First-principles molecular structure search with a genetic algorithm

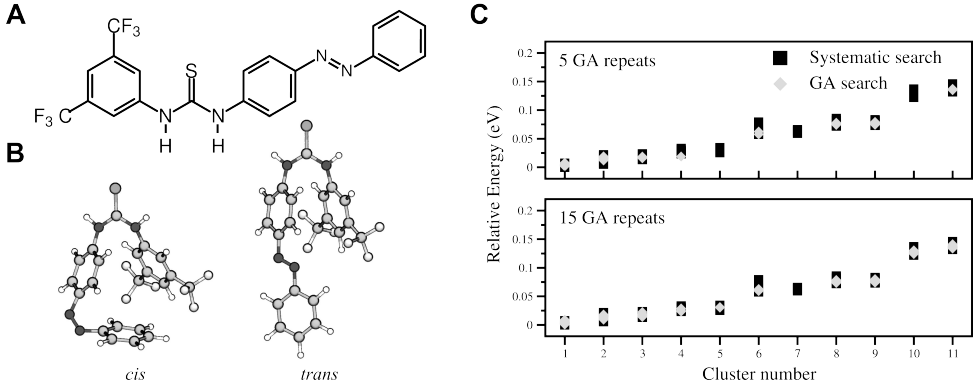
Adriana Supady[1], Volker Blum[1,2], Carsten Baldauf[1], and Matthias Scheffler[1]

[1] Fritz-Haber-Institut der MPG, Berlin, Germany, [2] MEMS, Duke University, Durham, U.S.A.

We present a genetic algorithm (GA) based framework for structural searches of complex molecules based on empirical or first-principles (density-functional theory, DFT) energy functions. The aim is not just to find the single global minimum structure, but also to identify all conformers that appear in the low-energy conformational hierarchy and thus could be experimentally relevant. The use of DFT gives access to rather accurate energy functions and avoids the problem of parameterization for specific classes of chemical compounds.

In our GA search, the geometry of a structure is encoded in a vector of torsional degrees of freedom (TDOF). The initial population of *N* individuals is randomly generated and evaluated by local geometry optimization. Two individuals are selected; the selection probability is a function of the energy. Next, genetic operations are applied: (i) crossing over exchanges parts of the encoding vectors and (ii) mutations randomly assign new values to selected TDOFs. The resulting candidate structures are again evaluated by local relaxation and eventually replace individuals of the previous generation with a higher energy. The algorithm proceeds with a new selection round until a predefined number of iterations or a convergence criterion is met. Generated geometries are first checked for steric clashes and for uniqueness (that is if they were computed already before) based on the root mean square deviation (rmsd) of Cartesian coordinates. This greatly reduces the number of unproductive or redundant calculations, which is especially important when using a first principles energy function.

We demonstrate the principle for an azobenzene-based molecule (the chemical structure is shown in Figure A), finding the conformational energy hierarchy for the *cis* and for the *trans* configurations (the global minima are depicted in Figure B). In order to verify if the conformational energy hierarchy from a systematic search can be reproduced by a GA search, the structures yielded by the systematic search were clustered into 11 clusters and the structures from GA repeats were sorted into these clusters. The accuracy of such a GA prediction is critically linked to the search settings, for example, the number of repeats (illustrated in Figure C). While the global optimum is found very reliably with only a few repeats, the reproduction of the hierarchy is more demanding. Similar testing was performed for seven amino acid dipeptides (Ala, Gly, Val, Leu, Ile, Phe, Trp).



Post-processing of the data, for example the evaluation of the geometrical similarity of the structures and taking into account their energetic relation, allows for visualization of the topology of the potential energy landscape in form of a graph. Such information can be further utilized to identify which pairs of states are likely to be connected by a low-energy barrier. We will use the described strategy to predict functional molecules (by adding a library of side groups) that are tuned for a specific use, e.g., as switchable catalysts of a target chemical reaction.