areas of interest have been highlighted an labelled. We record the difference between the minimum free energy values in the highlighted regions, shown in the middle plots. Plotting this metric in time allows us to gauge how much the free energy profile changes in those regions. In both cases we show that towards the last 10-20ns of the simulation fluctuations are as low as $1.5k_BT$ for the Ca^{2+} ion and $0k_BT$ for the cation, which gives us confidence that the CV space has been sampled sufficiently and the reported free energy profiles are reliable. Additionally, we report the positional standard error for each free energy profile for the last 20ns of the simulation. This calculation also shows that the CV space has been thoroughly explored as positional error is on average less than $1k_BT$ unit.

As a point of comparison, we also report the free energy profile for each ion at a flat, defect-free surface in Fig.16.

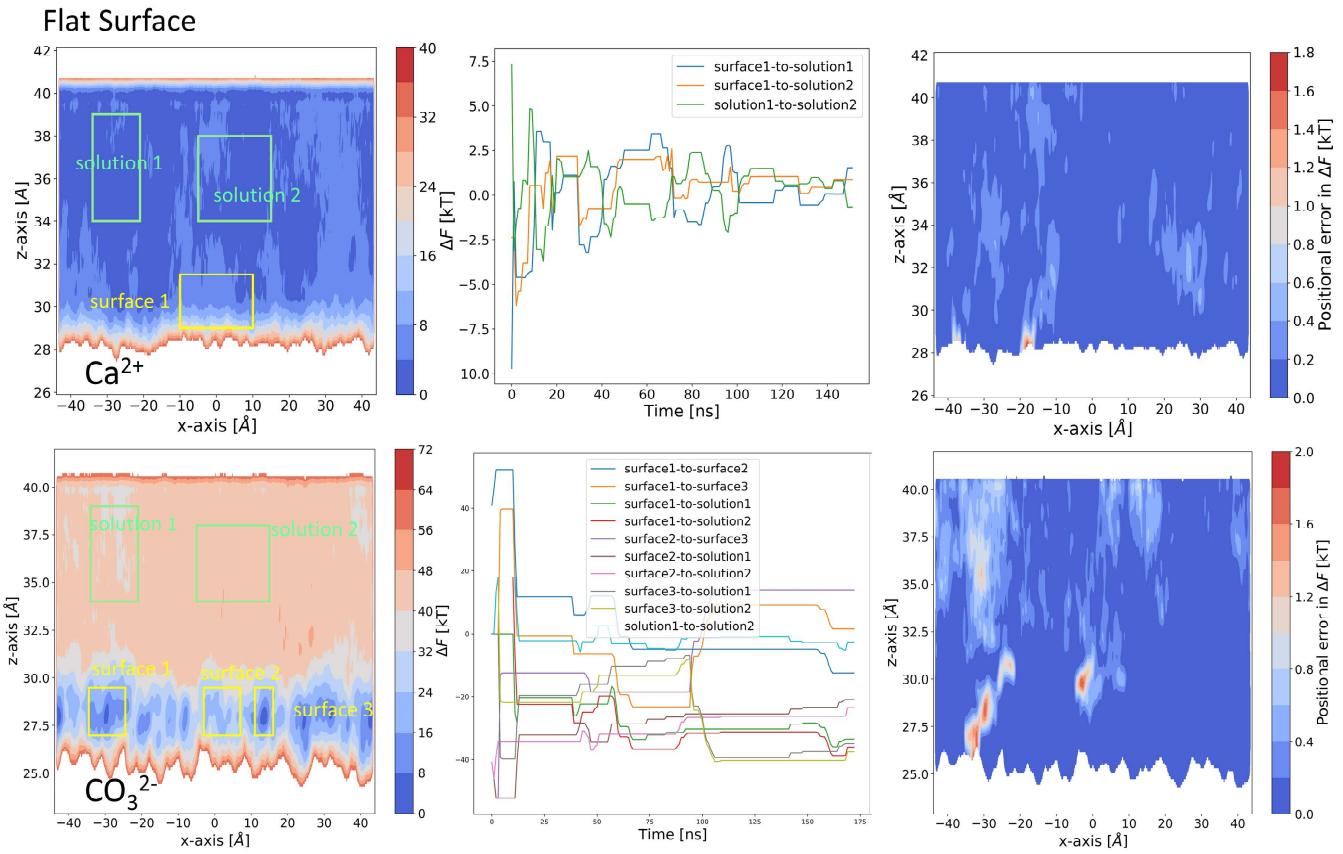

Fig. 16 From left to right: Free energy profile recovered from PLUMED; Convergence analysis in time based on the highlighted sections in the free energy profiles; Positional standard error in the free energy over the last 20ns of simulation time.

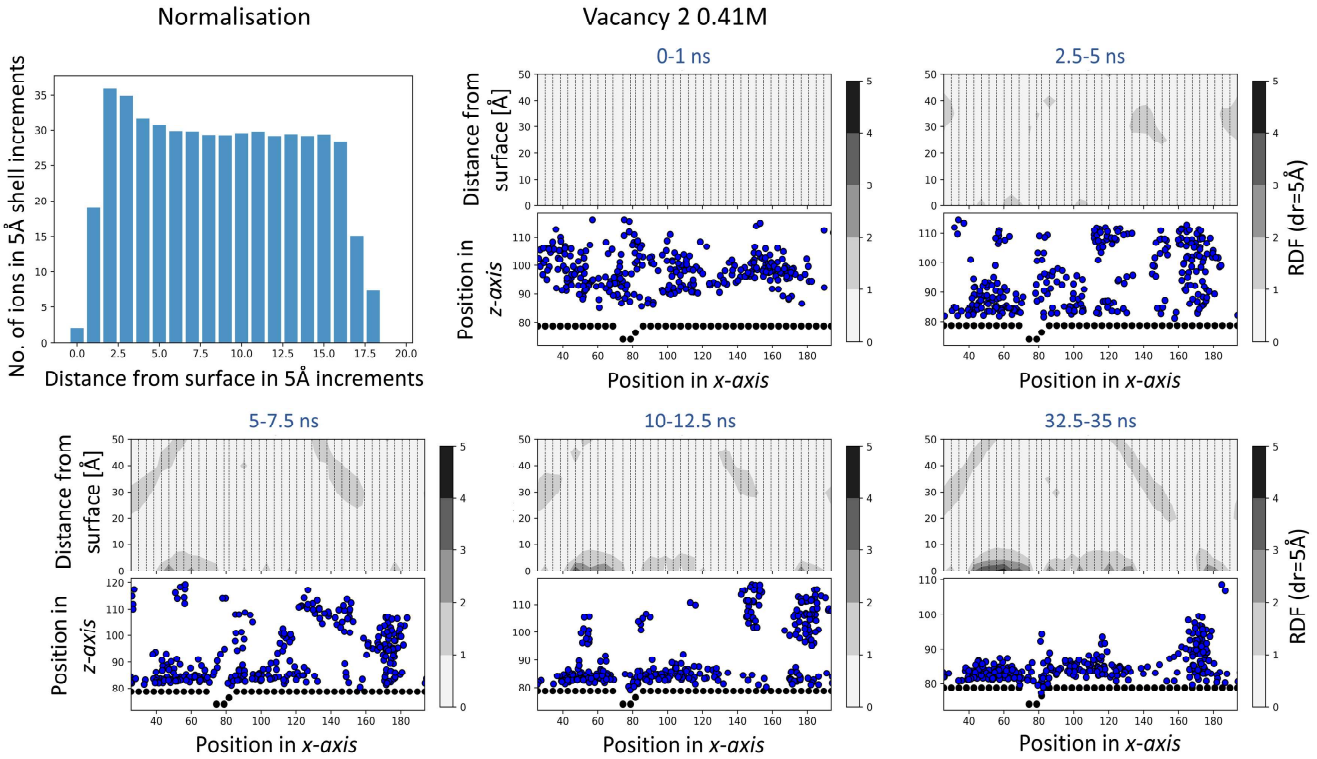

Fig. 17 Normalisation factor used for RDF analysis; Incremental time progression of the RDF calculated for the system with ionic concentration of 0.41M

5.5 Large-Scale Simulation RDF calculation

We calculate the Radial Distribution Function (RDF) between a unique -COOH functional group along the surface *xz*-plane cross-section and ions in solution.

Due to the specificity of the problem, we calculated a custom-made normalisation factor, reported in Fig.17. Ions onto a surface have a 2D distributional character which has to be accounted for when calculating their RDF profile. To do this, we generated 600 random configurations of 500 ions each using Packmol (corresponding to our higher concentration of ions) in a rectangular box of dimensions 130Å $\times$ 60Å $\times$ [10,12.5 and 15]Å. The dimensions were chosen to represent the typical surface coverage the typical thickness of the ionic layer in our simulations. To obtain a normalisation factor, we count the number of ions from the centre of the box in increments of 5Å in each configuration and average over the 600 configurations. The normalisation factor we used is reported in Fig.17.

In Fig.17 and Fig.18 we report the incremental time change in the RDF in the two sets of large-scale simulations we performed in order to demonstrate that in both cases we start from a random distribution of ions and the preferential clustering around the defects occurs naturally during the course of the simulation.

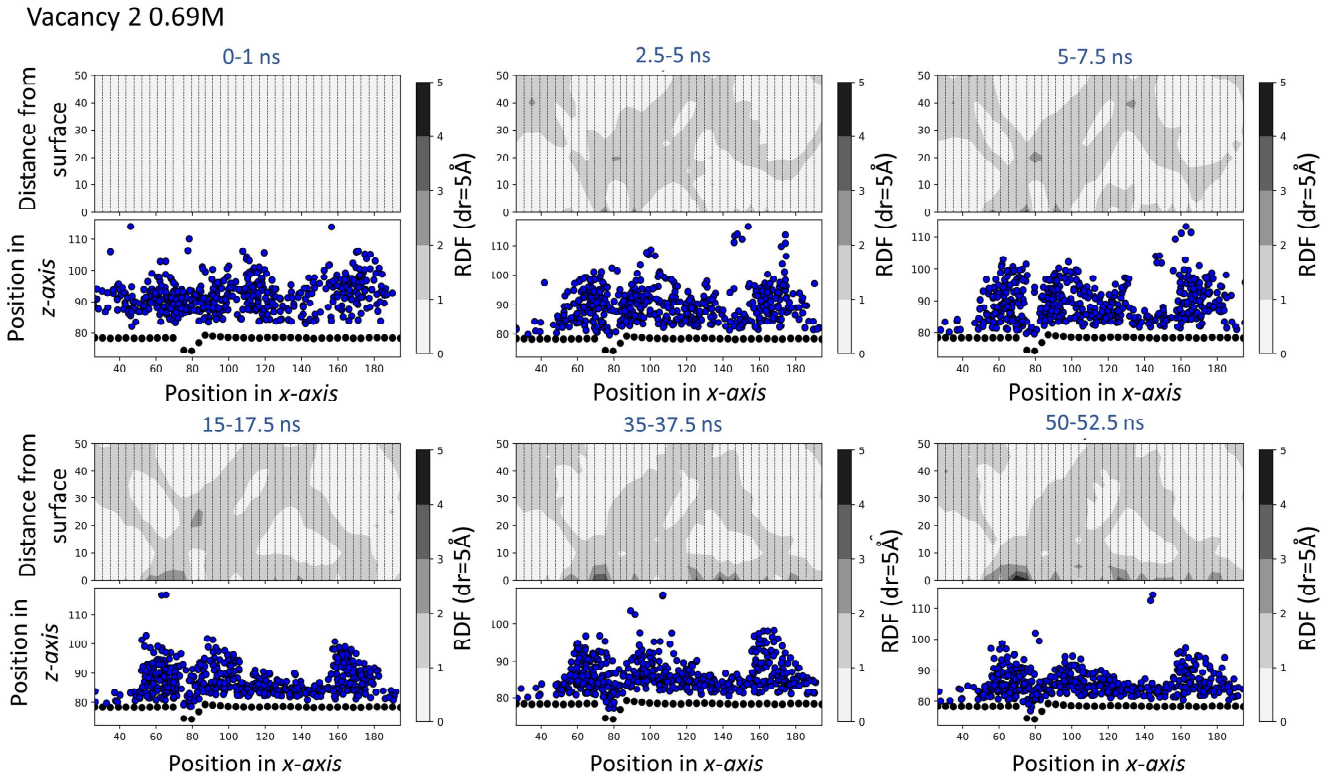

Fig. 18 Incremental time progression of the RDF calculated for the system with ionic concentration of 0.69M