Electronic Supplementary Information

Computational Modelling Workflows for Metal-Organic Polyhedra in The World Avatar

Patrick W.V. Butler,[a] Arravind Subramanian,[b] Simon D. Rihm,[ac] Ari F. Fischer,[bd] Tej S. Choksi,[bd] Aleksandar Kondinski,[e] Sebastian Mosbach,[ab] Jethro Akroyd,[ab] and Markus Kraft*[abf]

[a] *Department of Chemical Engineering and Biotechnology, University of Cambridge, Philippa Fawcett Drive, Cambridge, CB3 0AS, UK. E-mail: mk306@cam.ac.uk*
[b] *CARES, Cambridge Centre for Advanced Research and Education in Singapore, 1 Create Way, CREATE Tower, #05-05, Singapore, 138602.*
[c] *CMPG, GRIPS – Gründerinnenzentrum Pirmasens, Delaware Avenue 1–3, 66953, Pirmasens, Germany.*
[d] *School of Chemistry, Chemical Engineering and Biotechnology, Nanyang Technological University, 62 Nanyang Drive, Singapore, 637459.*
[e] *Institute of Physical and Theoretical Chemistry, Graz University of Technology, Stremayrgasse 9, Graz, 8010, Austria*
[f] *MIT, Chemical Engineering, 77 Massachusetts Avenue, Room E17-504, Cambridge, MA 02139 USA.*

S.1 Benchmark MOP Geometry Optimisations

S.1.1 Average Window Diameters

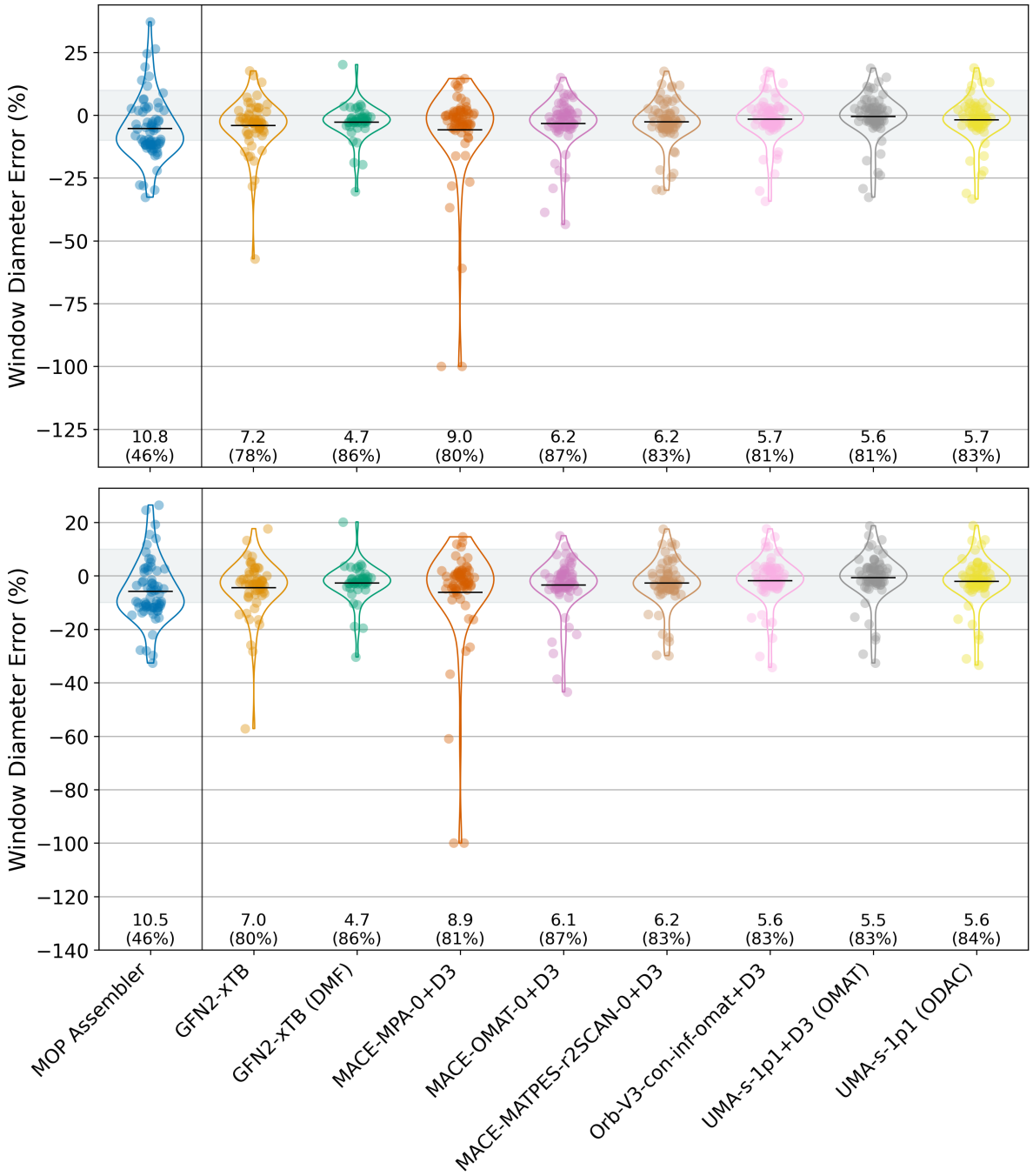

Fig. S.1 Violin plots for each method illustrating the percent error in terms of the average window diameter when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. The distributions include up to 85 structures with unconverged or failed optimisations omitted. The mean of the distributions are indicated by the black bars with the percent MAEs underneath each distribution. The proportion of data points within the shaded region is displayed in brackets below the MAE. The top plot excludes structures with window values less than 1.8 Å(removing 15 structures). To assess the sensitivity to the cutoff, the bottom plot shows the results excluding structures with window values less than 2.5 Å (removing 16 structures).

S.1.2 Parity Plots

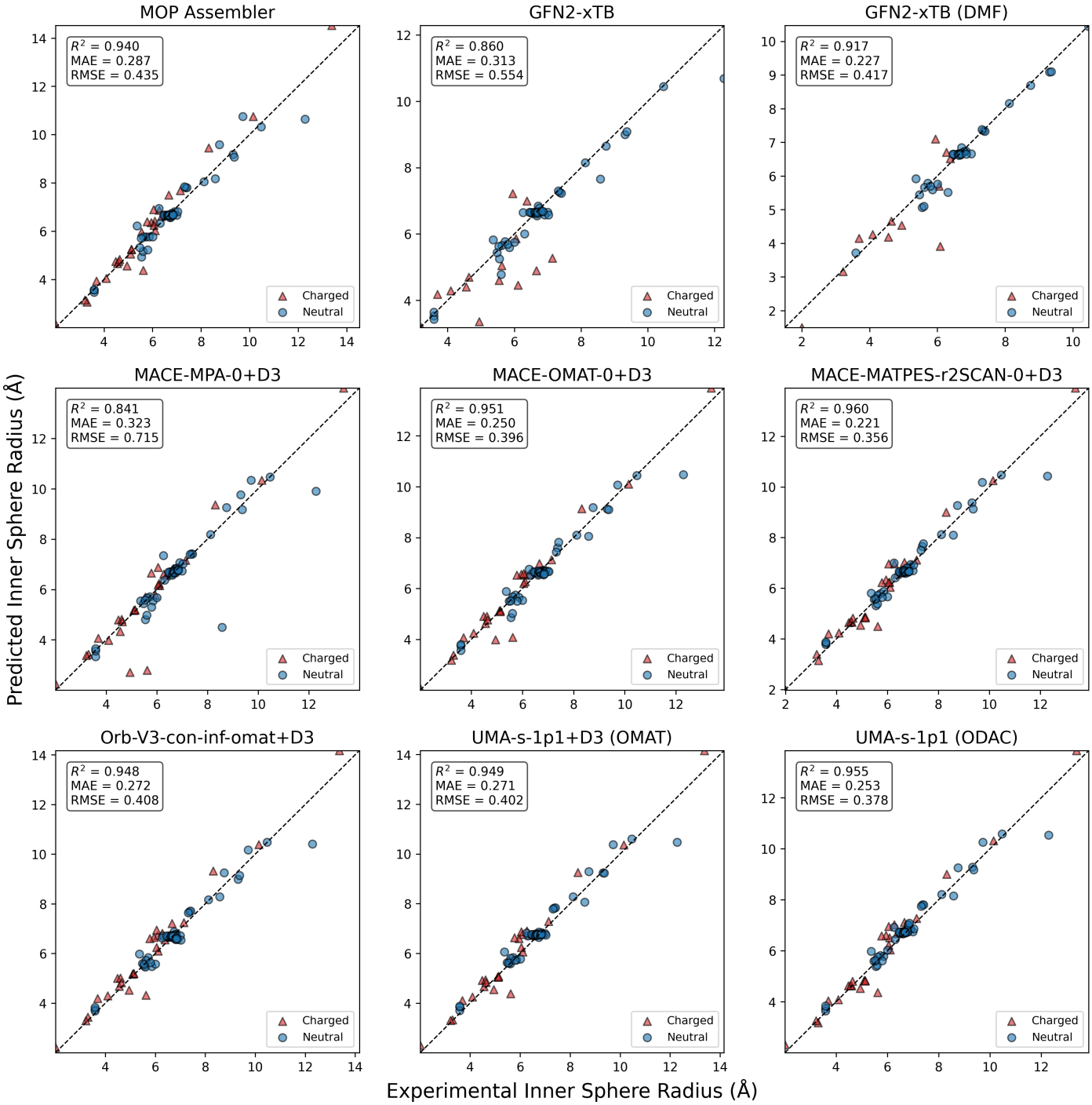

Fig. S.2 Parity plots of the experimental and predicted maximum inner sphere radii when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

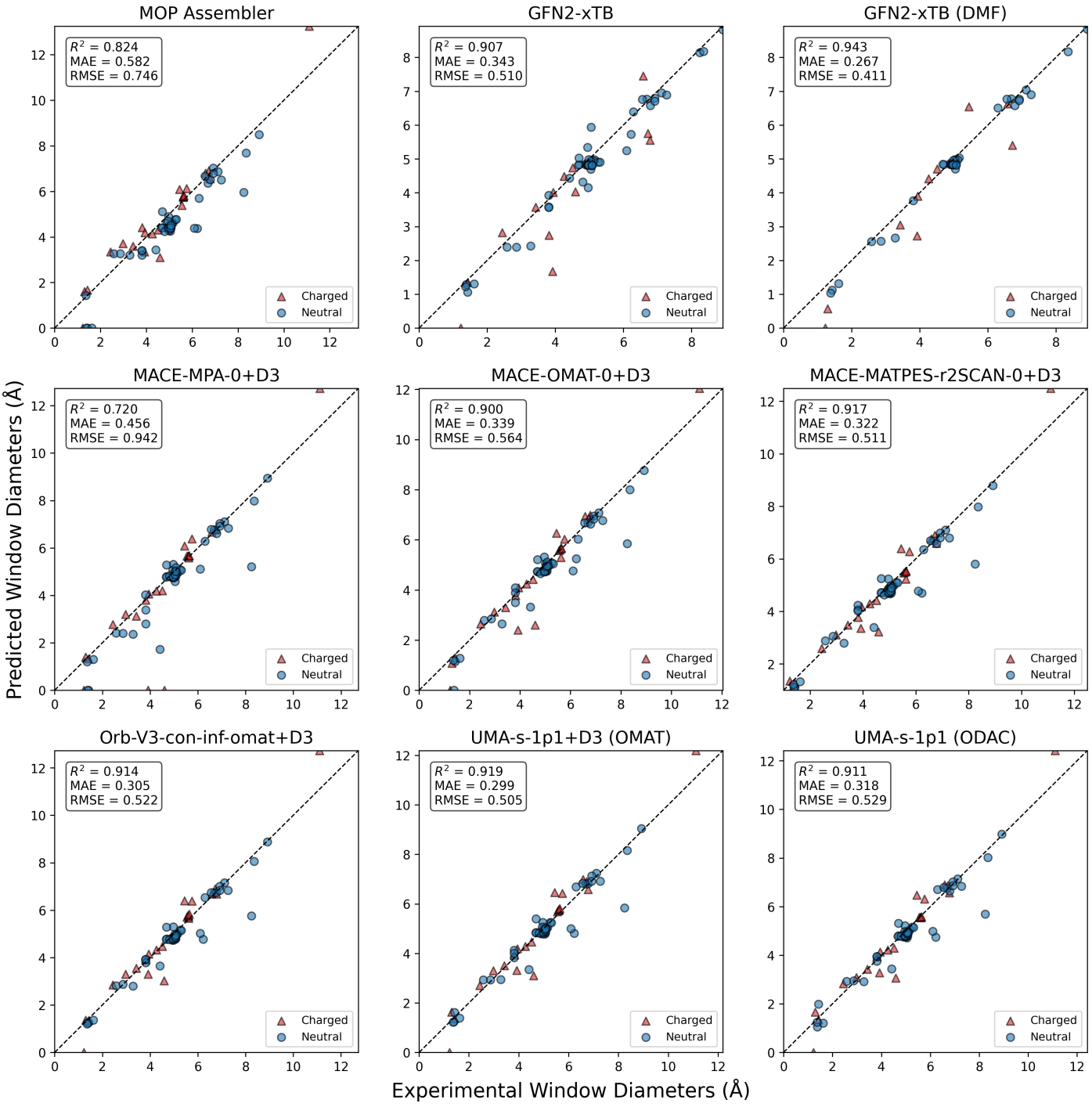

Fig. S.3 Parity plots of the experimental and predicted average window diameters (calculated by pywindow) when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

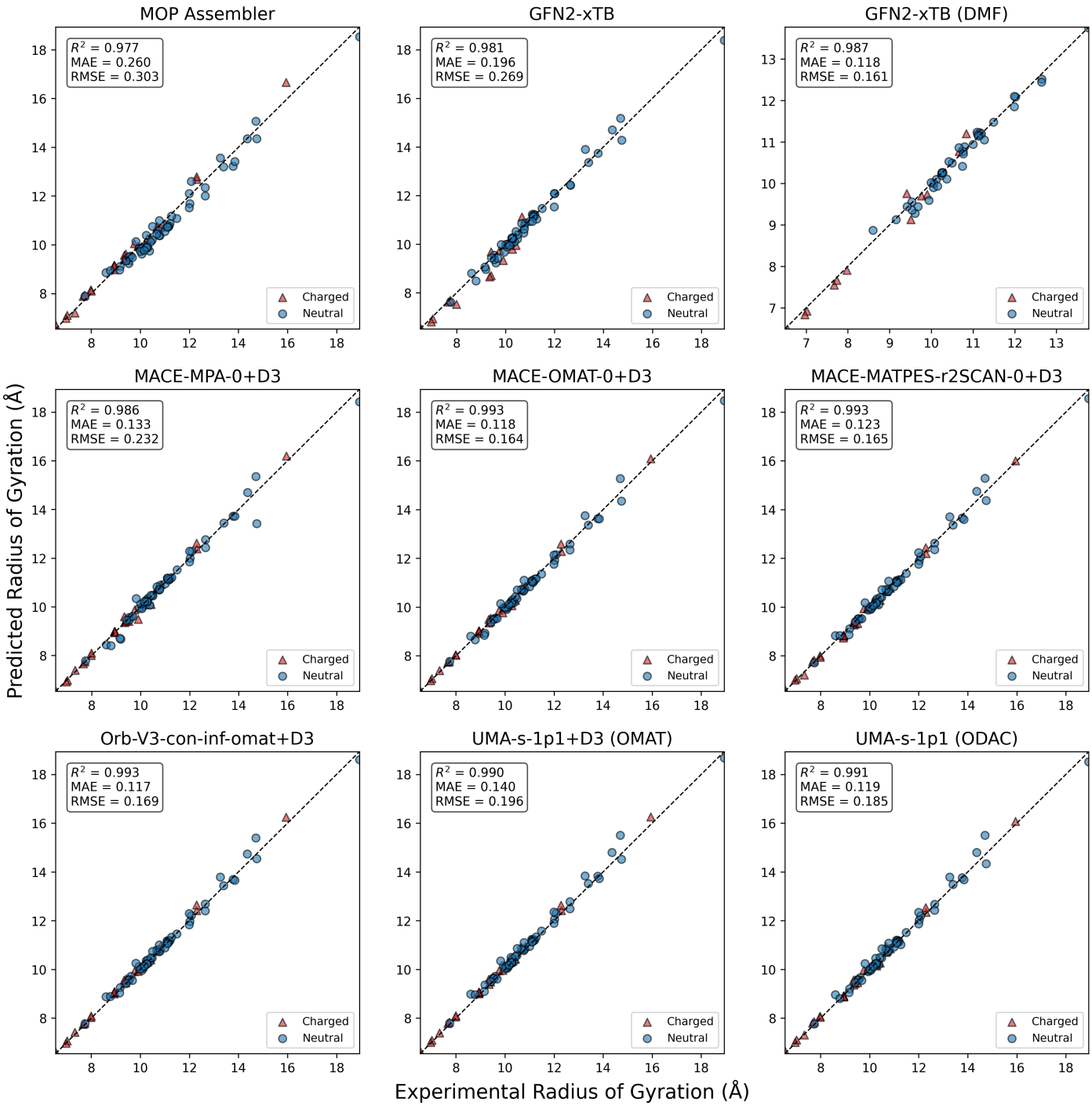

Fig. S.4 Parity plots of the experimental and predicted average radii of gyration when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

S.1.3 Statistics Summary

Table S.1 Summary of the absolute errors and correlation statistics for each of the methods and properties studied.

	Inner Radius (Å)			Window Diameter (Å)			Radius of Gyration (Å)		
Method	MAE	RMSE	R^2	MAE	RMSE	R^2	MAE	RMSE	R^2
MOP Assembler	0.287	0.435	0.940	0.582	0.746	0.824	0.260	0.303	0.977
GFN2-xTB	0.313	0.554	0.860	0.343	0.510	0.907	0.196	0.269	0.981
GFN2-xTB (DMF)	0.227	0.417	0.917	0.267	0.411	0.943	0.118	0.161	0.987
MACE-MPA-0+D3	0.323	0.715	0.841	0.456	0.942	0.720	0.133	0.232	0.986
MACE-OMAT-0+D3	0.250	0.396	0.951	0.339	0.564	0.900	0.118	0.164	0.993
MACE-MATPES-r2SCAN-0+D3	0.221	0.356	0.960	0.322	0.511	0.917	0.123	0.165	0.993
Orb-V3-con-inf-omat+D3	0.272	0.408	0.948	0.305	0.522	0.914	0.117	0.169	0.993
UMA-s-1p1+D3 (OMAT)	0.271	0.402	0.949	0.299	0.505	0.919	0.140	0.196	0.990
UMA-s-1p1 (ODAC)	0.253	0.378	0.955	0.318	0.529	0.911	0.119	0.185	0.991

S.1.4 Error Correlation with CBU Size

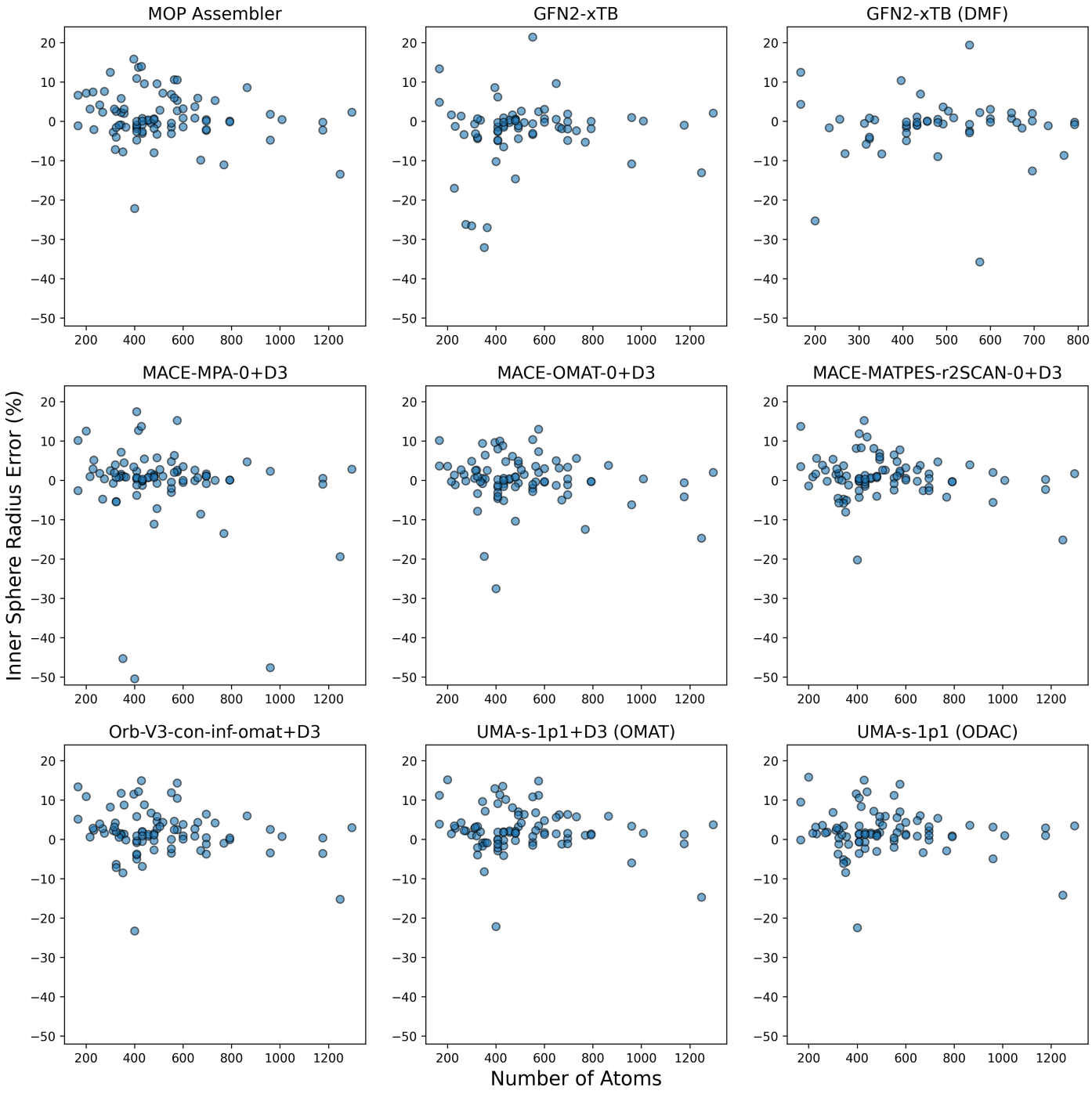

Fig. S.5 Plots of the number of organic CBU atoms against the percent error in maximum inner sphere radii when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

S.1.5 Error Distributions for Charged and Uncharged MOPs

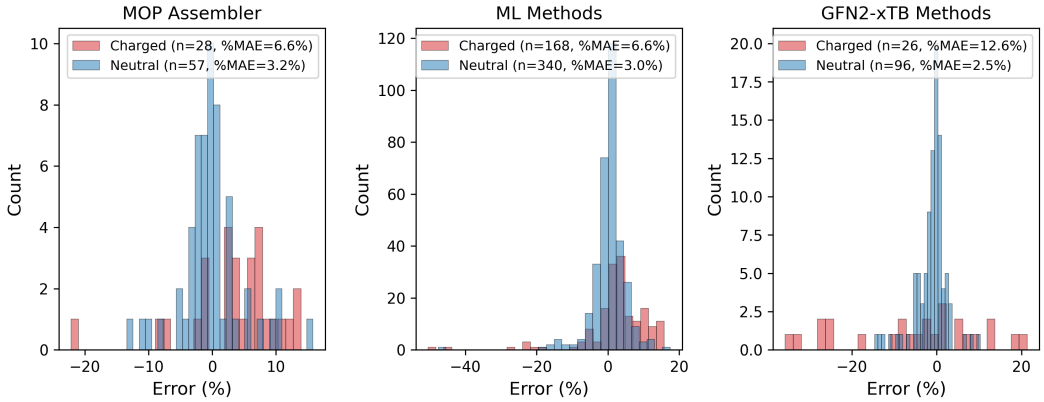

Fig. S.6 Distributions of the percentage error in maximum inner sphere radii when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

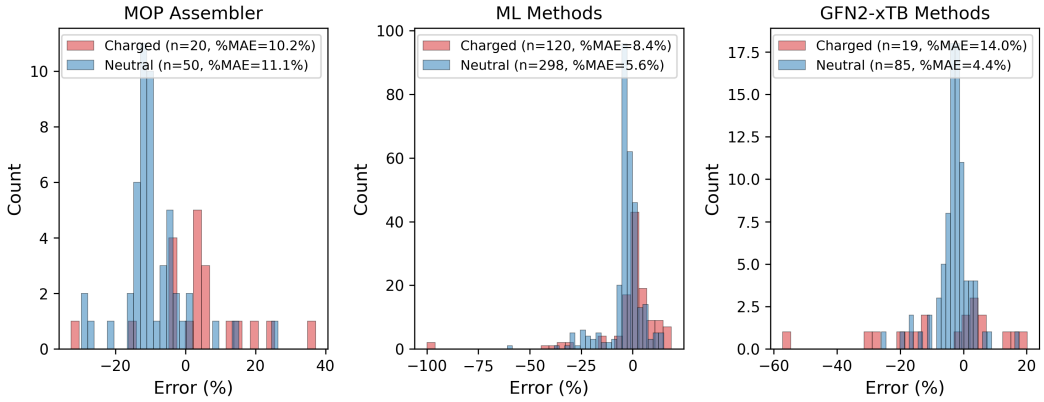

Fig. S.7 Distributions of the percentage error in window diameters when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

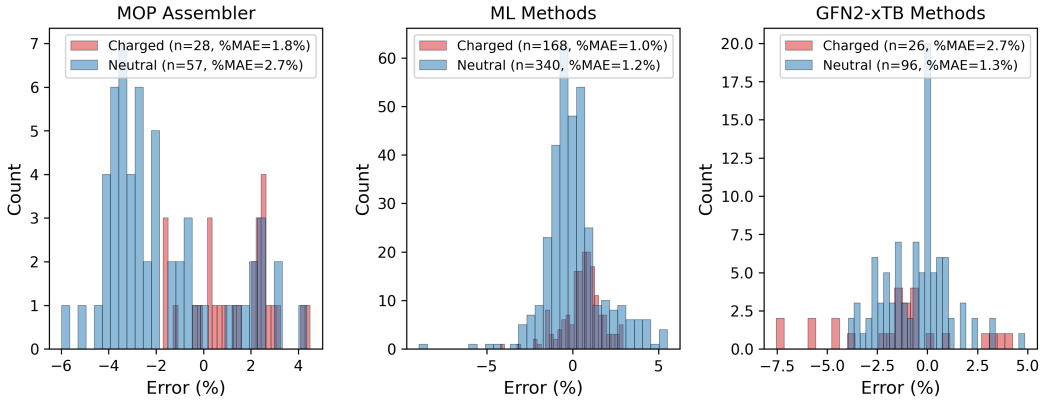

Fig. S.8 Distributions of the percentage error in radii of gyration when comparing the MOP Assembler structure optimised with the corresponding method against the experimental values. Structures with unconverged or failed optimisations have been omitted.

S.1.6 Comparison to Optimised Experimental MOP Structures

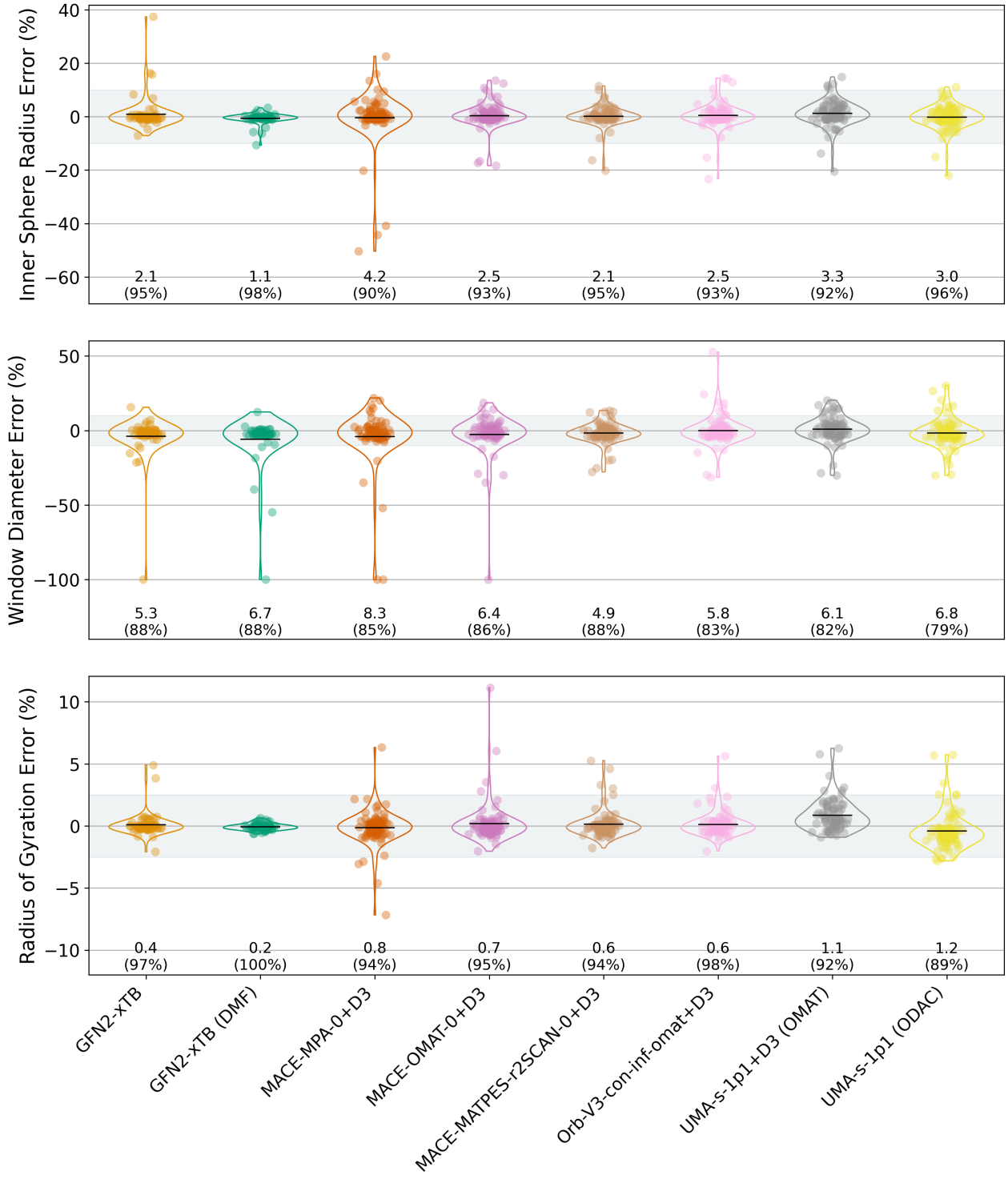

Fig. S.9 Violin plots for each method illustrating the percent error in terms of the inner sphere radius (top), average window diameter (middle), and mass-weighted radius of gyration (bottom) when comparing the MOP Assembler structure optimised with the corresponding method against the experimental structures optimised with the same methods. The distributions include up to 85 structures with unconverged or failed optimisations omitted. The mean of the distributions are indicated by the black bars with the percent MAEs underneath each distribution. The proportion of data points within the shaded region is displayed in brackets below the MAE.

S.2 Host-Guest Optimisations

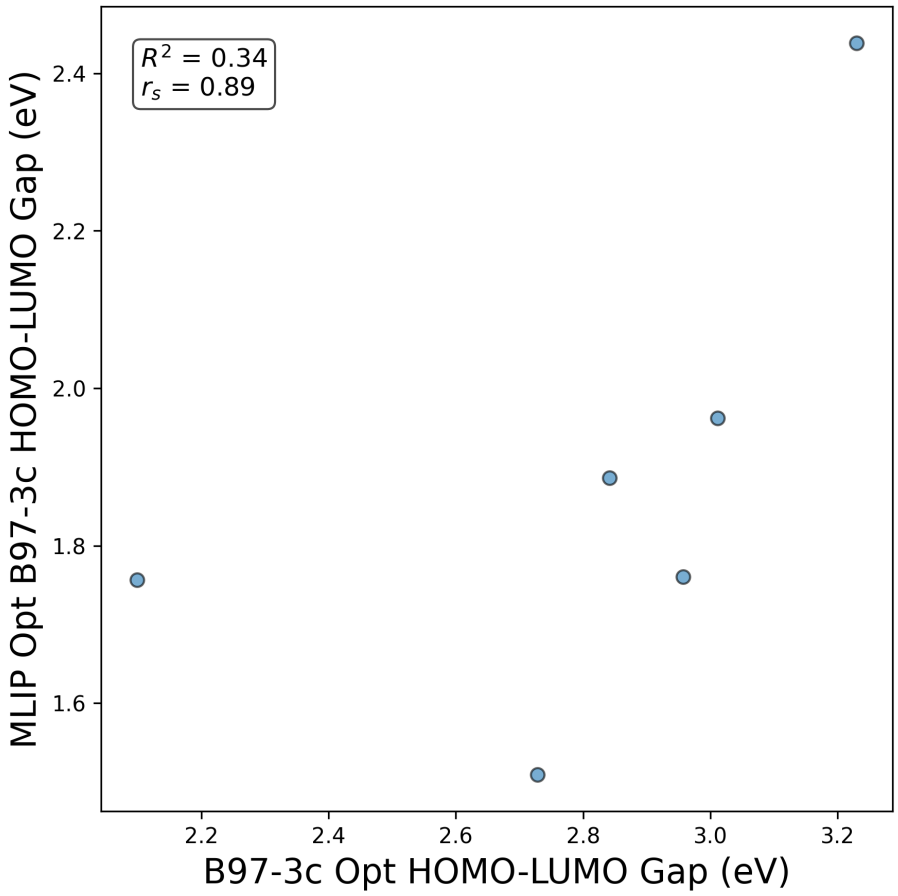

Fig. S.10 Correlation of HOMO-LUMO gaps calculated at the DFT level using B97-3c for at set of 6 Zr MOPs optimised at the B97-3c level against those claculated for the same MOPs optimised by MACE-MATPES-r2SCAN-0+D3. The correlation coefficient (R^2) and Spearman rank correlation coefficient (r_s) are shown.

S.3 Host-Guest Optimisations

The calculation of interaction energies was performed by translating the centre of mass of the guest molecule (i.e. urea) to the centroid of the host MOP cavity. The orientation of the guest, defined by the angles to each axis, was then optimised according to single point GFN2-xTB energies combined with the BFGS method as implemented in the scipy package (gtol $= 110^{-3}$, $\varepsilon = 110^{-2}$). Complexes that had final configurations exhibiting atom overlaps between the host and guest below 90% of the sum of the corresponding van der Waals radii were not carried forward. Complexes with no overlap were subsequently relaxed to a force convergence of 0.05 eV Å^{-1} using the FIRE algorithm and GFN2-xTB. The interaction energy was then calculated as $E_{\text{int}} = E_{\text{complex}} - (E_{\text{host}} + E_{\text{guest}})$, where the host and guest energies were determined by single point GFN2-xTB calculations of the host and guest fragments extracted from the optimised complex geometry without further relaxation. While this is an approximate ranking and a more accurate ranking could be achieved by optimising a variety of valid guest orientations and selecting the lowest, this approximation is likely less severe for the highest ranked hosts in this demonstration, which have restricted cavities and thereby only a few valid orientations.

S.4 Example MOPs

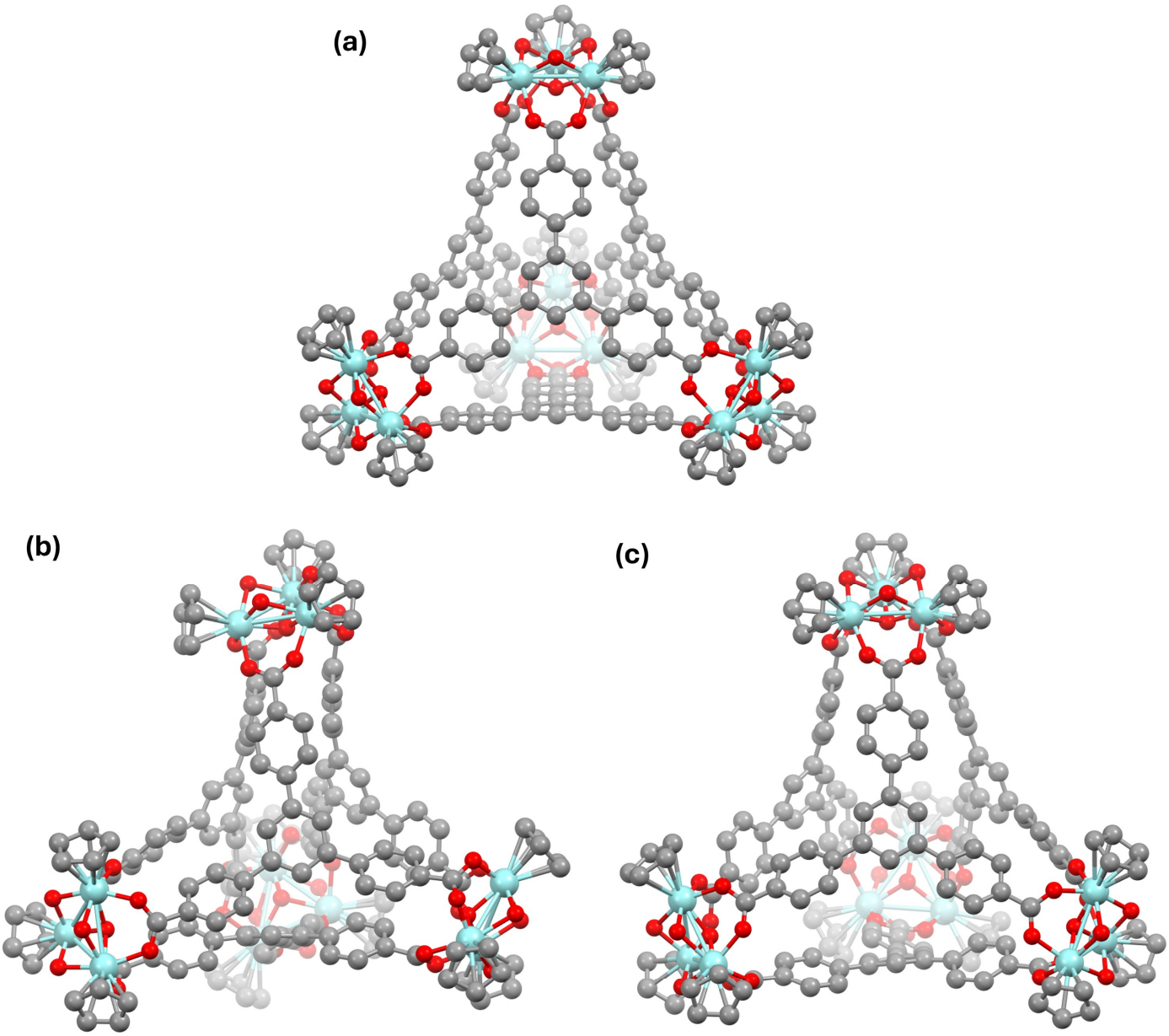

Fig. S.11 Structures of the MOP from CSD refcode XIYJUM assembled by the MOP Assembler (a) and following optimisation with GFN2-xTB (b) and GFN2-xTB(DMF) (c). Hydrogen atoms have been omitted for clarity.

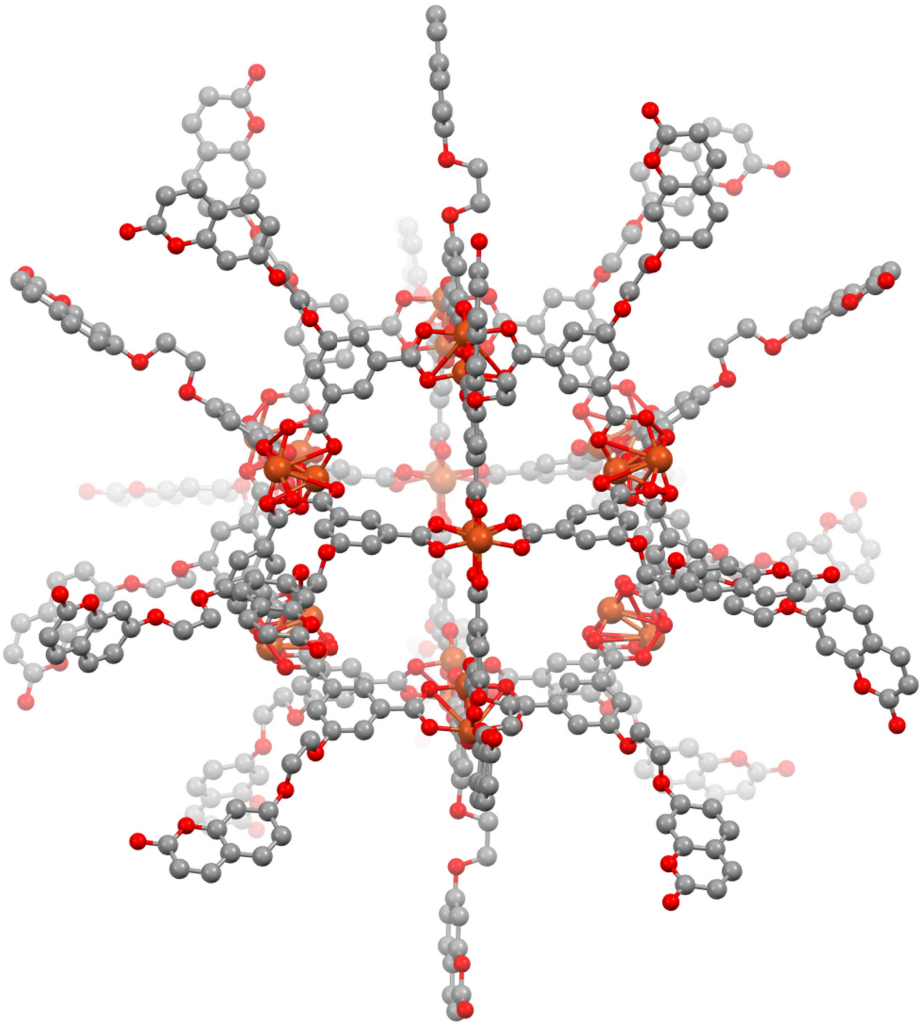

Fig. S.12 MOP Assembler structure of the MOP from CSD refcode GECQEN, which was unsuccessfully geometry optimised with 2 methods (GFN2-xTB and MACE-MPA-0+D3) and took more than 2500 optimisation steps with three other methods.

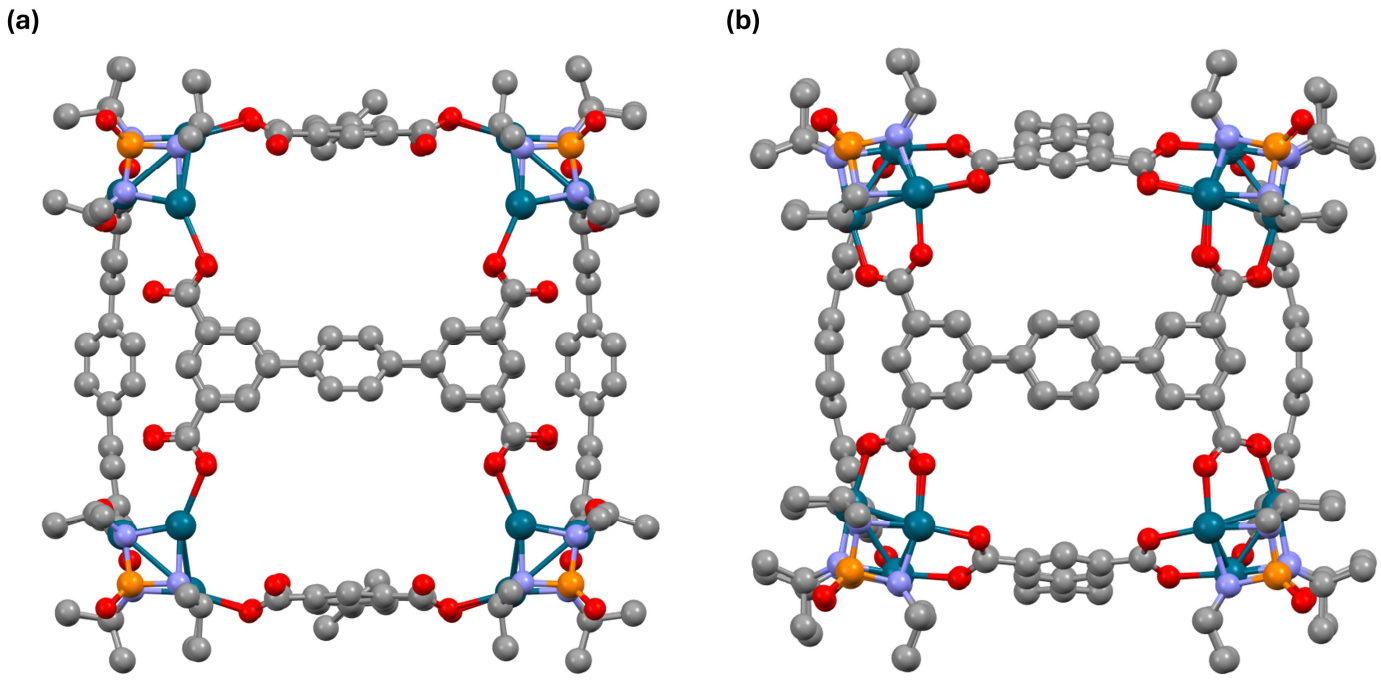

Fig. S.13 Example MOP structure with the largest change in radius of gyration before (a) and after (b) geometry optimisation with MACE-MATPES-r2SCAN-0+D3.

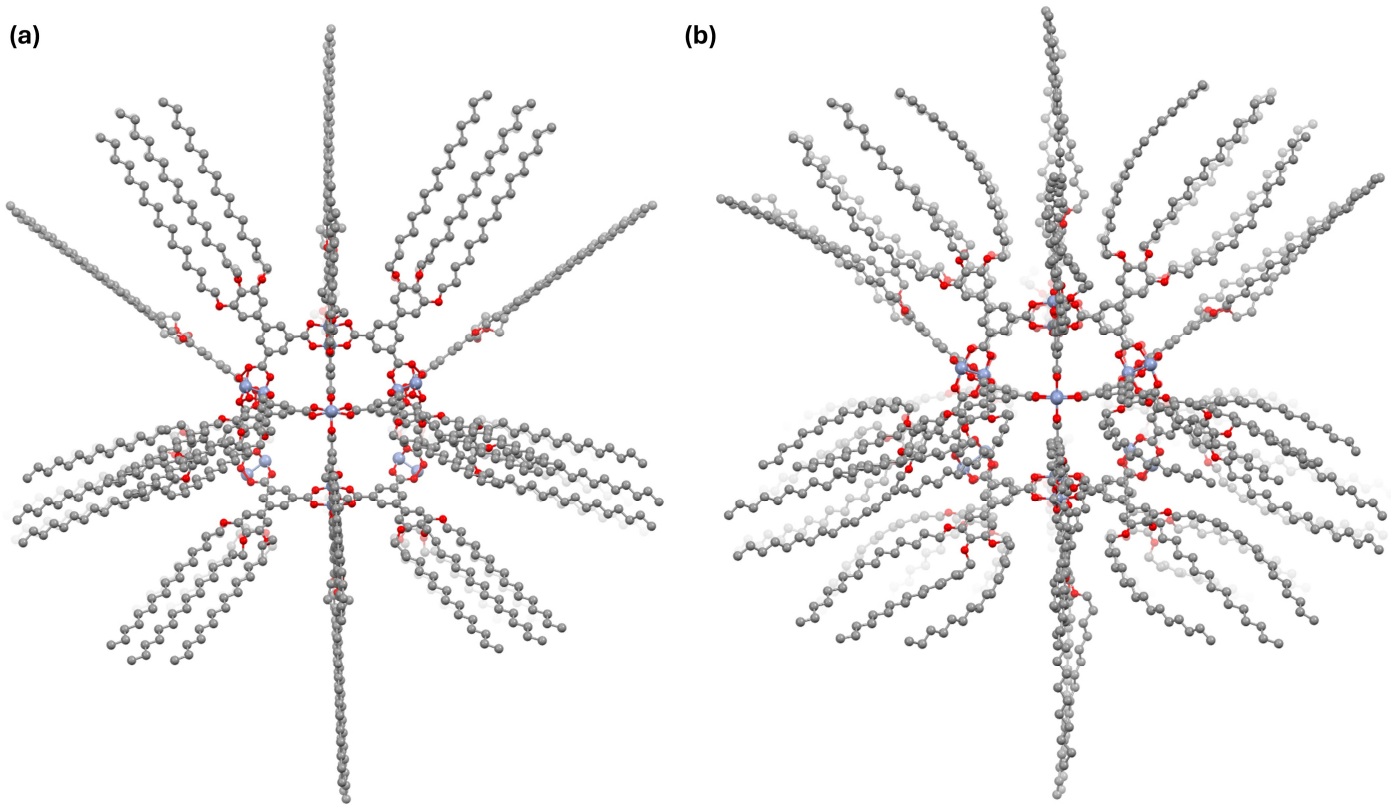

Fig. S.14 Example MOP structure with the largest RMSD in atomic positions before (a) and after (b) geometry optimisation with MACE-MATPES-r2SCAN-0+D3.