PCCP

Using Genetic Algorithms to Discover Novel Ground-State Triplet Conjugated Polymers

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-01-2023-000185.R1
Article Type:	Paper
Date Submitted by the Author:	17-Mar-2023
Complete List of Authors:	Abarbanel, Omri D; University of Pittsburgh, Chemistry Hutchison, Geoffrey; University of Pittsburgh, Department of Chemistry

Journal Name

ARTICLE TYPE

Cite this: DOI: 00.0000/xxxxxxxxxx

Using Genetic Algorithms to Discover Novel Ground-State Triplet Conjugated Polymers†

Abarbanel, Omri D.[a], and Hutchison, Geoffrey R.[a,b]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

Stable ground-state triplet π-conjugated copolymers have many interesting electronic and optoelectronic properties. However, the large number of potential monomer combinations makes it impractical to synthesize or even just use density functional theory (DFT) to calculate their triplet ground-state stability. Here, we present a genetic algorithm implementation that uses the semi-empirical GFN2-xTB to find ground-state triplet polymer candidates. We find more than 1400 polymer candidates with a triplet ground-state stability of up to 4 eV versus the singlet. Additionally, we explore the properties of the monomers of those candidates in order to understand the design rules which promote the formation of a stable ground-state triplet in π-conjugated polymers.

1 Introduction

Although organic π-conjugated polymers have been researched for their unique electronic properties and potential uses[1,2], a new subclass of π-conjugated organic polymers with a triplet ground-state has recently been introduced. Research into the discovery of these ground state triplet organic π-conjugated polymers has increased in the last few years, with works on molecular design and characterization. While many of these efforts are still ongoing, they all follow similar design rules, copolymers composed of electron accepting and electron donating monomer pairs[3–7].

By choosing the right combination of monomers a polymer with a small energy gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO), i.e. the HOMO-LUMO gap, can be achieved. This leads to degenerate or near-degenerate frontier molecular orbitals and allows for the unpairing of the normally paired HOMO atoms into two singly-occupied molecular orbital (SOMO) with a biradical nature[7,8].

Previous computational work has shown that the size of the HOMO-LUMO gap of singlet species directly correlates with the stability of the triplet species. That is, polymers with a small HOMO-LUMO gap also have a more stable triplet ground state. Additionally, the identity of the acceptor monomer was found to have a high correlation with the stability of the triplet ground state, where polymers that share the same acceptor monomer will have similar electronic properties[9].

Experimentally, it is highly impractical to create and characterize every monomer combination. Even computationally, this can require many resources in order to comb through a large number of monomer pairings. This calls for a more efficient method that can sort and eliminate unwanted monomers while preserving those that show promising results.

There are many approaches that can be used to accelerate the discovery of new materials. Machine learning (ML) is one method that is becoming increasingly popular in this field[10–12]. However, it requires a large data set, which does not exist in this case. Another method is the genetic algorithm[13] (GA) which is a non-deterministic optimization algorithm. The GA is an iterative method, where each generation a new set of offspring are created from the previous generation surviving parents, and those who survive the fit test go on to be the parents in the following generation. The algorithm also includes the possibility of a random mutation that can result in an increase or decrease in the survival rate.

In this work, we used a genetic algorithm to discover new ground-state triplet polymers. Our data set consists of 1226 monomers, which, if thoroughly combined with each other, would create over 1.5 million (1226^2) potential polymers to work with. This comprehensive method can be associated with high computational resources and time. The iterative GA method can efficiently sift through the large number of combinations and produce stable ground-state triplet candidates with high confidence. Moreover, multiple GAs can run in parallel, increasing the verity of potential candidates.

Additionally, from the GA we can also produce a list of the most common monomers that were used in each GA run. A highly common monomer means that it survived natural selection and passed to the next generation. Also, by promoting a low HOMO-LUMO gap, a monomer has a higher chance of creating offsprings.

[a] Department of Chemistry, University of Pittsburgh, 219 Parkman Avenue, Pittsburgh, Pennsylvania 15260, United States; Email: geoffh@pitt.edu
[b] Department of Chemical and Petroleum Engineering, University of Pittsburgh, 3700 O'Hara Street, Pittsburgh, Pennsylvania 15261, United States

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 00.0000/00000000.

From these common monomers we can gain insights into which monomer properties correlate with a stable triplet ground-state in the full polymer.

2 Methods

2.1 Correlation Between the Singlet HOMO-LUMO Gap and Stability of the Triplet State

In a previous study, we found a strong linear correlation between the HOMO-LUMO gap of the oligomer singlet state and the stability of its triplet[9]. That is, oligomers with a low HOMO-LUMO gap lead to a stable open-shell electronic structure due to the frontier molecular orbitals being closer energetically. Oligomers with a triplet ground-state tend to have a biradical electronic structure, in which each electron is found in a separate singly-occupied molecular orbital (SOMO).

Using these findings halves the number of potential calculations that needs to be performed to find if an oligomer has an open-shell electronic structure, as only the HOMO-LUMO gap of the singlet species is needed. This completely negates the necessity of calculating the electronic energy of the triplet state, and drastically accelerates the discovery of open-shell π-conjugated materials.

Furthermore, the oligomers in the aforementioned study were constructed from pairs of electron donor and electron acceptor monomers. However, we discovered that the stability of the triplet ground-state is based primarily on the identity of the acceptor monomer. By expanding the list of electron-accepting monomers, we can find better oligomers with a biradical nature.

2.2 Correlation between GFN2-xTB HOMO-LUMO gap and ωB97X HOMO-LUMO gap

In addition to the correlation between the singlet HOMO-LUMO gap and the stability of the triplet state, we also previously found a correlation between the singlet HOMO-LUMO gap calculated using density functional theory (DFT) and the HOMO-LUMO gap calculated using the semi-empirical GFN2-xTB method. While the correlation is imperfect, there is a clear trend. To strengthen the correlation, we calculated the HOMO-LUMO gap using both GFN2-xTB and ωB97X-D3 of randomly generated list of new oligomers from the expanded list of monomers used in this study (Figure S1). See Section 2.3.2 for how these calculations were performed.

Although both the logarithmic and radical functions can describe the correlation, the radical function slightly better (R^2 = 0.83) than the logarithmic (R^2 = 0.74), both show that a small HOMO-LUMO gap calculated with GFN2-xTB would correlate with a small HOMO-LUMO gap calculated using DFT. As GFN2-xTB is a semi-empirical method, it is much faster than DFT and can greatly accelerate the GA. We therefore use GFN2-xTB–calculated HOMO-LUMO gap in the GA, and find oligomers that minimize the HOMO-LUMO gap with every generation.

2.3 Computational Methods

2.3.1 The Genetic Algorithm

Genetic algorithms follows Charles Darwin's *Survival of the Fittest* idea, which describes how evolution works[14]. Parents with certain traits create offspring that end up with some combination of those traits. Offsprings with combinations of traits that help them survive in their environment can produce their own offsprings. This cycle can continue indefinitely or terminate by some external force. At certain points during this process a new, never seen before, trait has a chance to appear due to a random mutation. These mutations can either have no effect, help, or hinder the survival and reproduction chances of a off-spring.

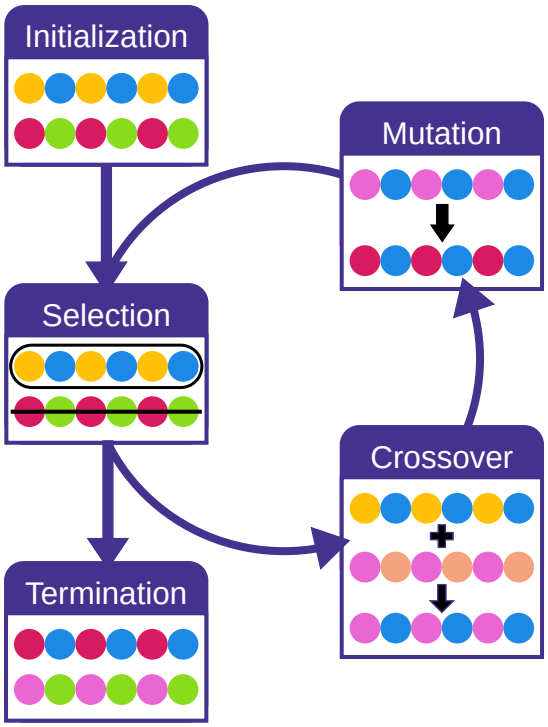

Fig. 1 The five steps of the genetic algorithm - initialization, selection, crossover, mutation, and termination. The selection-crossover-mutation cycle is repeated a set number of times before stopping at the termination step.

Our genetic algorithm follows several steps (Figure 1):

1. Initialization — which creates a starting population for the algorithm to work with. These are the "parents" in the GA.

2. Selection — which selects some of the population to survive and continue to reproduce, while it eliminates the others, based on some fitness function. This is the survival rate of the population.

3. Crossover — which creates a new population from a random combination of two parents from the population that survived the previous step. Here, successful parents create new candidates with a combination of their traits.

4. Mutation — in which, given some mutation rate, some of the current population has its traits changed to a randomly chosen one.

5. Termination — After repeating steps 2-4 for some number of cycles, the GA ends with the last surviving population.

In this study, the GA was initialized with a population size of 32 oligomers. Each oligomer was composed of a pair of monomers, repeated four times in an alternate fashion (i.e. ABABABAB, where A and B are the first and second monomers in the oligomer, respectively). During the *Selection* step, we use the GFN2-xTB–calculated HOMO-LUMO gap as the fitness function and eliminate half of the population, that is, 16 oligomers, with the largest HOMO-LUMO gap. In the *Crossover* step we create 16 new offsprings by randomly choosing two monomers from the surviving population, which brings the total population size back to 32. During the *Mutation* step, every oligomer has a 40% chance to have one of its monomers replaced by a random monomer from the entire list of possible monomers. We have used the same hyperparameters for the GA from a previous study done in our group as they have shown to be effective for similar molecular systems[15,16]. The GA terminates the *Selection–Crossover–Mutation* cycle after 40 generations, which we have found to be sufficient in finding the minimal calculated HOMO-LUMO gap (Figure 2). However, since random chance is an integral part of the GA, a single run of the GA can miss many potential oligomers with a low HOMO-LUMO gap—unless the GA is left to run indefinitely, which is an impossible task. To save run time and increase the chance of finding oligomers with a low HOMO-LUMO gap, we ran the GA ten times in parallel.

2.3.2 Geometry Optimization and Single Point Calculations

The GA was implemented using the Python programming language, version 3.8[17] using custom code which can be found on GitHub at https://github.com/hutchisonlab/oligomer-ga. We ran the GA on a list of 1226 different organic monomers[15,18,19]. The full list of monomers can be found in the **Supplementary Information**. Each monomer is numbered from 0 to 1225 in no particular order. The oligomers were 8 monomers long, that is, 4 monomer pairs (that is, AB-AB-AB-AB, where A and B are the different monomers in the oligomer), in order to match previous studies[7,9]. Additionally, this length was chosen because of the balance of a good approximation of the HOMO-LUMO gap of the long-chain polymer and the computational costs and time of running DFT calculations, which can take anywhere between a few days and a few weeks. Exploring the monomer sequence, i.e., in an alternating form, is beyond the scope of this study and can be the subject of future research. However, previous work in our group have studied the effect of the monomers sequence on the electronic properties of the oligomer, and can significantly tune HOMO-LUMO energetics[15,20–23].

In every step of the GA the oligomers were constructed from the SMILES strings of their respective monomers and followed by an initial force-field geometry optimization, and conformer search step with UFF[24] or MMFF94[25] using OpenBabel[26] version 3.1. A second geometry optimization step and the calculation of the HOMO-LUMO gap were done using GFN2-xTB[27] version 6.4.1. The GFN2-xTB output was parsed using a custom Python script.

The potential oligomers that were found by the GA were further analyzed underwent a third geometry optimization using the Density Functional Theory (DFT) B97-3c functional[28] followed by a single point calculation with the ωB97X-D3 functional[29,30] and the def2-SVP basis set[31] using ORCA version 4.2.0[32,33]. This process was repeated separately for both the singlet and triplet species of each oligomer. Single-point energy calculations using the dispersion-corrected CAM-B3LYP functional[34] were done in the same way. The energies and HOMO-LUMO gaps calculated by Orca were parsed using the cclib Python package[35].

The electronic energy difference between the singlet and triplet species of each oligomer or monomer is defined as

$$\Delta E_{T-S} = E_{Triplet} - E_{Singlet} \quad (1)$$

with $E_{Singlet}$ and $E_{Triplet}$ are the electronic energies of the singlet and triplet species, respectively. That is, when the ΔE_{T-S} if a certain oligomer is negative, its triplet ground-state is more stable, and vice versa.

3 Results and Discussion

3.1 The Genetic Algorithm

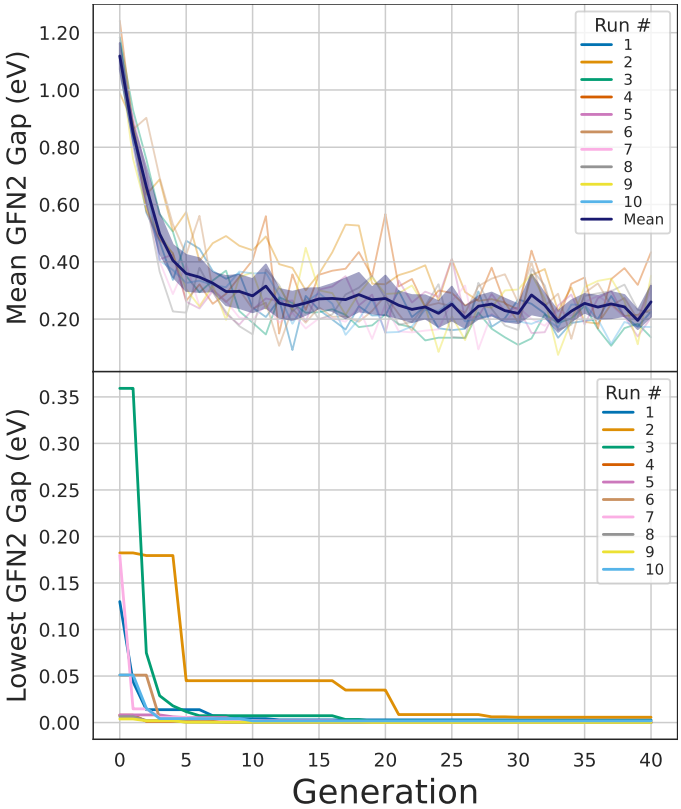

Fig. 2 Top: the mean GFN2-xTB–calculated HOMO-LUMO gap in each generation for each GA run, with the mean gap and standard deviation for each generation over all runs in dark blue. Bottom: the lowest GFN2-xTB–calculated HOMO-LUMO gap of every generation in each GA run.

As mentioned before, we ran the GA ten times in order to diversify the potential list of monomers with a small HOMO-LUMO

gap. The objective of our GA, that is, the value it was aiming to optimize, was a small HOMO-LUMO gap. As seen in Figure 2 the GA did do as intended and indeed minimized the GFN2-xTB–calculated HOMO-LUMO gap. The average HOMO-LUMO gap (Figure 2 Top) shows a significant decrease from the first few generations, while later it is subject to some randomness due to the nature of the GA. However the lowest HOMO-LUMO gap (Figure 2 Bottom) shows a clear trend where the GA does finds oligomers with a very small HOMO-LUMO gap.

From the bottom figure in Figure 2 it can be seen that after about 40 generations all the runs converged on a very low HOMO-LUMO gap, within the limitations of the GA. Although each run started with a random set of oligomers with different HOMO-LUMO gaps, they all ended with a set of oligomers that, on average, have a lower HOMO-LUMO gap and at least one oligomer with a GFN2-xTB–calculated HOMO-LUMO gap less than 0.01 eV. Based on our observed correlation between the GFN2-xTB and ωB97X-D3 gaps, these are roughly equal to a gap of 1.5 eV. 1426 copolymer candidates have been found to have a GFN2-xTB–calculated gap of less than 0.1 eV, which correspond roughly to a ωB97X-D3 gap of 2.8 eV.

3.2 Top Oligomers

The top 20 oligomers, that is, the oligomers with the smallest GFN2-xTB–calculated HOMO-LUMO gap found in any of the GA runs, were extracted for further analysis. The HOMO-LUMO singlet gap and the electronic energies of the singlet and triplet states were calculated using the ωB97X-D3 functional following the steps described in Section 2.3.2. Of the top 20, four oligomers encountered various problems during one or more of the DFT calculation steps and, therefore, were removed from further analysis. All of the top 20 oligomers share one of these two monomers, 35 or 642 (Figure S2). Both monomers share a similar molecular structure, as can be seen in Figure 3 Bottom, except that monomer 642 includes a vinyl bridge.

In Figure S3 it can be seen that the ωB97X-D3–calculated HOMO-LUMO gap and electronic energies agree with the GA and the GFN2-xTB–calculated values, as those oligomers indeed have a small singlet HOMO-LUMO gap and a stable triplet ground-state. This further shows that GFN2-xTB can be a good surrogate for DFT in finding oligomers with small HOMO-LUMO gaps with the GA. While most of the oligomers show similar electronic properties, two of them, one constructed from monomers 642 and 365 and another constructed from monomers 642 and 128, look like outliers. However, we expected to see some variation between the singlet HOMO-LUMO gap and the ΔE_{T-S} since the correlation is not perfect. On the contrary, those two outliers are the two data points closest to the best-fit line calculated in the previous study[9].

Moreover, the outlier 128-642 can be attributed to its conformation, since its lowest energy conformation found during the geometry optimization steps had the oligomer folded on itself instead of the flat linear conformation the other oligomers showed (Figure S4 (a)). We modified the conformation by manipulating the bond angles to create a more linear conformation using Avogadro2 version 1.95.1, followed by the same geometry optimization and single point calculations as described in section 2.3.2, The resulting geometry remained in the modified flat conformation (Figure S4 (b)), and the HOMO-LUMO gap of the oligomer decreased from 1.04 eV to 0.35 eV and its ΔE_{T-S} also decreased from -2.82 eV to -3.23 eV, and it got closer to the cluster of the other oligomers (Figure S3). The outlier 365-630 had the same flat and linear conformation as the other oligomers; however, its backbone includes a 7-membered ring, which breaks aromaticity and disrupts conjugation. We hypothesize that this may be the cause of this oligomer's properties. Nonetheless, we wish to reiterate that those two outliers still exhibit a stable triplet ground-state.

We would like to add that while we expect the conformation of the oligomers to be extended due to the rigidity that comes from the conjugated π-system , there is a possibility that the lowest energy conformation of an oligomer would be a nonlinear one, as seen above for oligomer 128-642. The conformation of the oligomer does affect the HOMO-LUMO gap, and we see that the HOMO-LUMO gap decreased when the oligomer conformation was intentionally extended. However, the HOMO-LUMO gap dispersion is relatively small compared to the scale of the triplet ground state stabilization[36–38].

In principal, quantum calculations for HOMO-LUMO gap and singlet-triplet energies should use a substantial conformational search, followed by a Boltzmann-weighted average of properties. In practice, given the size of the conjugated systems included, proper conformer sampling (e.g., with GFN2)[39] would significantly increase the run-time of the GA.

Additionally, the spin density plots of the top 20 oligomers (Figures 4 for an example and S6 for the rest of the oligomers) show their biradical nature by the delocalization of the two unpaired electrons. It can be seen, in some oligomers easier than others, that the oligomers show higher spin density towards the edges of the triplet ground-state oligomer. This matches with previous computational studies that showed a similar effect on a polymer that has been experimentally synthesized and characterized as having a triplet ground-state[7]. This effect occurs because of the Coulomb repulsion forces as the two unpaired electrons have the same spin in the triplet state.

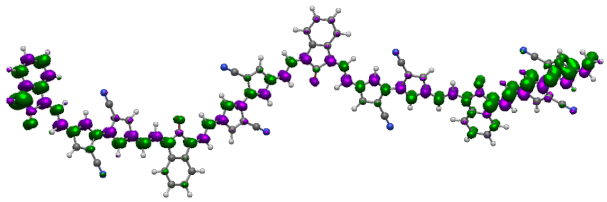

Fig. 4 Spin density plot of the oligomer constructed from monomers number 642 and 630. The purple and green orbital colors correspond to the α and β electrons, respectively. Isosurface value is 0.002 a.u.

40

To see if there are design rules that can be used to find other polymers with a triplet ground state, we have extracted the top oligomers with a GFN2-xTB–calculated HOMO-LUMO gap of

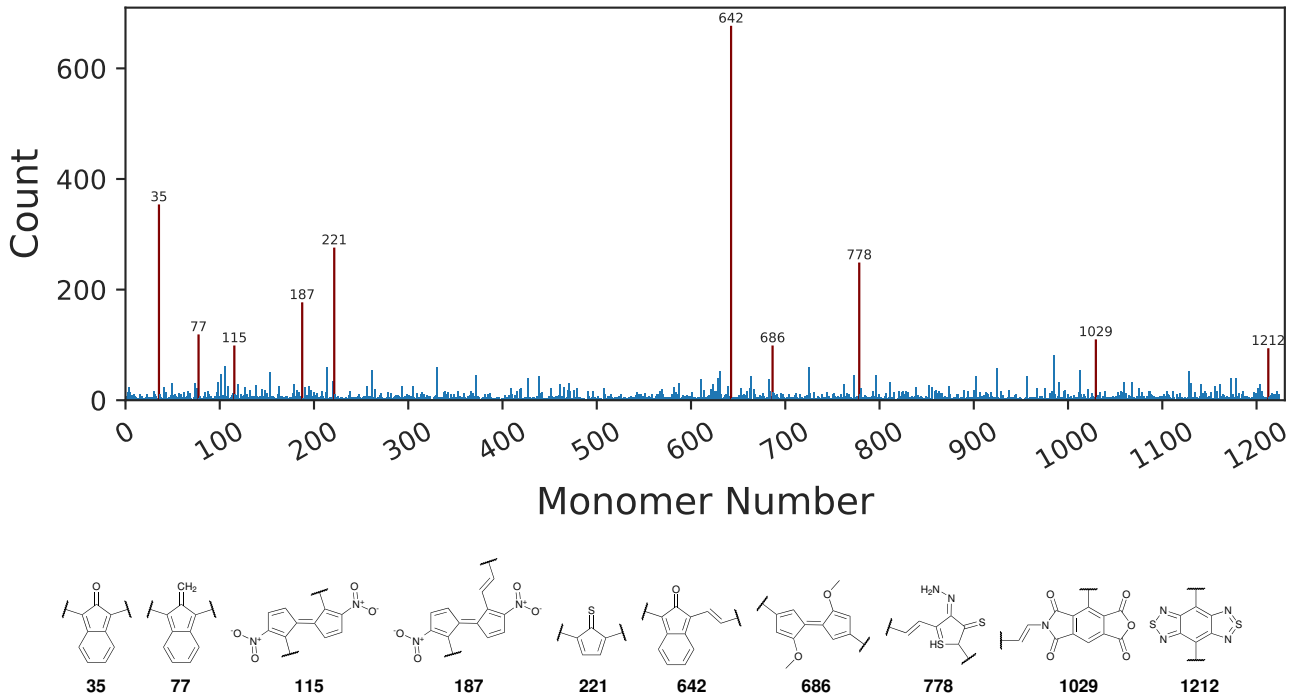

Fig. 3 Top: The number of times a monomer has been used in any of the ten GA runs. The top 10 most common monomers are boldly emphasized in red and have their monomer number above them. Bottom: The structures of the top 10 most common monomers.

less than 0.2 eV that were generated in the 10 GA runs (2024 oligomers), and compared them to all of the possible ($\sim$ 1.5 million) oligomer in our dataset. To help generalize the design rules we used RDKit to calculate various descriptors on the monomer pairs, instead of the full length oligomer . The descriptors include the molecular weight of the pair, the number of atoms, number of rotatable bonds, number of hydrogen-bond acceptors and donors, number of rings and aromatic rings, the partition coefficient (Crippen $logP$[41]), π-system size[18], and number of nitrogen, oxygen, sulfur, selenium, and halogen atoms in the pair (Figures S11 and S12). Although there is some noise due to the relatively small sample size in Figure S12, it is still possible to extract some potential design rules. For example, there are more oxygen atoms in the top monomer pairs (2.7 ± 2.1) compared to the full monomer combinations (1.7 ± 1.6), as well as a high proportion of monomer pairs with a π-system size of 12 atoms, while other comparisons can be attributed to the small sample size. However, this analysis shows that there are no generalized design rules, and that a search algorithm, like this GA, is needed in order to traverse the vase chemical space and find potential polymers with a triplet ground state.

3.3 Top Monomers

From the analysis of the top oligomers from the GA, two monomers have been shown to be ubiquitous, 35 and 642 (Figure S2). This further demonstrates that the stability of the triplet ground-state is frequently controlled by one of the two monomers in the oligomer. That is, some monomers induce a small HOMO-LUMO gap in many of the oligomers they are found in when paired with many different monomers, and we expect these monomers to be more common in the GA. To find which monomers in the set exhibit similar properties, a histogram of the number of occurrences of each monomer in all GA runs was constructed (Figure 3 Top). A higher number of instances in the GA would suggest a higher survivable rate throughout the GA cycle, due to it contributing to a small HOMO-LUMO gap relative to the rest of the population.

In fact, it appears that of the 1226 monomers in our data set, only a small subset has been captured by the GA to promote a small HOMO-LUMO gap. In Figure 3 the top 10 most common monomers are highlighted, along with their molecular structures at the bottom. At a first glance some of those monomers show a common electron-accepting motifs, such as monomers 115 and 187 with two highly electron withdrawing nitro groups, or monomer 1212 which is another common acceptor monomer used in various π-conjugated polymers[7,9].

Another common motif in the top 10 monomers is the vinyl bridge, also called a vinylene link. The inclusion of a vinyl bridge in the polymer backbone has been shown to lower the HOMO-LUMO gap by extending the conjugation of the π-system, leading to greater delocalization of πelectrons[42]. As a testament to this hypothesis, our GA found monomers with a vinyl bridge (e.g., monomers number 642 and 187) at a higher frequency than their derivatives without a vinyl bridge (monomers 35 and 115, respectively), as shown in Figure 3. Moreover, Cordaro and Wong have also commented that in their experience, in addition to drastically decreasing the HOMO-LUMO gap, a vinyl bridge also improves the solubility of polymers due to the increase in the polymer flexibility[42]. Therefore, polymers with a low HOMO-LUMO gap, and

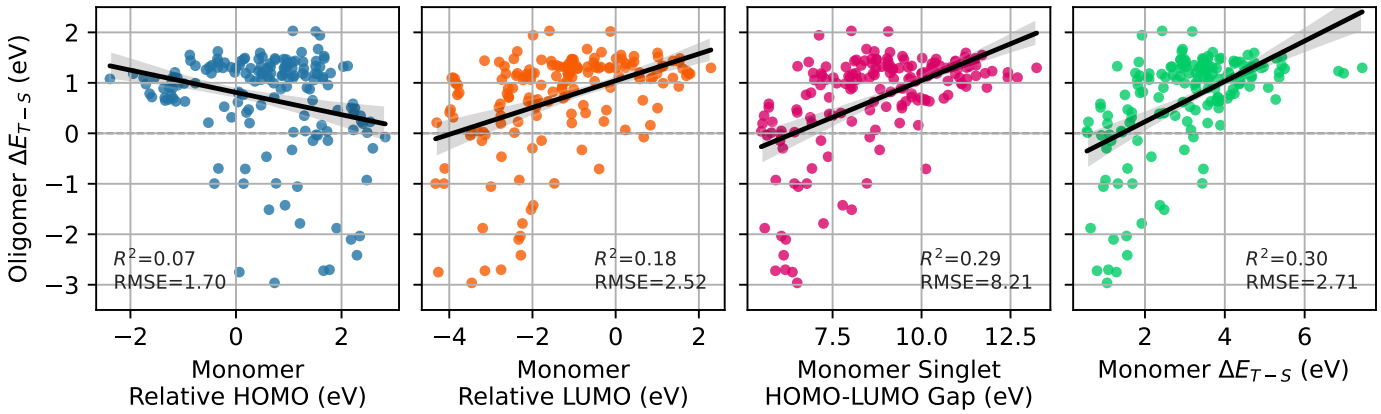

Fig. 5 Monomers' HOMO level (relative to thiophene's), LUMO level (relative to thiophene's), HOMO-LUMO gap, and their triplet ground-state stability (ΔE_{T-S}) versus the stability of the oligomer's triplet ground-state stability when paired with monomer 630. The linear best-fit line and standard deviation are shown in black line in each plot, as well as the R^2 and the RMSE (in eV).

potentially a stable triplet ground-state, will benefit from including a vinyl bridge in their backbone—both by lowering the intrinsic HOMO-LUMO gap compared to the non-bridged version and by potentially improving polymer solubility for synthesis, characterization, and application.

3.4 Monomer Properties

From looking at the list of the top 10 monomers, the first questions that should be asked are what are the electronic properties of those monomers have in common and whether, by discovering this property, we can find other monomers that share it and promote a stable open-shell electronic structure since similar monomers should have similar properties.

As mentioned before, we previously showed that the identity of the acceptor monomer has the highest correlation to a stable triplet ground-state[9]. However, the absolute classification of monomers between acceptor and donor is vague because these are relative terms. These are usually described by their HOMO levels, as a donor monomer will have a high HOMO level, while an acceptor will have a small one. However, an absolute scale is difficult to derive, as a monomer with a low HOMO level compared to its oligomer counterpart will behave as an acceptor (Figure S5). A higher difference between the HOMO levels would entail a stronger donor-acceptor pair and vice versa. f To classify the monomer into strong and weak acceptors and donors, we needed to set a relative scale because absolute HOMO eigenvalues highly depend on the DFT functional and the basis set used. Some have used thiophene as a "*spacer*" monomer in various π-conjugated polymers, as it is claimed to not affect the electronic properties significantly[43–46]. For the same reason, some have used thiophene as a reference monomer when comparing different donor and acceptor monomers. Therefore, we examine the relative HOMO and LUMO eigenvalues, as well as the HOMO-LUMO gap and the stability of their triplet state (ΔE_{T-S}) for all monomers (Figure S7). The single-point calculations using the ωB97X-D3 functional followed the same steps as the full oligomers, as described in Section 2.3.2.

For comparison and to reaffirm our results, we also performed single-point calculations on all monomers using the dispersion-corrected CAM-B3LYP functional (Figure S8). The results show similar distributions compared to the ωB97X-D3 single point calculations (Figure S7), showing that those results appear to be consistent across multiple functionals.

The top ten most common monomers in Figure 3 have small HOMO-LUMO gaps and small ΔE_{T-S}, as well as relatively low LUMO levels while their relative HOMO levels are more spread out (Table S1). However, these monomers are not all in the extreme ends of any category, and there are other monomers with small HOMO-LUMO gaps, for example, that did not show up as common in the GA as those top ten monomers. We can attribute this to several potential causes:

- Due to random chance in the GA. The GA is a stochastic optimization method, and by chance some potentially good monomers were not selected.
- Due to the electronic structure of the monomer. Some monomers with a small HOMO-LUMO gap, for example, have an antiaromatic electronic structure—like monomers with fused alternating 5- and 6-membered rings, such as s-indacene. This, we hypothesize, inhibits conjugation in the oligomer and does not promote a small HOMO-LUMO gap.
- Due to a human error with the SMILES string of the monomer. Some of the SMILES strings might have the wrong polymerization site which can break aromaticity and conjugation when the monomer is part of an oligomer.
- Due to inaccuracies in the GFN2-xTB calculations. While we have found a correlation between GFN2-xTB and ωB97X-D3 HOMO-LUMO gaps (Figure S1), this correlation is not as strong for small HOMO-LUMO gaps. It is possible that due to this some potentially good monomers did not survive the *Selection* step in the GA. This is a trade-off that we accept to greatly accelerate the GA.

Similarly to the full-length oligomers, a high correlation (R^2 = 0.86) was found between the monomers' singlet HOMO-LUMO

gap and the stability of the triplet ground-state (Figure S9). This agrees with previous studies that showed a biradical nature in molecules with a small HOMO-LUMO gap[47–51].

3.5 Other Potential Monomers

To find which property contributes the most to the stability of an open-shell electronic structure in the oligomer, as well as other monomers that the GA might have missed, we looked at four different monomer properties: relative HOMO level, relative LUMO level, HOMO-LUMO gap, and their triplet ground-state stability (ΔE_{T-S}). A representative selection of monomers with a range of values for each property were selected, and an oligomer was created for each monomer and monomer 630—which was paired with monomer number 642 in the oligomer with the second most stable triplet ground-state (Figure S10). We chose to use monomer 630 over monomer 365, which was paired with monomer 642 and had the most stable triplet ground state, since it contained a 7-membered ring in its backbone which broke its aromaticity and interrupted its conjugation to the πsystem. Monomer 630 has a highly conjugated and aromatic structure that includes a vinyl bridge, and we hypothesized that it will create more consistent and explainable results. The DFT singlet HOMO-LUMO gap and the singlet and triplet electronic energies for each oligomer were calculated as described in Section 2.3.2.

Figure 5 show the correlation between the monomers' property versus the oligomers' triplet ground-state stability. There is no strong correlation between each of the monomer properties and the stability of the whole oligomer, as they show heteroscedastic behavior. That is, monomers with low relative HOMO levels and high relative LUMO levels, HOMO-LUMO gap, and ΔE_{T-S} show low triplet state stability in the oligomer. On the other end, monomers on the opposite side of those properties do not show a clear-cut correlation between the property and the oligomers ΔE_{T-S}, at least when paired with monomer 630.

Interestingly, while oligomers with monomer number 630 were not as ubiquitous in the GA as other monomers were, Figure 5 show that many oligomers that include monomer number 630 did show a stable triplet ground-state—including monomers that are not in the top 10 most common monomers in the GA (Figure 3). For example, the oligomer constructed from monomers 630 and 261 showed a very strong ($\Delta E_{T-S} = -2.96$ eV) triplet ground-state stability, while not being found in any of the GA runs. This very low ΔE_{T-S} would be comparable to the top 20 oligomers found in the GA (Figure S3). See Table S2 in the **Supporting Information** for the full data.

3.6 Some Remarks

The finding above highlights a weakness in genetic algorithms as a whole, due to their non-deterministic nature and stochastic behavior. In other words, GAs can find a local optima, while sometimes missing the global one. There are ways to mitigate this behavior by tuning the GA's hyperparameters, such as the population size, mutation rate, and elitism rate[16]. However, even with well-tuned hyperparameters, there is still a chance that the GA misses the global optima. While there are other, deterministic algorithms that can find the global optima, they come with a greater computational cost[52]. In our case we tried to avoid this problem by running the GA 10 times, but even so it is evident that the GA did miss some potential candidates. The likelihood of this happening can be reduced by running the GA for more generations and more times, but then the return-on-investment (ROI) might not be favorable if this takes longer and has higher computational costs.

Another point we want to emphasize here, as Figure 5 shows, is that the identity of one monomer does not correlate with the oligomer triplet ground state stability, and it is the combination of the two monomers that overall dictates the oligomer's properties. While we presented here various monomers that were common in the GA (Figure 3), not every oligomer with them had a small HOMO-LUMO gap. For example, monomer 642 was the most common monomer in the GA and in the top 20 oligomers, but when combined with monomer 659 it had a GFN2-xTB HOMO-LUMO gap of 1.22 eV — which would correlate to $\sim$5.5 eV DFT HOMO-LUMO gap, according to Figure S1, and a more stable singlet ground state than a triplet one, according to Figure S3, by $\sim$2.0 eV.

4 Conclusions

Ground-state triplet polymers have a unique electronic structure and properties that have many possible uses in electronic devices. In this study we demonstrated how a Genetic Algorithm combined with GFN2-xTB, a fast semi-empirical method, can find unique and novel π-conjugated organic co-polymer candidates with a stable triplet ground-state. Those candidates exhibit a small HOMO-LUMO gap, which was previously shown to promote a triplet ground-state electronic structure due to the frontier molecular orbitals getting closer in energy. The spin densities show the biradical nature of those candidates, and the delocalization of the two unpaired electrons over two different singly-occupied molecular orbitals. DFT calculations show a triplet ground-state stabilization for up to 4 eV for the oligomer, and we expect this value to be similar or greater for the full-length polymer.

In addition, we have found that a small number of monomers have been found by the GA to promote a small HOMO-LUMO gap. All of those monomers exhibit small HOMO-LUMO gaps on their own, which helped promote a small HOMO-LUMO gap in the full oligomer. However, no other monomers with a small HOMO-LUMO gap were found by the GA, which shows that the GA has flaws. While the stochatstic nature of the GA imply that it can sometimes miss a potential candidate it is still a faster and more efficient method than an exhaustive search over the vast chemical space, particularly for finding top candidates and relevant motifs.

Data Availability

Raw data including calculated electronic values, optimized geometries and monomers' SMILES strings can be found at https://github.com/hutchisonlab/GST_GA.

Author Contributions

O. D. A. and G. R. H. conceived this study, O. D. A. performed all the computational studies, analyzed the results and wrote the

manuscript, G. R. H. has acquired the funding for this project and edited the manuscript.

Conflicts of interest

There are no conflict of interest.

Acknowledgements

We acknowledge support from Department of Energy-Basic Energy Sciences Computational and Theoretical Chemistry (Award DE-SC0019335) and the University of Pittsburgh Center for Research Computing through the computational resources provided.

Notes and references

1 T. Tadesse, *Review. J Phys Chem Biophys*, 2018, **8**, 263.
2 A. Facchetti, *Chemistry of Materials*, 2010, **23**, 733–758.
3 K. Wang, L. Huang, N. Eedugurala, S. Zhang, M. A. Sabuj, N. Rai, X. Gu, J. D. Azoulay and T. N. Ng, *Advanced Energy Materials*, 2019, **9**, 1902806.
4 Y. Huang and E. Egap, *Polymer Journal 2018 50:8*, 2018, **50**, 603–614.
5 L. Huang, N. Eedugurala, A. Benasco, S. Zhang, K. S. Mayer, D. J. Adams, B. Fowler, M. M. Lockart, M. Saghayezhian, H. Tahir, E. R. King, S. Morgan, M. K. Bowman, X. Gu and J. D. Azoulay, *Advanced Functional Materials*, 2020, **30**, 1909805.
6 X. Ji and L. Fang, *Polym. Chem.*, 2021, **12**, 1347–1361.
7 A. E. London, H. Chen, M. A. Sabuj, J. Tropp, M. Saghayezhian, N. Eedugurala, B. A. Zhang, Y. Liu, X. Gu, B. M. Wong, N. Rai, M. K. Bowman and J. D. Azoulay, *Science Advances*, 2019, **5**, eaav2336.
8 Y. Morita, S. Nishida, T. Murata, M. Moriguchi, A. Ueda, M. Satoh, K. Arifuku, K. Sato and T. Takui, *Nature Materials 2011 10:12*, 2011, **10**, 947–951.
9 O. D. Abarbanel, J. Rozon and G. R. Hutchison, *Journal of Physical Chemistry Letters*, 2022, **13**, 2158–2164.
10 Y. Liu, T. Zhao, W. Ju, S. Shi, S. Shi and S. Shi, *Journal of Materiomics*, 2017, **3**, 159–177.
11 K. Guo, Z. Yang, C. H. Yu and M. J. Buehler, *Materials Horizons*, 2021, **8**, 1153–1172.
12 R. Vasudevan, G. Pilania and P. V. Balachandran, *Journal of Applied Physics*, 2021, **129**, 070401.
13 M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, MA, USA, 1998.
14 C. Darwin, *On the Origin of Species by Means of Natural Selection*, Murray, London, 1859.
15 D. C. Hiener and G. R. Hutchison, *The Journal of Physical Chemistry A*, 2022, **126**, 2750–2760.
16 B. Greenstein, D. Elsey and G. Hutchison, *Best Practices for Using Genetic Algorithms in Molecular Discovery*, 2023.
17 G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*, CreateSpace, Scotts Valley, CA, 2009.
18 O. D. Abarbanel and G. R. Hutchison, *The Journal of Chemical Physics*, 2021, **155**, 054106.
19 B. L. Greenstein and G. R. Hutchison, *The Journal of Physical Chemistry Letters*, 2022, **13**, 4235–4243.
20 B. N. Norris, S. Zhang, C. M. Campbell, J. T. Auletta, P. Calvo-Marzal, G. R. Hutchison and T. Y. Meyer, *Macromolecules*, 2013, **46**, 1384–1392.
21 I. Y. Kanal, J. S. Bechtel and G. R. Hutchison, *ACS Symposium Series*, American Chemical Society, 2014, pp. 379–393.
22 S. Zhang, G. R. Hutchison and T. Y. Meyer, *Macromolecular Rapid Communications*, 2016, **37**, 882–887.
23 S. Zhang, N. E. Bauer, I. Y. Kanal, W. You, G. R. Hutchison and T. Y. Meyer, *Macromolecules*, 2016, **50**, 151–161.
24 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. G. III and W. M. Skiff, *Journal of the American Chemical Society*, 2002, **114**, 10024–10035.
25 T. A. Halgren, *Journal of Computational Chemistry*, 1996, **17**, 490–519.
26 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *Journal of Cheminformatics*, 2011, **3**, 33.
27 C. Bannwarth, S. Ehlert and S. Grimme, *Journal of Chemical Theory and Computation*, 2019, **15**, 1652–1671.
28 J. G. Brandenburg, C. Bannwarth, A. Hansen and S. Grimme, *The Journal of Chemical Physics*, 2018, **148**, 064104.
29 J.-D. Chai and M. Head-Gordon, *The Journal of Chemical Physics*, 2008, **128**, 84106.
30 Y. S. Lin, G. D. Li, S. P. Mao and J. D. Chai, *Journal of Chemical Theory and Computation*, 2013, **9**, 263–272.
31 F. Weigend, *Physical Chemistry Chemical Physics*, 2006, **8**, 1057–1065.
32 F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2012, **2**, 73–78.
33 F. Neese, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2018, **8**, e1327.
34 T. Yanai, D. P. Tew and N. C. Handy, *Chemical Physics Letters*, 2004, **393**, 51–57.
35 N. M. O'boyle, A. L. Tenderholt and K. M. Langner, *Journal of Computational Chemistry*, 2008, **29**, 839–845.
36 N. E. Jackson, B. M. Savoie, K. L. Kohlstedt, T. J. Marks, L. X. Chen and M. A. Ratner, *Macromolecules*, 2014, **47**, 987–992.
37 L. Wilbraham, E. Berardo, L. Turcani, K. E. Jelfs and M. A. Zwijnenburg, *Journal of Chemical Information and Modeling*, 2018, **58**, 2450–2459.
38 B. M. Savoie, N. E. Jackson, L. X. Chen, T. J. Marks and M. A. Ratner, *Accounts of Chemical Research*, 2014, **47**, 3385–3394.
39 D. Folmsbee and G. Hutchison, *International Journal of Quantum Chemistry*, 2020, **121**, e26381.
40 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, *Journal of Cheminformatics*, 2012, **4**, 17.
41 S. A. Wildman and G. M. Crippen, *Journal of Chemical Information and Computer Sciences*, 1999, **39**, 868–873.
42 B. M. Wong and J. G. Cordaro, *Journal of Physical Chemistry C*, 2011, **115**, 18333–18341.
43 K. Kobayashi, M. S. Mohamed Ahmed and A. Mori, *Tetrahedron*, 2006, **62**, 9548–9553.
44 W. I. Hung, Y. Y. Liao, C. Y. Hsu, H. H. Chou, T. H. Lee, W. S.

Kao and J. T. Lin, *Chemistry – An Asian Journal*, 2014, **9**, 357–366.

45 M. Paramasivam, A. Gupta, A. M. Raynor, S. V. Bhosale, K. Bhanuprakash and V. Jayathirtha Rao, *RSC Advances*, 2014, **4**, 35318–35331.

46 A. Shuto, T. Kushida, T. Fukushima, H. Kaji and S. Yamaguchi, *Organic Letters*, 2013, **15**, 6234–6237.

47 L. Salem and C. Rowland, *Angewandte Chemie International Edition in English*, 1972, **11**, 92–111.

48 M. Nakano, *Chemical Record*, 2017, **17**, 27–62.

49 Z. Sun, Q. Ye, C. Chi and J. Wu, *Chemical Society Reviews*, 2012, **41**, 7857–7889.

50 R. Rausch, D. Schmidt, D. Bialas, I. Krummenacher, H. Braunschweig and F. Würthner, *Chemistry - A European Journal*, 2018, **24**, 3420–3424.

51 A. Rajca, *Chemical Reviews*, 2002, **94**, 871–893.

52 M.-H. Lin, J.-F. Tsai and C.-S. Yu, *Mathematical Problems in Engineering*, 2012, **2012**, 756023.