The use of sunlight as an energy replacement for fossil fuels has become more common as the efficiency of these materials has improved. The community is optimistic that organic photovoltaic materials will provide the reduction of up-front cost in obtaining solar energy, which will allow for more widespread use of solar energy. Inverse design of molecular structures to meet particular properties is almost impossible since $\sim 10^{60}$ stable molecules are predicted to exist. We aim to use computational screening to efficiently find highly promising target polymers. A multi-step approach is taken to rapidly screen monomers to eliminate unlikely candidates and then employ more accurate methods as the screening progresses.

Our previous work, revealed hundreds of candidates with >8% estimated efficiency and dozens with >10% estimated efficiency with an initial set of 131 monomers and three possible sequences (ADAD, ADDA, and DAAD). Here we expand the number of monomers and sequences studied. The initial populations of candidate polymers were designated by combining two randomly selected monomers in a randomly chosen sequence. Candidates were mutated offering a 3 in 4 chance of monomer replacement with one of seven monomers most similar to it. Duplicate children of current population were discarded. Energy conversion efficiency was calculated based on the HOMO and lowest energy optical transition as computed by the INDO/S method. In each generation, some polymers were created by "crossover" where monomers were swapped between candidates. The new resulting combinations were generated with a randomly chosen sequence. After successive generations, top monomers emerge.

The original 131 monomer data set was run for 100 generations, at which point convergence had occurred. To efficiently expand the GA to larger studies, the number of generations to top monomer convergence was examined using the Spearman Rank Correlation. The results show that the GA performs in a quadratic manner for data sets up to ~1000 monomers at which point the rate at which the monomers are screened compared with an exhaustive search is reduced. We therefore suggest that data sets be broken into groups of ~1000 monomers in a tournament style screening approach.