Supplementary Information

Distributions of HALMO-EDA Components by GC Content

The different HALMO-EDA components and GC contents of the 4bp dsDNA structures in this study are depicted as histograms for comparing how the energies vary within the 25%, 50%, and 75% GC-content categories as distributions. The 0% and 100% GC-content categories are not depicted due to their low sample size of only 10 sequences. Since ONETEP EDA does not support implicit solvation, only HALMO EDA in vacuum is available.

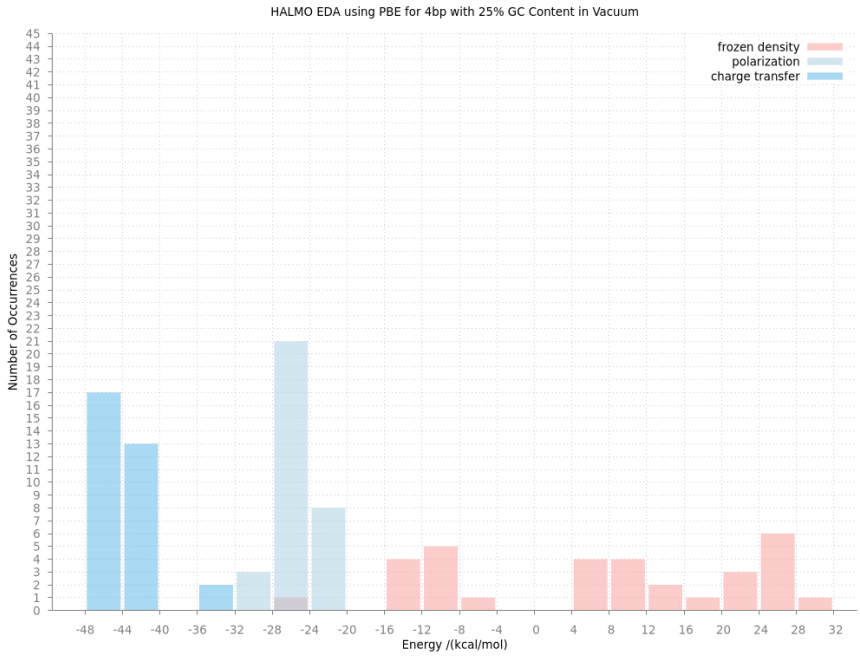

Fig. 1 HALMO-EDA frozen density (red), polarization (light blue), and charge transfer (blue) using PBE for 4bp dsDNA with 25% GC content in vacuum.

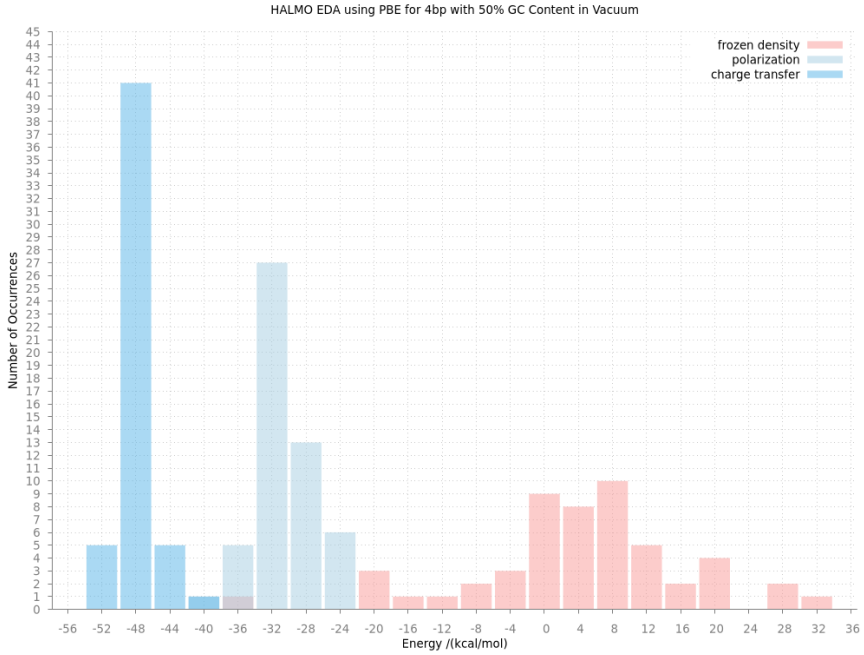

Fig. 2 HALMO-EDA frozen density (red), polarization (light blue), and charge transfer (blue) using PBE for 4bp dsDNA with 50% GC content in vacuum.

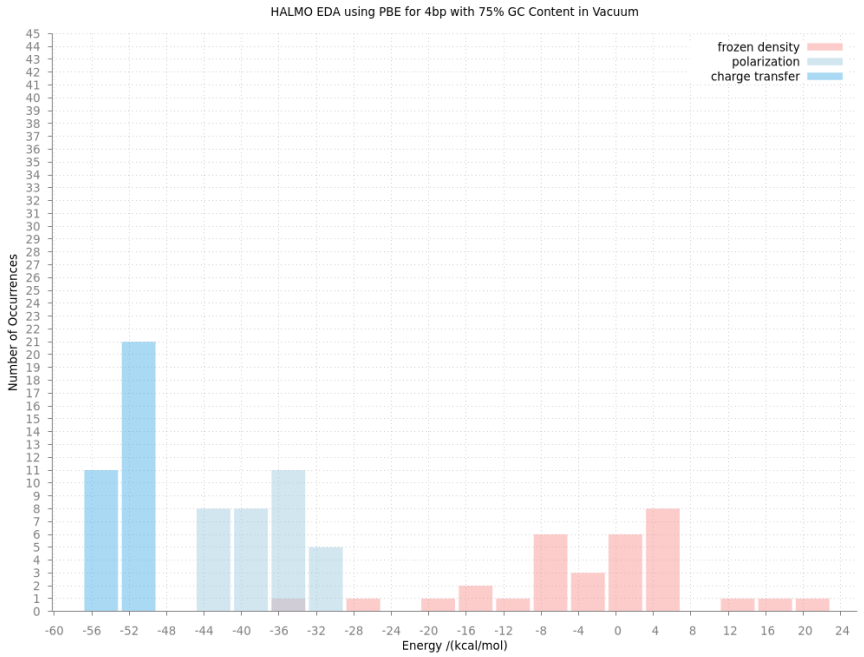

Fig. 3 HALMO-EDA frozen density (red), polarization (light blue), and charge transfer (blue) using PBE for 4bp dsDNA with 75% GC content in vacuum.

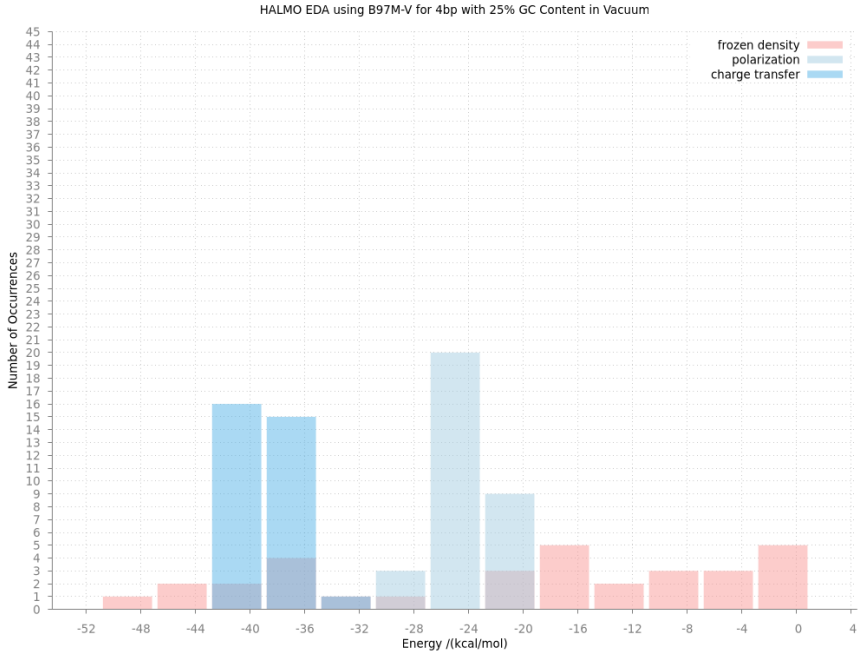

Fig. 4 HALMO-EDA frozen density (red), polarization (light blue), and charge transfer (blue) using B97M-V for 4bp dsDNA with 25% GC content in vacuum.

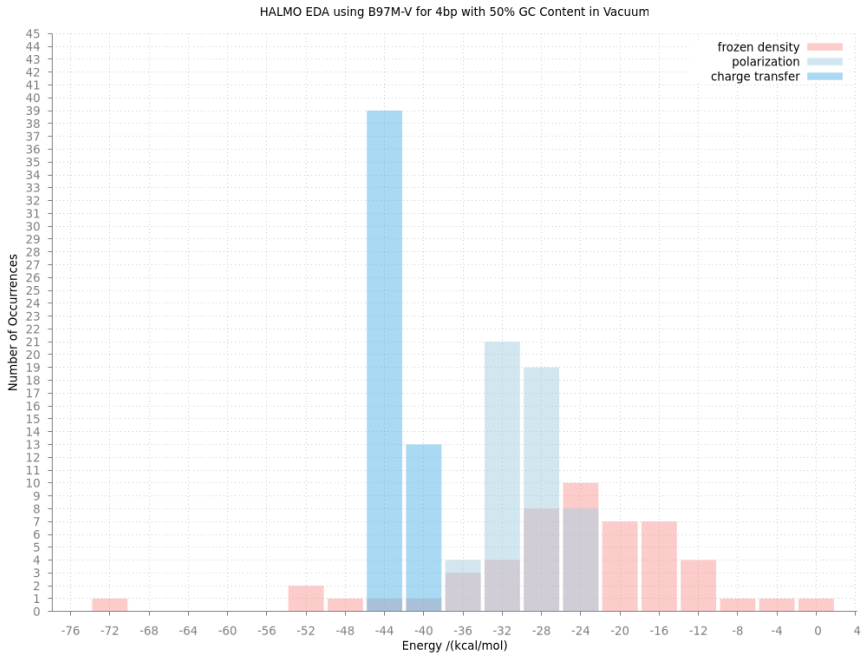

Fig. 5 HALMO-EDA frozen density (red), polarization (light blue), and charge transfer (blue) using B97M-V for 4bp dsDNA with 50% GC content in vacuum.

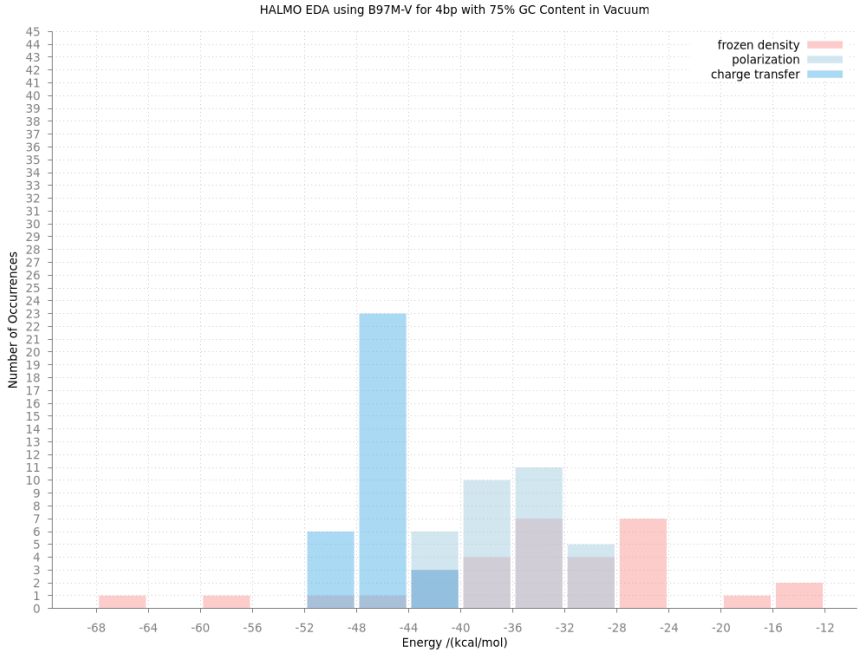

Fig. 6 HALMO-EDA frozen density (red), polarization (light blue), and charge transfer (blue) using B97M-V for 4bp dsDNA with 75% GC content in vacuum.

Comparisons with VdW2017

VdW2017 (updated from vdW2016) is a van der Waals (vdW) potential parameterized against SAPT[1] and can be used to calculate vdW energies between the two strands for each 4bp dsDNA (Table 1). Prior to calculating such energies, each atom was assigned an atom class according to the vdW2017 parameter set. The functional form used by vdW2017 is Buffered 14-7 (Buf-14-7)[2], with $\delta = 0.201303$ and $\gamma = 0.068272$. Combination rules for atomic radii and well depths are, respectively, cubic mean and Waldman-Hagler (W-H) rule.

Table 1 VdW energies for 4bp dsDNA in vacuum using HALMO EDA, BSC1 force field, and vdW2017 force field categorized by GC content. In HALMO EDA, vdW is defined either as the empirical dispersion correction (for PBE) or as the correlation component (for B97M-V). In BSC1, there is a dedicated component for vdW. For vdW2017, each atom is assigned an atom class, and then the vdW energy between the two DNA strands is calculated for each dsDNA structure. All energy values are in kcal/mol. Abbreviations: MAD, median absolute deviation

GC Content	PBE		B97M-V		BSC1		VdW2017	
	Median	MAD	Median	MAD	Median	MAD	Median	MAD
0%	-22.5	1.1	-9.2	0.4	-17.8	0.8	48.9	0.7
25%	-24.1	0.8	-21.1	12.4	-17.4	0.9	58.2	1.2
50%	-24.9	1.4	-10.1	0.6	-17.2	1.3	68.1	0.7
75%	-25.9	1.2	-10.4	0.5	-16.9	1.1	77.6	0.6
100%	-27.0	1.3	-11.0	0.7	-17.2	0.9	87.7	0.6

Notes and references

1 R. Qi, Q. Wang and P. Ren, *Bioorganic & Medicinal Chemistry*, 2016, **24**, 4911–4919.

2 T. Halgren, *Journal of the American Chemical Society*, 1992, **114**, 7827–7843.