Supplementary Information for Machine Learning Solvation Environments in Conductive Polymers.

Ioan B. Magdău* and Thomas F. Miller III*

Division of Chemistry & Chemical Engineering, California Institute of Technology,
Pasadena, California 91125, United States

E-mail: imagdau@caltech.edu; tfm@caltech.edu

Construction of Visual Molecular Representations (VMRs)

Visual molecular representations are helpful for interpreting the SE clusters and are obtained as follows. Each atom is first assigned to a molecular fragment (ProDOT-backbone, Hex-sidechain or PC). Within each fragment, every atom receives a unique index based on its symmetry. A list of closest $N_{nn} = 12$ nearest neighbours is built for every coordinate snapshot belonging to a given SE. Within this list, the most frequently occurring atoms are identified based on their symmetry index and fragment membership, and are labeled anchor atoms. The snapshots containing all anchor atoms within N_{nn} are kept, while the others are discarded. Neighbours that are not anchor are discarded, while atoms which are not immediate neighbors but belong to the anchor fragments are added. Finally, all collected coordinates are aligned with respect to the anchor atoms using the RMSD Trajectory Tool in VMD.[1] In PC liquid, the molecules cannot be aligned because local structure is not preserved, so a single snapshot is kept for illustration. Finally, a density map is build from the transformed Li^+ coordinates using VMD VolMap Tool. This approach, combines a sufficient number of SE-snapshots to construct a faithful visual molecular representation characteristic for each environment.

Binding Energy Trajectories for the Neat Crystalline Phase

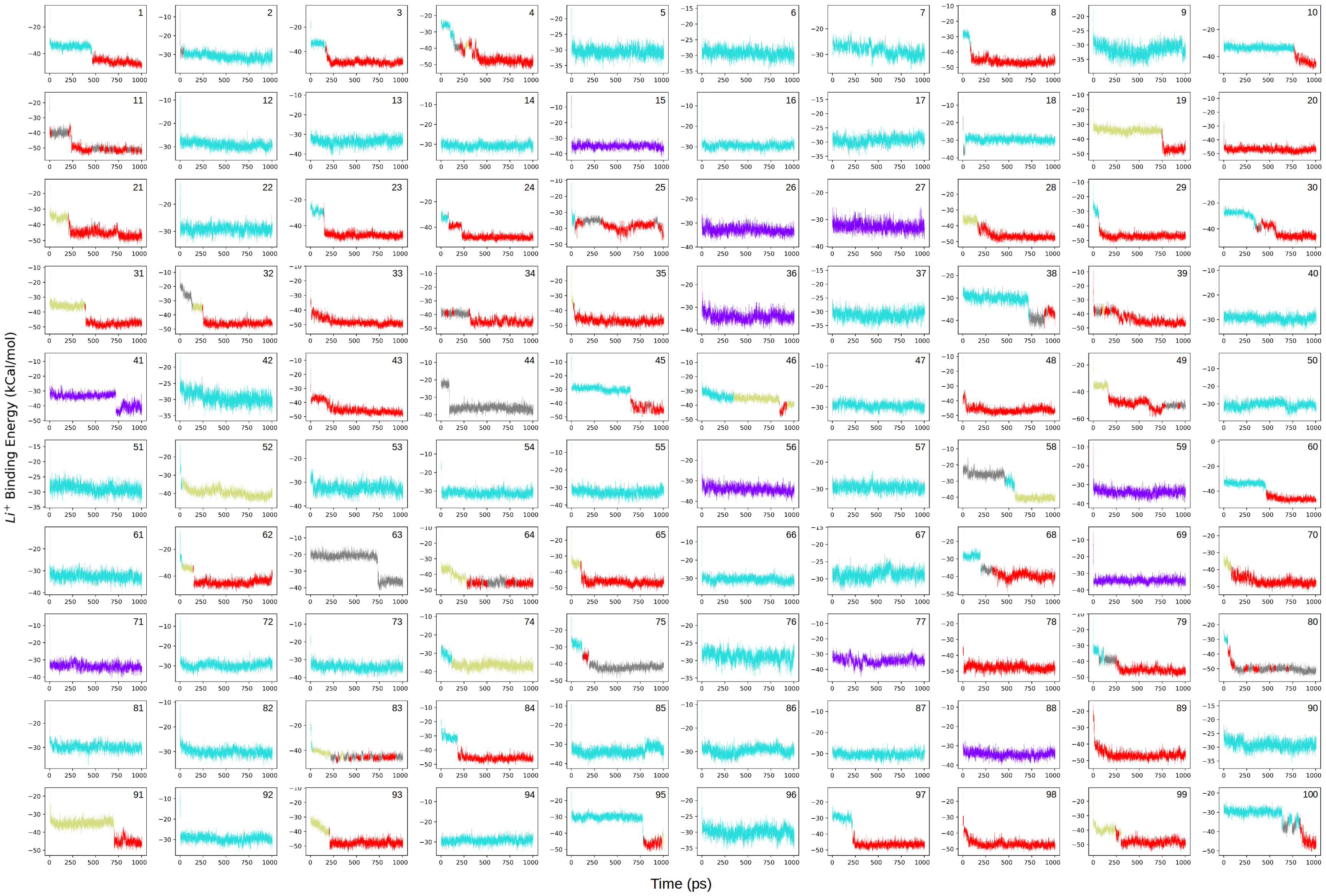

Figure 1: Li^+ binding energy as function of time for all trajectories ($N_{traj} = 100$) in neat crystalline ProDOT-2Hex. Colors are automatically assigned by SEC. Ttrajectories 1, 19, 58, 62 and 69 are reproduced in the main text.

Raw SEC Results for the Swollen Polymer

Provided here is the automatic clustering and labeling of SEs obtained through SEC for both the crystalline and the amorphous polymers. We show the latent space clustering in Figure 2, the Li^+ binding energy histograms in Figure 3 and the CDFs and VMRs in Figures 4 (for the crystalline phase) and 5 (for the amorphous phase). Here, the cluster colors and labels are automatically assigned by SEC for each swelling condition and are not consistent across different conditions. For the crystalline phase we manually relabeled the clusters based on CDFs and VMRs (Figure 4) to achieve consistency across swelling conditions (these results are presented in the main text).

References

(1) Humphrey, W.; Dalke, A.; Schulten, K., et al. VMD: visual molecular dynamics. *Journal of molecular graphics* **1996**, *14*, 33–38.

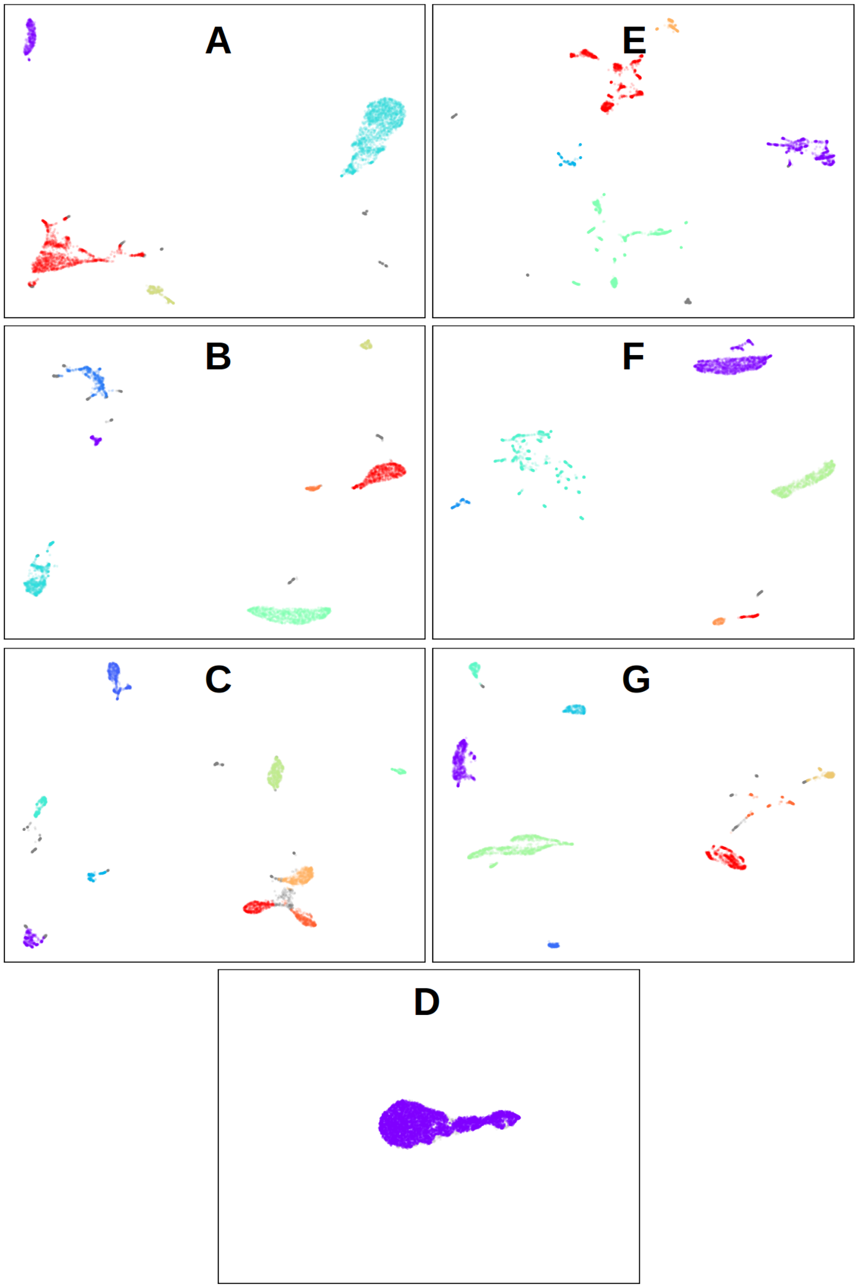

Figure 2: Latent space classification of SEs for neat and swollen polymers in both crystalline (A, B, C) and amorphous phases (E, F, G). Panels A, E: neat ProDOT-2Hex, Panels B, F: swollen ProDOT-2Hex with 17% PC, Panels C, G: swollen ProDOT-2Hex with 33% PC, and Panel D: pure PC liquid solvent.

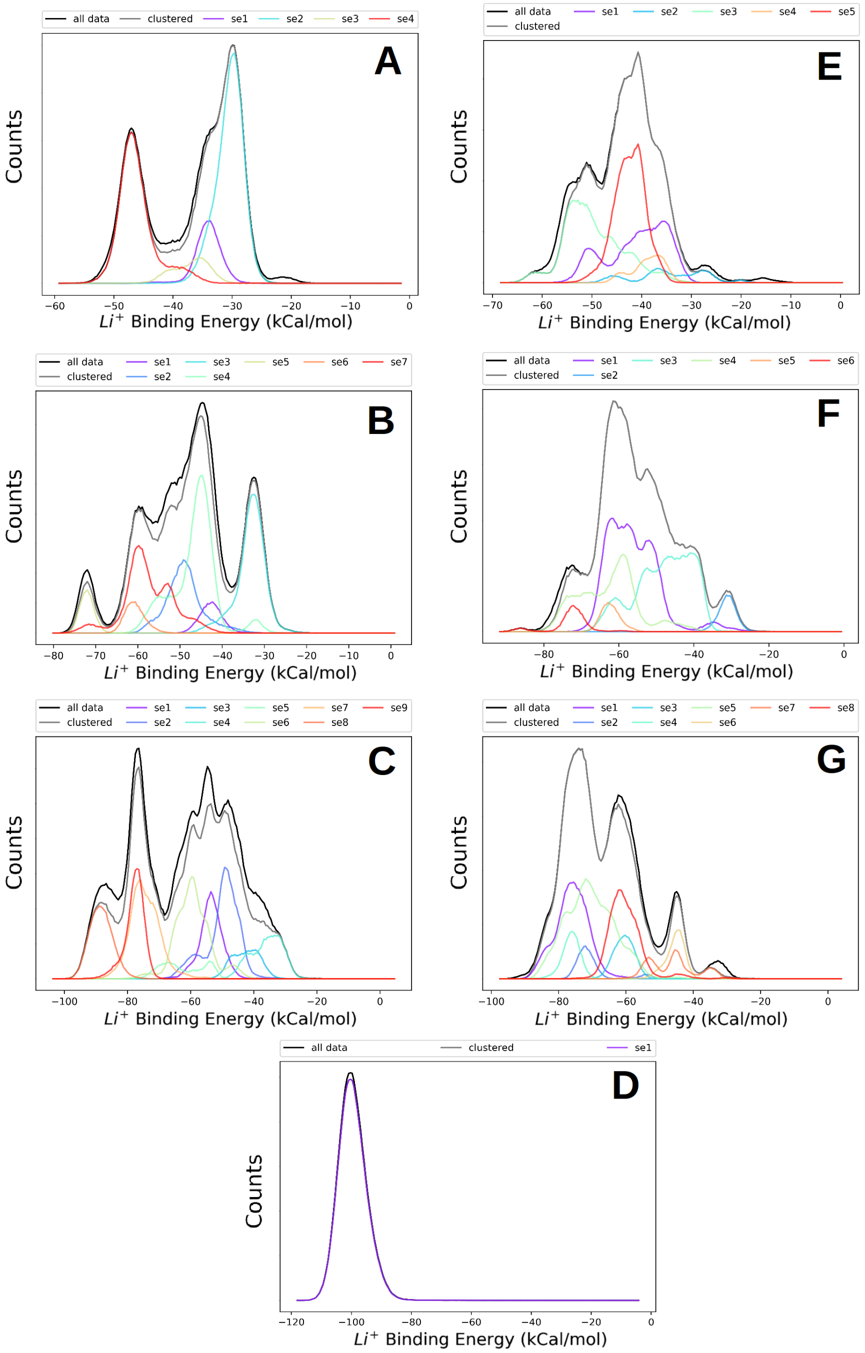

Figure 3: Li$^+$ binding energy per SE (color), overall (black) and for all successfully clustered data (gray). Mirroring Figure 2, Panels A, B, C show the distributions in the crystalline polymer (neat and swollen), Panels E, F, G in the amorphous polymer (neat and swollen), and Panel D in pure PC liquid solvent.

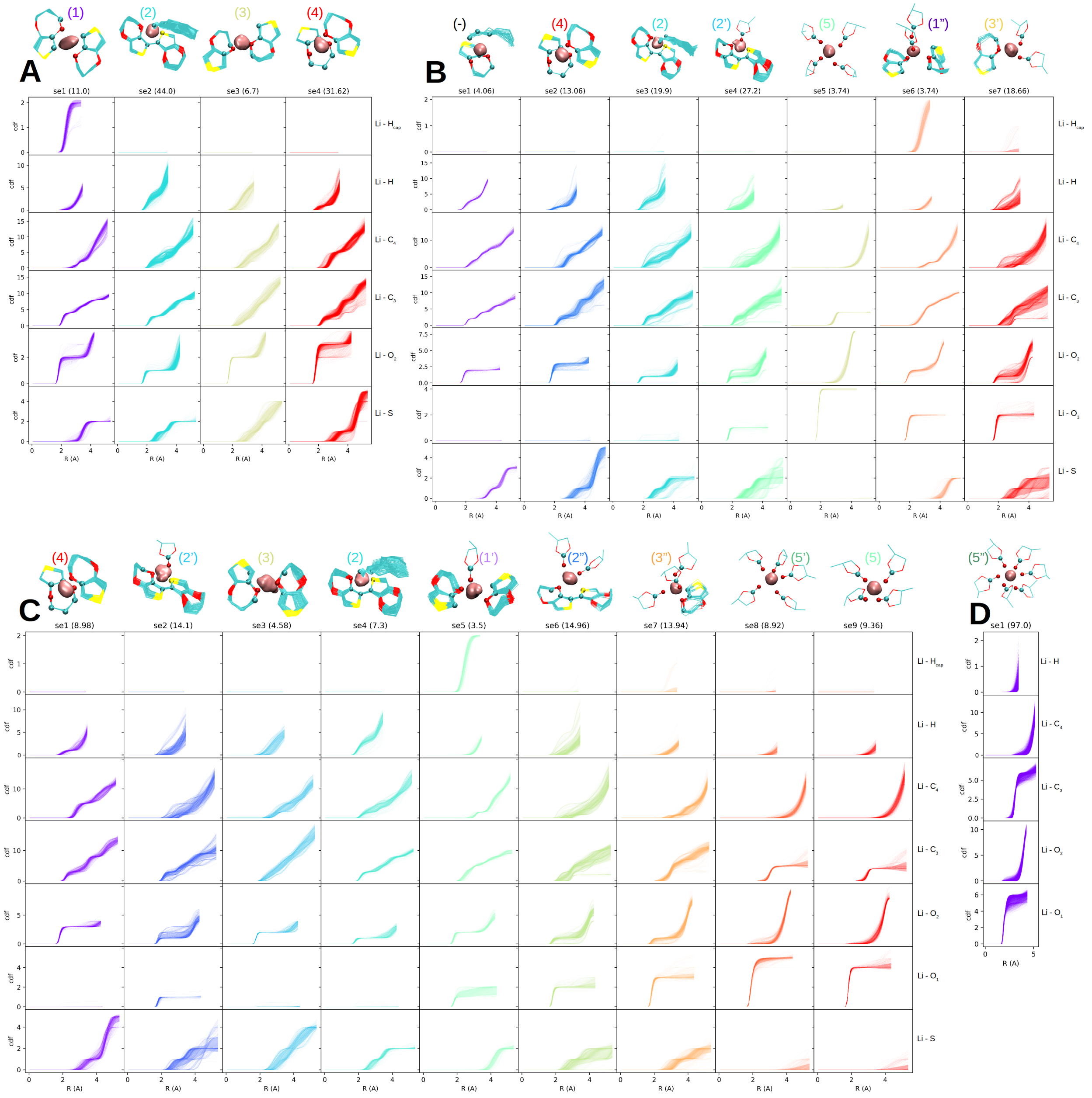

Figure 4: CDFs and VMRs for every SE in crystalline ProDOT-2Hex (A, B, C) and pure PC liquid solvent (D). Colors and panel labels mirror Figures 2 and 3. Above each VMR, in round brackets, is the relabeling of SEs for consistency across swelling conditions, as reported in the main text.

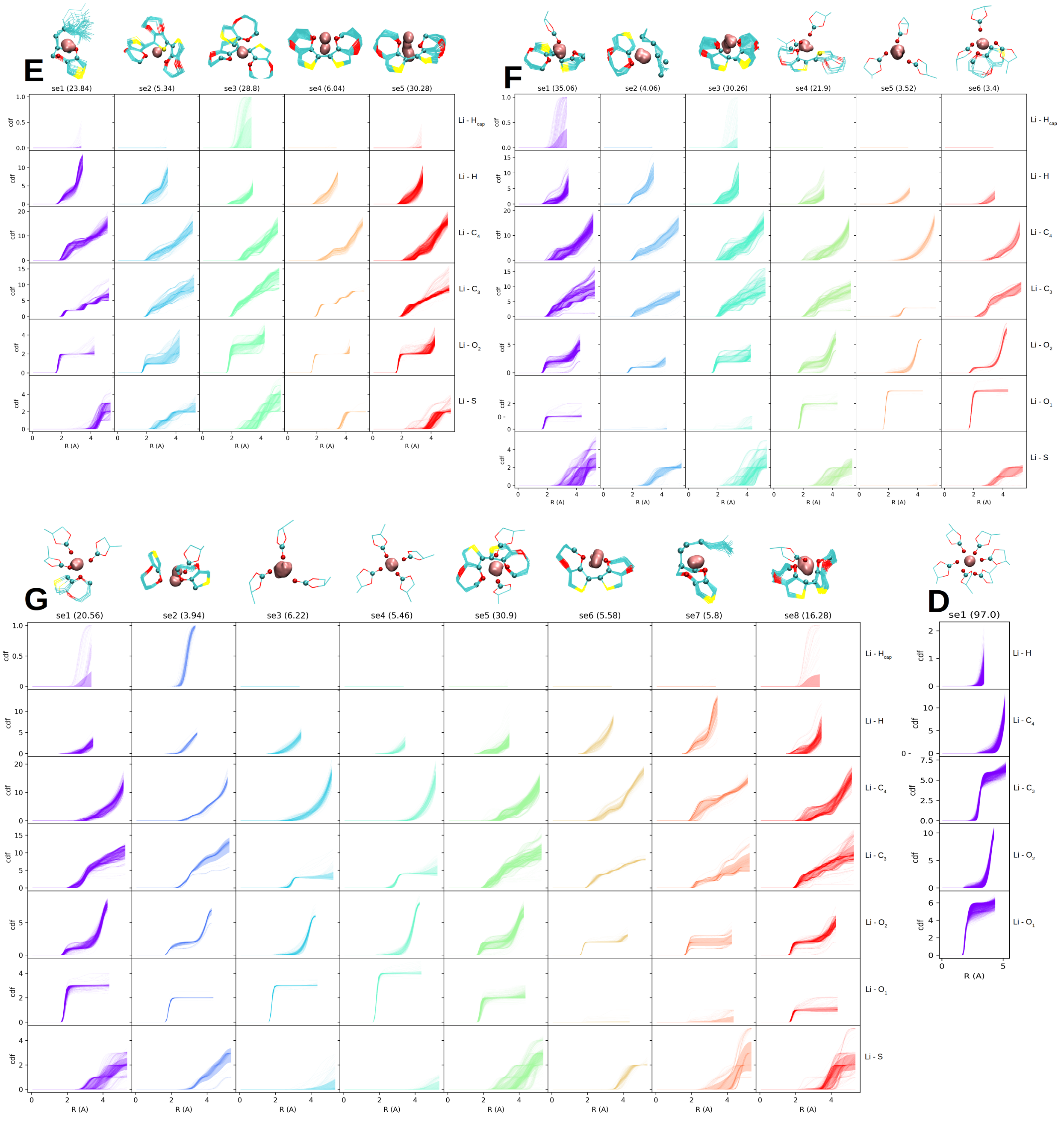

Figure 5: CDFs and VMRs for every SE in amorphous ProDOT-2Hex (E, F, G) and pure PC liquid solvent (D). Colors and panel labels mirror Figures 2 and 3.