Designing mechanosensitive molecules from molecular building blocks: a genetic algorithm-based approach

Matthias Blaschke, Fabian Pauly

Angaben zur Veröffentlichung / Publication details:

Blaschke, Matthias, and Fabian Pauly. 2023. "Designing mechanosensitive molecules from molecular building blocks: a genetic algorithm-based approach." *The Journal of Chemical Physics* 159 (2): 024126. https://doi.org/10.1063/5.0155012.

The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Designing mechanosensitive molecules from molecular building blocks: A genetic algorithm-based approach

Cite as: J. Chem. Phys. 159, 024126 (2023); doi: 10.1063/5.0155012
Submitted: 17 April 2023 • Accepted: 19 June 2023 •
Published Online: 12 July 2023

Matthias Blaschke and Fabian Pauly[a)]

AFFILIATIONS

Institute of Physics and Centre for Advanced Analytics and Predictive Sciences, University of Augsburg, 86135 Augsburg, Germany

[a)]Author to whom correspondence should be addressed: fabian.pauly@uni-a.de

ABSTRACT

Single molecules can be used as miniaturized functional electronic components, when contacted by macroscopic electrodes. Mechanosensitivity describes a change in conductance for a certain change in electrode separation and is a desirable feature for applications such as ultrasensitive stress sensors. We combine methods of artificial intelligence with high-level simulations based on electronic structure theory to construct optimized mechanosensitive molecules from predefined, modular molecular building blocks. In this way, we overcome time-consuming, inefficient trial-and-error cycles in molecular design. We unveil the black box machinery usually connected to methods of artificial intelligence by presenting all-important evolutionary processes. We identify the general features that characterize well-performing molecules and point out the crucial role of spacer groups for increased mechanosensitivity. Our genetic algorithm provides a powerful way to search chemical space and to identify the most promising molecular candidates.

I. INTRODUCTION

As the miniaturization of silicon-based electronic devices reaches the limits of what is technically feasible,[1,2] the vision of molecular electronics becomes increasingly relevant: to build functional electronic units at the molecular scale and to synthesize electronic circuitry through appropriate chemistry bottom-up instead of conventionally top-down using lithography techniques.[3] Electronic functionality in the form of diodes, transistors, and switches has been realized at the single-molecule level early on.[4,5] Switches can function through a multitude of mechanisms,[6,7] including conformational, electrochemical, and spin switching. The external stimuli for switching may vary from light to electric fields, temperature, current, or mechanical control by electrode separation.[6] While the precise switching mechanism may not be clear in every experiment, a recently emerging research theme is that of mechanosensitive molecules.[8–10] Such molecules change their conductance strongly by some orders of magnitude for tiny changes in electrode displacement. They may be viewed as mechanically controlled switches. The functionality may ultimately be used in ultrasensitive quantum distance sensors.

Mechanical control of electrode separations in single-molecule junctions may lead to changes at the metal–molecule interface or internally inside the molecule. We have recently studied a class of mechanosensitive molecules, where molecule-internal changes cause huge conductance modulations.[8,11–13] Molecules may hence be viewed as the functional unit, and they all feature two decks of π-electron systems that can be shifted with respect to each other. The underlying switching mechanism is a destructive quantum interference (DQI) that is established or lifted by mechanical manipulation. The emergence of DQI depends critically on the energetic alignments and shapes of the wave functions of the molecular frontier orbitals.[8,14–16] Seemingly small changes in the molecular structure, such as para-connectivity vs meta-connectivity, may lead to the vanishing of DQI in the vicinity of the equilibrium, ground state molecular geometry and thus to the vanishing of mechanosensitivity.[11] Even if a molecule yields a DQI that can be mechanically lifted, its observation may be hampered by a low overall

electrical conductance or a small switching ratio of highest to lowest conductance values. Finally, molecules should not be too stiff to allow for a mechanical manipulation by the typically employed, rather soft metal electrodes. These factors differ for the molecules studied so far.[8,9,11–13] Thus, while in the class of π-stacked molecules the underlying mechanisms of mechanosensitivity are well understood, design rules for molecular structures with enhanced mechanosensitivity are lacking. Development processes so far involve time-consuming trial-and-error cycles of molecular design.

The idea of inverse molecular design allows overcoming such inefficient trial-and-error cycles in development processes.[17,18] Genetic algorithms are nowadays routinely used in *de novo* molecular design studies.[19–23] With the help of genetic algorithms, approximate solutions can be found on an affordable time scale. While much work has been carried out in the field of drug design,[24–27] genetic algorithms can also target other applications, for instance, photovoltaics.[28,29]

In order to discover optimal mechanosensitive molecules, we present here a genetic algorithm that searches molecular candidates according to predefined criteria. We combine methods of artificial intelligence with electronic structure theory to provide a fast and systematic way to find well-performing molecules at a high level of accuracy. The candidates should yield (i) a DQI with a large switching factor, (ii) the average conductance outside of the DQI dip should be high to facilitate measurements, and (iii) the molecules should not be too rigid to afford mechanical manipulation through rather soft metal electrodes. A drawback of inverse design approaches can be the poor synthesizability of proposed molecules.[30] We overcome this issue by constructing them from synthesizable, modular building blocks in the multifactorial optimization of electrical and mechanical properties.

II. THEORETICAL APPROACH

In this section, we describe the theoretical approach that we are using. This involves the genetic algorithm that we developed, the genetic encoding of molecules, the evolution loop, and the fitness calculation. The fitness evaluation requires methods to analyze the electrical and mechanical properties of the molecules. Some technical details are deferred to Appendixes A and B, including an accuracy assessment of the estimate of the conductance and the precise fitness evaluation. The full program code is publicly available at Zenodo,[31] and the discussion in this section may be directly followed at the code.

A. Genetic algorithm

A genetic algorithm is a randomly directed search algorithm based on evolutionary theory.[32] It is driven by natural selection and survival of the fittest. Evolution is simulated by means of a population, consisting of multiple individuals in each generation. The best individuals from a population are chosen to produce offspring. We make this choice using a selection scheme, which can be regarded as deterministic k-tournament selection.[33–35] The k fittest individuals are chosen in each generation to fill a mating pool, from which the parents are chosen. The basis for the selection process is the so-called fitness value. This measure can take into account several properties, which should be weighted and optimized in an appropriate form. We define measures for the DQI, average conductance, and rigidity of the molecules, according to the criteria (i) to (iii) mentioned above. The genetic information of the parents is inherited by the offspring to form the new generation. The individuals in the next generation may be better with an increased chance of survival due to a higher fitness value. Poorly performing combinations are instead penalized by a low fitness value. This process is iterated in an evolution loop until a convergence criterion is reached or until the evolution is stopped for other reasons. The most important components of a genetic algorithm are the genetic encoding, the evolution loop with the corresponding selection scheme and genetic operators, and the fitness evaluation.[36] We will describe these aspects in the following subsections.

B. Genetic encoding

The genetic encoding of individuals depends on the nature of the studied problem[37] and needs to capture the important adjustment screws. In recent years, a lot of research has been carried out on the representation of molecules for machine learning techniques.[18,26,38–40] Due to the special problem studied here, i.e., the contacting of a molecule by macroscopic electrodes, we developed a tailored representation. We use a fragment-based, gradient-free approach,[26,41] where we generate new molecules from relevant substructures. Usually, fragmentation algorithms are applied to large molecular databases in order to determine these components in similar studies.[42,43] For the molecules considered here, this is not expedient because of the complexity and specificity of the features sought.

Based on previous analysis and literature results, we have identified specific building blocks and their coupling configurations as essential for creating mechanosensitive molecules. The molecular building blocks are depicted in Fig. 1. The interaction of π-electron systems[44,45] has been found to be a key component of mechanosensitive molecules.[8,9,11,46] Thus, we build candidates from modular molecular blocks that include such π-stacked components (see blocks 3–7 in Fig. 1). In addition, we added simple spacer blocks, such as acetylene, benzene, and naphthalene, to provide more flexibility to the molecules (see blocks 0–2 in Fig. 1). These spacers can separate π-stacked units from each other or increase the distance of the functional stacks from the electrodes.

The importance of spacers for DQI features has been investigated in the past,[47–49] but to the best of our knowledge it has not yet been studied much in the context of mechanosensitivity. Tsuji and co-workers[50] showed that such spacers cannot lead to a vanishing of DQI in the energy-dependent transmission of π-conjugated hydrocarbons but that they can alter the transmission behavior. Therefore, we include those spacers as an additional degree of freedom.

Apart from the building blocks, the coupling among themselves and to the electrode anchors is important. [2.2]Paracyclophane (block 3 in Fig. 1), for example, exhibits an experimentally observable DQI for the para configuration but no DQI for meta-coupling.[11] For the sake of simplicity, we concentrate here exclusively on para- and meta-substitution patterns.[51]

The geometric degrees of freedom must be reproduced by the genetic encoding, and an example is given in Fig. 2. Each molecule is encoded from the left to the right in an alternating list of couplings and molecular blocks. The length of the molecule is determined by

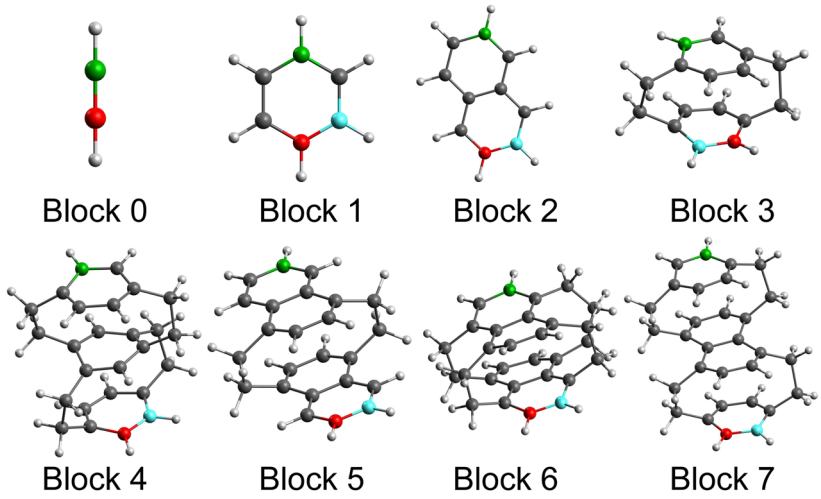

FIG. 1. Molecular building blocks used to construct mechanosensitive molecules. Green carbon atoms indicate the "left" coupling point. Red carbon atoms mark the para-coupling position on the "right," while cyan carbon atoms mark the meta-coupling position. Molecular blocks with just a red carbon atom on the "right," see block 0, do not differ in para- and meta- connections. Blocks 0–2 can be regarded as spacers and the remaining blocks 3–7 as π-stacked units.

the integer L, see the top of Fig. 2. The electrode anchors consist of thiol groups because they form strong covalent bonds to gold metallic electrodes and hence stable gold–molecule–gold junctions in experiments.[8,52]

The number of possible combinations scales exponentially with respect to the molecular length L. The basis is determined by the number of molecular blocks n_B and the number of couplings n_C that are available. Thus, the total number of possible combinations N is

$$N = \sum_{L=L_{\min}}^{L_{\max}} n_C n_B{}^{L} n_C. \tag{1}$$

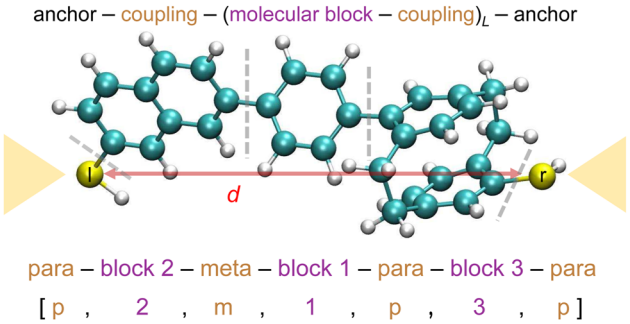

FIG. 2. Genetic encoding of molecules. The general scheme is shown in the top row. The encoding consists of alternating molecular blocks and couplings. The example demonstrates the encoding of a molecule with a length L = 3. Brown and purple colors in the encoding distinguish couplings from molecular building blocks. Anchors are omitted since they are always sulfur atoms, more precisely, SH groups. Dashed gray lines indicate the encoded part of the molecular structure. Hydrogen atoms of the molecular building blocks in Fig. 1 are removed at connection points. The red arrow displays the stretching axis. Gold tips of the metal–molecule–metal junction are represented by faint yellow triangles, and d is the distance between the sulfur atoms marked l and r.

We consider eight building blocks (see Fig. 1), two coupling types (see Figs. 1 and 2), and a molecular length between 1 and 7 building blocks, i.e., $L_{\min} = 1 \le L \le L_{\max} = 7$. The restriction to a maximum length of seven building blocks is motivated by the exponential decay of the conductance with molecular length in the typical off-resonant transport regime. The overall conductance should not be too low for the molecules to be experimentally detectable. Our setting leads to a search space of about $N \approx 5.73 \times 10^8$ molecules. Thus, a brute-force approach is not feasible, and a systemic way like the genetic algorithm is needed instead to identify the most appropriate individuals in the ensemble of all N possibilities.

C. Evolution loop

The evolution loop is the central part of the algorithm. It is depicted in Fig. 3. The evolutionary process starts by initializing a population with randomly generated individuals. Each individual represents a molecule, and the population consists of multiple molecules. Individuals are fully characterized by their genetic encoding. After this initialization, the evolution cycle starts. In the beginning, the fitness of the individuals is evaluated. Various properties can be taken into account in the fitness. We focus on the DQI behavior, the median conductance, and the mechanical stiffness, as detailed further below. In order to form a new population for the next generation, the k fittest individuals are selected to form the mating pool. Selected candidates produce offspring by so-called crossover schemes. For each crossover, the two parents are chosen through uniform sampling from the mating pool. Afterward, the newly generated candidates can be altered through mutation. We treat the best N_{elite} individuals from the previous generation separately: They are directly transferred to the next generation. This is done to preserve the best combinations and is referred to as elitism in the literature.[36] To increase computational efficiency, the new population is filtered so that only unique individuals occur. Genomes

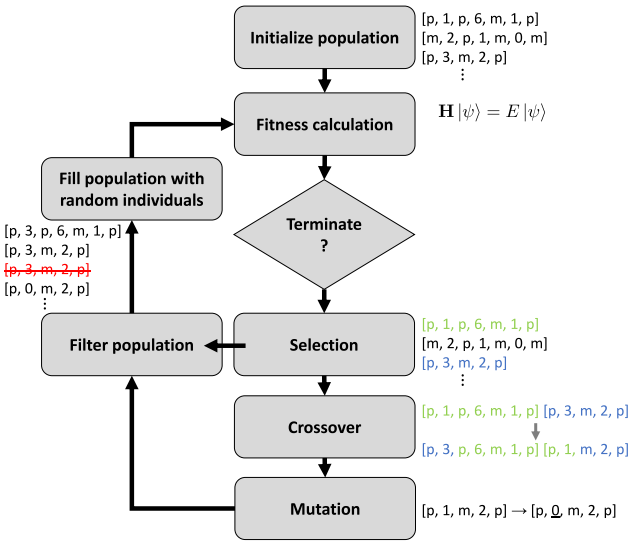

FIG. 3. Evolution loop of the genetic algorithm. Text and figural representations adjacent to the flowchart elucidate individual steps. In the initialization step, random individuals are created. Blue and green colors in the selection step indicate the two selected individuals, which create two offspring in the crossover step. The offspring may subsequently be changed by a mutation at the underlined position. The precise action of all genetic crossover and mutation operators is presented in Fig. 4. In the filtering step, the red-marked individual, which occurs twice, is removed from the population. The direct connection between the selection step and the filtering step visualizes the elitism.

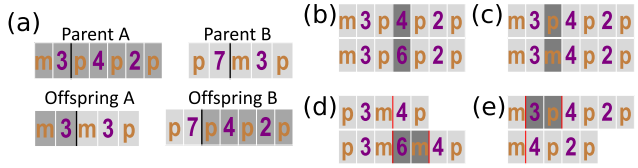

FIG. 4. Genetic operators. The top row in each panel shows the initial genetic information according to Fig. 2, and the outcome of each operation is given in the lower line. Brown and purple letters in the encoding string distinguish couplings from molecular building blocks. (a) Single-point crossover: A random position in the shorter parent is chosen, here parent B. The longer parent, parent A, is cut at the same point. Offsprings A and B are produced by combining heads and tails of the two parents. (b) and (c) Block mutation and coupling mutation: A molecular block or a coupling is chosen randomly and replaced by a new one. (d) Insert mutation: A randomly chosen pair of molecular block plus coupling are inserted at some point of the genome, if the maximum length $L \leq L_{max}$ is not exceeded. (e) Truncate mutation: A pair of a molecular block plus coupling are chosen randomly and removed from the string, as long as the length does not fall below one molecular block, i.e., $L \geq L_{min}$.

that are sorted out are replaced by randomly generated individuals to keep the population size constant and to increase genetic diversity. The whole process is repeated until a convergence criterion is reached or a maximum number of generations has been processed. Finally, we define the result of the genetic algorithm as the best individual of the last generation, even if the whole population has been optimized.

The genetic operators that we use are depicted in Fig. 4. In the single-point crossover, a single random cut in each parent string is chosen, producing two tail and two head sections. The offspring is created by combining the heads and tails of the two parents. To preserve the lengths of the parents, we choose the cut at some point of the shorter string, and the string of the other parent is cut at the same point. In this manner, one offspring inherits the length from parent A and the other offspring from parent B. The procedure is depicted in Fig. 4(a). We employ this length-preserving crossover to avoid large length fluctuations. Without this restriction, the number of building blocks for one offspring could potentially increase to twice the length of the longest molecule in the population in this single step. Furthermore, we use four different mutation operators, which are illustrated in Figs. 4(b)–4(e). In the mutation step, one of the four operators is selected with an equal probability of 1/4 and applied with a probability p_m. Thus, the probability for a specific mutation operation is $p_m \sim 4$. Combined with the N_{elite} individuals from elitism, the new generation is then formed by genomes, produced by crossover and mutation from the selected parents.

We use a population size of $N_{pop} = 26$ in this work. This is small compared to other approaches using a genetic algorithm[53] but is attributed to the computationally demanding fitness calculations. From the N_{pop} individuals in the population, we directly transfer the best $N_{elite} = 2$ to the next generation, leaving $N_{pop} - N_{elite} = 24$ remaining places in the next generation to be filled. For this purpose, we select the $k = 5$ fittest from the full population of N_{pop} entries to form the mating pool. We now fill the remaining 24 places by selecting two parents from the mating pool. Since our crossover scheme produces two offspring A and B from the two parents A and B, see Fig. 4, we select $N_{pop} - N_{elite} \sim 2 = 12$ couples with uniform probability $k^{-2} = 1 \sim 25$ from the $k^2 = 25$ possible combinations of parents. The 24 individuals, generated by crossover of the 12 pairs of two parents, may then be modified by mutation.

D. Fitness calculation

A crucial and potentially time-consuming part of the evolution loop is the fitness calculation. Each individual is ideally rated accurately within the shortest possible time to work with a large population. Simplified predictions of desired properties are often used,[23,54,55] hampering the validity of the computational approach. For this reason, we rely here on high-level calculations. A well-performing mechanosensitive molecule should have the following properties: It should exhibit a DQI, which is tunable by stretching, it should be highly conductive on average, it should not be too rigid to allow for the mechanical control of DQI through rather soft metal electrodes. The fitness value rates the performance of each candidate and crucially determines the predictive power of the genetic algorithm.

To determine the fitness value, we simulate a mechanical modulation cycle for each candidate, starting from the relaxed structure. The aim is to mimic recent experiments,[8,11] where a molecule is incorporated into an electrode–molecule–electrode junction and where the electrical conductance is measured while the junction is stretched or compressed. We transfer the genetic encoding to a molecular input file, compatible with quantum chemistry software, as illustrated in Fig. 2. The displacement direction is aligned with the sulfur–sulfur axis of the anchors. We optimize the molecular geometry with GFN2-xTB[56,57] under mechanical stretching and compression, i.e., for different given sulfur–sulfur distances d. In the

procedure, d is varied around the equilibrium distance by ±2.4 ‰ in steps of $\Delta d = 0.1$ ‰ Subsequently, we determine the electronic structure of the relaxed molecule for each d through a self-consistent density functional theory (DFT) calculation, yielding displacement-resolved data. The DFT calculations are performed using TURBOMOLE,[58] employing the def-SV(P) Gaussian basis set[59] for all atoms and the PBE exchange–correlation functional.[60]

Based on the mechanical and electronic properties obtained in the previous steps, we model the charge transport characteristics. We assume that the molecular candidate is connected to two gold metallic electrodes at the sulfur anchors. Transport through this single-molecule junction is then described in the phase-coherent regime via the Landauer–Büttiker approach.[61,62] Assuming low temperatures, the electrical conductance G can be expressed through the energy-dependent transmission $\tau(E)$, evaluated at the Fermi energy E_F, as

$$G = G_0 \tau(E_F). \tag{2}$$

Here, $G_0 = 2 \times e^2/h$ is the quantum of conductance. The transmission $\tau(E)$ depends on the retarded and advanced Green's functions of the central part of the junction $G^r_{CC}(E) = [G^a_{CC}(E)]^\dagger$ and the corresponding linewidth broadening matrices $\Gamma_L(E)$ and $\Gamma_R(E)$ due to the coupling to left and right electrodes via[61,62]

$$\tau(E) = \mathrm{Tr}\left[\Gamma_L(E) G^r_{CC}(E) \Gamma_R(E) G^a_{CC}(E)\right]. \tag{3}$$

Using the wide-band limit, the linewidth broadening matrices become independent of energy. Furthermore, they are assumed to be structured such that they have nonvanishing entries only on the sulfur atoms, which are directly bonded to the left and right electrodes.[63]

We thus estimate the conductance in Eq. (2) with the help of the following expression, where the indices l and r denote the anchoring sulfur atoms, as visualized in Fig. 2:

$$G_{\mathrm{est}}(d) = G_0 \gamma_0^2 \,\mathrm{Tr}\left[G^{(0)r}_{l,r}(E_F, d)\, G^{(0)a}_{r,l}(E_F, d)\right]. \tag{4}$$

Note that, in this expression, we explicitly indicate the dependence on the sulfur–sulfur distance d (see Fig. 2), and the trace runs over the atomic orbitals (or more precisely atom-centered basis functions) on sulfur atom l. The retarded Green's function in Eq. (4) has been further approximated to zeroth order by neglecting embedding self-energies. By using the spectral representation,[64] we can express it as

$$G^{(0)r}_{l,r}(E, d) = \sum_k \frac{\tilde{C}^l_k(d)\, \tilde{C}^r_k(d)}{E + i\eta - \varepsilon_k(d)}. \tag{5}$$

The vectors $\tilde{C}^l_k(d)$ and $\tilde{C}^r_k(d)$ contain the expansion coefficients of molecular orbital k in terms of atomic orbitals at the left and right sulfur atoms l and r at each distance d, and $\varepsilon_k(d)$ are the molecular orbital energies. All these quantities $\tilde{C}^l_k(d)$, $\tilde{C}^r_k(d)$, and $\varepsilon_k(d)$ are extracted from the DFT calculations. Finally, in Eq. (4) we use $G^{(0)a}_{r,l}(E, d) = [G^{(0)r}_{l,r}(E, d)]^\dagger$ and set $\gamma_0 = 1$ hartree to obtain the correct units, and $\eta = 10^{-12}$ hartree in Eq. (5) is a small real value. From Eq. (4), it is obvious that the estimated conductance depends on γ_0^2. By using a constant scalar γ_0, the coupling of the sulfur atoms to the gold electrodes is assumed to be identical for all molecules, and $G_{\mathrm{est}}(d)$ in Eq. (4) thus characterizes genuine molecular transport properties. While the precise value of the prefactor γ_0^2 in Eq. (4) plays a role for estimating the absolute value of the median conductance, it does not matter for a comparison of different molecules.

Equation (2) states that we need to evaluate the transmission at the Fermi energy of the connected gold electrodes of the molecular junction to determine the zero bias-voltage conductance in the low temperature regime. The proper level alignment between molecular orbitals and the metallic electrodes is, however, challenging to describe within DFT.[65,66] For reasons of computational efficiency, we hence neglect the metallic electrodes and concentrate only on the molecular part. It is reasonable to assume that the Fermi energy of gold lies inside the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of all rather short hydrocarbons, which are built by our encoding.[64,67] In this way, the molecules stay rather charge neutral when contacted by the macroscopic electrodes. We thus choose the Fermi energy to be positioned in the middle of the HOMO–LUMO gap of each isolated molecule, i.e., $E_F = E_{\mathrm{LUMO}} - E_{\mathrm{HOMO}} \sim 2$. The alignment of molecular orbitals with respect to E_F is certainly crucial for obtaining quantitative results but would also depend on the precise contact geometry. Since we are interested in DQI features that cross the entire gap,[8,11] the mid-gap approximation appears to be justified.

To validate the simplified calculation scheme, we have compared its results for selected molecules with and without DQI behavior to full DFT quantum transport calculations with metallic electrodes.[62] The example of block 3, modified with two different spacers, is discussed in Appendix A. We find that the approximations made are sufficiently accurate to identify molecules exhibiting mechanosensitivity and to compare their performance. The scheme presented above thus constitutes a reasonable compromise to assess the electrical conductance response of each candidate in a mechanically controlled single-molecule junction experiment in a fast and accurate way.

The final step is the assignment of a fitness value. We consider measures for the DQI, f_{DQI}, the median conductance, f_{medG}, and the mechanical stiffness of the molecule, $f_{\mathrm{stiffness}}$. They are evaluated independently, as described shortly further below and in greater detail in Appendix B. We define the overall fitness f of a molecule through the product

$$f = f_{\mathrm{DQI}} \times f_{\mathrm{medG}} \times f_{\mathrm{stiffness}}. \tag{6}$$

The DQI is evaluated through displacement-resolved data, using $G_{\mathrm{est}}(d)$ of Eq. (4). We fit a function to $G_{\mathrm{est}}(d)$, mimicking a possible DQI dip. The median of the estimated conductance over all studied displacements d is used to optimize for highly conductive molecules. In this way, we avoid difficulties in actual experiments if, for example, the conductance would drop below measurable limits. The stiffness is rated in a similar way as the DQI. The molecule is assumed to behave like a classical spring in the harmonic limit with the dimensionless spring constant $a_{\mathrm{stiffness}}$.[8] By fitting a parabola to the total DFT energy vs displacement d, we extract the constant $a_{\mathrm{stiffness}}$. The standard deviations of the fit parameters for the DQI and stiffness measures are used to sort out unsuitable candidates. These may be candidates, where no DQI feature is detected or where

the energy vs displacement data do not show a quadratic behavior. If the standard deviation of the fit parameters exceeds a certain threshold, the corresponding fitness is set to zero. Further details and explanations with examples can be found in Appendix B.

III. RESULTS

A. Evolution runs

To study the performance of the genetic algorithm, we analyzed a large number of evolution runs. The evolution of fitness values in two runs are presented in Fig. 5.

Each run in Fig. 5 is characterized by a sharp increase in mean fitness value during the first generations. This mean fitness is defined as the average of fitness values of all individuals in a population. For higher generations, the fitness tends to saturate. Regarding the distribution of fitness values, evolved individuals, indicated by crosses in Fig. 5, generally feature high fitness. The randomly generated individuals, marked by dots, in contrast show below-average performance in most of the cases. This confirms the adequacy of our genetic algorithm. It is visible that the best two individuals of a population remain the same for several generations. This is caused by the elitism that transfers the best two individuals directly to the next generation. Figure 5 furthermore shows a substantial fraction of randomly generated individuals in late generations. This means that a larger number of individuals, produced by crossover and mutation, are identical and thus filtered out.

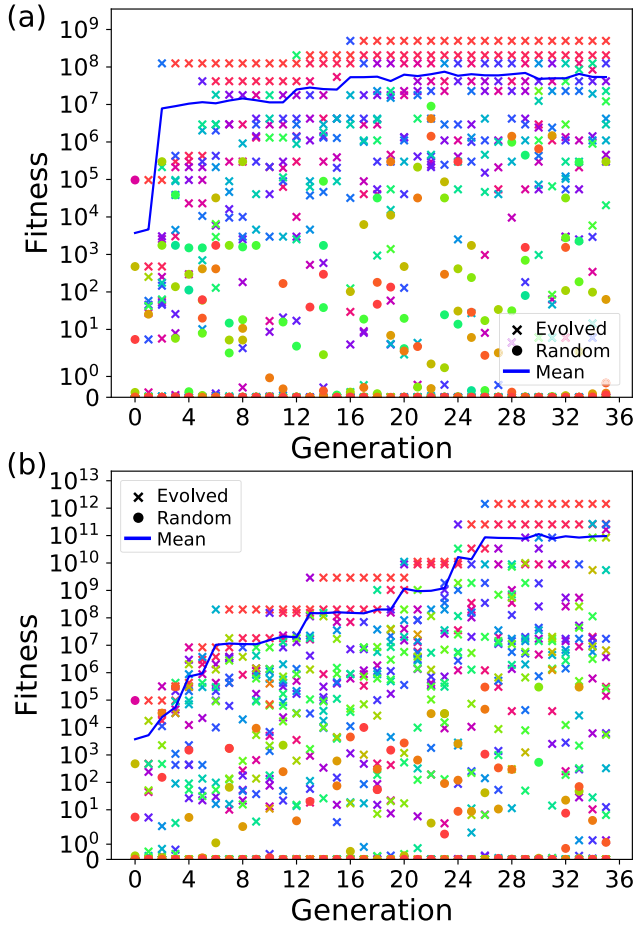

FIG. 5. Fitness values of two evolution runs as a function of the generation. The runs start from the same initial population in generation 0. The parameters used in these fitness functions are $\alpha = 2.5$, $\beta = 1.2$, $\gamma = 2.0$, and $\delta = 0.005$ (see Appendix B), the population size is 26, the generation limit 35, and mutation probabilities are (a) $p_m = 0.5$ and (b) $p_m = 0.9$. For the tournament selection, the best $k = 5$ individuals are taken to generate offspring. Crosses indicate individuals evolved by selection, crossover, and mutation, while dots mark randomly generated individuals. The blue line shows the mean value, obtained as the average fitness of the full population in each generation. We color-coded the candidates in the plot to improve the distinction of individuals with similar fitness values.

The initial population is the same in both evolution runs, shown in Fig. 5, but probabilities for mutation differ. While both evolutions show a significant increase in fitness for the first few generations as common feature, the run in Fig. 5(b) yields a better performance for higher generations. The life spans of the best individuals are shorter, and the fitness value of the best individual after 35 iterations is significantly larger than for Fig. 5(a). We define the life span as the number of generations, in which the best individual, and therefore the highest fitness value, does not change. The lower the life spans of the best individuals, the more desirable evolutionary processes occur. Due to the low probability of mutation, the evolution gets stuck for the run in Fig. 5(a). We thus conclude that sufficiently high mutation probabilities are needed for creating the next generation from a given population in order to prevent a quick saturation of the evolution, to achieve short life spans and an overall high fitness value.

We stop the evolution after 35 generations. Although this does not fulfill a defined convergence criterion, this appears to be an adequate number of generations, since no major increase in fitness is observed over several generations. For each run, the best individual of the last generation is considered as the final result of the whole evolution.

The performance of the best individuals for the run in Fig. 5(b) is illustrated in Fig. 6. We only focus on the electrical response, as characterized by $G_{est}\ d$. In generation 0, a suitable candidate is found within the random population, but the DQI dip is only weak. The performance gradually increases, and the DQI dips get deeper for higher generations. Behavior similar to that seen in generations 2, 3, or 13 might not be desirable due to the lack of a sharp DQI dip but gets rewarded by the chosen fitness function. These intermediate stages however help to produce even better offspring. Such molecules are found in generations 24 and 26. These molecules at the end of our evolution exhibit a DQI dip, where $G_{est}\ d$ drops by several orders of magnitude in a narrow displacement range. This is combined with a relatively high average conductance, which is the desired behavior. In the end, we find an excellent candidate, which outperforms the best molecules in our test set and which was unknown before.

We observe that the fittest individuals with a deep DQI dip and high switching ratio between highest $G_{est}\ d$ value and lowest $G_{est}\ d$ consist of π-stacked units (blocks 3–7 in Fig. 1) combined with spacers (blocks 0–2 in Fig. 1). The importance of spacer blocks for the performance of mechanosensitive molecules was not studied extensively before and is pointed out as a general principle by our genetic algorithm. We analyze the influence of spacer blocks

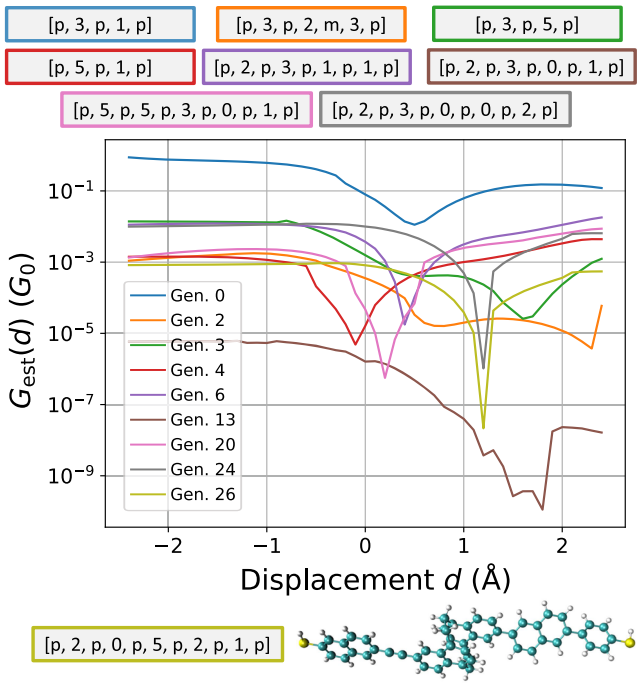

FIG. 6. G_{est} d plotted against the sulfur-to-sulfur displacement d for the best individuals of the evolution run, shown in Fig. 5(b). The distance axis is resolved in steps of Δd = 0.1 ‰ We present the best individuals only for those generation, in which they first occur. Molecular encodings are shown above the figure in the same color as in the legend, while the best individual is depicted both with its encoding and molecular structure at the bottom.

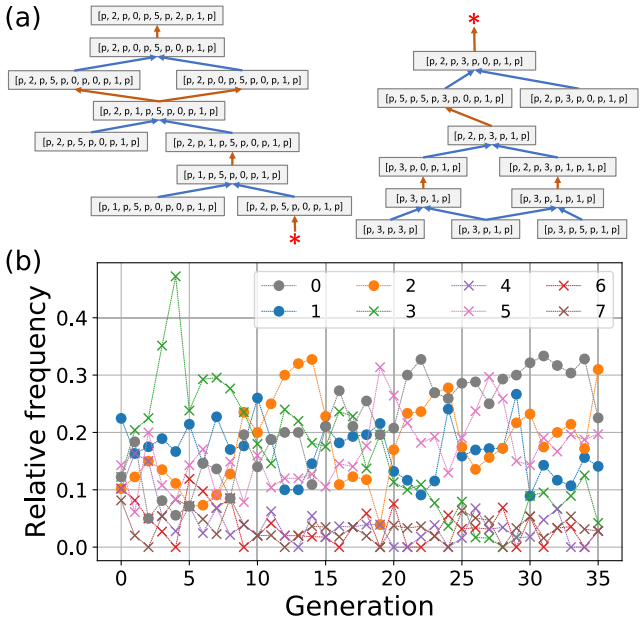

FIG. 7. Analysis of the evolution run in Fig. 5(b). (a) Family register of the best individual of Fig. 5(b), whose geometry is plotted in Fig. 6. Each box shows the genetic encoding of a molecule. Blue arrows indicate crossover operations and orange arrows represent mutation operations. The family register is split into two parts at the red star for illustrative purposes. (b) Relative frequency of molecular blocks for each generation in the run of Fig. 5(b). Different markers are used to distinguish spacer blocks (dots) from π-stacked building blocks (crosses). Encoding and block numbering follow Fig. 2.

in more detail in Subsection III B, where we demonstrate that the steepness of the DQI dip and switching ratio can be enhanced by spacers.

Unlike many other approaches of artificial intelligence, our genetic algorithm is not a black box. Apart from the optimized molecules, we can provide insight into the evolution processes. As an example, we study in Fig. 7(a) the family register of the best individual of the evolution run, shown in Fig. 5(b). Crossover and mutation operations can clearly be seen as the driving mechanism for the evolution. The best candidate of Fig. 5(b) is obtained after seven crossover and eight mutation operations. A similar analysis of the best individual of the evolution in Fig. 5(a) is represented in Fig. 8(a). In this case, the best performing molecule is found after four crossover and two mutation operations. The run in Fig. 5(b) yields a higher total fitness, compared to Fig. 5(a). Due to the increased mutation probability for evolution in Fig. 5(b) and the higher number of mutations that lead to the best individual, one might assign a higher importance to mutation than to crossover operations. However, it should be noted that the offspring is identical to that of the parents for several crossover operations in Figs. 7(a) and 8(a). For instance, p, 3, p, 1, p at the bottom of Fig. 7(a) is reproduced by a crossover of p, 3, p, 3, p and p, 3, p, 1, p . This is important because only the best two individuals are transferred directly to the next generation by elitism. The remaining candidates have to reproduce themselves by crossover. Otherwise, the genome, on which the mutation operations work, would not be part of the considered search space anymore. In summary, both crossover and mutations are important and their interplay is essential for the evolution.

The idea of natural selection can be seen in the evolutions as well. For this purpose, we analyze the relative frequency of the building blocks in each generation. The results for the evolution

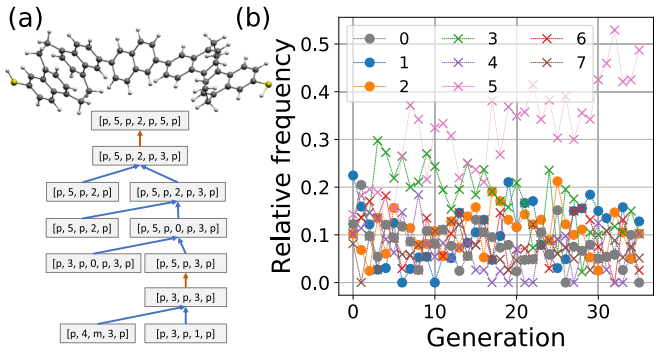

FIG. 8. (a) and (b) Same as Fig. 7 but for the evolution run shown in Fig. 5(a). Additionally, panel (a) displays the structure of the best performing individual of the corresponding evolution, i.e., [p, 5, p, 2, p, 5, p].

run in Fig. 5(b) are depicted in Fig. 7(b). In the beginning, building blocks are uniformly distributed. Small deviations arise from the limited population size. We distinguish between spacer blocks (dots) and π-stacked building blocks (crosses). The distribution shows that the spacer blocks become dominant in higher generations. The frequency of blocks 1 and 2 fluctuates roughly between 10% and 30%. An upward trend can be observed for block 0, consistently reaching more than 19% after generation 15. For the π-stacked blocks, block 3 initially dominates, but around generation 20 block 5 becomes prevalent. The doubly π-stacked building blocks 4 and 7 are rather unimportant, similar to the stiff block 6, featuring in total four ethyl bridges between upper and lower naphthalene units. Overall, the individuals of this run mainly consist of spacer blocks. The π-stacked building blocks occur significantly less frequently. However, data analysis corroborates that in the well-performing individuals, at least one π-stacked block occurs. A similar behavior can be observed for the evolution run in Fig. 5(a), whose analysis is shown in Fig. 8(b). The spacer blocks have a higher relative frequency compared to most of the π-stacked blocks. However, block 5 exhibits the highest relative frequency. The fact that the spacer blocks are not predominant in this case is mainly due to the shorter length of the evolved molecules [cf. the family registers in Figs. 7(a) and 8(a)].

Additionally, it should be pointed out that we observe almost exclusively para-couplings toward the end of the two evolution runs shown in Fig. 5. However, this is not because DQI cannot occur when meta-configurations are part of the molecule (e.g., [m, 1, p, 3, p, 1, m] is mechanosensitive, corresponding to the ps-*para-meta* [2.2]paracyclophane of Ref. 11) but rather because the average conductance is too low, the DQI feature is not pronounced enough, or mechanical manipulation of π-stacked units is hampered.

Our analysis of the relative frequency of building blocks reveals an effective reduction of the chemical search space. A reduction from eight blocks to five relevant ones, i.e., the spacers 0, 1, 2 and π-stacks 3 and 5, combined with just the para-configuration as relevant coupling, reduces the search space according to Eq. (1) from 5.73×10^8 to around 9.77×10^4 possible combinations.

A drawback of the results presented so far may be the lack of symmetry with respect to the central building block. In this context, we note that our encoding always consists of an odd number of entries, starting and ending with couplings and featuring a block in the middle. Since a symmetry with regard to the central block is no criterion of the fitness function, strings are generally asymmetric [cf. the family registers in Figs. 7(a) and 8(a)]. Since the synthesizability of the molecules may be facilitated by symmetric building blocks, let us finally discuss how a symmetrization step can be added to the action of the genetic operators. After crossover and mutation operations, we symmetrize candidates with respect to the central block. The building blocks on the left side of the center are dominant and mirrored to the right part. The couplings remain unchanged.

An evolution run with this additional block symmetrization step is depicted in Fig. 9. The fitness values in Fig. 9(a) show a similar behavior as the runs displayed in Fig. 5: A sharp increase in fitness in the first generations is followed by a saturation of the mean value, and evolved candidates outperform the randomly generated individuals. In contrast to the runs in Fig. 5, we stop the evolution after 41 generations because of the major changes in generation 34. The performance of the best candidate is depicted in Fig. 9(b). A DQI

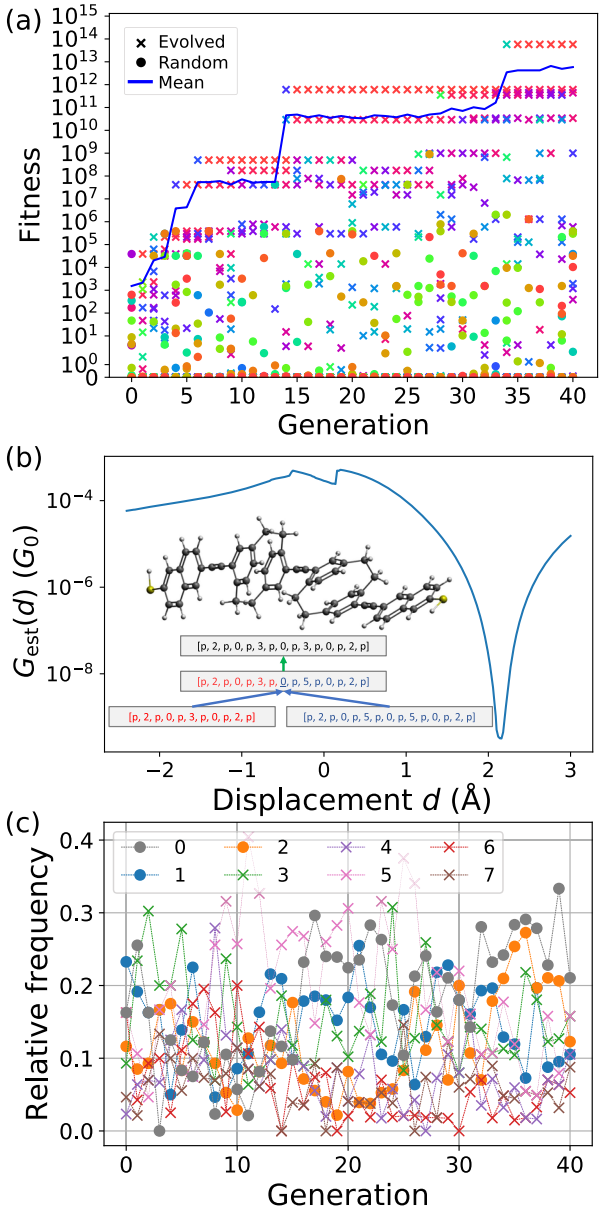

FIG. 9. (a) Fitness values of an evolution run with symmetrization of the molecules as a function of the generation. Parameters used in the fitness functions are $\alpha = 2.5$, $\beta = 1.2$, $\gamma = 2.0$, and $\delta = 0.005$, the population size is 26, the generation limit of 40, and the mutation probability is set to $p_m = 0.5$. For the tournament selection, the best k = 5 individuals are taken to generate offspring. Crosses indicate individuals evolved by selection, crossover and mutation, while dots mark randomly generated individuals. The blue line shows the mean value. (b) G_{est} d plotted against the sulfur-to-sulfur displacement d for the best individual of the evolution run, shown in panel (a). For the purposes of presentation, the stretching resolution was increased to 0.02 ‰per step compared to Δd = 0.1 ‰during the evolution. Additionally, a part of the family register of the best individual is plotted in the inset. Blue arrows indicate crossover operations, while the green arrow denotes symmetrization. The geometric structure is depicted above the family register. (c) Relative frequency of molecular blocks for each generation in the run of panel (a). Different symbols are used to distinguish spacer blocks (dots) from π-stacked building blocks (crosses). Encoding and block numbering follow Fig. 2.

dip of about 5 orders of magnitude is visible. Solely, the median of $G_{\mathrm{est}}\ d$ is rather low. The behavior is graded by a fitness value of around 10^{14}, which is even higher than those of the final candidates shown in Fig. 5. The inset of Fig. 9(b) shows an excerpt of the family register of the best individual. The best candidate is found by a crossover operation, where the head is taken from the left parent and the tail from the right parent. After the symmetrization, the blocks from the left parent are mirrored, while the central block and the length stem from the right parent. Thus, genetic information of both parents is transferred to the offspring, justifying the symmetrization. The geometry of the best molecule of the evolution, displayed above the family register of Fig. 9(b), contains two paracyclophane blocks, block 3, interlinked by acetylene, block 0. This central structure is connected to the anchors via naphthalene spacers, similar to previous runs (see, e.g., Fig. 6). Although not part of the symmetrization procedure, which acts only on the blocks, all couplings are in a para-configuration. As in Figs. 7 and 8, an analysis of frequently occurring blocks shows that blocks 3 and 5 are apparently the most suitable π-stacks, and spacers 0, 1, and 2 all occur with around 10% or more. Among the spacers, block 0 is the most favored motive. Together with these most relevant building blocks, symmetrization reduces the chemical search space further.

B. Influence of spacer blocks

The genetic algorithm has corroborated spacers on π-stacked blocks as a general feature of well-performing mechanosensitive molecules. For this reason, we study their role in more detail in this subsection.

The depth of DQI dips can be enhanced by spacers. Although it is obvious that the minimal conductance drops with molecular length due to the exponential decay in the tunneling regime,[68] it turns out that spacers typically enhance the contrast of the dips, i.e., the ratio between highest and lowest conductance values during stretching.

Let us analyze a simple model with paracyclophane (block 3 in Fig. 1) as the central unit. As illustrated in Figs. 10(a) and 10(b), either we equip it with phenyl spacers (block 1 in Fig. 1) on both sides before attaching it to two semi-infinite one-dimensional gold chains or we connect the gold chains directly without spacers. The system studied in Fig. 10(a) resembles those of Ref. 11. For describing the electronic structure of the molecules in Figs. 10(a) and 10(b), we use a tight-binding model, which follows Ref. 69 and considers just a single p orbital on each carbon atom. The leads are modeled by monoatomic gold chains, and coupling self-energy contributions are treated according to Refs. 14 and 69. During the stretching, the gold chains and covalently connected molecular parts are displaced.

To isolate the influence of phenyl spacers, we study the model of Fig. 10(a) in two different settings: In the first case, we consider electronic interactions between all the parts, marked L, C, and R, as they emerge from the tight-binding parameterization. In the second case, we restrict the interactions between the center C and the spacers L or R to just nearest neighbors, i.e., only the interaction between the covalently connected atoms l and r is described by a nonvanishing matrix element β in the single-particle Hamiltonian. For simplicity, this coupling β is assumed to have the same value as the nearest-neighbor coupling in a ring, corresponding to benzene rings

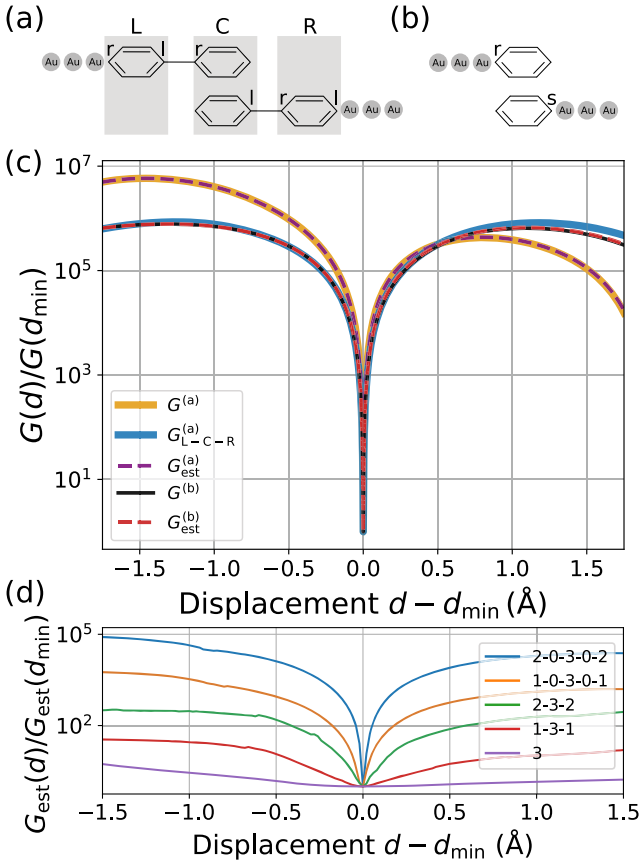

FIG. 10. (a) Paracyclophane with phenyl spacers connected to semi-infinite monoatomic gold chains. The gray boxes and the line connections indicate the restricted electronic interaction range, assumed in the setting referred to as L-C-R. Letters r and l indicate those atoms of the phenyl rings that are covalently linked to another ring or to the electrodes. (b) Paracyclophane directly connected to semi-infinite monoatomic gold chains. (c) Conductance G d or conductance estimate G_{est} d plotted against the electrode displacement d − d_{min}. Conductances are normalized to the respective minimum, which occurs at the distance d_{min}. (d) G_{est} d ~G_{est} d_{min} as a function of displacement d − d_{min} for paracyclophane, equipped with various spacers. The displacement resolution is set to Δd = ±0.02 ‰per step. To save space, the encoding, used in the legend, omits couplings, which are all in para-configuration.

of spacers and the paracyclophane, which are aligned in a plane. Consistent with Eq. (2) and previous sections, we evaluate the transmission probability at the Fermi energy, which we define through molecular HOMO and LUMO levels as $E_{\mathrm{F}} = E_{\mathrm{HOMO}} + E_{\mathrm{LUMO}}$ ~2.

The electrical response of all model systems is summarized in Fig. 10(c) by plotting the distance-dependent conductance, normalized to the respective minimal conductance, $G\ d$ ~$G\ d_{\mathrm{min}}$. Due to normalization, we find that the conductance–distance curve for the model of Fig. 10(a) with only nearest-neighbor L-C-R block interactions is basically identical to the paracyclophane directly attached to the gold electrodes in Fig. 10(b). If we include the full interactions of the central paracyclophane and its spacers L and R for the system in Fig. 10(a), we find however that the switching ratio between the highest and lowest conductance is increased by an order

of magnitude. The enhancement originates mainly from a higher conductance for negative displacements $d - d_{\min}$. A reason might be the increasing overlap of upper and lower decks for decreasing displacements.

In Fig. 10(c), we also show the conductance–distance curves, determined by Eq. (4). Since the factor γ_0^2 drops out, we observe that normalized curves $G_{\text{est}}\ d \sim G_{\text{est}}\ d_{\min}$ show an excellent agreement with the full conductance calculations of the tight-binding model. This holds for both systems depicted in Figs. 10(a) and 10(b). The observations thus justify again our approximations for $G_{\text{est}}\ d$ since the relative depth of conductance dips is well reproduced.

To corroborate the aspects further, we study in Fig. 10(d) the paracyclophane block 3 from Fig. 1 with different spacers attached. We determine $G_{\text{est}}\ d \sim G_{\text{est}}\ d_{\min}$ in this case from DFT electronic structure calculations. For these more advanced calculations, involving multiple orbitals on each atom, the same qualitative behavior is apparent. The switching ratio increases with increasing length of the spacers connected to the core structure. Although this effect provides a more pronounced dip, the overall conductance decreases because of the longer molecules. Therefore, a reasonable trade-off between switching ratio and median conductance needs to be found, which may depend on the particular measurement resolution. Overall, our theoretical modeling shows the crucial influence of the spacer blocks for the contrast of the DQI dip.

IV. FURTHER DISCUSSION

Predicting DQI features in energy-dependent transmission curves based on frontier orbitals of isolated molecules has become a rather straightforward task.[14,64] In order to produce a clearly detectable experimental signal, the DQI dip may be tuned to cross the Fermi energy, and we concentrate here exclusively on mechanical tuning. On the theoretical side, exploration of mechanosensitivity requires the simulation of stretching traces, involving nonequilibrium molecular geometries. Mechanically controlled DQI may however be masked by low conductance or occur at displacements beyond reach by soft metallic electrodes.[11]

Standard calculations of quantum transport through molecular junctions include parts of the electrodes[62] (see also Appendix A). This renders structural relaxation and subsequent transport calculations time-consuming. With our approach, costly trial-and-error cycles for the molecular design are cut short. We provide a systematic search in the chemical space using a fast and accurate fitness evaluation, which considers the most important measures. The fitness evaluation of a molecule takes about 2 h, depending on its complexity and using on average 12 cores of an Epyc-7742 CPU. Massive parallelization within a generation leads to fast evaluation of the search space. An evolution run takes typically about three days using the presented techniques and settings. The extension to a larger database is straightforward.

All optimized molecules show a pronounced DQI dip in the $G_{\text{est}}\ d$ curves, combined with a high mechanical flexibility. Even though some of the optimized structures might have a conductance below the experimentally accessible range, our approach allows the identification of robust mechanisms. Dividing building blocks into π stacks and spacers, we find π stacks to be necessary for the desired mechanosensitive response. Here, paracyclophane (block 3), which is known from previous studies,[8,11] and the stacked naphthalene (block 5) are the most promising elements. Well-performing candidates, involving blocks 3 and 5, were consistently constructed from para-couplings. As another robust feature, we could show that spacer blocks lead to an enhanced mechanosensitive response. This is evident from statistical analysis of block distributions during the evolution [see Figs. 7(b), 8(b), and 9(c)] as well as the model results [see Figs. 10(c) and 10(d)].

Despite the good performance of our genetic algorithm, there are some limitations. Results depend on the fitness function and its parameters, such as $\alpha, \beta, \gamma, \delta$ (see Appendix B), and also the population size and mutation probabilities matter. We constructed the fitness function based on a test set, containing a limited number of candidates from the literature or from internal considerations. We attempted to reward the desired behavior and penalize poorly performing molecules, defining balanced measures to avoid getting stuck in particular regions of chemical space. The paracyclophane-based structure of Ref. 8 was the best performing molecule in our internal test set, showing outstanding performance both experimentally and theoretically. The corresponding fitness value was around $f = 2.9 \times 10^{10}$ for the tuning parameters used in this work (see, e.g., Fig. 5). Despite the excellent performance of this reference molecule, our genetic algorithm was able to find molecules with increased fitness [see Figs. 5(b) and 9(a)] without prior knowledge. Only the evolution in Fig. 5(a) got stuck, creating molecules with slightly smaller fitness values than in the test set. These encouraging results let us assume that our fitness function has sufficient predictive power and avoids overfitting. Based on the chosen building blocks, we cannot expect to significantly outperform the test set.

V. CONCLUSIONS AND OUTLOOK

In this work, we presented a genetic algorithm for the design of mechanosensitive molecules to be applied in single-molecule junction experiments. We considered important degrees of freedom of the overall molecular structure in the genetic encoding through modular molecular building blocks. The fitness evaluation represents a compromise between computational efficiency and accuracy. It involves electronic structure calculations to assess the candidates during the evolution with high precision. The developed fitness function rewards good candidates and penalizes deviations from the desired behavior.

The set of modular molecular building blocks that we explored can be divided into spacer units and π-stacked units. The blocks offered a choice in chemical space that allows us to go beyond what has been synthesized so far. The genetic algorithm revealed that at least one π-stacked unit is needed for molecules to be mechanosensitive. The most favorable π-stacks turn out to be blocks 3 and 5, i.e., the paracyclophane or a stack of two naphthalenes. Additionally, the genetic algorithm highlighted the importance of spacers. The mechanosensitivity of a central π-stacked core can be optimized by attaching blocks 0–2, i.e., acetylene, benzene, or naphthalene, while the direct molecule–electrode connection leads to a broader and less deep DQI dip. We have demonstrated this general property of spacers at a simple tight-binding model. Among the spacers, acetylene is particularly well suited due to its short length and good electronic overlap to attached building blocks.

In contrast to many approaches of artificial intelligence, we can understand molecular design mechanisms in great detail. Evolution

emerges from crossover and mutation. Well-performing molecular blocks prevail, while unsuitable ones are suppressed. This can be seen as the survival of the fittest. Thus, the search space, consisting of around 5.73×10^8 combinations, is automatically reduced to a small fraction of the original size. In this way, the important parts of chemical space can be explored efficiently.

To conclude, using a genetic algorithm we have developed a powerful tool for the design of functional molecules, targeted at single-molecule junction experiments. We proved the performance at the example of mechanosensitivity and discussed the chemical insight that the method offers. Our method can easily be applied to more or other molecular blocks. By a more flexible genetic encoding, for instance through SMILES or SELFIES notations,[70,71] the chemical search space may be substantially extended. By choosing appropriate fitness functions, the methodology can furthermore be adapted to optimize other relevant properties. These might be more sophisticated DQI features, like double dips,[12] or the molecular thermal conductance, which is of significant interest recently[72] and where general design rules are lacking at the moment.

ACKNOWLEDGMENTS

We thank our experimental colleagues Marcel Mayor, Herre van der Zant, Diana Dulić, Nicolas Agraït, and their groups for many stimulating discussions in regular meetings, where we developed experimental and theoretical concepts for mechanosensitive molecules.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Matthias Blaschke: Conceptualization (equal); Investigation (equal); Software (equal); Visualization (equal); Writing – original draft (equal). **Fabian Pauly**: Project administration (equal); Supervision (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: CONDUCTANCE ESTIMATE

To validate the simplified calculation scheme for the conductance, explained in Sec. II D, we compare here the estimated conductance to full quantum transport calculations on the DFT level.[62] As well-studied reference molecules we use ps-*para*-*para* PCP[11] [orange box in Fig. 11(a)] and the paracyclophane core with longer oligophenylene-ethynylene (OPE) spacer groups, long PCP, presented in Ref. 8 [blue box in Fig. 11(a)]. The conductance data for long PCP are reused,[8] but those of ps-*para*-*para* PCP are recalculated. To build the latter junction, we followed the technique described in Ref. 73. In the extended central cluster, shown in Fig. 11(a), the outermost

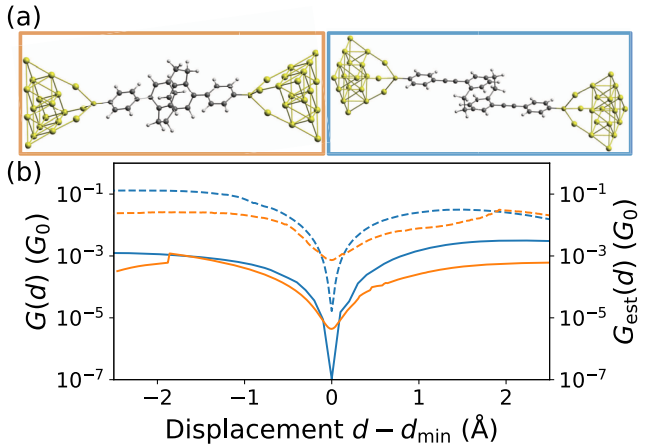

FIG. 11. (a) Single-molecule junction geometries containing ps-para-para PCP (orange)[11] and long PCP (blue),[8] immobilized between two gold electrodes. (b) Comparison of full conductance–distance curves $G(d)$ (solid lines) and estimates $G_{est}(d)$ based on Eq. (4) (dashed lines). Electrodes for the full quantum transport calculations or the sulfur–sulfur distance in the case of the estimate are displaced in steps of ±0.02 ‰

two layers on each side of the electrodes are kept fixed, while the rest of the inner atoms are relaxed. For the stretching, the fixed gold electrodes are separated in steps of ±0.02 ‰, followed by a geometry optimization. Finally, the conductance $G(d)$ is calculated in every step from the DFT electronic structure via Eqs. (2) and (3). We construct the embedding self-energies according to Ref. 62 and use a Fermi energy $E_F = -5.0$ eV. In the DFT calculations with TURBOMOLE,[58] we employ the def-SV(P) Gaussian basis set[59] for all atoms, the PBE exchange–correlation functional,[60] converge energies to better than 10^{-6} hartree ("$scfconv 6"), and relax geometries until the gradient norm is below 10^{-3} a.u. ("gcart 3").

The full conductance–distance curves, $G(d)$, are plotted in Fig. 11(b) together with the corresponding conductance estimate $G_{est}(d)$ for the isolated molecules. Both show the same qualitative behavior. Quantitatively, for both examples the estimate is around two orders of magnitude higher than the full conductance calculations including part of the electrodes in the extended central cluster. However, it should be kept in mind that the junction conductance depends on the precise binding geometry.

For our purposes, it is important that $G_{est}(d)$ reproduces relative changes in the overall conductance between different molecules and that the mechanosensitivity is reliably detected. Figure 11 confirms that this is indeed the case. In the end, molecules are selected based on their fitness, which is a relative measure. It is thus important that the DQI feature, the median conductance and mechanical stiffness are weighted appropriately so that the genetic algorithm ultimately yields molecules with the desired properties.

APPENDIX B: FITNESS CALCULATION

We designed the general form of the fitness function by manually assessing the performance of selected molecules, including molecules from Refs. 8 and 11. On the one hand, a particular test set is needed for the design. On the other hand, the fitness function

should be general enough to detect the desired properties reliably. This means that it must have predictive power because the evolved molecules might outperform the known candidates. The overall fitness in Eq. (6) is split into three parts: The factors f_{DQI} and $f_{\mathrm{med}G}$ describe the electrical conduction behavior, while $f_{\mathrm{stiffness}}$ characterizes the mechanical elasticity. In this way, all desired properties of good mechanosensitive molecule are addressed. Let us now discuss the evaluation of the independent fitness measures f_{DQI}, $f_{\mathrm{med}G}$, and $f_{\mathrm{stiffness}}$ in detail.

For mechanosensitive molecules, the DQI should show up as a sharp dip in $G_{\mathrm{est}}(d)$. We use a fit function to characterize the DQI performance by the extracted fit parameters. Based on our reference calculations, the fit function

$$\frac{g(d)}{G_0} = \frac{G_{\mathrm{est}}(d_{\max})}{G_0} - \exp\left(-b\frac{|d-d_{\min}|}{d_0}\right)\frac{\Delta G_{\mathrm{est}}}{G_0} \tag{B1}$$

turns out to be a suitable choice. As introduced in Fig. 2, d denotes the sulfur–sulfur displacement, $d_{\min}$ $d_{\max}$ is the displacement at the minimum (maximum) value of $G_{\mathrm{est}}(d)$ in the calculated distance range, $\Delta G_{\mathrm{est}} = G_{\mathrm{est}}(d_{\max}) - G_{\mathrm{est}}(d_{\min})$, and $d_0 = 1$ Å. Thus, the DQI is assessed by the single dimensionless fit parameter b. Denoting the median conductance over all computed d as $\mathrm{med}\,G_{\mathrm{est}}(d)$, we define the corresponding fitness value through

$$f_{\mathrm{DQI}} = \left(\frac{b}{\mathrm{med}\,G_{\mathrm{est}}(d) - G_0}\right)^{\alpha}. \tag{B2}$$

The component $f_{\mathrm{med}G}$ of the fitness evaluation is intended to penalize candidates with low overall conductance. However, there is no need to strongly distinguish between molecules, whose conductance surmounts a certain threshold. A suitable measure for the desired behavior resembles the Fermi function given by

$$f_{\mathrm{med}G} = \frac{1}{1+\exp\left(-\beta\ln\left(\mathrm{med}\,G_{\mathrm{est}}(d) - G_0\right) - \gamma\right)}. \tag{B3}$$

We determine the mechanical stiffness through a dimensionless spring constant $a_{\mathrm{stiffness}}$. It is extracted by a harmonic fit to the DFT energies of the form

$$E_{\mathrm{DFT}}(d) = \frac{1}{2}\gamma_0 a_{\mathrm{stiffness}}\frac{d^2}{d_0^2}. \tag{B4}$$

The smaller the spring constant $a_{\mathrm{stiffness}}$, the higher should the fitness $f_{\mathrm{stiffness}}$ be. Therefore, we choose the reciprocal relationship

$$f_{\mathrm{stiffness}} = \frac{1}{a_{\mathrm{stiffness}} + \delta}. \tag{B5}$$

The dimensionless parameter δ ensures that candidates are not overrated for being very soft.

There are several tuning parameters appearing in Eqs. (B2), (B3), and (B5), namely α, β, γ, and δ. Their values are fixed before we start our search with the genetic algorithm. Throughout this paper, we use $\alpha = 2.5$, $\beta = 1.2$, $\gamma = 2.0$, and $\delta = 0.005$.

Figure 12 studies the conductance–distance behavior and corresponding fitness values of four different molecules. Blue and red candidates in Fig. 12(a) are taken from the evolution in Fig. 6 and show desired mechanosensitive properties for $G_{\mathrm{est}}(d)$, whereas the

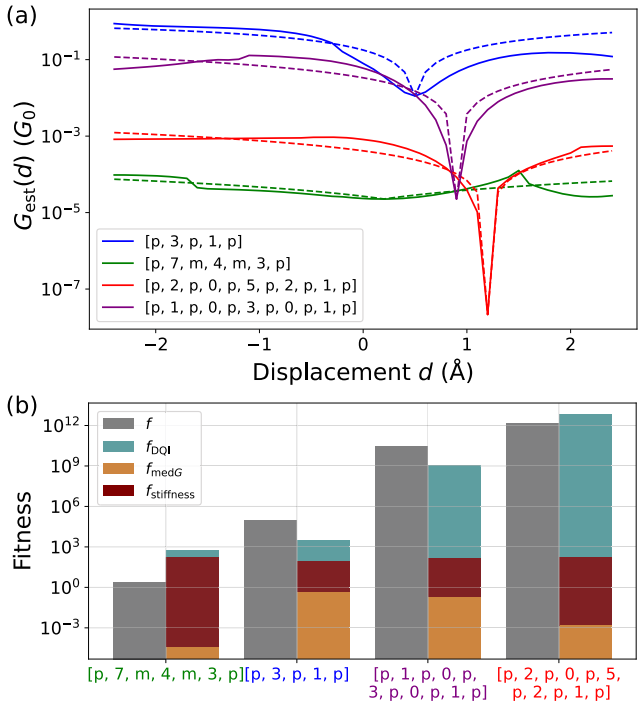

FIG. 12. Analysis of the fitness function for four different molecules. (a) Computed conductance–distance curves $G_{\mathrm{est}}(d)$. Dashed lines show the corresponding fit function (B1) for each molecule. Red, blue, and green candidates originate from the evolution in Fig. 5(b), while the purple candidate is a reference molecule from an experimental study.[8] (b) Overall fitness f shown together with the individual fitness contributions f_{DQI}, $f_{\mathrm{med}G}$, and $f_{\mathrm{stiffness}}$ of the four candidates. The molecules are evaluated using the tuning parameters $\alpha = 2.5$, $\beta = 1.2$, $\gamma = 2.0$, and $\delta = 0.005$.

green candidate features no DQI dip. Additionally, the molecule from Ref. 8 is depicted in purple. The fitness values f_{DQI}, $f_{\mathrm{med}G}$, $f_{\mathrm{stiffness}}$, and f are visualized in Fig. 12(b). The worst overall performance is observed for the green candidate without a DQI dip. The fitness value f_{DQI} is very low, and the low median conductance is penalized further through $f_{\mathrm{med}G}$. Solely, the stiffness contribution is large due to the low spring constant. A better performance is observed for the blue candidate. A weak DQI dip is visible in the $G_{\mathrm{est}}(d)$ plot of Fig. 12(a), which leads to an enhanced factor f_{DQI} in Fig. 12(b). The high median of $G_{\mathrm{est}}(d)$ leads to the largest $f_{\mathrm{med}G}$ in the test set, while the stiffness part is substantially smaller than for the green candidate. The molecule from Ref. 8 is analyzed as a reference for an experimentally available system. A DQI dip of about three orders of magnitude is rewarded by a high f_{DQI}. The high median conductance $\mathrm{med}\,G_{\mathrm{est}}(d)$ leads to a low median penalty. The stiffness measure is comparable to the previous candidate. Importantly, the molecule is outperformed by the red candidate, which is the best individual of the evolution in Fig. 5(b). It is characterized by a high value of f_{DQI} due to the pronounced DQI dip of about four orders of magnitude. The low median conductance is penalized, but the mechanical softness is acknowledged by a high $f_{\mathrm{stiffness}}$. Overall, the performance is rated very high, and it

exceeds even the best performing candidates in our test set for the construction of the fitness evaluation.

A balanced trade-off between DQI dip, median conductance, and mechanical stiffness is important. Tuning parameters or more generally the whole construction of the fitness evaluation can easily be adapted to the needs of a particular design study or experimental requirements.

REFERENCES

[1] I. Osborne, M. Lavine, and R. Coontz, "Looking beyond silicon," Science **327**, 1595 (2010).
[2] S. Wang, X. Liu, and P. Zhou, "The road for 2D semiconductors in the silicon age," Adv. Mater. **34**, 2106886 (2022).
[3] A. Nitzan and M. A. Ratner, "Electron transport in molecular wire junctions," Science **300**, 1384 (2003).
[4] M. Ratner, "A brief history of molecular electronics," Nat. Nanotechnol. **8**, 378 (2013).
[5] S. V. Aradhya and L. Venkataraman, "Single-molecule junctions beyond electronic transport," Nat. Nanotechnol. **8**, 399 (2013).
[6] J. L. Zhang, J. Q. Zhong, J. D. Lin, W. P. Hu, K. Wu, G. Q. Xu, A. T. S. Wee, and W. Chen, "Towards single molecule switches," Chem. Soc. Rev. **44**, 2998 (2015).
[7] D. Xiang, X. Wang, C. Jia, T. Lee, and X. Guo, "Molecular-scale electronics: From concept to function," Chem. Rev. **116**, 4318 (2016).
[8] D. Stefani, K. J. Weiland, M. Skripnik, C. Hsu, M. L. Perrin, M. Mayor, F. Pauly, and H. S. J. van der Zant, "Large conductance variations in a mechanosensitive single-molecule junction," Nano Lett. **18**, 5981 (2018).
[9] J. Li, P. Shen, S. Zhen, C. Tang, Y. Ye, D. Zhou, W. Hong, Z. Zhao, and B. Z. Tang, "Mechanical single-molecule potentiometers with large switching factors from ortho-pentaphenylene foldamers," Nat. Commun. **12**, 167 (2021).
[10] A. Vezzoli, "Mechanoresistive single-molecule junctions," Nanoscale **14**, 2874 (2022).
[11] K. Reznikova, C. Hsu, W. M. Schosser, A. Gallego, K. Beltako, F. Pauly, H. S. J. van der Zant, and M. Mayor, "Substitution pattern controlled quantum interference in [2.2]paracyclophane-based single-molecule junctions," J. Am. Chem. Soc. **143**, 13944 (2021).
[12] W. M. Schosser, C. Hsu, P. Zwick, K. Beltako, D. Dulić, M. Mayor, H. S. J. van der Zant, and F. Pauly, "Mechanical conductance tunability of a porphyrin–cyclophane single-molecule junction," Nanoscale **14**, 984 (2022).
[13] C. Hsu, W. M. Schosser, P. Zwick, D. Dulić, M. Mayor, F. Pauly, and H. S. J. van der Zant, "Mechanical compression in cofacial porphyrin cyclophane pincers," Chem. Sci. **13**, 8017 (2022).
[14] K. Yoshizawa, T. Tada, and A. Staykov, "Orbital views of the electron transport in molecular devices," J. Am. Chem. Soc. **130**, 9406 (2008).
[15] G. C. Solomon, J. Vura-Weis, C. Herrmann, M. R. Wasielewski, and M. A. Ratner, "Understanding coherent transport through π-stacked systems upon spatial dislocation," J. Phys. Chem. B **114**, 14735 (2010).
[16] D. Nozaki, A. Lücke, and W. G. Schmidt, "Molecular orbital rule for quantum interference in weakly coupled dimers: Low-energy giant conductivity switching induced by orbital level crossing," J. Phys. Chem. Lett. **8**, 727 (2017).
[17] V. Venkatasubramanian, K. Chan, and J. M. Caruthers, "Evolutionary design of molecules with desired properties using the genetic algorithm," J. Chem. Inf. Comput. Sci. **35**, 188 (1995).
[18] B. Sanchez-Lengeling and A. Aspuru-Guzik, "Inverse molecular design using machine learning: Generative models for matter engineering," Science **361**, 360 (2018).
[19] A. Nigam, R. Pollice, and A. Aspuru-Guzik, "Parallel tempered genetic algorithm guided by deep neural networks for inverse molecular design," Digital Discovery **1**, 390 (2022).
[20] T. Sousa, J. Correia, V. Pereira, and M. Rocha, "Combining multi-objective evolutionary algorithms with deep generative models towards focused molecular design," in *EvoApplications* (Springer, 2021).
[21] D. Douguet, E. Thoreau, and G. Grassy, "A genetic algorithm for the automated generation of small organic molecules: Drug design using an evolutionary algorithm," J. Comput. Aided Mol. Des. **14**, 449 (2000).
[22] J. H. Jensen, "A graph-based genetic algorithm and generative model/Monte Carlo tree search for the exploration of chemical space," Chem. Sci. **10**, 3567 (2019).
[23] C. Kim, R. Batra, L. Chen, H. Tran, and R. Ramprasad, "Polymer design using genetic algorithm and machine learning," Comput. Mater. Sci. **186**, 110067 (2021).
[24] G. Chalmers, "Introducing ligand GA, a genetic algorithm molecular tool for automated protein inhibitor design," Sci. Rep. **12**, 20877 (2022).
[25] J. O. Spiegel and J. D. Durrant, "AutoGrow4: An open-source genetic algorithm for de novo drug design and lead optimization," J. Cheminf. **12**, 25 (2020).
[26] J. Meyers, B. Fabian, and N. Brown, "*De novo* molecular design and generative models," Drug Discovery Today **26**, 2707 (2021).
[27] J.-F. Zhu, Z.-K. Hao, Q. Liu, Y. Yin, C.-Q. Lu, Z.-Y. Huang, and E.-H. Chen, "Towards exploring large molecular space: An efficient chemical genetic algorithm," J. Comput. Sci. Technol. **37**, 1464 (2022).
[28] M. Sumita, X. Yang, S. Ishihara, R. Tamura, and K. Tsuda, "Hunting for organic molecules with artificial intelligence: Molecules optimized for desired excitation energies," ACS Cent. Sci. **4**, 1126 (2018).
[29] I. Y. Kanal and G. R. Hutchison, "Rapid computational optimization of molecular properties using genetic algorithms: Searching across millions of compounds for organic photovoltaic materials," arXiv:1707.02949.
[30] W. Gao and C. W. Coley, "The synthesizability of molecules proposed by generative models," J. Chem. Inf. Model. **60**, 5714 (2020).
[31] M. Blaschke (2023), "ga_mechano_sens," Zenodo, v.1.0.1, Dataset https://doi.org/10.5281/zenodo.7830800.
[32] I. Gridin, *Learning Genetic Algorithms with Python: Empower the Performance of Machine Learning and AI Models with the Capabilities of a Powerful Search Algorithm* (BPB Publications, 2021).
[33] Y. Fang and J. Li, "A review of tournament selection in genetic programming," in *Advances in Computation and Intelligence*, edited by Z. Cai, C. Hu, Z. Kang and Y. Liu (Springer, 2010).
[34] B. Miller and D. Goldberg, "Genetic algorithms, tournament selection, and the effects of noise," Complex Syst. **9**, 193 (1995).
[35] U. K. Chakraborty, K. Deb, and M. Chakraborty, "Analysis of selection algorithms: A Markov chain approach," Evol. Comput. **4**, 133 (1996).
[36] M. Affenzeller, S. Wagner, S. Winkler, and A. Beham, *Genetic Algorithms and Genetic Programming: Modern Concepts and Practical Applications* (CRC Press, 2009).
[37] M. Srinivas and L. M. Patnaik, "Genetic algorithms: A survey," Computer **27**, 17 (1994).
[38] S. Reeves, B. DiFrancesco, V. Shahani, S. MacKinnon, A. Windemuth, and A. E. Brereton, "Assessing methods and obstacles in chemical space exploration," Appl. AI Lett. **1**, e17 (2020).
[39] D. S. Wigh, J. M. Goodman, and A. A. Lapkin, "A review of molecular representation in the age of machine learning," Wiley Interdiscip. Rev.: Comput. Mol. Sci. **12**, e1603 (2022).
[40] J. Xiong, Z. Xiong, K. Chen, H. Jiang, and M. Zheng, "Graph neural networks for automated de novo drug design," Drug Discovery Today **26**, 1382 (2021).
[41] X. Q. Lewell, D. B. Judd, S. P. Watson, and M. M. Hann, "Recap retrosynthetic combinatorial analysis procedure: A powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry," J. Chem. Inf. Comput. Sci. **38**, 511 (1998).
[42] N. C. Firth, B. Atrash, N. Brown, and J. Blagg, "MOARF, an integrated workflow for multiobjective optimization: Implementation, synthesis, and biological evaluation," J. Chem. Inf. Model. **55**, 1169 (2015).
[43] P. Polishchuk, "CReM: Chemically reasonable mutations framework for structure generation," J. Cheminf. **12**, 28 (2020).
[44] S. Wu, M. T. González, R. Huber, S. Grunder, M. Mayor, C. Schönenberger, and M. Calame, "Molecular junctions based on aromatic coupling," Nat. Nanotechnol. **3**, 569 (2008).

[45]S. T. Schneebeli, M. Kamenetska, Z. Cheng, R. Skouta, R. A. Friesner, L. Venkataraman, and R. Breslow, "Single-molecule conductance through multiple $\pi-\pi$-stacked benzene rings determined with direct electrode-to-benzene ring connections," J. Am. Chem. Soc. **133**, 2136 (2011).
[46]R. Frisenda, V. A. E. C. Janssen, F. C. Grozema, H. S. J. van der Zant, and N. Renaud, "Mechanically controlled quantum interference in individual π-stacked dimers," Nat. Chem. **8**, 1099 (2016).
[47]D. Q. Andrews, G. C. Solomon, R. P. Van Duyne, and M. A. Ratner, "Single molecule electronics: Increasing dynamic range and switching speed using cross-conjugated species," J. Am. Chem. Soc. **130**, 17309 (2008).
[48]D. Q. Andrews, G. C. Solomon, R. H. Goldsmith, T. Hansen, M. R. Wasielewski, R. P. V. Duyne, and M. A. Ratner, "Quantum interference: The structural dependence of electron transmission through model systems and cross-conjugated molecules," J. Phys. Chem. C **112**, 16991 (2008).
[49]G. C. Solomon, D. Q. Andrews, R. H. Goldsmith, T. Hansen, M. R. Wasielewski, R. P. Van Duyne, and M. A. Ratner, "Quantum interference in acyclic systems: Conductance of cross-conjugated molecules," J. Am. Chem. Soc. **130**, 17301 (2008).
[50]Y. Tsuji, T. Stuyver, S. Gunasekaran, and L. Venkataraman, "The influence of linkers on quantum interference: A linker theorem," J. Phys. Chem. C **121**, 14451 (2017).
[51]A. D. McNaught and A. Wilkinson, *Compendium of Chemical Terminology* (Blackwell Science, 1997).
[52]H. Häkkinen, "The gold–sulfur interface at the nanoscale," Nat. Chem. **4**, 443 (2012).
[53]S. G. B. Rylander and B. Gotshall, "Optimal population size and the genetic algorithm," Population **100**, 900 (2002).
[54]Y. Kwon, S. Kang, Y.-S. Choi, and I. Kim, "Evolutionary design of molecules based on deep learning and a genetic algorithm," Sci. Rep. **11**, 17304 (2021).
[55]Z. Zhang and Y.-G. Wang, "Molecular design of dispersed nickel phthalocyanine@nanocarbon hybrid catalyst for active and stable electroreduction of CO_2," J. Phys. Chem. C **125**, 13836 (2021).
[56]S. Grimme, C. Bannwarth, and P. Shushkov, "A robust and accurate tight-binding quantum chemical method for structures, vibrational frequencies, and noncovalent interactions of large molecular systems parametrized for all spd-block elements ($Z = 1$–86)," J. Chem. Theory Comput. **13**, 1989 (2017).
[57]C. Bannwarth, S. Ehlert, and S. Grimme, "GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions," J. Chem. Theory Comput. **15**, 1652 (2019).
[58]S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, and J. M. Yu, "TURBOMOLE: Modular program suite for ab initio quantum-chemical and condensed-matter simulations," J. Chem. Phys. **152**, 184107 (2020).
[59]A. Schäfer, H. Horn, and R. Ahlrichs, "Fully optimized contracted Gaussian basis sets for atoms Li to Kr," J. Chem. Phys. **97**, 2571 (1992).
[60]J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett. **77**, 3865 (1996).
[61]J. C. Cuevas and E. Scheer, *Molecular Electronics*, 2nd ed. (World Scientific, 2017).
[62]F. Pauly, J. K. Viljas, U. Huniar, M. Häfner, S. Wohlthat, M. Bürkle, J. C. Cuevas, and G. Schön, "Cluster-based density-functional approach to quantum transport through molecular and atomic contacts," New J. Phys. **10**, 125019 (2008).
[63]C. J. O. Verzijl, J. S. Seldenthuis, and J. M. Thijssen, "Applicability of the wide-band limit in DFT-based molecular transport calculations," J. Chem. Phys. **138**, 094102 (2013).
[64]K. Yoshizawa, "An orbital rule for electron transport in molecules," Acc. Chem. Res. **45**, 1612 (2012).
[65]M. Strange, C. Rostgaard, H. Häkkinen, and K. S. Thygesen, "Self-consistent GW calculations of electronic transport in thiol- and amine-linked molecular junctions," Phys. Rev. B **83**, 115108 (2011).
[66]L. A. Zotti, M. Bürkle, F. Pauly, W. Lee, K. Kim, W. Jeong, Y. Asai, P. Reddy, and J. C. Cuevas, "Heat dissipation and its relation to thermopower in single-molecule junctions," New J. Phys. **16**, 015004 (2014).
[67]M. J. S. Dewar, "The molecular orbital theory of organic chemistry," Sci. Prog. **40**, 604 (1952).
[68]B. Mann and H. Kuhn, "Tunneling through fatty acid salt monolayers," J. Appl. Phys. **42**, 4398 (1971).
[69]X. Li, A. Staykov, and K. Yoshizawa, "Orbital views on electron-transport properties of cyclophanes: Insight into intermolecular transport," Bull. Chem. Soc. Jpn. **85**, 181 (2012).
[70]D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," J. Chem. Inf. Comput. Sci. **28**, 31 (1988).
[71]M. Krenn, F. Häse, A. Nigam, P. Friederich, and A. Aspuru-Guzik, "Self-referencing embedded strings (SELFIES): A 100% robust molecular string representation," Mach. Learn.: Sci. Technol. **1**, 045024 (2020).
[72]L. Cui, S. Hur, Z. A. Akbar, J. C. Klöckner, W. Jeong, F. Pauly, S.-Y. Jang, P. Reddy, and E. Meyhofer, "Thermal conductance of single-molecule junctions," Nature **572**, 628 (2019).
[73]M. Bürkle, J. K. Viljas, D. Vonlanthen, A. Mishchenko, G. Schön, M. Mayor, T. Wandlowski, and F. Pauly, "Conduction mechanisms in biphenyl dithiol single-molecule junctions," Phys. Rev. B **85**, 075417 (2012).