RAPID COMPUTATIONAL DISCOVERY OF π-CONJUGATED MATERIALS

by

Ilana Yocheved Kanal

B.S. University of Pittsburgh, 2002

Submitted to the Graduate Faculty of

the Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2017

UNIVERSITY OF PITTSBURGH

DIETRICH SCHOOL OF ARTS AND SCIENCES

This dissertation was presented

by

Ilana Yocheved Kanal

It was defended on

March 30, 2017

and approved by

Geoffrey R. Hutchison, Professor, Department of Chemistry

Daniel S. Lambrecht, Professor, Department of Chemistry

Tara Meyer, Professor, Department of Chemistry

John A. Keith, Professor, Department of Chemical and Petroleum Engineering,

Swanson School of Engineering, University of Pittsburgh

Dissertation Director: Geoffrey R. Hutchison, Professor, Department of Chemistry

RAPID COMPUTATIONAL DISCOVERY OF π-CONJUGATED MATERIALS

Ilana Yocheved Kanal, PhD

University of Pittsburgh, 2017

The focus of this thesis is conjugated polymer properties for improved computational discovery of π-conjugated materials. Combination of these materials in differing orders alter the electronic structure and in tetramers, on average, an energy effect is seen. When expanded to hexamers, it became apparent that a more complicated effect exists that depends on block length and placement of that block or sequence within a hexamer. Sequence effects were applied both computationally and experimentally for combinations of benzothiadiazole and phenylene vinylene monomers to confirm importance of sequence in both solar cell performance as sequence can affect intrinsic and bulk properties orthogonally, such as HOMO-LUMO gap.

In addition to sequence study, inverse design of conjugated polymers from computed electronic structure properties demonstrate that while it is unreliable to predict polymer properties from the monomer properties alone, it is very reliable to make predictions from simple models. These models allow for better polymer property predictions without costly polymer calculations.

A large scale computational investigation assessing the utility of common classical force fields for computational screening of low energy conformers provided us with insight for the most reliable methods to use when screening molecules. Using statistical analyses on the energies of up to 250 diverse conformers of 700 different molecular structures, we find that energies and geometries from widely-used classical force fields show poor energy correlation with semiempirical and DFT energies calculated at PM7 geometries. In contrast, semiem-

pirical (PM7) energies show better correlation with DFT calculations. With these results, we make recommendations for more reliably carrying out conformer screening.

Sequence effect, models for polymer predictions and assessment of classical force field methods for low energy conformer predictions are combined to produce our genetic algorithm to rapidly, computationally select materials. Optimization of our genetic algorithm shows that with relatively few calculations, millions of molecules can be screened with a significant speedup compared with brute force calculation of those same molecules.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

When I began this program, not only did I doubt my own ability to go back to school after nearly ten years, but chemistry faculty members doubted me as well, telling me that there was no way that I would be able to complete the requirements to receive a PhD with family responsibilities. After six years, I am preparing to receive my PhD and am thankful for those people who believed in me and helped me attain my goal instead of discouraging me from continuing.

Before beginning the program, I spoke with Dr. Weber, who was very encouraging and made me feel that earning a PhD was a dream that I could fulfill. During my first semesters in the program, Dr. Golde and George Bandik were understanding that I had a family and worked to assign me to teach sections at times of the day when I had childcare available. I would like to thank my advisor, Dr. Geoffrey Hutchison, who has taught me most of what I have learned in graduate school. He provided me with just the right amount of guidance to allow me to learn things on my own without too much frustration about not making progress. I know that many of the early tasks that I performed were just meant as exercises for me to learn basic programming and at the time, these were truly challenges for me, but with time, I have become proficient in scientific programming and have come to really enjoy debugging programs. Our group meetings have also been extremely beneficial in teaching me to convey the knowledge that I have learned to others. There is no comparison between my first group meeting presentations and my current presentations, not only in the slides, but also in my comfort level for discussion of the research topics. Thank you for believing in me!

My family has been my biggest support throughout this journey. Although I can not mention all of you, I do thank all of you for your support. My grandmother, who received

her PhD at the age of 65, inspired me to come back to school to work on my PhD at a *slightly* younger age. My father has been helpful with working through the best way to handle problems as they arose within the university framework. My mother and my sisters have babysat for my children for countless hours, sometimes at inconvenient times, enabling me to complete what was required of me at the university. My five children, Reuven, Yonatan, Yehudis, Shmuel and Betzalel, have shown me that they are around of my accomplishments. Additionally, they have been patient with me as I put in long hours and late nights to complete different deadlines throughout school from courses to teaching to comprehensive exam and research proposal. Finally, I thank Elli, my life partner and husband who encouraged me to begin this journey and has alway supported me each step of the way.

Thank you all for helping me complete my schooling and I am looking forward to the next journey that lies ahead!

1.0 INTRODUCTION

The text and figures in this chapter has been adapted from Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R., Efficient Computational Screening of Organic Polymer Photovoltaics, *The Journal of Physical Chemistry Letters* 2013, 1613-1623.[1] The author's contributions include part of the literature searches performed for the review portion of the paper and the calculations and analysis for the donor-acceptor study.

1.1 BACKGROUND

As the global population increases and world economies develop, energy consumption is increasing at an alarming rate. Currently, the United States derives most of its energy from nonrenewable fossil fuels such as coal, oil and natural gas. Since there is a finite supply of these materials, other renewable energy sources have become of interest. While many forms of renewable energy are currently being investigated, including solar, wind, biomass, geothermal and hydropower, according to the National Renewable Energy Laboratory (NREL) most of our renewable energy comes either directly or indirectly from the sun. The sun, which is always shining on the earth, provides abundant energy that if we are able to use for our needs shows considerable promise for solving the worlds energy issues. For example, in 2004, NREL calculated that to satisfy the United States demand for electricity with solar power, covering only 0.4% of US land mass (10 million acres) with typical commercial solar panels would be needed.[2] Yet in 2011, only 9% of U.S. energy consumption came from renewable sources, 2% of that solar.[3] The economic challenge with existing photovoltaic devices made of inorganic materials is the up-front cost, and the several year lag for a return on investment.[4]

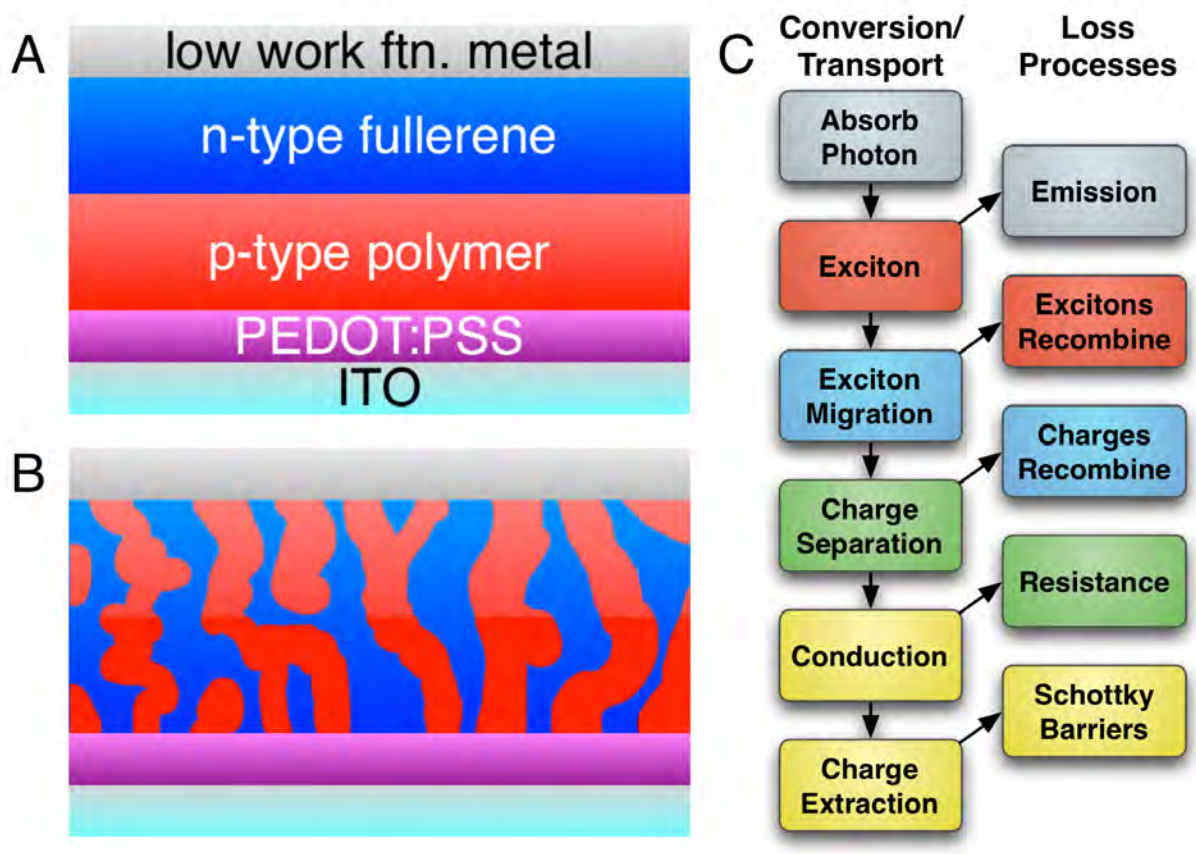


Figure 1.1: (a) Schematic of a simple planar heterojuction cell compared with a (b) conventional bulk heterojuction cell. Note that in reality, multiple mixed phases typically occur at the organic-organic interfaces. (c) Schematic of energy conversion, transport, and loss processes.

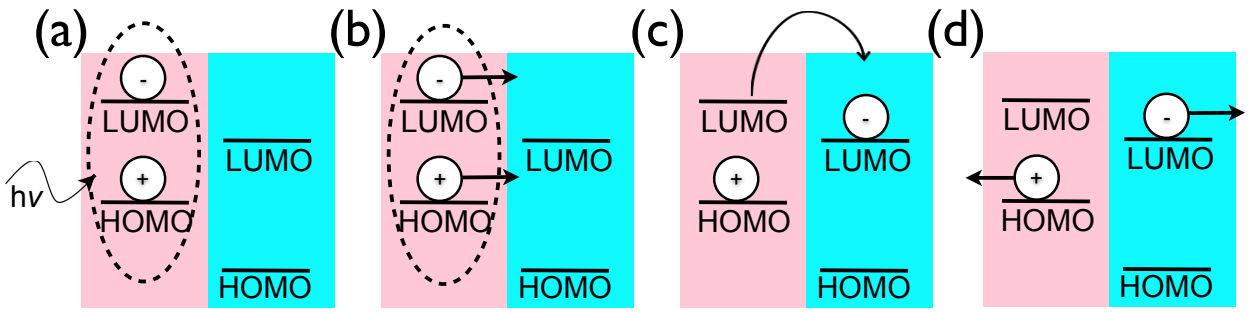


Figure 1.2: Charge separation in OPV materials

1.2 ORGANIC PHOTOVOLTAICS

Organic materials offer the promise of reduced cost through roll-to-roll processing and the high tailorability of synthetic organic chemistry in organic field-effect transistors (OFETs), organic light-emitting diodes (OLEDs), and organic photo-voltaics (OPVs).[5] P3HT and PCBM are common organic photovoltaic materials, but many other promising materials exist. OPVs when many energy conversion and charge transport processes work together, as shown in Figure 1.1c. As shown in Figure 1.2, after a photon is absorbed by the system (a), an electron in the donor material is excited from its ground state to an excited state, generating a hole electron pair (b). These charges want to relax to their ground states, so when the LUMO of the accepting material is higher than the HOMO of the donor material, the electron can relax to the accepting material's HOMO energy level (c). Once charges have separated to different materials, charges can generate energy assuming other loss processes (emission, charge recombination, etc.) do not interfere with the process (d).

The first OPV was reported in 1986 by Tang with an efficiency of 1% and the first bulk heterojunction (BHJ) cell was reported in 1995 by Heeger (Figure 1.1a,b).[6,7] Over the last few years, efficiencies of single-junction and tandem devices have increased slowly, with recent reports in the 8-10% range.[8,9] Multiple factors act as limits on OPV efficiency (Figure 1.1c).[10] While efficiencies continue to increase, truly transformative improvements will likely

require new strategies such as rational design of new materials.[11]

Organic photovoltaic devices typically comprise a p-type conjugated molecule or polymer and an n-type material, usually a substituted fullerene. In this work, we will refer to the p-type phase as a polymer, since such devices predominate, but p-type small molecules are also used. The p-type polymer is typically a co-polymer designed by combining donor (easily oxidized) and acceptor (easily reduced) monomers, providing a narrow band gap and good energy level alignment with the n-type fullerene.[12] The n-type fullerene is often phenyl-C61-butyric acid methyl ester (PC61BM) or phenyl-C71-butyric acid methyl ester (PC71BM). The p-type and n-type materials are then mixed into planar heterojunction or BHJ films (Figure 1.1a,b).

Initial OPVs were fabricated in the planar heterojunction architecture with discrete p- and n-type layers sandwiched between two electrodes. A BHJ cell replaces the separate p-type and n-type layers with a single, bicontinuous network of the polymer and fullerene with domain sizes similar to the polymer exciton diffusion length.[7] By keeping the domains in the active layer of a BHJ around this size, an efficient pathway for charge transport and collection is formed, allowing for increased film thickness.[13,14] For OPVs to generate current, the polymer (and sometimes the fullerene) absorbs photons, generating bound electron-hole pairs (excitons). The excitons diffuse toward the p-n junction where they separate into free charge carriers.[15–18] Electrons are transferred from the polymer to the fullerene through a charge transfer state. In some cases where the fullerene has a significant optical extinction coefficient, holes are transferred from the fullerene to the polymer.[19] The BHJ architecture provides the necessary interpenetrating network of the p- and n-type materials for efficient charge separation (Figure S1). Without appropriate domain sizes, it is possible for excitons to undergo emission or recombination before they are able to migrate to the interface where they can separate.[20–23] There are also many other loss mechanisms that terminate the electron-hole dissociation pathway (Figure 1.1c) including interaction with impurities, defects, and charge traps in the device.[24–27] Due to these many factors, the improving OPV efficiency is a complicated matter.x

1.3 COMPUTATIONALLY DRIVEN MATERIALS DESIGN

Over the last few years, there has been an increasing interest in computationally driven design of new materials, showcased by the Materials Genome Initiative. Announced in 2011, the initiative seeks to foster both scientific and technological advancements and decrease the time required for breakthrough materials to become commercially available.[28] The use of computational screening methods allows "virtual synthesis" and exploration of properties prior to synthesis. Combinatorial materials science is thus a promising technique for quickly surveying a wide array of variables by creating libraries in silico, similar to strategies used in the pharmaceutical industry.[29] Computational exploration of various conjugated thiophenes uncovered a new compound with extremely high hole mobility.[30] Similar computational exploration has also uncovered other experimentally-verified trends affecting charge mobility.[31,32]

Not surprisingly, since the realm of materials research is vast, many different approaches have appeared to efficiently tackle computationally driven materials design and the "Materials Genome" challenges. For example, computational screening approaches have recently been used to explore perovskite metal oxides[33,34] and oxynitrides for water splitting photocatalysts,[35] pseudocapacitive electrodes,[36] refrigerant fluids,[37] chromophores for dye-sensitized solar cells,[38] and rational design of porous metal-organic frameworks.[39–43] Inverse design strategies attempt to locate materials with certain reactivity or other properties by starting with ideal target values and use an algorithm that works backward to find new materials matching the target.[44] Other groups have used machine learning to rapidly estimate properties that would otherwise require computationally-intensive calculations.[45] Finally, when an inorganic or molecular structure is known, often alchemical methods allow efficient optimization of elemental composition,[46–52] although these methods require a fixed scaffold. Following the approach of generating a large in silico library for screening, the Harvard Clean Energy project began with a molecular library of 2.6 million conjugated molecules and corresponding density functional theory calculations, attempting to find materials with ideal optoelectronic properties.[53] A related study used a set of 50 training molecules with known current-voltage device characteristics, and a set of physicochemical descriptors, to

fit empirical models for current-voltage parameters, fill factor, and power conversion efficiency.[54] There are limitations to this simple approach, since the descriptors used may not accurately describe the underlying physics in solar cells, but nonetheless this can be used as a guide for finding new OPV materials, since the empirical descriptor models can quickly filter out poor targets before time-consuming quantum calculations or experiments are performed. Throughout all of these efforts, a key (and often rarely-discussed) issue is the reliable automation of computational resources and analysis across large materials libraries.[55]

One of the most difficult problems in this context of finding ideal p- and n-type materials for OPVs is the size of molecular space, estimated to contain over $\sim 10^{60}$ molecules.[56,57] A search for ideal molecular and polymeric materials is thus similar in spirit, if not application, to computationally driven drug design.[58] One hindrance to rational synthetic design is the inverse design problem; that is, solving for a molecular structure given a particular set of target parameters. Many approaches have been used to solve this problem, including QSARs[59,60], inverse-QSAR,[44,61], combinatorial materials science[29,46–52], data mining libraries[54,62], linear combination of atomic potentials (LCAP)[48,51], Monte Carlo simulations[49] and genetic algorithms (GA).[63]

1.3.1 QSAR

A commonly used modeling technique, first introduced by Corin Hansch[64], QSAR (quantitative structure activity relations) mathematical models relate quantitative parameters (descriptors) derived from the chemical structure to a quantitative measure of a chemical property.[59,60] Due to the widespread use of QSAR modeling for pharmaceutical and biological molecule discovery, QSAR capability is provided in many software packages. As with any method, there are many challenges to the development of reliable predictors with QSAR including failure to consider data heterogeneity, inclusion of confounded descriptors or non-interpretable descriptors, use of incomplete data or overfitting of data which can lead to poor transferability of models to other systems. Additionally, it is essential for researchers to check that the input data is correct and not flawed as this would deem all conclusions from the models flawed. QSAR modeling includes a range of techniques including linear regression,

Random Forest and neural networks which take varying amounts of analysis to provide chemical insight.[65] Inverse design (Inverse-QSAR) strategies attempt to pinpoint materials with certain reactivity or sought after properties by using an algorithm which works backward to find new materials matching the target molecule or chemical properties.[44,61]

1.3.2 Combinatorial Materials Science

As the number and complexity of the molecules of interest to the scientific community grows, the previous techniques of examination of moleulces 'one by one' is increasingly difficult to impossible. Combinatorial materials science is a useful method to rapidly survey a wide array of variables by creating libraries with a very large number of compounds then identifying useful components of the libraries for future moluecules. This approach is similar to drug discovery methods in the pharmaceutical industry.[29] When an inorganic or molecular structure is known, an alchemical method allows efficient optimization of elemental composition when a fixed scaffold is applied.[46–52]

1.3.3 Data Mining

Data mining libraries generate large groups of possible molecules which can then be used to determine which have the chemical properties which correspond to a specific group of properties. The main benefit of this method is that if a library is large enough, it should contain the ideal candidate, but depending on the size of the library, finding that ideal candidate molecule will be difficult. The more expensive problem with the generation of large libraries is the tremendous computational time required to perform all necessary calculations on each molecule within the entire library. The Harvard Clean Energy project solved the problem of immense computational time by having many individuals run the calculations on their personal computers.[54] Other groups have created libraries as well,[62] but the problem of sifting through these large libraries to find the useful information remains challenging.

1.3.4 LCAP

Linear combination of atomic potentials (LCAP) technique transforms the molecular optimization from a discrete optimization problem into a continuous optimization problem and makes a search for a molecule through a library more rapid by using specific chemical properties to search molecular space, thus avoiding evaluation of each molecule in the library.[48,51] Monte Carlo simulations, which rely on random sampling to obtain results, can be combined with the LCAP method to optimize nonlinear optical properties in a class of donor-acceptor substituted benzene and porphyrin frameworks.[49]

1.3.5 Genetic Algorithm

A genetic algorithm is a stochastic method for global optimization problems which uses the concepts of evolution from the biological sciences, specifically natural selection. While other stochastic search methods, such as simulated annealing, threshold acceptance, identify a single solution to a given problem, a genetic algorithm instead uses a population of solutions.

To begin to implement a genetic algorithm, an initial population of chromosomes of the desired size is randomly generated, most often from a random seed. Each chromosome of the initial population is then evaluated to determine how well it fits with the desired requirements. This is referred to as fitness testing. After fitness testing, "bad" chromosomes, as defined by chromosomes a large distance from our desired parameters, are discarded. To prepare for the next generation, new individuals are created by crossover in which aspects of our selected individual chromosomes are combined to get the best characteristics. Genetic algorithms also allow for mutation, in which small changes to the chromosomes can occur. The new population is now complete and can be tested, evaluated, selection, crossed over and mutated until the genetic algorithm selection process is complete and the "best" set of chromosomes has been identified. The genetic algorithm is run multiple times starting with different initial populations by varying the random seed to ensure that the final chromosomes selected represent a global minima or overall best set of chromosomes and not a local minima. The genetic algorithm will converge and identify the same final chromosomes independent of the initial population once sufficient generations have elapsed.

1.4 MULTI-STAGE SCREENING FOR ORGANIC PHOTOVOLTAICS

In our work, we take a slightly different approach. Since the optimization of device efficiency requires a complex interplay of multiple properties (i.e., optical absorption, energy level offsets, exciton dynamics, charge separation and recombination, and charge transport), and subtle changes in polymer structure can lead to huge variations in performance, we believe a multi-step screening and refinement process is required.[66] Importantly, to find optimal or nearly optimal targets for device applications, we seek to survey a large, diverse fraction of molecular space, and eventually establish a set of ∼100-200 highly promising targets. We also hope that at each step of refinement, beyond simply locating promising targets, we will learn structure/function correlations and "design rules" which may have applications outside of solar electricity generation.

To efficiently sample millions of molecules without generating the entire set of structures, the early steps in our pipeline must be as fast as possible. Indeed, approaches such as branch-and-bound and genetic algorithms (GAs) can eliminate unproductive categories or subpopulations of candidate structures entirely - the fastest possible computational method is doing nothing at all. Furthermore, as seen in Figure 1.3, we seek to tackle easier problems first, for example the prediction of optoelectronic properties.[63] Using fast semiempirical quantum methods, such as AM1,[67] PM6,[68] and ZINDO/S,[69] we can quickly compute low-energy conformers, optical absorption spectra, notably the lowest energy optical excitation, and ground-state energy levels. Such methods, while not state-of-the-art, nevertheless compute optical excitation energies in conjugated polymers accurate to ±0.3 eV,[70] and ionization potentials to 0.2 eV[63] in a fraction of the time required for DFT and TD-DFT methods. The GA, in turn, spends several generations sampling lower-efficiency oligomers, followed by increased sampling of oligomers predicted to have near-optimal optoelectronic properties. On average, only 4% of the entire pool of oligomers is sampled by the GA, but 60-70% of all oligomers with near-optimal properties, a considerable acceleration over brute force approaches.

As noted in our previous work,[63] the GA functions by selecting oligomers (i.e., varying the molecular structure) to minimize the geometric distance to the optimal HOMO and

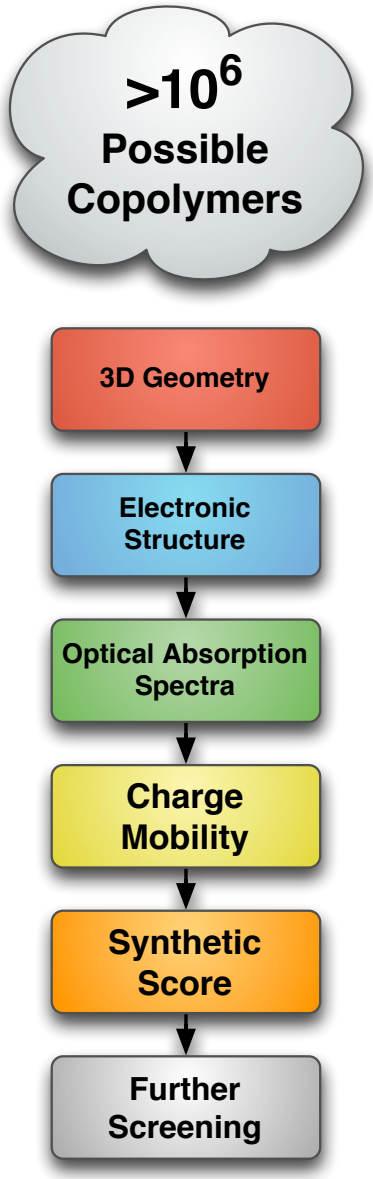


Figure 1.3: Schematic of multi-stage screening pipeline for organic photovoltaics. Note that faster, more reliable methods are performed earlier in the process (i.e., predicting the 3D geometry and conformation of a polymer) and subsequent steps involve more complicated, slower evaluations.

excitation energies, based on the predicted efficiencies by Scharber, et. al.[71] Other optimization targets can be used, as discussed below. Beyond these criteria, the optimization function rewarded oligomers with large computed oscillator strength for the lowest energy singlet transition. In this way, the optimizing function sought oligomers with the desired optoelectronic properties as well as reasonable oscillator strength (and thus high extinction coefficients). An initial population of 64 oligomers was generated from random dimers among the library of 130 monomers. Crossover occurred by selecting two oligomers at random from the fittest set, and swapping their component monomers. Children in subsequent generations were mutated by selecting randomly among the monomers most similar by electronic structure (i.e., HOMO and LUMO orbital energies). This similarity measure was introduced to significantly speed convergence of the GA. In all cases, we ran the GA through 100 generations, although typically 5-6 generations were sufficient to obtain convergence. Oligomers generated after convergence all came from a population with high oscillator strengths and predicted efficiencies (by the Scharber criteria) $8\% \pm 2\%$.

1.5 COMPUTATIONAL METHODS

1.5.1 SMILES

Computational chemical calculations require a method for the computer program understand the three dimensional chemical structure. Several such methods have been introduced, but the most successful and widely used method is Simplified Molecular Input Line Entry System (SMILES). SMILES allows a molecule to be represented through a string of letters and numbers, as shown in Figure 1.4. All atoms are represented using their atomic symbols with the designation that atoms used in compounds without resonance are written as upper case letters and those with resonance are written in lower case letters. This is shown in Figure 1.4 such that the carbon and sulfur atoms within the ring are represented by lower case letters and the carbon and oxygen atoms that are not in the ring is represented by upper case letters.The bond type is an important characteristic to chemicals and thus must be

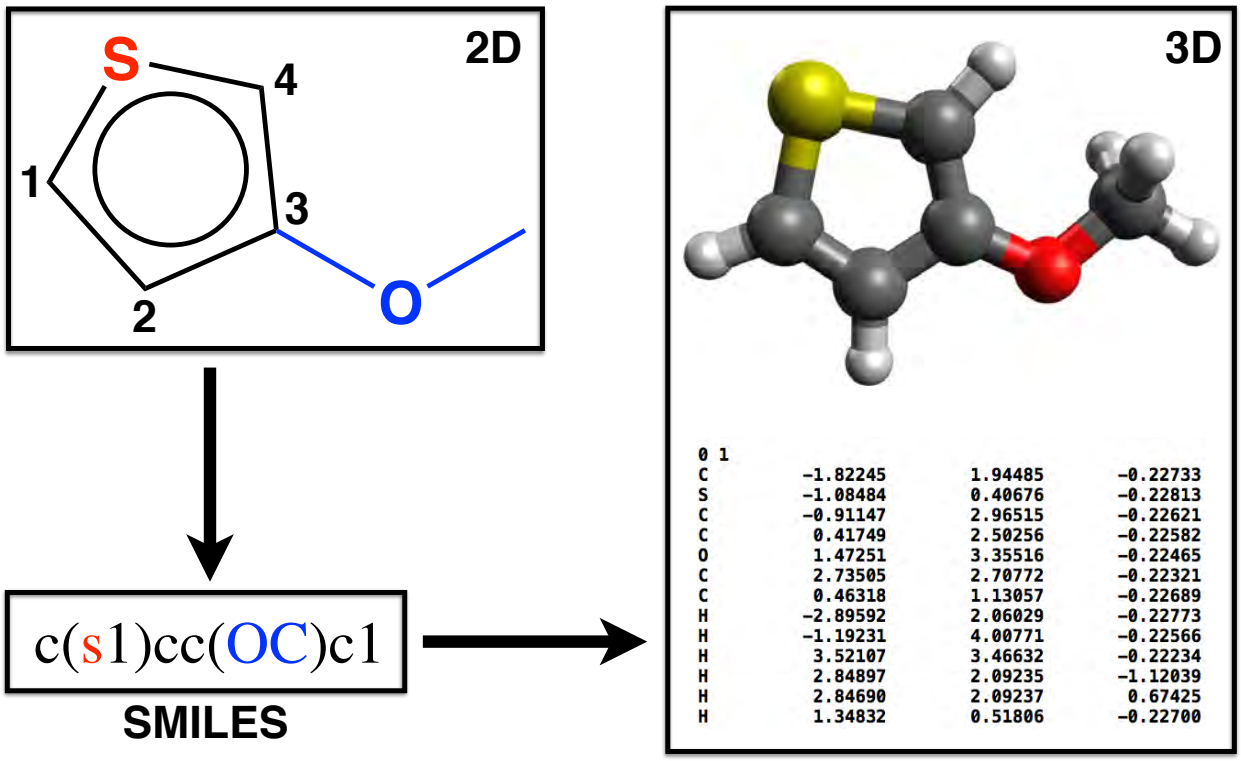


Figure 1.4: SMILES are used to depict a 2D image of a molecule for the computer to be able to generate a 3D structure using programs such as OpenBabel.

identified using symbols that can be read by the computer. Double bonds are represented with the equals sign (=), triple bonds by the pound sign (#), and aromatic systems are designated by using lower case letters (instead of upper case) and the bonds are inserted automatically, thus in the SMILES in Figure 1.4, no bonds are indicated. To efficiently use SMILES, a molecule is examined to find the longest linear string of atoms and branches on the chemical system are written in parentheses next to the atom from which the branch grows.[72] In the systems which are studied in this research, the polymerization sites are important, so it is crucial begin coding each monomer at one polymerization site and end at the other. In addition, numbers are used to represent the beginning and ending atoms used to form rings.

1.5.2 OpenBabel

While SMILES allows a user to denote a chemical structure as string of letters, numbers and symbols, the three-dimensional chemical coordinates and structure are not contained in the SMILES. OpenBabel, an open-source software chemical toolbox, allows the user to interconvert between formats and thus converts a string of letters to a 3D structure as illustrated in Figure 1.4. Within the work described here, many functions of OpenBabel have been used and are cited as needed.

1.5.3 Conformer Searching

Many small molecules are able to adopt a variety of low-energy conformations by rotating around single bonds within that molecule. For computational methods to be reliable, the chosen conformer must match the actual structure of the experimentally determined structures. In solution, a molecule may have many conformers in equilibrium with one another[73] while in the crystal form, a single, lowest energy structure is determined. Many software packages are available to generate conformers some of which are freely available (CONFAB,[74] FROG2,[75] RDKit[76]) and others which are commercial (MOE[77]). Studies have been done to identify tools that most accurately reproduce experimentally determined structures, to examine the diversity of the generated conformational set and to measure the computational

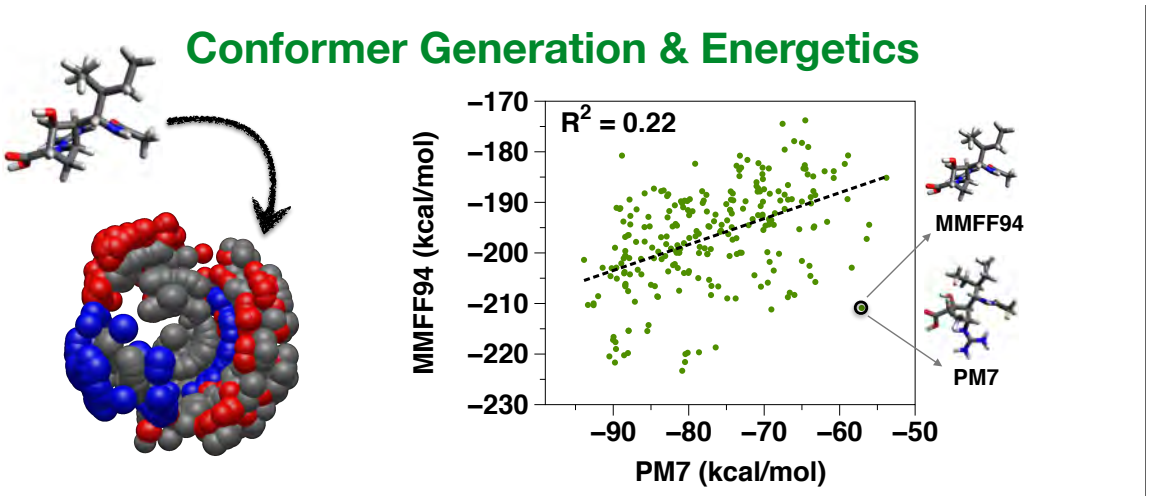


Figure 1.5: Conformers provide different possible structures of a single molecule. Figure from table of contents image produced for the article presented in Chapter 6.

time expended. From one study it was determined that RDKit is a valid free alternative to commercial software to provide similar results at a fraction of the cost.[78]

In order to perform a conformational search, the first step is identification of single, rotatable bonds within the molecule of interest. Since the single bonds are identified, the number of conformers generated will increase with the number of single bonds. Next, the molecule is expanded into conformational space with the chosen algorithm. This step usually varies to torsion angles and keeps the bond lengths and the bond angles fixed. Since many conformers have been produced by this point, it is helpful to check for similar structures. This similarity check is usually performed by comparing the root mean square deviation (RMSD) values. The RMSD value can additionally be used to calculate the similarity between a computationally generated conformer and the experimentally determined structure.[79]

1.5.4 Force Fields

Force fields aim to reproduce molecular properties such as molecular geometries, energies (conformational, stereoisomeric, intermolecular-interaction), vibrational frequencies and

heats of formation in both gas and condensed phase simulations. To this extent, the force field is described as total energy of a system such that:

$$E_{total} = E_{bonded} + E_{nonbonded} = (E_{stretch} + E_{bend} + E_{torsion}) + E_{nonbonded}$$

Many force fields have been established each of which has been optimized for use with different systems. Three common force fields that have been used in this work are UFF, GAFF and MMFF94. The Universal Force Field (UFF) is a general force field which can be used for any element on the periodic table. UFF estimates force field parameters using general rules based on the element, its hybridization and its connectivity, but produces larger errors compared with other available force fields.[80]

General AMBER force field (GAFF)[81] is a force field which was designed for rational drug design. GAFF has parameters for almost all the organic molecules made of C, N, O, H, S, P, F, Cl, Br and I and it is appropriate to use GAFF to study large amounts of molecules in an automatic manner. The error is significantly reduced from the error in calculations using UFF.

The Merck Molecular Force Field, MMFF94, has been parameterized for many organic and bio-organic systems and used high quality data for these parameterizations allowing it to reproduce computational data used in the parameterization very accurately.[82–86] MMFF94 is more accurate that other available force fields and provides a fast first step in a geometry optimization.

1.5.5 Semi-empirical

Semi-empirical quantum chemistry methods are based on the Hartree-Fock formalism but make approximations to decrease the computational time for larger systems.[87] To explain this approximation, imagine a two electron system such as the H_2 molecule. The Schrödinger equation takes into account all interactions between the two electrons and therefore, the number of dimensions on which the wave function depends is three times the number of electrons (3N). In the two electron case, the wave function therefore depends on six dimensions (x^2, y^2, z^2, xy, xz, yz). Although the Schrödinger equation is solvable for the two electron system,

it quickly becomes too large to solve due to electron coupling. The Hartree-Fock method attempts to simplify the Schrödinger equation by assuming that the electrons do not interact with one another (i.e. no coupling), reducing the wave function to a product of orbitals (or one electron wave functions) and adding a correction term which assumes that each electron sees the other electrons as an average field[88]. The Hartree-Fock (HF) equation, used in semi-empirical methods is:

$$\left\{-\frac{1}{2}\sum_{i=1,2}\nabla^2 + v_s(\mathbf{r})\right\}\phi(\mathbf{r}) = \epsilon\phi(\mathbf{r})$$

where:

$$v_s(\mathbf{r}) = \mathbf{v_{ext}}(\mathbf{r}) + \frac{\mathbf{1}}{\mathbf{2}}\int \mathbf{d^3r'}\frac{\mathbf{n(r')}}{\mathbf{|r-r'|}}$$

The term added to the external potential term (v_{ext}) is the correction term which accounts for ignoring the electron coupling. For small systems, the HF equations accurately predict the minimum position for the molecule, but underbinds the energy by failing to include the correlation energy.

Semi-empirical methods approximate the the molecular integrals which reduces the computational time from N^4 for HF methods to N^2.[89] Although many semi-empirical methods exist, three have been employed in this work, PM6, PM7 and ZINDO. ZINDO, is a development of the Intermediate Neglect of Differential Overlap Approximation (INDO) method which was further developed by Michael Zerner. INDO neglects two-centre two-electron integrals which are not Coulombic as well as those neglected by the NDDO approximation, which include all products of basis functions that depend on the same electron coordinates when located on different atoms and to correct for this, the remaining integrals are made into parameters with their values assigned based on calculations or experimental data.[90] The main improvement in ZINDO (over INDO) is the inclusion of parameters for many more elements than the elements from boron to fluorine, which were included in the INDO parameterization. PM6 is the result of changes made to the NDDO interaction term which led to the parameterization of 70 elements and lowered the average unsigned error between the calculated and reference heats of reaction for many compounds. The most dramatic improvements are observed in compounds containing H, C, N, O, F, P, S, Cl, and Br, although

other elements have seen improvements as well. In addition, slight modifications were made to the core-core interaction term, which increases the accuracy of the approximation when atoms are more than three angstroms apart, better estimates the hydrogen bonding and therefore shows improvement in the prediction of molecular geometries compared with earlier methods. PM7 improves on the work of PM6 with a major improvement in the removal of a large number of non-bonding interactions and improved parameters for certain types of chemical interactions.[68]

1.5.6 Density Functional Models (DFT)

The novel idea in DFT models is to use electron density instead of one electron wave functions in calculations which ignores electron correlation. The sum of the exchange and correlation energies of a uniform gas can be calculated by knowing its density. Three general types of exchange/ correlation functionals are used which include Local (Spin) Density Models (L(S)DA), Generalized Gradient Approximation (GGA) and hybrid models. L(S)DA are based on the assumption that the electron density is constant throughout all space. GGA improves on the LDA model by including the dependence of the density of gradient. Hybrid models incorporate a fraction of the exact exchange energy from the Hartree-Fock model with a fraction of approximated exchange energy, thus improving on the GGA model.[88]

Many models have been developed which build on these principles, including: EDF1, BP and BLYP models which improve on the LDA by specifically accounting for non-uniformity in electron distributions and the B3LYP model, a hybrid model that includes the exchange energy from the Hartree-Fock model and has three adjustable parameters. The B3LYP functional provides better dihedral angles and energies than EDF1, BP and BLYP models, which make it a commonly used functional. New developments in functionals include dispersion correction, which makes corrections to the binding curve produced by B3LYP and range separated hybrid density functional (RSH) which uses different hybrids at different distances as a correction to local density approximation that eliminates the dominant (1/r) part of the long-range self repulsion.

It would be recommended to use DFT for equilibrium and transition state geometries,

calculations involving inorganic and organometallic systems for which other methods are unacceptable and for thermochemical calculations which involve net bond making and breaking and absolute activation energy calculations. One weakness of DFT methods are the linger computational times compared to semi-empirical methods and similar or slightly higher cost than HF methods.

Time-dependent DFT (TDDFT) is used to calculate excitations and construct a spectrum from non-solid systems or to extract the ground state exchange correlation from examining the sum of the excited states.

1.5.7 Basis Sets

Basis sets are used in HF and DFT calculations. As basis sets become more complex, computational time and accuracy increase. Some basis sets use slater type orbitals which consider atomic orbitals while others use linear combinations of gaussian orbitals. Several common basis sets include STO-3G, 3-21G, 6-31G, 6-311G, 6-31G*, 6-31G**, 6-311G*, 6-311G**, 6-31+G*, 6-31+G**, 6-311+G*, 6-311+G**, cc-pVDZ, cc-pVTZ and cc-pVQZ. STO-3G is a slater-type orbital which is described by three Gaussian functions and is known as the minimal basis set, is the simplest atomic orbital representation overall spherical symmetry is maintained by considering orbitals as the functions. This basis set does not describe aspherical molecular environments or electron distributions between atoms (bonds) well. 3-21G, 6-31G and 6-311G are a split valence basis set, which represents the core atomic orbitals by one set of functions and the valence atomic orbitals by two sets of functions. For instance, the representation of 6-31G explains that the core electrons are represented by six gaussians and the valence orbitals split into three and one Gaussian components. Since there is additional valence shell splitting, the result should be higher flexibility. 6-311G splits the valence functions into three parts instead of two parts. Similarly, 6-31G*, 6-31G**, 6-311G* and 6-311G** are a split valence basis set which takes polarization into effect which provides for d-type function on main group elements. The * signifies that heavy (d-type functions) have been included and the ** indicates that p-type polarization functions for hydrogen have been included as well. 6-31+G*, 6-31+G**, 6-311+G* and 6-311+G** incorporate

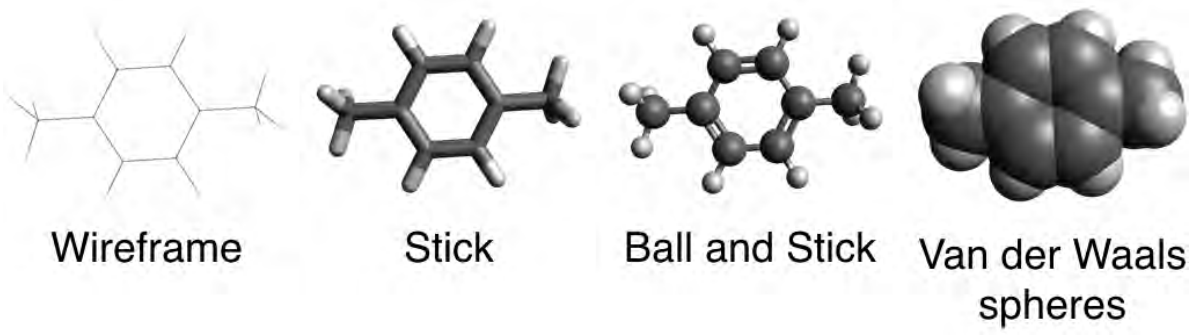


Figure 1.6: Avogadro display types

diffuse functions which provide better results for calculations of anions. The ++ annotation signifies that diffuse functions have been added to hydrogens. Finally, cc-pVDZ, cc-pVTZ and cc-pVQZ: Correlation consistent polarized basis sets which systematically converge with the complete basis set limit. The 'V' stands for valence only basis set, D means double, T indicates triple, Q represents quadruple Zeta. Aug-cc-pVDZ indicates that augmented versions of the other basis sets are used with added diffuse functions. These yield the lowest configuration interaction slater determinants (CISD) ground state atom energies and does a good job of calculating most of the correlation energy in free atoms, although not as computationally efficient as less complex basis sets.

1.5.8 Avogadro

Avogadro is computer program which acts as a molecule editor and visualizer. The software is free and easy to install and allows for each visualization of molecules in many display types such as wireframe, stick, ball and stick and Van der Waals spheres, all of which are useful depending on the application of the desired project. Any molecule, such as 1,4-Dimethylbenzene, as shown in Figure 1.6, can be visualized using different methods. In addition, once a molecule has been generated in Avogadro, it it simple to alter bond lengths, select certain molecules, generate input files to run geometry optimizations or single point calculations and many other features. Built in force fields for geometry optimization include

MMFF94 to be used for organic molecules and drug-like molecules, UFF that can be used on all atoms on the periodic table and Ghemical to be used on simple organic molecules. Visualization of molecular orbitals from an .fchk file is generated is a method employed by these researchers and is described in Chapter 3.[91]

1.6 STATISTICAL METHODS

1.6.1 Analysis of Variance (ANOVA)

Pairwise comparison of groups of numbers is required for many analyses. The Student t which is used to compare two groups of numbers, but performing pairwise Student t tests on many groups would be confusing, if not impossible, depending on the number of groups to be compared. Analysis of variance (ANOVA) is used to compare three or more groups of values to determine whether they come from statistically the same group or if the differences between the groups are statistically significant. The key result of an ANOVA test is the p-values, which provides insight into the population of the data. Generally, if the p-value is less than 0.05, a valid conclusion is that the differences of the means are statistically significant.

1.6.2 Stepwise Regression

Stepwise regression is a method of fitting regression models such that the predictive variables of the model are determined by an automatic process. To perform a stepwise regression, parameters are identified as possible predictive variables. In the forward selection approach, the process begins with no variables in the model and variables are added in during each step to attain a statistically better result. The process of adding variables continues until none of the provided parameters gives a statistically improved result. Conversely, stepwise regression can be performed by starting with all provided parameters and eliminating parameters during each step to find the most statistically significant set of parameters. Most programs allow for bidirectional elimination, combining the forward selection and the backward elimination

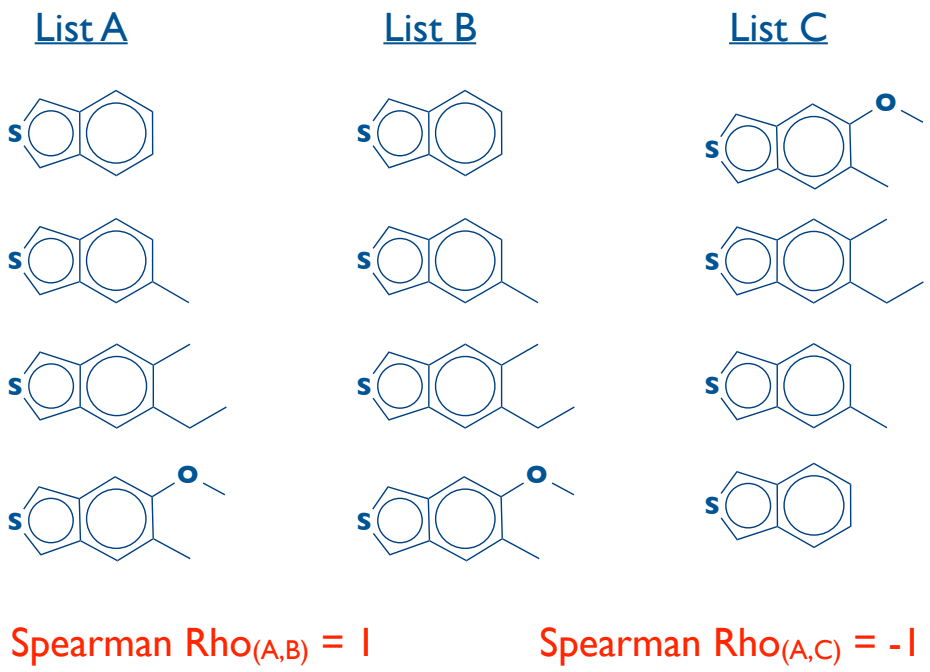


Figure 1.7: Explanation of Spearman Rank Correlation

methods allowing each step of the process to add or subtract parameters until the statistically "best" set of parameters is determined.

1.6.3 Spearman Rank Correlation

The Spearman rank correlation coefficient is a nonparametric measure of how well two variables relate to one another in a monotonic manner. The values of the Spearman correlation range from -1 to 1. A perfect Spearman correlation is +1, and a perfect inverse correlation is -1. If 'Run A' with 'Run B' in Figure 1.7 are compared to one another, it is obvious that these two lists are identical, meaning that they contain the same four monomers that are listed in the same order and therefore, these two lists have a Spearman correlation of +1. If, on the other hand, "Run A' is compared with 'Run C', it is seen that the two runs contain the exact same list of monomers, but the lists are ordered in the reverse order of one another, resulting in a Spearman correlation of -1. In actual data analysis, comparison between two

sets of data rarely give the identical list or the exact order and therefore, decimal values are expected for Spearman correlations.

1.7 PROJECT DESCRIPTION

In this section we give a brief summary of all the work performed and discussed in this thesis. The work has focused on organic photovoltaic materials from selection of computational methods, selection of potential materials and screening through large numbers of materials and sequence combinations in hopes to find the ideal materials with use of a genetic algorithm. In our research, we utilize a GA approach identify likely OPV molecule candidates for improved solar cell efficiency. The main goal of this project was to improve the GA used in our earlier work[63] by increasing the pool of available monomers available for mutation within the GA, allowing novel types of mutations within the GA (e.g., permitting new element substitutions) and screening increasingly large libraries of molecules. Finally, we note that several of the chapters are adaptations of previously published works.

1.8 OVERVIEW

The first four chapters examine polythiophene property prediction, beginning with the inverse design of polymers from computed electronic structure properties. This discussion is continued in the following chapter describing prediction of polythiophene electronic structures though statistical model screening to aid in the efficient discovery of OPV materials. Not only do polythiophene electronic structures help determine polythiophene properties, but fine tuning these properties through sequence tailoring is essential as well. Described in the next two chapters, sequence is an important aspect, the following chapter focuses on a purely theoretical study and the next chapter continues the sequence study by including both computational and experimental work affirming the sequence dependence on electronic properties. The next chapter examines results from our genetic algorithm and its optimiza-

tion to find the amount of time required for reliable results based on data set size. The final chapter examines classical methods for low energy conformer predictions.

2.0 SEQUENCE MATTERS: DETERMINING THE SEQUENCE EFFECT OF ELECTRONIC STRUCTURE PROPERTIES IN π-CONJUGATED POLYMERS

The text in this chapter has been adapted from Kanal, I. Y.; Bechtel, J. S.; Hutchison, G. R., Sequence Matters: Determining the Sequence Effect of Electronic Structure Properties in π-Conjugated Polymers; American Chemical Society: Washington, DC, 2014; pp 379-393 .[92] The author's contribution to the work include performing all chemical calculations, much of the tetramer analysis and all of the hexamer analysis.

2.1 INTRODUCTION

Organic π-conjugated oligomers and polymers have gained both scientific and technological interest for a wide range of potential applications.[93,94] Notably, many properties of these molecules can be tailored to adjust optoelectronic properties, solid-state packing, solubility, and many others, allowing optimization for particular needs. For example, the first organic photovoltaic (OPV) was reported in 1986 by Tang,[6] and device efficiencies have improved significantly by tuning orbital energies, optical band gap, and other properties, even though important challenges remain.[8–10] Current organic photovoltaic polymers employ the donor-acceptor approach in which electron-poor acceptor and electron-rich donor monomers are mixed to create copolymers with the desired optoelectronic properties.[12,95–98] Much effort has been made on finding novel monomers or side-chains to tailor the properties of the resulting copolymers.[99] To facilitate this effort, computational screening methods, including those from our group, have allowed rapid development of both sets of target monomers and new

design principals.[54,63] Nature, on the other hand, creates biopolymers with a fairly limited set of monomers, but instead create complex function with sequence-controlled polymers for protein translation[100] and photosynthesis.[101] Little effort has been made to target sequenced patterns in copolymers for OPV or other organic semiconductor applications.[12,95,96] Typically, these materials involve random order (ADDADA), alternating order (ADADAD), or in simple blocks (AAADDD). Our motivation for this project is to determine whether a sequence effect exists which could allow fine-tuning of HOMO-LUMO band gap without creating complicated monomers. Recent experimental results, suggest that sequence is useful as a strategy to tailor properties of π-conjugated polymers.[102] Using the sequence effect to tune the band gap of polymers for OPVs is an alternative to the standard methods such as modification of monomers.[103] In previous work, we sampled the sequence effect across over a thousand tetramers using density functional theory (DFT) calculations.[1] In this work, we seek to understand the effects of sequence in larger oligomers, notably analyzing over four thousand hexamers. We will use the same set of initial monomers to allow comparisons of the sequence effect as a function of oligomer length with our previous work. We will compare sequence effects in hexamers to those found in tetramers, discuss how these changes can be considered as a function of the proportion or the block-length of the two constituent monomers, and use simplified models such as particle-in-a-box and Hückel theory to explain and predict the effects of monomer sequence on optoelectronic properties.

2.2 METHODS

Undoubtedly, large sequence effects can be found in specific cases with well-chosen monomers, but an important goal is to determine the average sequence effect expected in general. From a pool of 670 monomers, a group of 1,948 polymeric repeat units was generated since many of these fragments could potentially polymerize through multiple sites. Twelve monomers were chosen at random from the homotetramers in the complete monomer set. To ensure sample diversity, the homotetramers were imagined to be in four quadrants (as demarcated with red lines in Figure 2.1(a)) and three monomers were chosen from each section (Figure

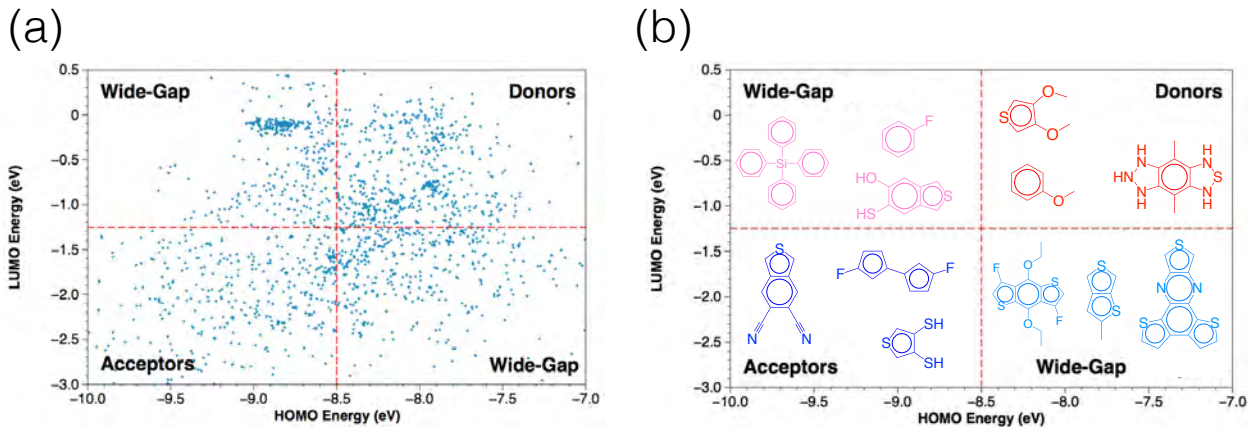


Figure 2.1: (a) Monomer diversity (b) Chosen monomers for sequence study

2.1(b)). Each of these twelve monomers was paired with every other monomer, resulting in 66 monomer pairs. A Python script created all permutations from the SMILES[72] string for each monomer. Each possible tetramer and hexamer sequence was formed, yielding 1,056 tetramer sequences and 4,244 hexamer sequences. From the resulting SMILES string for the co-oligomers, 3D structures were generated using a multistep process. An initial 3D structure was generated using Open Babel 2.3.2[104] (accessed through its Python interface Pybel[105]) and minimized using the MMFF94 force field[82–86] (500 steps using steepest descent minimization, convergence at 1.0 kcal/mol). Next, a weighted-rotor search (MMFF94, 100 iterations, 20 geometry optimization steps) was carried out to find a low-energy conformer. This was then further optimized using MMFF94 (500 steps of conjugate gradient optimization, 1.0 kcal/mol convergence). Finally, Gaussian09 was used to optimize the structure using the PM6[68] semi-empirical method. The Python library cclib[106] was used to extract the HOMO and LUMO eigenvalues. Statistical analysis was performed using RStudio.[107,108] Although HOMO and LUMO eigenvalues are non-physical and do not directly correspond with the oxidation potential or electron affinity, they are common parameters for screening optoelectronic properties in conjugated oligomers.[63,109] In our previous study of the sequence effect, tetramer calculations were calculated using DFT (B3LYP functional),[110,111] but when

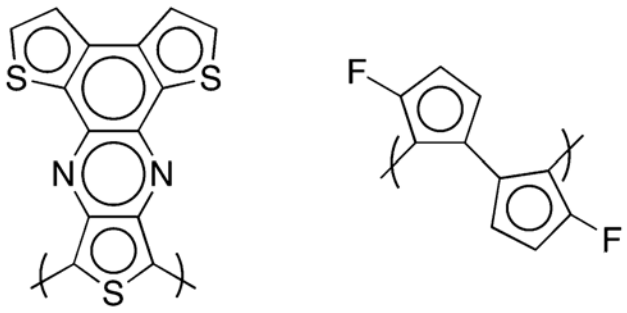


Figure 2.2: Molecules shown to have (left) low and high (right) energies

progressing to the hexamers, the computational time and number of molecules increased drastically, so the semiempirical PM6 method was used for both tetramers and hexamers. The relative errors in the hexamer and tetramer calculations should be approximately equivalent for the comparisons examined. After extracting the PM6-computed orbital eigenvalues, the data set was arranged with each sequence (AAAAAA, AADDAD, DDDDDD, etc.) in a vertical row, while horizontal rows represent a particular monomer combination. Each of these rows was averaged, and the analysis below discusses the average offsets that is, the expected average effect of a particular sequence regardless of the monomer combination.

2.3 RESULTS

2.3.1 Tetramers

In our previous work,[1] trends were established regarding certain monomers effects on the associated tetramer HOMO-LUMO energy band gaps. Regardless of its coupling agent, a monomer with a five fused aromatic ring structure, trithieno[3,4-b:2,3- f:3,2-h]- quinoxaline (Figure 2.2, left), consistently produced the lowest band gap across all sequence pat-

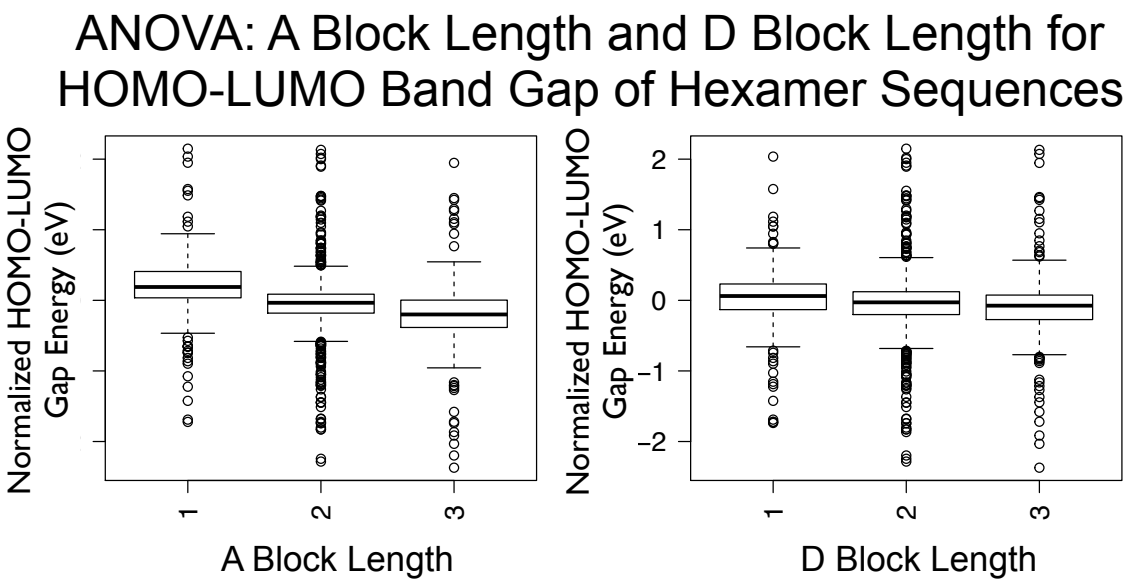


Figure 2.3: ANOVA plots for A and D block lengths for HOMO-LUMO band gap energies.

terns, independent to which monomer it was coupled. Contrarily, coupled 4,4-difluoro-[1,1-bi(cyclopentane)]-1,1,4,4-tetraene (Figure 2.2, right) tended to raise the band gap of its associated tetramers. This was an interesting finding, but demonstrated an already established idea: certain monomers make better photovoltaic devices than others.[99,112]

Beyond the effects of particular monomers on tetramer properties, sequence-dependent phenomena were also examined to identify patterns that can predict average energies of monomer sequences. Calculated HOMO, LUMO, and band gap values for each tetramer were averaged across the 66 monomer combinations. We then take the individual sequences (e.g., ADDA, DADD) and compute the normalized offsets compared to the average HOMO, LUMO, and band gap, expected from the particular monomer combinations. For example, we find that sequences with only single "A" monomers (i.e., a block length of one) have slightly higher normalized HOMO-LUMO gap energies than those with AAA blocks (Figure 2.3). We will use these normalized offsets as measures of the sequence effects below.

Several clear patterns emerge from the normalized offsets. For example, LUMO and HOMO-LUMO band gap energies show, on average, a drastic jump between the ADDD and

	Tetramer	Hexamer
Avg HOMO Energy Spread (eV)	0.11 ± 0.03	0.18 ± 0.06
Avg LUMO Energy Spread (eV)	0.30 ± 0.06	0.39 ± 0.07
Avg HOMO-LUMO Gap Energy Spread (eV)	0.40 ± 0.08	0.57 ± 0.09
50% 'A', 'D' HOMO Energy Spread (eV)	0.10 ± 0.02	0.16 ± 0.05
50% 'A', 'D' LUMO Energy Spread (eV)	0.28 ± 0.05	0.34 ± 0.06
50% 'A', 'D' HOMO-LUMO Gap Energy Spread (eV)	0.36 ± 0.06	0.49 ± 0.07

Table 2.1: Range of HOMO, LUMO and HOMO-LUMO gaps as a function of all sequences

	p-value	
Equal A/D Composition Data	**Tetramer**	**Hexamer**
HOMO	0.0064	1.4×10^{-7}
LUMO	3.0×10^{-6}	2.1×10^{-11}
HOMO-LUMO Gap	9.0×10^{-7}	1.9×10^{-14}
HOMO vs A Block Length	0.00026	7.9×10^{-13}
HOMO vs D Block Length	0.95	0.14
LUMO vs A Block Length	1.3×10^{-7}	$<2 \times 10^{-16}$
LUMO vs D Block Length	0.51	0.0015
HOMO-LUMO Gap vs A Block Length	3.4×10^{-8}	$<2 \times 10^{-16}$
HOMO-LUMO Gap vs D Block Length	0.62	0.0024
HOMO vs A+D Block Length	A: 2.5×10^{-4} D: 0.17	A: 7.2×10^{-13} D: 0.016
LUMO vs A+D Block Length	A: 1.3×10^{-7} D: 0.24	A: $<2 \times 10^{-16}$ D: 0.21
HOMO-LUMO Gap vs A+D Block Length	A: 3.2×10^{-8} D: 0.15	A: $<2 \times 10^{-16}$ D: 0.054

Table 2.2: Statistical p-values for some of the patterns studied

DAAA sequences. HOMO energies show the opposite trend with an energy decrease between the ADDD and DAAA sequences. DDDD sequence values are consistently higher for the HOMO, LUMO, and band gap energy values. These observations support the hypothesis that sequence order affects the HOMO, LUMO, and band gap energies. To confirm our hypothesis, an analysis of variance (ANOVA) was performed for the sequenced averaged band gap energies, which showed a statistically significant difference in means among the sixteen sequence permutations (Tables 2.1, 2.2).

Since properties will depend on the fractions of A and D, the six sequenced tetramers with equal composition of A and D monomers (i.e., AADD, ADAD, ADDA, DAAD, DADA, and DDAA) were compared. An increase in HOMO energy of approximately 0.2 eV between the ADDA and DAAD sequences and an associated decrease in LUMO and band gap energy of approximately 0.2 eV between ADDA to DAAD is observed. Note that because not all of the monomers are symmetric, sequences such as AADD and DDAA are not necessarily geometrically equivalent. Additionally, the A block length changes the HOMO, LUMO, and HOMO-LUMO gap energies, but the D block length has little effect on the energies, as shown in Figure 2.3. With the tetramer set, however, it is not possible to separate between a block-length effect and a general position-dependent sequence effect.

2.3.2 Hexamers

Hexamers were produced using the same method as described for tetramers with the same twelve monomers to provide meaningful comparison. The number of sequence combinations in the analysis increases from 16 to 64, which in turn increases the number of calculations from 1,056 (tetramers) to 4,244 (hexamers). This provides more meaningful statistical results due to a significantly larger sample size. An important question is whether the observed effect was a sequence effect or a block length effect. As with the tetramer sequences, hexamer sequences with equal A and D composition were examined. As shown in Table 1, analysis of tetramers and hexamers with fifty percent composition suggest that the energy spreads of HOMO eigenvalues increase slightly from tetramers (0.10 $\pm$ 0.02 eV) to hexamers (0.16 $\pm$ 0.05 eV). LUMO eigenvalue ranges showed a similar trend (tetramer: 0.28 $\pm$ 0.05 eV;

hexamer: 0.34 ± 0.06 eV). The HOMO-LUMO gap showed the greatest difference, with the tetramer spread 0.36 ± 0.06 eV increasing to 0.49 ± 0.07 eV for the hexamers. These results were found to be statistically significant (Table 2.2).

This shows that on average it should be possible to tune the hexamer band gap by about 0.5 eV by varying the order in which the monomer units are combined in sequences, even with equal amounts of the two different monomers. The tetramers and hexamers with the highest band gaps are ADDA and ADDADA, respectively. The tetramers and hexamers with the lowest band gaps are DAAD and DDAAAD, respectively. In order to determine if the observed variation in HOMO, LUMO and HOMO-LUMO band gap energies are derived from a sequence effect, rather than a simple block-length effect, we studied the dependence of tetramer and hexamer energy on block length (A vs AA vs AAA, etc). This correlation was verified by examining block length for all species (AAAA, AAAD, AADA, AADD, etc.) and observing that, for all possible combinations of tetramers and hexamers, this relationship still exists. The analysis indicates that the length of the acceptor (A) chain is a statistically significant factor in HOMO, LUMO, and HOMO-LUMO band gap energy values.

It is not possible from the tetramer study alone to determine if there is a sequence effect independent of a block length effect, but the hexamers, with more combinations to consider, suggest that the block length and its placement within the hexamer (end, between alternating units, middle, etc.) both affect the energy. Surprisingly, the donor (D) block length is not correlated in a statistically significant manner to the HOMO, LUMO, or HOMO-LUMO energy values.

2.4 DISCUSSION

In most cases, π-conjugated polymers and oligomers are considered excellent examples of the simple one-dimensional particle-in-a-box model, also known as the free-electron molecular orbital (FEMO), model with an infinite barrier height, and the potential inside the box is zero. Kuhn expanded on FEMO by introducing a one dimensional potential inside the box with a sine function as the potential energy, taking into account the difference in bond

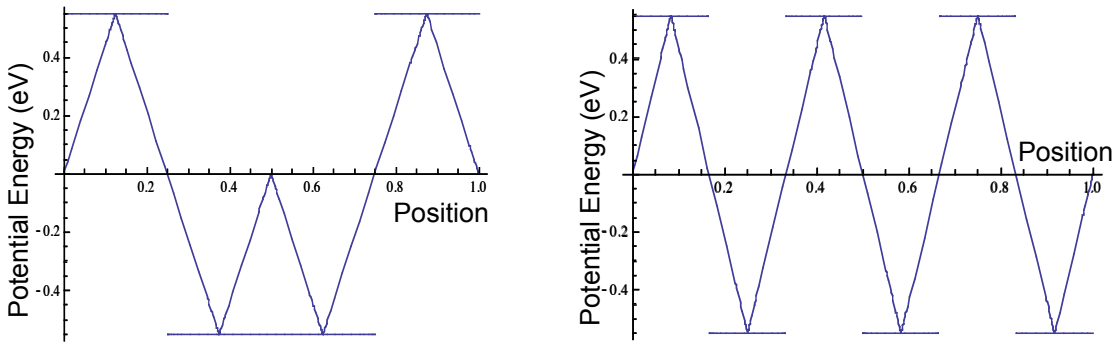


Figure 2.4: DAAD (left) and DADADA (right) step and triangle potential functions, overlaid.

length between single and double bonds and explaining the affinity for electrons to not be equally distributed.[113] The Kuhn model effectively captures the effective conjugation length, saturation of electronic properties and the finite band gap.[114]

Rather than a sine potential, one can imagine conjugated sequenced oligomers as a particle in a box with the donor monomer represented by a positive potential and acceptor monomers by a negative potential (i.e., accepting electrons). Two piecewise functions, either step potentials or triangle potentials, were applied as V(x) in the Schrödinger equation, and their shapes are depicted in Figure 2.4 for both tetramers and hexamers. With defined potential functions and boundary conditions (where the solution is zero on the boundaries), the eigenvalue value problem was solved for tetramers with equal A and D composition. When the two potential systems are compared, it is concluded that their results are consistent with one another, but with the triangle potentials showing slightly smaller effects, as shown in Figure 2.5. In these idealized PIB simulations, the sequences are exactly symmetric, even though as discussed earlier in PM6 calculations, sequences such as AADD and DDAA are not necessarily geometrically equivalent due to asymmetric monomers or conformational effects. The PIB model shows that different perturbations of the first energy level are observed with varying A and D arrangements (Figure 2.5). Each of the different sequences shows a slightly different first energy level perturbation and the wave function shows greater perturbation at the ends than in the middle. The tetramers show that the sequence had an effect (Figure

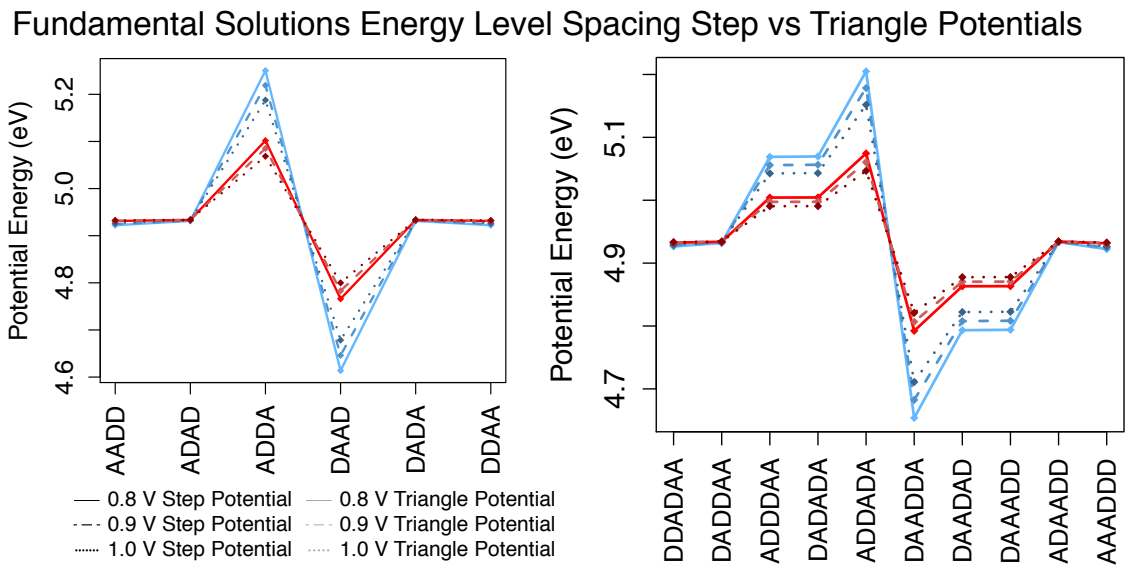


Figure 2.5: Energy level comparison of solutions for step (blue) and triangle (red) functions.

2.5). The tetramer with the highest energy in the PIB model is ADDA. Interestingly, the hexamer with the highest energy is ADDADA, which has the same ADDA sequence as seen in the tetramer in the terminal position of the sequence and exhibits no other AA or DD pairing. Other hexamer sequences that have the ADDA sequence are AADDAD and ADDAAD, in which there is either additional AA pairing, decreasing the energy, or the ADDA sequence is not in the terminal position and therefore exhibits lower energy. This result confirms the conclusion from the PM6 data that showed that the AA pairing is statistically influential to the energy of the sequence. In addition, the PM6 hexamer sequence that shows the highest energy is the ADDADA sequence, which further confirms the result since each of these methods is different, but estimate that the sequence which has the highest energy as the same sequence. The tetramer with the lowest energy in the PIB model is DAAD and the hexamer with the lowest energy is DAADDA. Similar to the hexamer highest energy case, the DAAD sequence is in the terminal position, but in contrast to the highest energy case, this hexamer contains a DD pair. The other sequences that contain the DAAD chain are DAADAD, with no additional pairing and thus has slightly higher energy, and DDAADA,

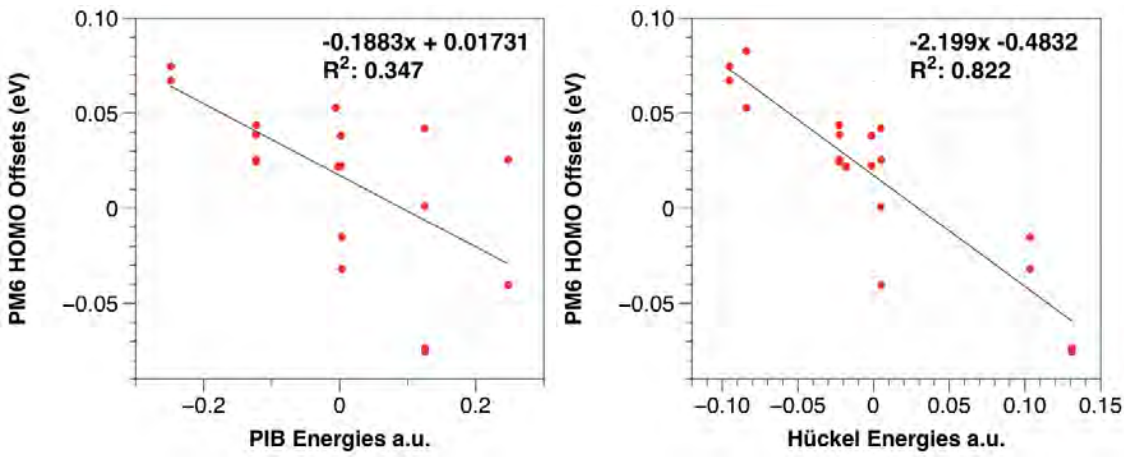


Figure 2.6: Correlation between PIB & Hückel model energies relative to PM6 energy offsets

which does not have the DAAD sequence in the terminal position and the energy is not the lowest. These relative sequence orderings from the particle-in-abox model do not match the lowest energy sequence from the PM6 calculations; the lowest energy sequence was DDAAAD. The PIB model trends do match the general trend from the PM6 data in that the D block length does not statistically affect energy values. This suggests that most of the trends between the PM6 calculated values and the PIB model values will show a weak correlation (i.e., an R^2 value of 0.35 as in Figure 2.6). This led to creation of a Hückel model to better mimic the trends seen in the hexamer data.

2.5 THE HÜCKEL MODEL

Beyond a simple PIB model, π-conjugated polymers are frequently treated using a π-electron Hückel model. Based on the poor correlation between the PIB treatment of the sequenced hexamers and the PM6 results, a similar Hückel treatment was used, with each site reflecting an A or D monomer. Since the Hückel model requires α (site energy) and β (electronic coupling) parameters, we extracted the average difference in HOMO energy between A and

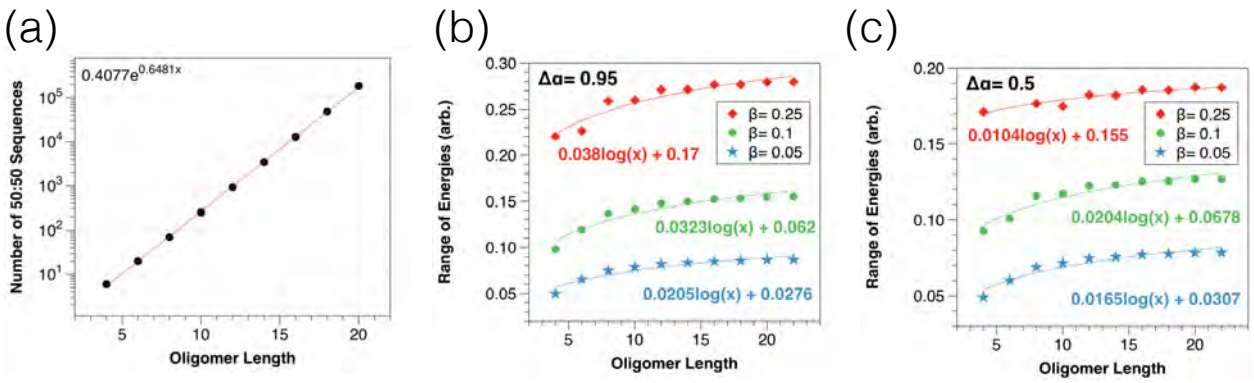


Figure 2.7: Hückel calculation results with respect to oligomer length

D monomers for all 66 combinations from the PM6 calculations (0.95 eV) and the coupling for homodimers AA or DD for all 12 monomers (0.25 eV). With these parameters, we find an outstanding correlation between the PM6-computed average HOMO offsets for each hexamer sequence and the Hückel predictions (Figure 2.6).

Since the Hückel model only requires solving a system of linear equations, we can use this computationally efficient method to explore the effects of sequence on longer oligomers. Scanning from tetramers to 24-mers, while the number of 50:50 sequences increases exponentially to over 2.6 million candidates, the range of HOMO energies increases only logarithmically (Figure 2.7). Thus, for optoelectronic properties, there is little reason to synthesize sequences beyond 6-8 monomer units, since the variations in HOMO energies will saturate. A pattern also emerges for the sequences with the most negative Hückel-computed HOMO energy (i.e., hardest to oxidize) and the least negative HOMO energy (i.e., easiest to oxidize) as compiled in Table 2.3. Remember that the slope of the PM6-Hückel correlation is negative, so the most stable, easiest to oxidize sequence would be the pattern (DD)x(AAAA)2x(DD)x where the subscripts sum to the total number of monomers (n). Since the most delocalized wavefunction in the PIB picture or Hückel model will have highest amplitude in the center of the -conjugated oligomer, the A monomers will have the largest contribution in the center of the wavefunction, and the D monomers will have the smallest perturbation on the edges.

Oligomer Size	Lowest Eigenvalue Sequence	Highest Eigenvalue Sequence
4	DAAD	ADDA
6	DDAAAD	ADDADA
8	DDAAAADD	ADADDADA
10	DDAAAAADDD	ADADADDADA
12	DDDAAAAAADDD	ADADADDADADA
14	DDDAAAAAAADDDD	ADADADADDADADA
16	DDDDAAAAAAAADDDD	ADADADADDADADADA

Table 2.3: Compilation of sequences with lowest and highest Hückel-computed eigenvalues

Indeed, the hexamer sequence DDAAAD is found to have one of the two highest averaged PM6-computed HOMO energies. The Hückel model, since it neglects atomistic detail is imperfect, and the highest PM6-computed HOMO energy is found instead for AAADDD (i.e., the ordering is reversed between PM6 and Hückel).

Conversely, the least stable, hardest to oxidize sequence would be the pattern (ADAD)x (DADA)x, again with the subscripts summing to the total number of monomers. The alternating pattern of A and D monomers introduces frequent barriers that disrupt the delocalized HOMO, and the sequence inverts, creating a kink exactly at the midpoint of the wavefunction. Indeed, the tetramer ADDA and hexamer ADADDA are found to be extrema from the averaged PM6-computed HOMO energies.

Finally, the Hückel model allows exploring different ranges of parameters. For example, the average ΔE for the PM6-computed monomer HOMO energies was 0.95 eV, which was used for the difference in β site energies in the Hückel results described above (with β = 0.25). This is clearly a large difference in orbital site energies, and in Figure 2.7, the range of orbital energies created by sequence variation decreases with decreasing β parameter. In other words, for a large energy difference between A and D monomers, a large β (i.e., highly delocalized) is needed to obtain significant variations in orbital energies from the sequence effect. On the other hand, for smaller differences in β (0.5 eV) between A and D sites, a smaller β (i.e., more localized) yields the largest variation in orbital energies (Figure 2.7).

This suggests two different regimes to obtain maximal variations of electronic structure from the sequence effect: for a large difference in monomer orbital energies (e.g., strong donor, strong acceptor) attempt to maximize the delocalization and electronic coupling between the monomers, but for smaller variations in monomer orbital energies, attempt instead to maximize the localization between the monomers.

When the study was expanded to hexamers, a more complicated effect, which depends on block length and placement of that block or sequence within a hexamer emerged. Block length and placement impact the energy as shown through the general PIB model and the more detailed Hückel model and verified by the PM6 calculations. The result is encouraging since the PIB model does not take into account any details about the system. Two potential sequences of interest to study experimentally are ADDA and DAAD, which when arranged in the terminal or middle position would verify whether the changes in energy are seen in experiment. This would lead to new ways of exploring polymers in the hope of finding an ideal OPV material. In addition, these sequences can be applied to the screening project to further verify the results (with other monomers) and help with the mutation of polymers in the search for the ideal band gap for effective charge transfer.

2.6 CONCLUSION

The results from the tetramers suggested that the sequence in which the monomers are arranged on average has an effect on the energy. When expanded to hexamers, it is apparent that there is a more complicated effect that depends on block length and placement of that block or sequence within a hexamer. The fact that the block length and placement seem to have an impact on the energy as shown through the general PIB and Hückel models and verified by the PM6 calculations is encouraging since neither model takes into account the detailed molecular structure of the monomers or the conformational preferences in particular species. For future applications of π-conjugated polymers, both monomer design and sequence control will yield the most significant control over electronic structure properties. Since sequence also controls polymer conformation and packing motifs, we believe further

investigation is needed to predict and control charge transport and other related solid-state properties.

2.7 ACKNOWLEDGEMENTS

This work was supported by the University of Pittsburgh Center for Energy and Department of Chemistry. Computational resources were provided in part by the University of Pittsburgh Center for Simulation and Modeling.

3.0 SEQUENCE EFFECTS IN DONOR-ACCEPTOR OLIGOMERIC SEMICONDUCTORS COMPRISING BENZOTHIADIAZOLE AND PHENYLENE VINYLENE MONOMERS

The text in this chapter has been adapted from Zhang, S.; Bauer, N. E.; Kanal, I.Y.; You, W.; Hutchison, G.R. and Meyer, T.Y., *Sequence Effects in Donor-Acceptor Oligomeric Semiconductors Comprising Benzothiadiazole and Phenylene Vinylene Monomers*; Macromolecules, 2017, 50(1), pp 151-161 (DOI: 10.1021/acs.macromol.6b02215).[115] The author's contribution to the work include performing all theoretical chemical calculations and analysis of computational work.

3.1 INTRODUCTION

The power and potential of conjugated organic materials stems from their rich diversity and ease of tailoring key properties including optical band gap, absorption and emission intensities, packing, and charge transport properties.[116–119] Applications include photovoltaics, efficient organic light-emitting displays, photocatalytic systems, polymer batteries and supercapacitors, and more.[12,120–126] While there has been a dominant focus on polymeric systems, more recent scientific efforts have demonstrated that oligomers, with complete control over chain length, chain ends, and chemical purity, offer unique advantages.

Controlling monomer sequence is an approach that is increasingly used to engineer properties in copolymers, but has not been widely exploited in conjugated systems.[63,92,102] Instead, researchers have largely focused on designing increasingly sophisticated repeat units,[8,116] tailoring side-chains,[99,127] and combining electron-rich and electron-poor monomers

(donor-acceptor strategy).[12,95–97,128,129] Some efforts have also focused on the use of end-group modification to control p- and n-type majority carrier transport, oxidation and reduction potentials, and also optical properties.[124,130] Sequence remains largely unexplored in these conjugated materials, however, despite increasing evidence from non-conjugated materials that controlling monomer order is possible and that sequence-based differences in properties can be documented.[100,131–140]

We are interested in the application of the sequence strategy to conjugated oligomers and polymers and in studying the effects of sequence on properties related to the use of these materials in photovoltaic devices. While rare, there have been some promising examples of sequence effects reported previously.[132,141–144] Liang and coworkers reported, for example, that two isomeric conjugated oligomers with different sequences exhibited power conversion efficiencies that were significantly different, 4.53% vs. 1.58%.[144] Sequence-based differences in morphology were also observed by Palermo, et al. in their investigation of thiophene and selenophene polymers with gradient sequence, block, and random structures.[142] The influence of sequence on properties, particularly photophysics, was also established by Noonan, and coworkers for periodic copolymers comprising sequences of furan, thiophene, and selenophene.[145]

These intriguing reports have inspired our own interest in developing a more detailed understanding of the influence of sequence on copolymer properties through the systematic preparation, characterization, and modeling of sequenced conjugated oligomers and polymers. In a prior study we described the response of oxidation potentials, HOMO energies, and band gaps on the order of two monomers: an un-substituted and a dialkoxy-substituted phenylene vinlyene.[102] In tetramers, we found that the optical band gaps could be tuned over a range of 0.2 eV, based only on sequence and the coupling of sequence with end group effects. Interestingly, the sequence was found to be important despite the fact that both monomers are electronically similar.

In the present investigation, we explore the role of sequence in determining the electronic properties in a more electronically differentiated donor-acceptor pair of monomers: dialkoxy-substituted phenylene vinylenes (electron-rich, P) and benzothiadiazole vinylenes (electron-poor, B). While these monomers have been widely investigated for applications in OLED

and solar cells,[146–151] the effects of monomer order have not been probed outside of our preliminary communication.[152] In this early report, we characterized two trimers prepared from P and B units and incorporated these trimers into polymeric structures. Herein, we extend the study to tetrameric oligomers, examine carefully the complex relationship of end group effects, and explore the effect of sequence on hole-conductivity and solar cell performance.

3.2 EXPERIMENTAL

3.2.1 General materials

Br-P-CHO, Phos-P-CN, Br-P-Br, Phos-B-CN, Br-PP-CHO, Br-PB-CN, Br-PB-CHO, Br-BP-CN, Br-BP-CHO, Br-PBP-Br, C8-PBP-C8, Br-BPP-Br, and C8-BPP-C8 were synthesized as described previously.[102,152] Phos-B-Br and Br-B-CHO were prepared according to the method of Jorgenson, et al[153] and Lin, et al[130], respectively. DIBAL-H (1.0 M in hexanes) was purchased from Aldrich and dispensed using air-sensitive techniques. LiCl was stored in a 120 °C oven for at least 24 h before use. Dry THF from Sigma Aldrich was used for all reactions. CH_2Cl_2 was dried by passage through an alumina-packed column. All other reagents and solvents were used as received. Column chromatography was carried out on standard grade silica gel (60 $\AA$ pore size, 40-63 mm particle size), which was purchased and used as received.

3.2.2 Spectroscopy

3.2.2.1 NMR Spectroscopy 1H (400 and 500 MHz) and 13C (100, 125 and 150 MHz) NMR spectra were recorded on Bruker spectrometers. Chemical shifts were referenced to residual 1H or 13C signals in deuterated solvents (7.26 and 77.0 ppm, respectively, for $CDCl_3$ and 5.32 and 54.0 ppm, respectively, for CD_2Cl_2).

3.2.2.2 Mass Spectrometry

High resolution mass spectra were recorded on EI-quadrupole or ESI-TOF instruments in the Mass Spectrometry Facility of the University of Pittsburgh. MALDI spectra were recorded on Voyager-DE PRO instrument.

3.2.2.3 Optical Spectroscopy

Solution ($CHCl_3$) UV/VIS absorption spectra were recorded on a Perkin Elmer Lambda 9 UV/VIS/NIR spectrometer. UV/VIS absorption spectra of films on glass substrates were recorded on an Ocean Optics HR2000+CG-UV-NIR high-resolution spectrometer. Solution ($CHCl_3$) emission spectra were recorded on a Varian Cary Eclipse Fluorimeter.

3.2.3 Electrochemistry

Cyclic voltammetry (CV) and differential pulse voltammetry (DPV) were performed on a CHI Electrochemical Workstation Model 430a (Austin, TX) collected using a three electrode system consisting of a glassy carbon disk (3 mm diameter) as working electrode, a non-aqueous Ag/Ag+ reference electrode (1 mM $AgNO_3$ in acetonitrile), and a Pt-wire as auxiliary electrode in 0.1 M Bu_4NPF_6 in dry THF. CV were recorded at 100 mV/s. DPV parameters were as follows: scan rate of 25 mV/s, pulse amplitude 0.05 V and pulse period 0.16 seconds.

3.2.4 Computational Methods

Each possible trimer and tetramer sequence permutation was generated with a python script from the monomer SMILES.[72] An initial 3D structure was generated using Open Babel 2.3.0[104] (accessed through Pybel[105]) and was minimized using the MMFF94 force field[82–86] to find a low energy minima conformation. Final geometries were optimized using Gaussian 09[154] with density functional theory (DFT) B3LYP/6-31G*.[110,111] To compare computational results with electrochemical experiments, redox potentials were determined using a combination of orbital energies (i.e., vertical ionization potential and electron affinity) and the ΔSCF procedure, taking the adiabatic energy difference between the optimized geometries of neutral and charged species using the conductor polarizable continuum model (C-

PCM) model for tetrahydrofuran (THF).[155] To compare with optical absorptions, excitation energies and oscillator strengths were computed using ZINDO[69] and TDDFT using the optimized solution geometry of the neutral species using the C-PCM solvation model[156] for $CHCl_3$. Images of molecules and orbitals were prepared using Avogadro.[91]

3.3 RESULTS

3.3.1 Synthesis

A series of conjugated oligomers with varying sequences were prepared by connecting two units, benzothiadiazole (B) and 2,5-dihexylalkoxy-substituted phenylene (P), with vinylene linkers (Figure 3.1a). The oligomers comprised dimers, trimers and tetramers, based on the total number of P/B units, and bore either two bromo (Br) end groups, one Br and one cyano (CN) end group, or two α-olefinic alkyl groups (C8). Species with reactive end groups including aldehyde (CHO) and dimethyl phosphonate (Phos) were also prepared as synthetic intermediates. Oligomers are named throughout by listing their P/B sequence and end groups, e.g., Br-PB-CN. Oligomers were assembled from a set of building block monomers by sequential Horner-Wadsworth-Emmons (HWE) reactions as described previously (Figure 3.1b).[152] Nitrile-terminated oligomers were prepared for subsequent additions by reductive conversion to the aldehyde functionality. Using this approach, two dimers, six trimers and six tetramers were prepared (Table 3.1).

3.3.2 Optical and Electronic Properties

The optical and electrochemical properties of the sequenced oligomers were determined and are presented in Table 3.1 and Figures 3.2 and 3.3. Unsurprisingly, the absorption maxima show red-shift with increasing oligomer length, from dimers (429-450 nm), to trimers (458-479 nm), and lastly, to tetramers (490-530 nm). Emissions likewise shift towards longer wavelengths; and band gaps, both optical and electrochemical, narrow as expected with increasing conjugation length. Although it is challenging to deconvolute the end group effects

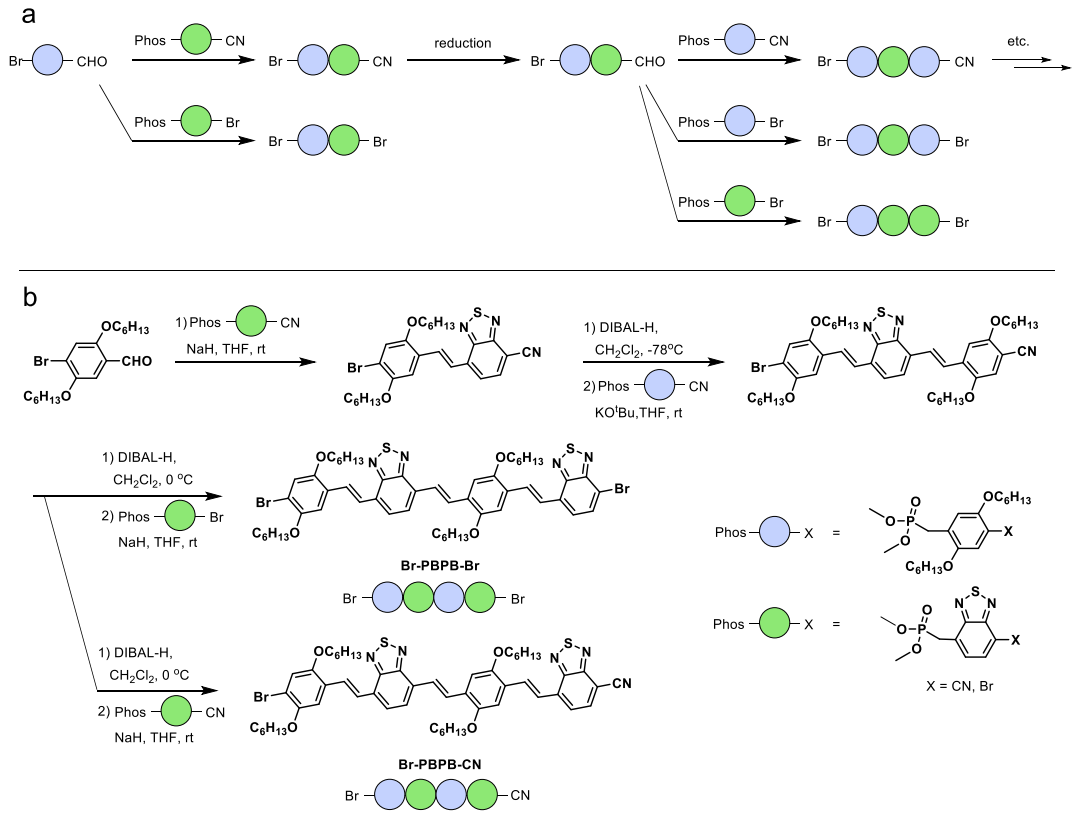


Figure 3.1: Optical and electrochemical data for sequenced oligomers

Oligomer [e]	λ_{max}^{abs} [a] / nm	λ_{max}^{em} [a] / nm	E_{gap}^{opt} [b] / V	E_{peak}^{ox} [c] / V	E_{peak}^{red} [c] / V	ΔE_{gap}^{ec} [d] / eV
Br-PB-Br	432	576	2.97	1.05	-1.50	2.55
Br-PB-CN	450	593		-	-	-
C8-PB-C8	429	585	2.98	-	-	-
Br-BPB-Br	479	-	-	-	-	-
Br-BPP-Br	464	628	2.29	0.77	-1.44	2.21
Br-PBP-Br	476	594	2.28	0.89	-1.47	2.36
Br-PBP-CN	467	583	2.29	-	-	-
Br-PPB-CN	498	658	2.14	-	-	-
Br-BPP-CN	458	609	2.33	-	-	-
C8-BPP-C8[47]	448	613	-	-	-	
C8-PBP-C8[47]	489	618	-	-	-	-
Br-BPPB-Br	493	639	2.19	0.65	-1.45	2.10
Br-PBPB-Br	507	613	2.15	0.75	-1.44	2.19
Br-PBPB-CN	523	702	2.07	-	-	-
Br-PPBB-Br	508	637	2.10	0.71	-1.31	2.02
Br-PPBB-CN	530	707	1.99	-	-	-
Br-PBBP-Br	512	595	2.13	0.82	-1.31	2.13

[a] Measured in $CHCl_3$ (1.0×10^{-5} M); [b] Determined at the onset of absorption spectra; [c] Potential vs. Ag/Ag+, 240 mM in 0.1 M Bu_4NPF_6 in THF; [d] Determined as $\Delta E_{gap}^{ec} = e(E_{peak}^{ox} - E_{peak}^{red})$; [e] B: benzothiadiazole unit, P: 2,5-dihexylalkoxy substituted phenylene units, Br: bromo end group, CN: cyano end group.

Table 3.1: Optical and electrochemical data for sequenced oligomers

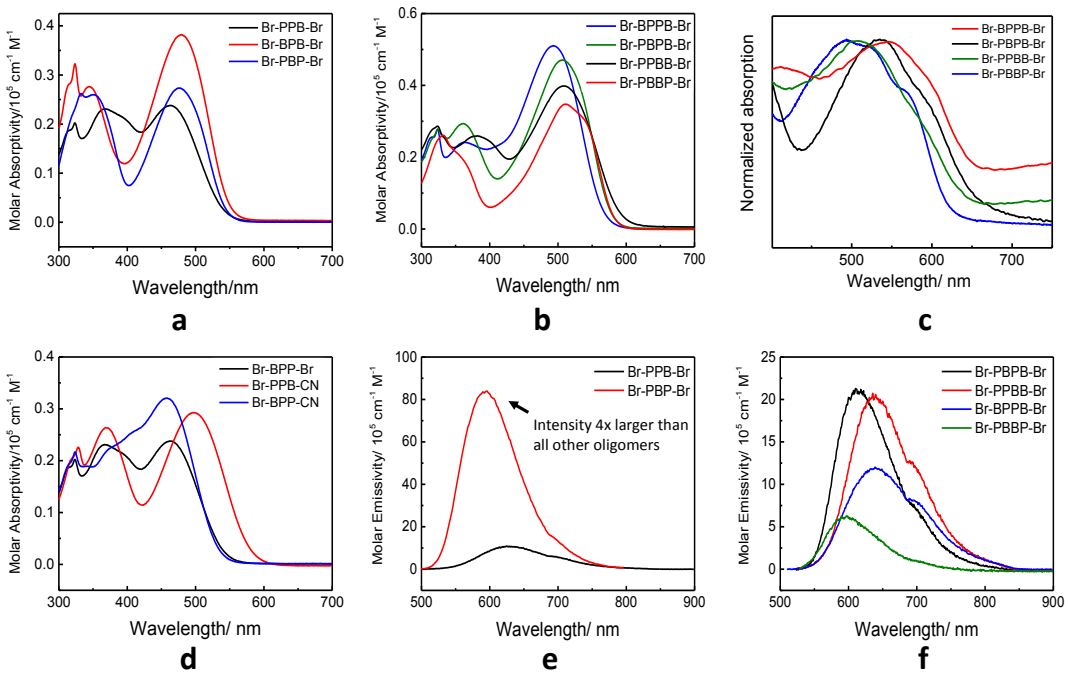


Figure 3.2: Absorption and emission spectra in $CHCl_3$: (a) absorption spectra for all dibromo trimers; (b) absorption spectra for all dibromo tetramers; (c) film absorption spectra of PB tetramers, cast from chloroform solution (d) absorption spectra for BPP trimers bearing cyano and bromo end groups; (e) emission spectra for selected trimers; (f) emission spectra for dibromo tetramers.

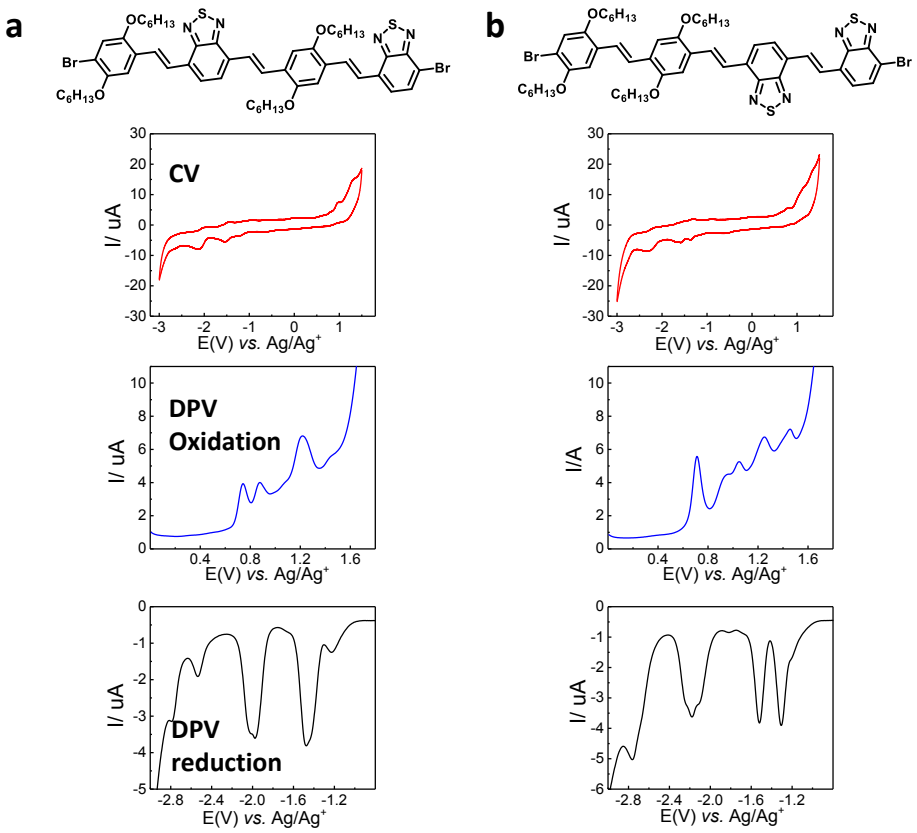


Figure 3.3: Sample cyclic voltammograms and differential pulse voltammograms of (a) Br-PBPB-Br and (b) Br-PPBB-Br

from the sequence effects in these oligomeric structures, by studying a range of examples we were able to understand the trends and focus our attention on bromo end groups which exhibited a minimal perturbation. In considering end group effects it is important to understand that terminal monomers are distinct from internal monomers due to the neighboring free space, independent of the identity of the functional end group. As tetramers comprise 50% terminal monomers and 50% internal monomers, many changes in sequence will necessarily involve changes in the terminal monomers as well.

Consistent with our earlier studies on sequenced phenylene vinylene oligomers, the effect of the unsaturated, electron-withdrawing cyano substituent was profound and depended significantly on the identity of the terminal monomer to which it was attached. Comparing two oligomers that have the same inherent sequence, BPP, but reversed end groups, Br-BPP-CN vs. Br-PPB-CN (=CN-BPP-Br), it was observed that the λ_{max} red-shifted nearly 40 nm. Adding the cyano end group to a B monomer created a much stronger electron-withdrawing unit, due to conjugation. Oligomers with -CN attached to a P monomer absorbed at a slightly higher energy than any dibromo analogues studied, while oligomers with the -CN located on a B-monomer absorbed at lower energies than the dibromo-terminated sequences.

Bromo and C8 end groups appeared to exert only a modest influence on the optical properties, especially when compared to the highly perturbing -CN. That being said, the same pattern of dependence on the identity of the terminal monomer which was noted for -CN was also observed for these two end groups. The C8 (C8 = -(CH2)6CH=CH2) group would be expected to be only a mild σ-donor while the bromo group should be modestly σ-withdrawing and π-donating. In solution, a red shift of 13 nm was observed when changing the electron-withdrawing Br to an electron-donating C8 on P units in the PBP analogues, Br-PBP-Br ($\lambda_{max} = 476$ nm) and C8-PBP-C8 ($\lambda_{max} = 489$ nm). The effect of the interaction of the end-group with the attached monomer can also be seen in the comparison of Br-PB-Br ($\lambda_{max} = 429$ nm) vs C8-PB-C8 ($\lambda_{max} = 432$ nm) and Br-BPP-Br ($\lambda_{max} = 464$ nm) vs C8-BPP-C8 ($\lambda_{max} = 448$ nm). Based on these data, we hypothesize that when a Br attached to a B unit is replaced with a C8 the blue shift of the λ_{max} is partly canceled by the red shift due to the C8 substitution of the Br on the P unit. As these effects were relatively modest relative to those observed with the -CN group, we elected to focus our sequence comparison

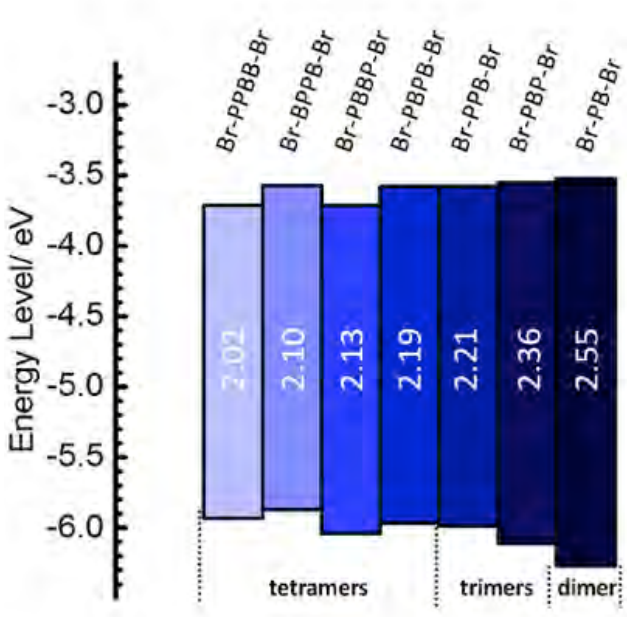


Figure 3.4: Sample cyclic voltammograms and differential pulse voltammograms of (a) Br-PBPB-Br and (b) Br-PPBB-Br

studies on the dibromo-substituted oligomers.

In examining the Br-terminated oligomers, we did indeed find evidence for sequence effects in both the trimer and tetramer series (Figure 3.4). Focusing only on the two trimers with the same 2:1 ratio of P:B and bromo end groups, Br-PBP-Br and Br-BPP-Br, differences in absorption maxima (Δ = 10 nm), oxidation potential (Δ = 0.12 V), and electrochemical gap (Δ = 0.15 V) were observed. The reduction potentials were, however, similar (Δ= 0.03 V) suggesting that they are determined primarily by the single B-unit.

Unambiguous sequence effects are also clearly seen in the dibromo-terminated tetramer series all of which have the same 1:1 P:B ratio. Most persuasively, the two bromo-terminated tetramers Br-PPBB-Br and Br-PBPB-Br, exhibited the largest difference in the magnitude of their electrochemical gaps (0.17 V). Since both of these oligomers have exactly one P-Br and one B-Br interaction, the difference must be attributed to sequence alone. Br-PPBB-Br exhibited both a less positive reduction and less negative oxidation potential than the alternating sequence isomer (Br-PBPB-Br). In examining the other two oligomers in the

series, it becomes clear the presence of a BB-pairing defines the reduction potential: both Br-PPBB-Br and Br-PBBP-Br were reduced at -1.31 V. The oligomers, Br-BPPB-Br and Br-PBPB-Br exhibited more negative reduction potentials of -1.45 and -1.46 V, respectively. The trend in oxidation potentials appears to depend more on the distance between P units. Those oligomers with PP-pairing, Br-BPPB-Br and Br-PPBB-Br, exhibited lower oxidation potentials than those with separated P units. The trend is gradual, however, not binary as was the case for the reduction potentials vs. BB-pairings.

We also observe some intriguing sequence effects in the solution phase absorption and emission spectra, especially in absorption/emission intensities. For the trimers with a 2:1 P:B ratio, the absorption intensities at 10^{-5} M in chloroform are similar (ca. 0.3×10^5 cm^{-1} M^{-1}) but the emission intensities are dramatically different (Figure 3.2e). In particular, the intensity of the emission for Br-PBP-Br of 80×10^5 cm^{-1} M^{-1} is at least 4x larger than that for all other oligomers characterized. Within the 1:1 P:B tetramer series, the absorption intensities are modestly different (range $0.35\text{-}0.5 \times 10^5$ cm^{-1} M^{-1}) with Br-BPPB-Br > Br-PBPB-Br > Br-PPBB-Br > Br-PBBP-Br which is inversely related to the increase in absorption wavelength. The emission intensities for these tetramers exhibited larger differences (range $5\text{-}20 \times 10^5$ cm^{-1} M^{-1}) but follow the order Br-PBPB-Br ≈ Br-PPBB-Br > Br-BPPB-Br > Br-PBBP-Br which does not appear to correlate with the changes in emission wavelength.

Absorption data for thin films were also collected for those tetramers that were selected for incorporation in devices (Figure 3.2c). The λ_{max} of films cast from chloroform solutions followed the trend Br-BPPB-Br (λ_{max} = 546 nm) > Br-PBPB-Br (λ_{max} = 536 nm) > Br-PPBB-Br (λ_{max} = 510 nm) > Br-PBBP-Br (λ_{max} = 494 nm). Notably this trend is opposite to their absorption maxima in solution Br-BPPB-Br (λ_{max} = 493 nm) < Br-PBPB-Br (λ_{max} = 507 nm) < Br-PPBB-Br (λ_{max} = 508 nm) < Br-PBBP-Br (λ_{max} = 512 nm) (Figure 3.2b). The fact that these sequences exhibit a different pattern of absorption in the solid state suggests that the interchain interactions and short-range order are also sequence-dependent, with Br-BPPB-Br exhibiting the strongest aggregation-based absorption shift. Also consistent is the fact that we observe larger sequence-based differences in the λ_{max} absorptions in the solid state (52 nm) than in solution (19 nm).

A selection of these oligomers were incorporated into solar cells with the goal of under-

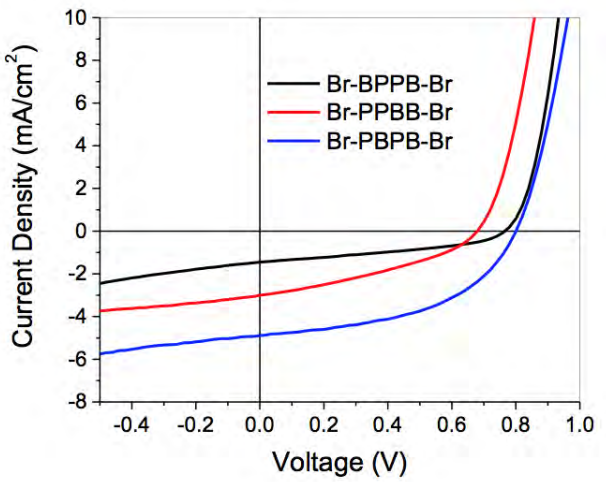


Figure 3.5: Representative J-V output of photovoltaic devices based on oligomers.

Oligomers	Thickness / nm	J_{sc} / mA·cm^{-2}	V_{oc} / V	FF / %	PCE / %
Br-PBP-Br	131	0.02±0.01	0.455±0.109	27.3±1.4	0.00±0.00
Br-BPP-Br	152	0.96±0.04	0.824±0.009	35.1±0.4	0.28±0.01
Br-BPPB-Br	85	1.45±0.11	0.770±0.016	41.2±5.8	0.47±0.10
Br-PPBB-Br	84	3.16±0.16	0.717±0.075	34.5±1.7	0.79±0.15
Br-PBPB-Br	89	4.85±0.42	0.768±0.036	49.4±2.9	1.85±0.26

Table 3.2: Device characteristics of BHJ solar cell with oligomers: PCBM (1:1)

standing sequence effects (Figure 3.5, Table 3.2). While we expected, based on literature reports of related molecules and the relatively short conjugation lengths, only modest power conversion efficiencies for these materials,[123] we hypothesized that any observed differences in performance would give us insight into the effect of sequence on the multiplicity of properties that contribute to device performance. To investigate these properties, bulk heterojunction (BHJ) solar cells were fabricated with the structure ITO/PEDOT:PSS/oligomer:PC61BM (1:1)/Ca/Al for Br-PBP-Br, Br-BPP-Br, Br-PBPB-Br, Br-BPPB-Br, and Br-PPBB-Br. The tetramer, Br-PBBP-Br was not included due to synthetic challenges (extremely poor solubility of intermediates) that precluded the preparation of the quantities necessary for these studies.

The first sequence-based difference was observed in the trimer series with the same 2:1 P:B ratio (Table 3.2). Br-PBP-Br did not give any measurable performance in the solar cell, while Br-BPP-Br exhibited a small but reproducible power conversion efficiency (PCE) of 0.28%. BPP analogs with different end groups (Br-BPP-CN and Br-PPB-CN) were also studied. The differences in PCE (0.28% - 0.37%) between all three BPP analogs were negligible, therefore no reliable conclusion about end group effects on solar cell performance can be drawn from these data. Increasing the conjugation length from trimer to tetramer increased the overall performance of the materials as would be expected.[157] For the 1:1 P:B ratio tetramers, the measured efficiencies ranged from 0.47% for Br-BPPB-Br to 1.86% for Br-PBPB-Br, a difference of ca. 3x. Devices prepared with Br-PPBB-Br exhibited an intermediate PCE of 0.79%. Please note that all three devices had similarly thin active layers ($\sim$ 85 nm) such that the observed device performance can be directly correlated with the optoelectronic properties of these oligomers. To provide more insight into the reasons for these differences, the hole mobilities of the BHJ blends were measured via the space charge limited current (SCLC) method by fabricating hole-only devices with the structure ITO/PEDOT:PSS/Oligomer:$PC_{61}BM$ (1:1)/MoO_3/Al. The hole mobilities follow the trend Br-BPPB-Br (5.94×10^{-5} cm^2 $V^{-1}s^{-1}$) > Br-PBPB-Br (2.87×10^{-5} $cm^2V^{-1}s^{-1}$) > Br-PPBB-Br (1.58×10^{-5} cm^2 $V^{-1}s^{-1}$). The relatively low hole mobilities are consistent with the modest PCEs exhibited by these oligomers; high fill factors are normally associated with mobility values of $\sim 10^{-3}$ $cm^2V^{-1}s^{-1}$.[158–161]

Film topologies of neat tetramer films were further characterized by tapping mode atomic force microscopy (AFM). Distinct topologies in spin-cast neat films of three tetramers were observed. Particularly, the root mean squared (RMS) roughness of the Br-PBPB-Br film is much smaller than that of other two sequences (0.843 nm vs 14.2 nm and 17.2 nm). However, no obvious topology differences were observed across films of photoactive layers (tetramers/$PC_{61}BM$), with consistent RMS roughness (ranging from 1.2 nm to 2.3 nm).

3.3.3 Computational approach

Computational methods provide a fast and relatively inexpensive mechanism to screen optoelectronic properties of π-conjugated materials. Several studies have found a high degree of correlation between density functional theory (DFT) computed orbital eigenvalues, vertical ionization potentials and electron affinities[162,163] though these calculations yield nonphysical results.[164,165] In addition, DFT calculations provide accurate predictions of optical band gaps.[70] In solution electrochemistry, redox potentials can be predicted based on the free energy change.[166,167] The adiabatic difference in total energy between the neutral and positively or negatively charged systems (ΔSCF) provides oxidation or reduction potentials, respectively.

Since our objective was to reliably and accurately screen for targeted properties of sequenced oligomers, we chose to extend these regression techniques by use of a "consensus model" to minimize both systematic and random errors, i.e., to improve accuracy and correlation. The consensus model used here combines two different computational predictions of an experimental property using multivariate regression, e.g., oxidation potential. For redox potentials, calculated HOMO or LUMO eigenvalues and adiabatic total energy differences (ΔSCF) were both used, and to predict optical absorption energies, ZINDO and time-dependent DFT (TDDFT) methods were combined with the HOMO-LUMO difference.

The computational method was parameterized on the trimer and tetramer compounds that were synthesized. The electronic properties of all possible dimer, trimer and tetramer sequences were then predicted based on the derived models. (Table 3.3) When palindromic sequences were examined (i.e. Br-PPB-Br and Br-BPP-Br), energy differences in predicted

Oligomer	Predicted E^{ox}_{peak}/ V	Predicted E^{red}_{peak}/ V	ΔE^{comp}_{gap}/eV
Br-PB-Br[a]	1.06	-1.47	2.55
Br-PB-CN	1.20	-1.53	-
Br-BP-CN	1.27	-1.43	-
Br-BPB-Br	0.84	-.143	3.23
Br-PBP-Br	0.82	-1.44	3.36
Br-PBP-CN	0.96	-1.41	3.31
Br-PPB-Br[a]	0.81	-1.46	3.34
Br-PPB-CN	0.83	-1.45	3.12
Br-BPP-CN	0.99	-1.44	3.41
Br-PBB-CN	0.96	-1.40	3.09
Br-BPB-CN	0.94	-1.42	3.13
Br-BPPB-Br	0.66	-.144	3.09
Br-PBPB-Br[a]	0.76	-1.41	3.07
Br-PBPB-CN	0.63	-1.41	3.11
Br-PPBB-Br[a]	0.71	-1.36	3.05
Br-PPBB-CN	0.80	-1.40	3.11
Br-PBBP-Br	0.76	-1.37	3.10
Br-PBBP-CN	0.70	-1.38	3.11
Br-BPPB-CN	0.65	-1.44	3.19
Br-BPBP-CN	0.63	-1.41	3.11
Br-BBPP-CN	0.64	-1.41	3.04

[a] average of values for two conformations

Table 3.3: Consensus model predicted oxidation, reduction and gap energies for dimers, trimers and tetramers.

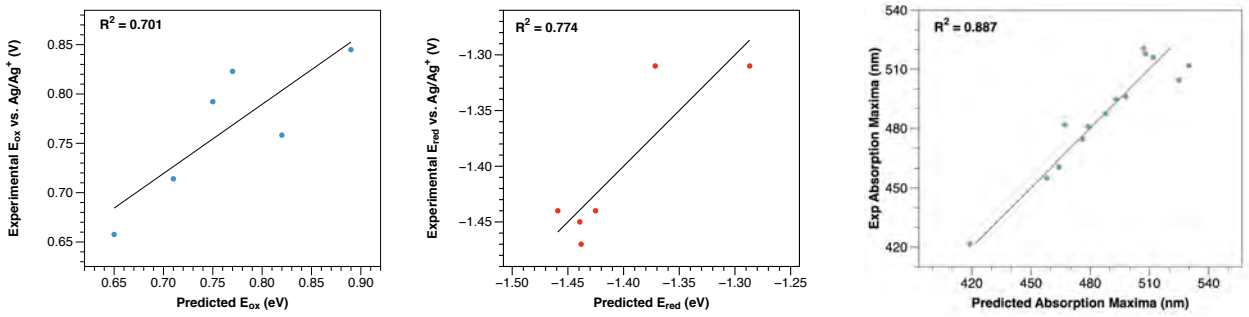


Figure 3.6: Correlations between computed first oxidation potential, first reduction potential and optical excitation energies with experimental counterparts. For all predicted properties, a consensus model yields small residual errors compared with experimental counterparts.

oxidation potentials (∼0.04 V), reduction potentials (∼0.01 V) and optical absorption energies (∼0.03 eV) were observed due to conformational differences.[148]

In general, computed and experimental parameters show only small residual errors compared to their experimental counterparts (Figure 3.6). We find mean unsigned errors (MUE) between computed and experimental parameters after the linear regression analysis to be very low, with 0.03 V MUE for oxidation potentials ($R^2 = 0.70$), 0.04 V MUE for reduction potentials ($R^2 = 0.77$), and 9 nm MUE for optical absorption maxima ($R^2 = 0.89$). The high degree of agreement is not surprising because the sequenced oligomers define a closely analogous series, and the consensus technique minimizes systematic and random errors. With the limited number of experimental electrochemical measurements, the correlation coefficient R^2 is deceivingly poor. Orbital shapes for each of the oligomers prepared were computed and are plotted in Figure 3.7.

As the MUEs between experiment and computed properties were low, we extended the calculations to longer oligomers to explore the role of sequence and PP/BB pairings. The electronic structure of all hexamers with 50:50 B:P ratios were computed (Table 3.4). Since conformational effects can be significant, we again computed low energy conformers for both palindromic orders (e.g., Br-PBPBPB-Br and Br-BPBPBP-Br) to estimate the variations

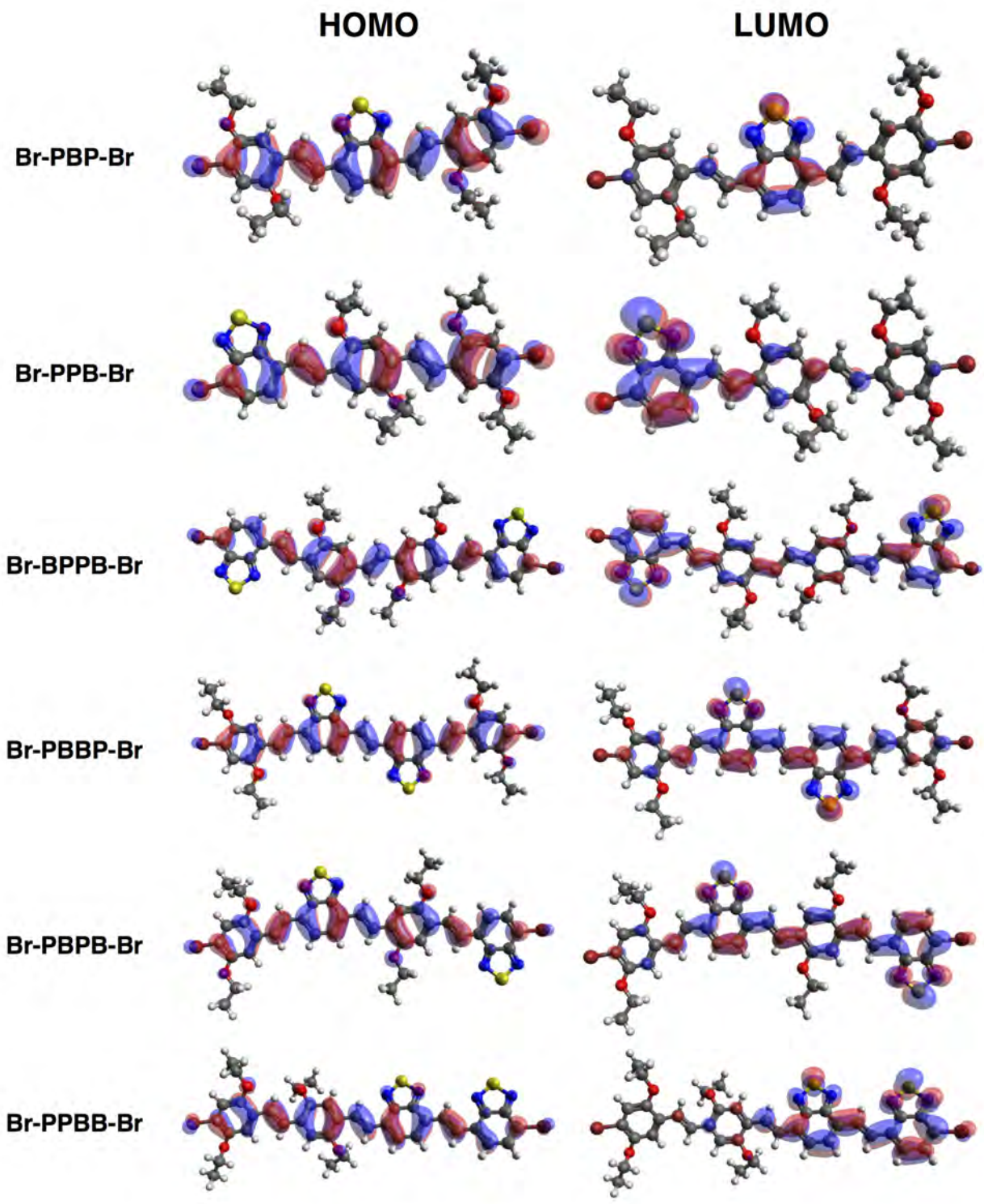


Figure 3.7: Computed orbital shapes for trimers and tetramers studied

Oligomer	Computed HOMO[a] /eV	Computed LUMO[a] /eV	ΔE_{gap}^{comp} [a]/ eV
Br-PPPBBB-Br	-4.79	-2.99	1.80
Br-PPBPBB-Br	-4.80	-2.91	1.89
Br-PBPPBB-Br	-4.81	-2.87	1.94
Br-BPPPBB-Br	-4.77	-2.82	1.95
Br-BPBBPP-Br	-4.81	-2.88	1.93
Br-BPPBPB-Br	-4.83	-2.74	2.09
Br-BPPBBP-Br	-4.70	-2.78	1.91
Br-PBBBPP-Br	-4.86	-3.01	1.85
Br-PBPBBP-Br	-4.87	-2.92	1.96
Br-PBPBPB-Br	-4.66	-2.70	1.96

[a]average of values for two conformations

Table 3.4: Computed HOMO, LUMO and gap eigenvalues for hexamers.

due to conformational local minima. We find the variation to be ∼0.1 eV, on par with other estimates.[148]

3.4 CONCLUSIONS

We find that sequence is important in both solar cell performance and in related properties. In addition to PCE, we find that absorption, emission, solid-state packing, hole mobilities, and HOMO-LUMO energy levels, are sequence dependent. We also demonstrate that using calculations we can explore sequence-space to increase our understanding of structure/function correlations and to direct synthesis. Although not measured for these materials, it seems likely that other characteristics that are important to device performance, including domain size, thermal stability, etc. will likewise exhibit sequence dependence. The current work highlights one of the most important potential advantages of sequence engineering which is the idea that sequence has the potential to affect intrinsic and bulk properties orthogonally. Amongst sequences that exhibit a targeted intrinsic property, such as HOMO-LUMO gap, a range of bulk properties could be exhibitedsome sequences might pack well while others do not. The inverse is also possiblea range of sequences could be identified that exhibit a particular morphological trait and then refined on a desired intrinsic property, such as HOMO level. Future efforts will aim to correlate intermolecular interactions, packing, film morphology, and interfacial organization with sequence effects. The result should allow for combined computational and synthetic rational design of materials that can fulfill the complex set of properties necessary for highly efficient organic solar cells and other applications.

4.0 INVERSE DESIGN OF CONJUGATED POLYMERS FROM COMPUTED ELECTRONIC STRUCTURE PROPERTIES: MODEL CHEMISTRIES OF POLYTHIOPHENES

The text in this chapter has been adapted from a manuscript submitted to the The Journal of Physical Chemistry, written in conjunction with Steven G. Owens, Andrey B. Sharapov, Geoffrey R. Hutchison and submitted to The Journal of Physical Chemistry, Part C. The author's contribution to the work is most of the analysis and discussion.

4.1 INTRODUCTION

Polythiophene-derived materials continue to gather interest for both fundamental scientific study and technological applications, including field-effect transistors[168], light-emitting diodes[169], lithium-ion batteries[170], flexible electronic devices[171], and organic photovoltaics[172]. Arguably most of the important parameters for these applications derive from fundamental electronic structure parameters such as band gap, oxidation and reduction potentials[173,174]. Consequently, optimizing these parameters is a key problem for the scientific community to solve.

The wide array of applications for polythiophene-based materials continues to drive both fundamental and application-based research. Many applications including field-effect transistors (FETs), light-emitting diodes (LEDs), batteries, flexible electronics, and organic photovoltaics (OPVs) have drawn considerable research interest. OPV applications of polythiophenes have become a major focus due to their easy processability and the effect relatively cheap photovoltaics would have on solar energy technology. The fundamental electronic

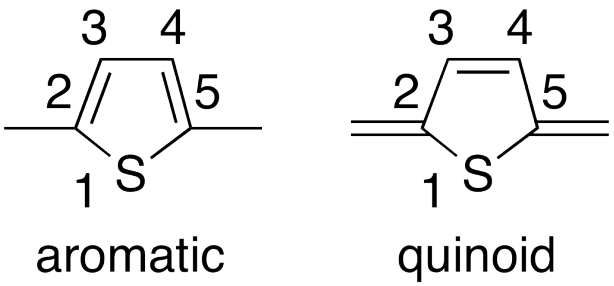


Figure 4.1: Bond length between carbons 3 and 4 used to determine aromaticity of a compound

structure of the molecule including properties like band gap and oxidation and reduction potentials play a large part in the properties of these materials. In order to better understand and improve the technological applications, these key-underlying factors that impact the properties of the materials must be investigated.

There are tens of thousands of possible thiophene monomers with varying functional group substitution (electron donating/withdrawing and steric bulk), aromaticity/quinoidal character, and heteroatom substitution. These differences in monomer structure ultimately affect the electronic and structural properties of their polymers. Through a computational study of 100 thiophene monomers, we demonstrate how monomer properties determine the electronic and structural properties of the polymer. We seek to understand which key factors determine these important properties by investigating trends in properties such as HOMO, LUMO, band gap, dihedral angle, ligand width, van der Waals size of ligand, ligand type, average bond length, and bond length between carbon 3 and 4 which estimates aromaticity of the thiophene rings (Figure 4.1).

By studying monomers and oligomers we can observe the relationship between monomer, oligomer, and ultimately polymer properties. The goal would be to trivially predict the properties of the extended polymer from the known or computed properties of the monomers greatly reducing the computational and experimental time required. In the last several years, while large automatic searches and machine learning have become a major research

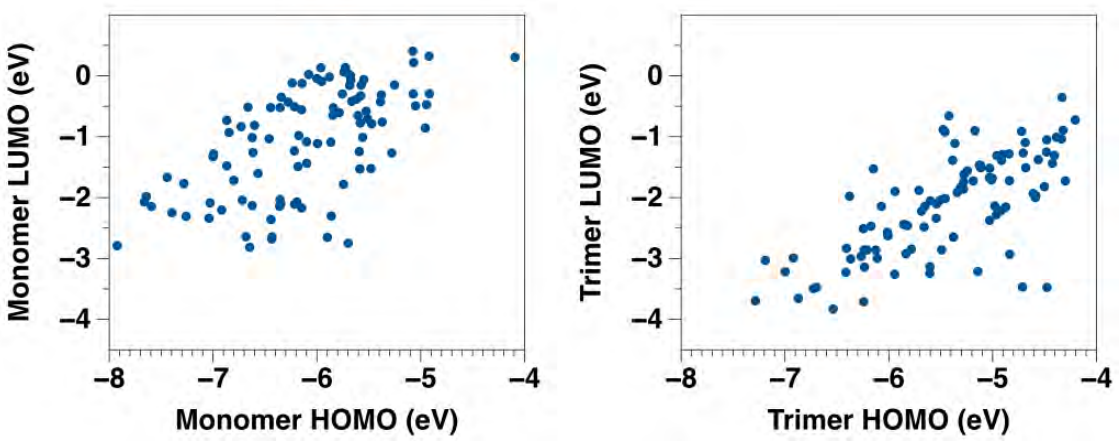


Figure 4.2: Monomer and trimer diversity demonstrated by a 3 eV range in both the computed B3LYP HOMO and LUMO energy values. Note that both donor (electron rich, less negative HOMO/LUMO) and acceptor (electron poor, more negative HOMO/LUMO) monomers exist along a spectrum of properties.

focus,[29,44,46–52,61 62 48,51,54 63] it is also important to carefully construct manual studies to test structure-property relationships.[175] In this study, we follow the later approach by examining a diverse set of polythiophenes, as shown in Figure 4.2, which shows a 3 eV range for both HOMO and LUMO monomer and trimer values, including a range of both electron rich donors and electron poor acceptors. Our goal is to accurately predict properties of the homopolymers from monomers or small oligomers.

4.2 COMPUTATIONAL METHODS

4.2.1 Monomer Data Set

The 100 monomers in this study were selected by composing a list of potential thiophene substitutions, including simple hand-picked functional groups, nonaromatic and aromatic fused rings. None of the chosen functional groups were sterically bulky (e.g. t-butyl, mesityl),

to ensure relatively planar, conjugated oligothiophenes as a consistent set. For this study, monomers were limited to species containing C, H, N, O, S, Se, F, Cl, and Br. Figure 4.3 shows the structures of the monomers studied, Table A1 shows the IUPAC names of the monomers and Tables A2-A11 of the supporting information list the names of each of the monomers and literature references. SMILES are given in Table A12.

4.2.2 Generation of Optimized 3D Structures

For each oligomer, the 3D structure of a low-energy conformer was generated starting from the SMILES[72], generating the 3D structure using Open Babel[104] and minimized using the MMFF94 force field.[82–86] Next a weighted-rotor search (MMFF94, 100 iterations, 20 geometry optimization steps) was carried out with Open Babel to find a low-energy conformer. This was then further optimized using 500 steps of conjugate gradients and MMFF94. Finally, Gaussian09[154] was used to optimize the structure using DFT with the B3LYP[110,111] functional and the 6-31G* basis set. Although HOMO and LUMO eigenvalues from density functional methods cannot be formally taken as either the ionization potential or electron affinity, respectively, previous studies have shown that B3LYP-derived eigenvalues compare favorably with experimental electron affinities[162,176–178], ionization potentials,[162] and band gaps[179]. The idealized infinite polymer HOMO, LUMO, and HOMO-LUMO gap eigenvalues were determined based on linear regressions from the corresponding oligomer values versus the reciprocal oligomer length (i.e., 1/N)[113,114,173].

4.2.3 Statistical Methods

As discussed above, this set of data contains 100 diverse thiophene monomers with different substitutions. The experimental goal of discovering a set of descriptor properties to be used to predict polymer properties by computing the monomer, or perhaps small oligomers such as dimer and trimer electronic structure, thus reducing computational time compared to the full polymer[45,180] was accomplished by performing applicable statistical tests and generate plots using R[108]. First, a set of distinct, unique, carefully chosen parameters were selected, including the values of elements 3 and 4, monomer ligand 3-4 width, element 3 van der Waals

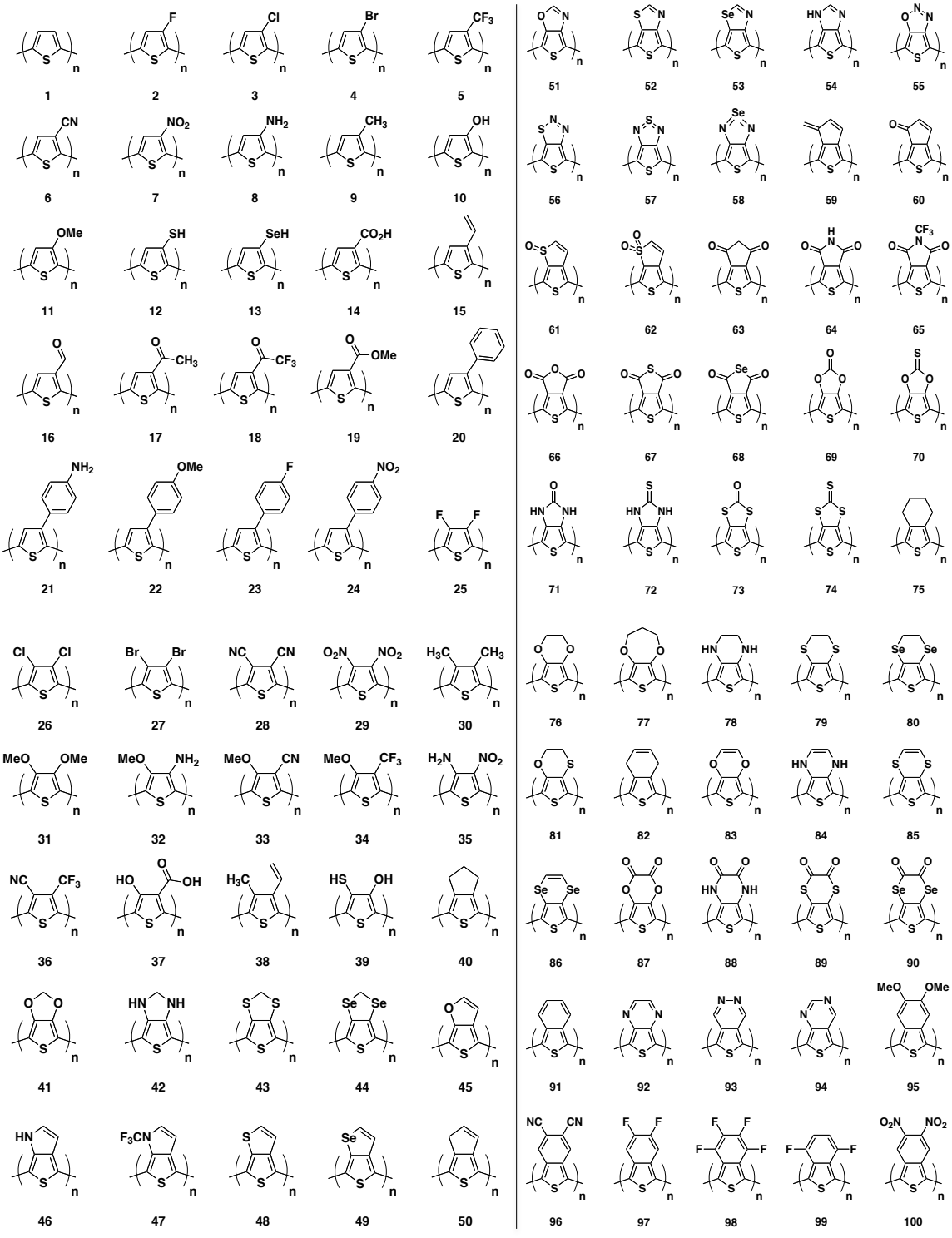


Figure 4.3: Chemical structure of the diverse oligothiophenes studied, including a range of mono- and di-substituted species, including a wide range of fused ring structures.

interaction, element 4 van der Waals interaction, ligand 3- 4 van der Waals interaction, charge of monomer 3, charge of monomer 4, monomer absolute charge, monomer average charge, HOMO, LUMO and HOMO-LUMO gap values of the monomer, dimer, trimer, tetramer and pentamer, dimer bond length alternation, tetramer bond length alternation, pentamer bond length alternation, monomer and pentamer length of the bond between elements 3 and 4, monomer and pentamer S charge, dihedral angles of the dimer, trimer, tetramer and pentamers. Care was taken to ensure that several parameters were not measuring similar properties that would skew the model results. A stepwise regression function, as discussed below, was used to predict the best set of predictors for the model by adding and subtracting different descriptors until the most predictive set was chosen without overfitting. Once this set of descriptors was chosen, it was defined as a particular model (e.g., examples described below). Still, many models that appear predictive, suffer from significant overfitting by incorporating too many descriptors. To test for overfitting, each model was tested for reliability using both bootstrap and cross-validation tests.[181,182] Both of these methodologies provide information for how predictive a model will be when applied to new, unknown, data sets. Cross-validation is most reliable when studying large data sets since the data is divided into two parts, one for model development and one for model testing. In small data sets, each of these groups is not sufficiently large to allow for predictive model development. Hence, in our calculations, the bootstrap models consistently give better results since the data set only contains 100 species.

4.3 RESULTS

Electronic properties were calculated for monomers and oligomers from two to five repeat units for all 100 thiophenes of interest in the study, and results were examined. Linear regressions and R^2 values for the infinite polymer HOMO, LUMO, and HOMO-LUMO gap values are shown in Tables A13-A14. The expected results from these calculations were negative HOMO slopes (i.e., increasing oligomer lengths yield less negative HOMO eigenvalues due to delocalization) and positive LUMO (i.e., increasing oligomer lengths yield more negative

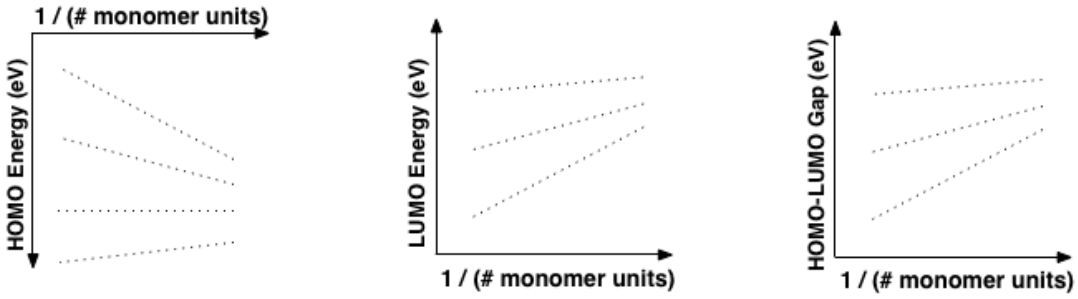


Figure 4.4: HOMO and LUMO expected energies as a function of the inverse of the number of monomer units, which demonstrates that the HOMO slopes are expected to be negative (i.e. increasing number of monomer units to the left results in higher HOMO energies) and LUMO slopes are expected to be positive.

LUMO eigenvalues due to stabilized, delocalized electron affinities) and HOMO-LUMO gap slopes, as shown in Figure 4.4.

A flat slope suggests that the electrons are not delocalized in a species (i.e., the electronic properties do not change as a function of oligomer length) while a high slope shows significant changes in the properties as the chain length changes. Most of the compounds follow these expected trends, with compounds **80, 86, 89, 90** and **100** as exceptions to the trends. In addition, the R^2 values were expected to show high correlation which also was demonstrated, with the exception of compounds **26, 28, 29, 34, 44, 67, 80, 84, 86, 89, 90** and **100**, which show low R^2 statistics; while the properties are nearly linear, they are horizontal. Thus, calculations for these molecules suggest orbitals which are not delocalized, despite appearance as aromatic structures (and typical aromatic bond lengths, e.g., Figure 4.1). The plots, therefore, appear to rise to a certain point and then flatten out as they switch from being aromatic to no longer being aromatic.

A natural question is what properties (e.g., geometric or chemical) alter the electronic properties such as HOMO, LUMO, and gap. Considering highly conjugated oligomers and polymers are typically planar, the average dihedral angle should have a significant effect on these properties. In Figure 4.6, we have graphed the effect of the steric crowding caused by

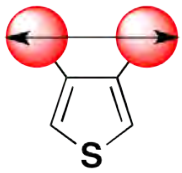


Figure 4.5: Effect of steric crowding caused by the thiophene substituent ligand width, as measured by the distance between substituents. (Black arrows)

the thiophene substituent, as measured by the distance between substituents (as illustrated in Figure 4.5). Although as the chain size increases from dimer to pentamer, the R^2 increases from 0.32 to 0.44, there is no improvement when the chain size increases from the tetramer to the pentamer. Based on the modest R^2 values, torsional twisting caused by steric crowding does occur, but other effects also dictate dihedral angles in oligothiophenes. Moreover, we can see in Figure 4.4, that dimers do not yet reflect the dihedral angles of longer oligomers.

4.3.1 Computationally Efficient Models for Predicting Polymer Properties

The main goal of this study is to determine if there is a simple way to predict the HOMO and LUMO values of a polymer from properties of the monomer, dimer, and trimer to create more efficient photovoltaics or other optoelectronic materials, by utilizing targets with a small optical gap and stable LUMO.

4.3.1.1 HOMO The HOMO eigenvalues of monomers are only weakly correlated with the extrapolated HOMO energy of the infinite polymer, as summarized in Table 4.1 (i.e., R^2 only 0.58 and mean unsigned errors of ~0.5 eV). Increasing to dimers significantly improves the correlation and decreases mean unsigned errors (MUE), but as shown in Table 4.1, Figure 4.7 and elaborated in Figures A1-A2, the trimer HOMO values reliably predict the polymer with R^2 of 0.94 and MUE < 0.2 eV. In order to ensure the significance of the model terms, an analysis of variance (ANOVA) test was run and the p-values show statistical significance in our models (i.e., $p < 2.2 \times 10^{-16}$ showing that the results are highly significant). In order

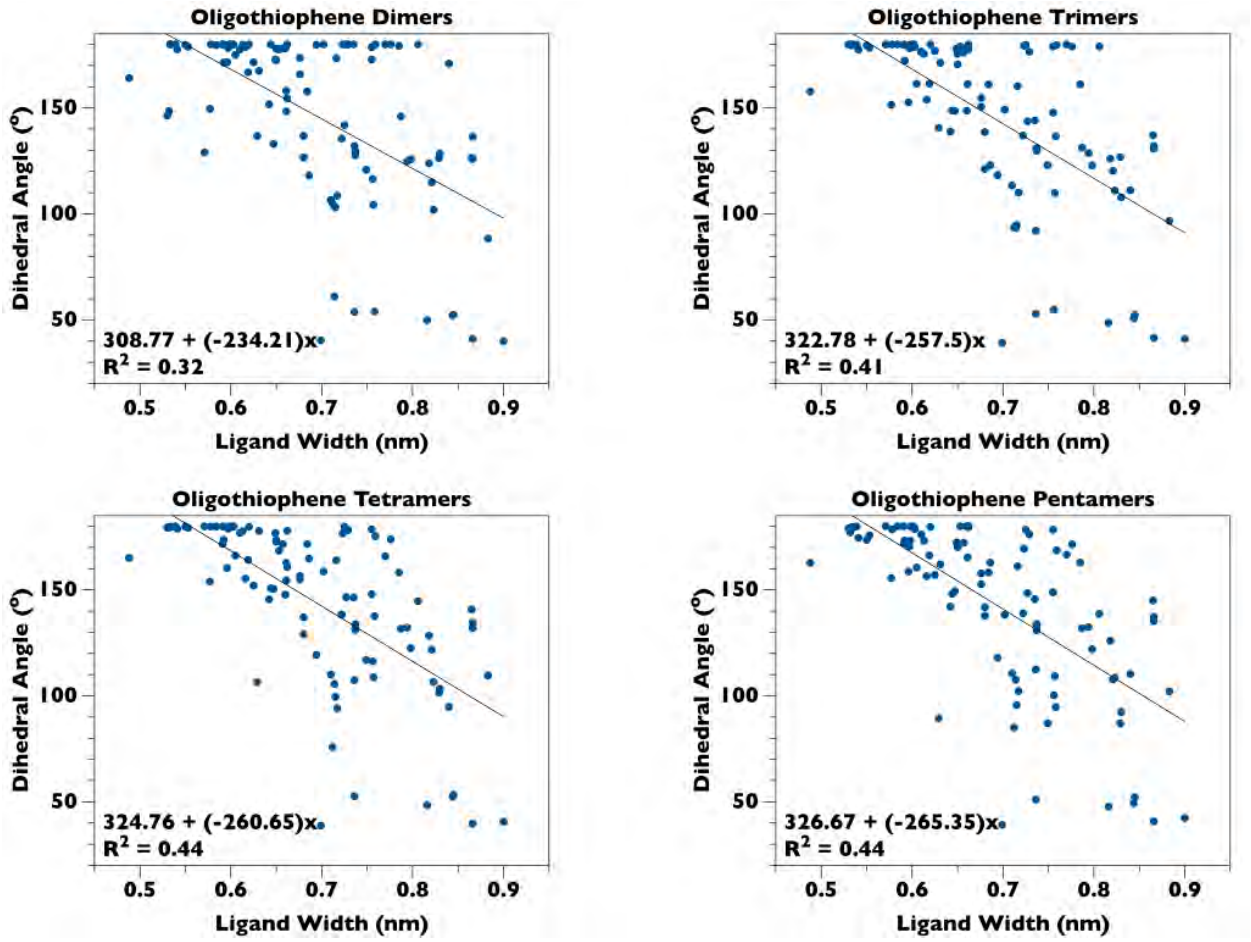


Figure 4.6: Ligand width correlation with the dihedral angle as shown with dimers, trimers, tetramers and pentamers.

Model	Adjusted R^2	Mean Unsigned Errors (eV)
HOMO ~ Monomer HOMO	0.58	0.54
HOMO ~ Dimer HOMO	0.87	0.29
HOMO ~ Trimer HOMO	0.94	0.19
HOMO ~ HOMO Model	0.97	0.11
LUMO ~ Monomer LUMO	0.74	0.36
LUMO ~ Dimer LUMO	0.89	0.24
LUMO ~ Trimer LUMO	0.93	0.18
LUMO ~ LUMO Model	0.96	0.12
Gap ~ Monomer Gap	0.11	0.66
Gap ~ Dimer Gap	0.53	0.49
Gap ~ Trimer Gap	0.77	0.33
Gap ~ Gap Model	0.87	0.23

Table 4.1: Summary of statistical correlation and predictive power for HOMO, LUMO and HOMO-LUMO gap models

to achieve lower mean absolute errors, models were prepared using other descriptors and stepwise regression to select optimal terms. The HOMO model combines the value of the monomer HOMO energy (MonHOMO), the trimer HOMO energy (3HOMO) and the value of the dihedral angle of the trimer (3Dihedral) as shown in the following equation:

$$\text{HOMO} = 0.82 - 0.28(\text{MonHOMO}) + 1.44(\text{3HOMO}) + 0.002(\text{3Dihedral}) \quad (4.1)$$

This model achieves slightly higher predictability as demonstrated by the HOMO R^2 value increasing from 0.94 to 0.97, while lowering MUE to 0.11 eV, which is less than the known error from DFT (~0.2 eV).

4.3.1.2 LUMO Much like the HOMO eigenvalues, the monomer LUMO energies are modestly correlated with the polymer LUMO (R^2 0.74) As shown in Table 4.1, Figure 4.7 and elaborated in Figures A1-A2, the trimer LUMO values reliably predict the polymer with R^2 0.93 and MUE under 0.2 eV. From the ANOVA test, the p-values for the trimer LUMO as a predictor of the polymer LUMO are 2.2 x 10^{-16}, showing that the results are statistically

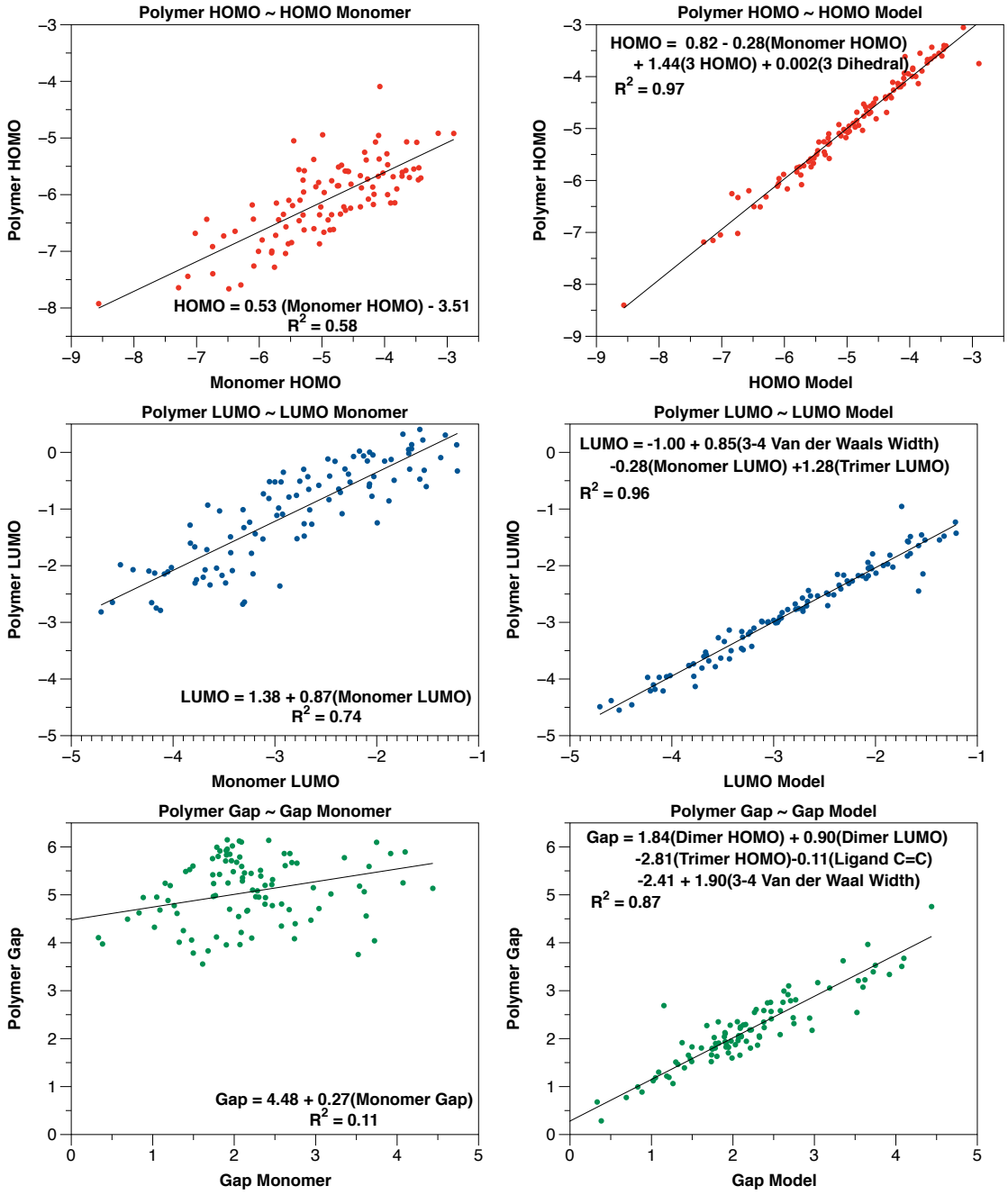


Figure 4.7: Linear regression models for predicting polymer HOMO, LUMO, and HOMO-LUMO gap from monomers (left column) and multivariate fits

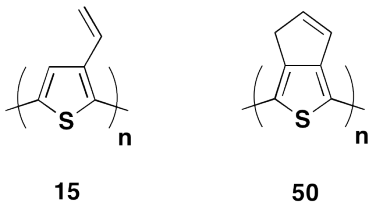


Figure 4.8: Examples of substituents with carbon-carbon double bonds

significant. An improved LUMO model combines the values of the monomer and trimer LUMO energies with the van der Waals width of the molecule in the following equation:

$$\text{LUMO} = -1.00 - 0.28(\text{MonLUMO}) + 1.28(\text{3LUMO}) + 0.85(\text{vdW Width}) \tag{4.2}$$

As with the HOMO case, this increases the probability of correctly calculating the polymer LUMO energy since the R^2 increases from 0.93 to 0.96, and MUE $\sim$0.12 eV. Including the steric width of the substituents likely indicates some effect of the dihedral angle on the LUMO energies.

4.3.1.3 HOMO-LUMO Gap Despite the accuracy of models for the HOMO and LUMO, the HOMO-LUMO band gap correlations were surprisingly less accurate. Unlike the HOMO and LUMO monomer energies, which demonstrated some correlation with the polymer HOMO and LUMO energies, the gap of the monomer shows very low correlation with the band gap of the polymer as indicated by the R^2 value of 0.11. The low R^2 value can be attributed to the slopes of the HOMO and LUMO energies not correlating to one another as shown through their respective coefficients. The band gap of the trimer shows improvement over the monomer by increasing the R^2 to 0.77 and decreasing MUE to $\sim$0.3 eV. After deriving multivariate models for the HOMO and LUMO polymer energies, the expectation was that the trimer gap would be selected in the model for the gap, as the HOMO and LUMO trimer energies are part of the HOMO and LUMO polymer models, respectively. The most predictive model for the HOMO-LUMO gap of the polymer, instead combines the dimer

	Adjusted R^2	Mean Unsigned Error (eV)
λ Pentamer ~ λ Monomer	0.09	0.16
λ Pentamer ~ λ Dimer	0.43	0.11
λ Pentamer ~ λ Trimer	0.78	0.06
λ Pentamer ~ λ Model	0.89	0.04

Table 4.2: Summary of models for internal reorganization energies for hole transport

HOMO and LUMO energies, the trimer HOMO energy, the 3-4 van der Waals width and if the ligand attached to the molecule has a C=C (Figure 4.8):

$$\text{Gap} = -2.41 + 1.84(\text{2HOMO}) + 0.90(\text{3HOMO}) + 1.90(\text{vdW Width}) - 0.11(\text{Ligand C=C}) \tag{4.3}$$

This equation shows improvement by increasing the R^2 to 0.87 and decreasing MUE to ~0.2 eV (Table 4.1). In conclusion, the most predictive multivariate models involve the HOMO and LUMO trimer energies, and the correlation increases when other descriptors are added to the HOMO and LUMO trimer energies to slightly increase the predictability of the model.

4.3.2 Reorganization Energies

Beyond orbital energies and HOMO-LUMO gap values, an important factor in organic electronics are the molecular Marcus reorganization energy.[173,174] Much like the orbital eigenvalues, one might seek a rapid predictor of polymer reorganization energies on the basis of monomer or small oligomer properties. A similar trend to the one described for prediction of polymer HOMO, LUMO, and HOMO-LUMO gap, as described above was found for the reorganization energies (Table A15). Low correlation is shown from the monomer λ compared with higher correlation from dimer or trimer reorganization energies. This encouraged a continued search to discover an accurate multivariate model to predict the reorganization energy. The dihedral angle and bond length alternation were found to have little correlation

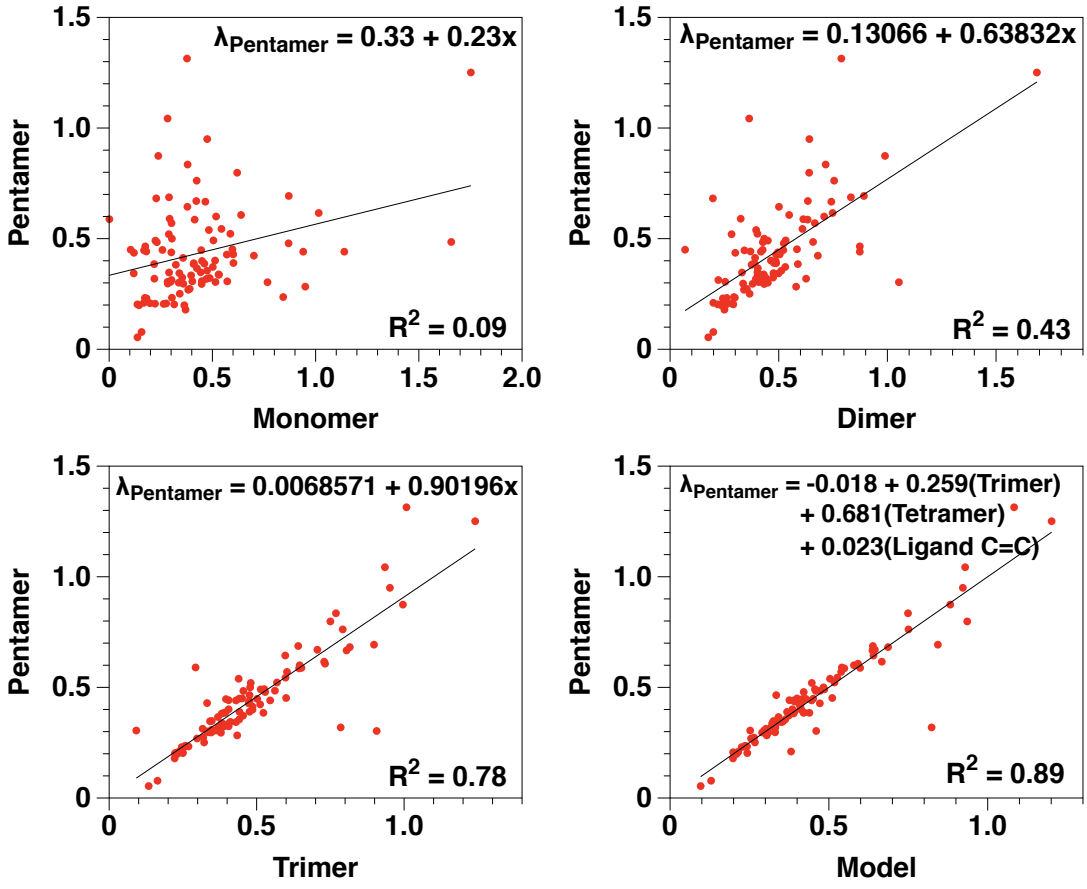


Figure 4.9: Models for internal reorganization energies for hole transport

with the reorganization energy, as verified by a model with an R^2 of 0.22. The reorganization energy also shows little correlation with the polymer HOMO or LUMO, even though the HOMO and LUMO slopes are indicators of the polymer energies and delocalization. The final model:

$$\lambda_{\text{Pentamer}} = -0.018 + 0.259\lambda_{\text{trimer}} + 0.681\lambda_{\text{tetramer}} + 0.023(\text{Ligand C=C}) \tag{4.4}$$

combines the lambda of the trimer and tetramer with whether the ligand has a carbon-carbon double bond for slightly better correlation (0.78 to 0.89) and MUE $<$ 0.05 eV. (Table 4.2, Figure 4.9)

During hole transport, the reorganization energy reflects an activation barrier to charge transfer. Consequently, large reorganization energy can be considered as a filter for organic electronic materials. That is, oligomers or polymers with high computed reorganization energies are not likely to have high charge mobility. Compounds 5, 16, 19, 22, 24, 26, 27,29, 30, 31, 33, 34, 35, 36, 38, 47, 55, 68, 75, 79, 82, 88, and 89 (Figure 4.10) were found to have large computed hole reorganization energies and are not ideal targets. The following trends that emerge suggest that large, bulky groups decrease mobility. Monomer 5 compared with similar monomers (1-4 and 6-14) suggests that the CF3 group with its three fluorine atoms decreases the mobility compared with the other monomers which have less bulky substituents. Monomers 22 and 24 compared to similar monomers (20, 21, 23) indicates that a thiophene with a fused benzene substituent loses mobility when the benzene is substituted with OMe and NO2. Monomers 26 and 27 when compared with monomer 25 show that chlorine and bromine, show lower mobility, potentially due to larger steric bulk compared to fluorine, producing a more highly twisted oligomer backbone.

On the other hand, compounds 4, 6, 40, 45, 46, 48-54, 56-58, 60, 80, 92 and 94 (Figure 4.11) were found to small reorganization energies ($<$ 0.07 eV) and are likely to have high relative mobilities and are good candidates for efficient charge transport. Although few trends are obvious, monomer 40 compared with similar monomers 41-44 suggests that nonaromatic five-membered rings made solely from carbon, have much higher activation energy than heterocycles with oxygen, nitrogen, sulfur or selenium.

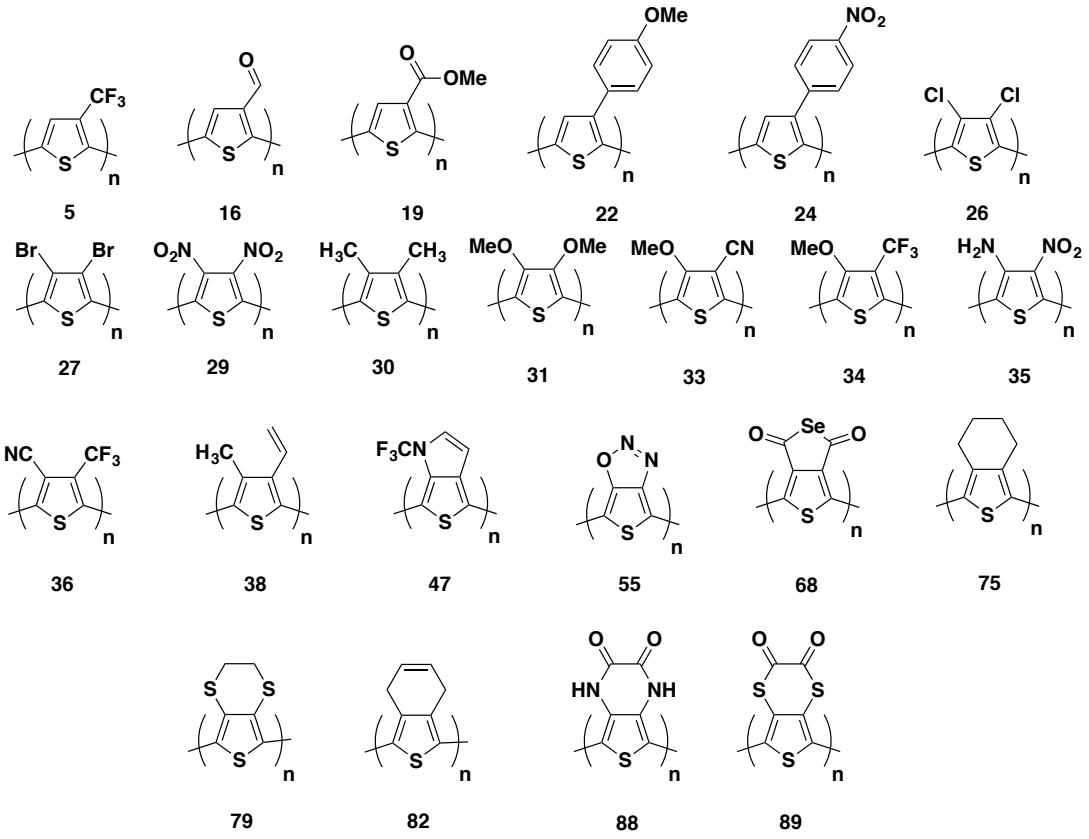


Figure 4.10: Compounds with large computed hole reorganization energies, predicted to show low relative hole transport mobility at 300 K

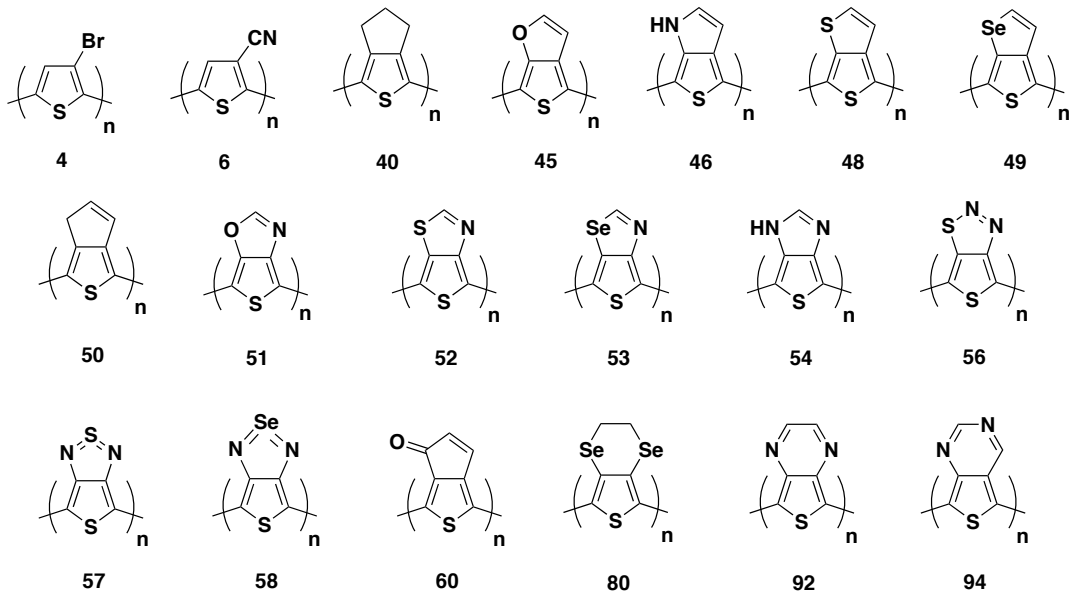


Figure 4.11: Compounds with small computed hole reorganization energy (<0.07 eV) expected to have high charge mobility

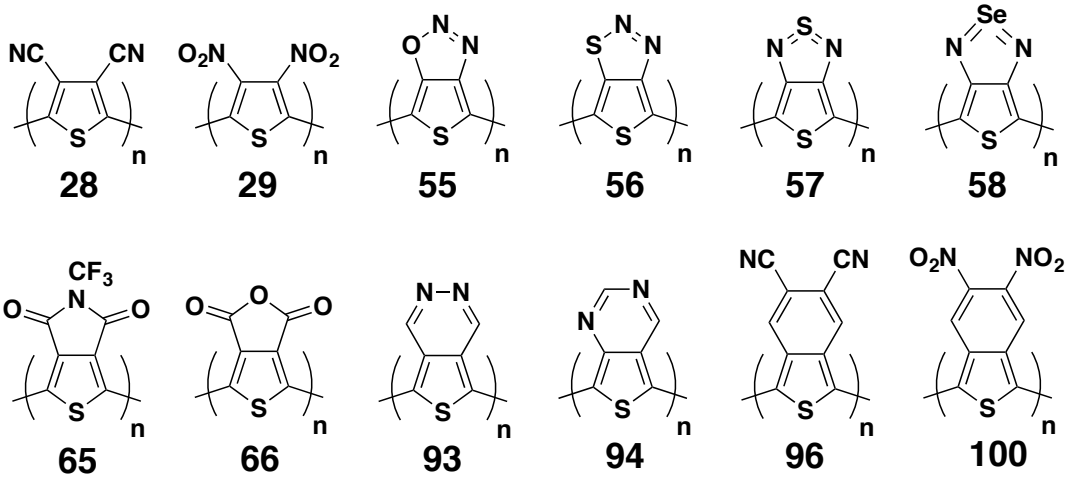


Figure 4.12: Molecules which are potential n-type or acceptor materials

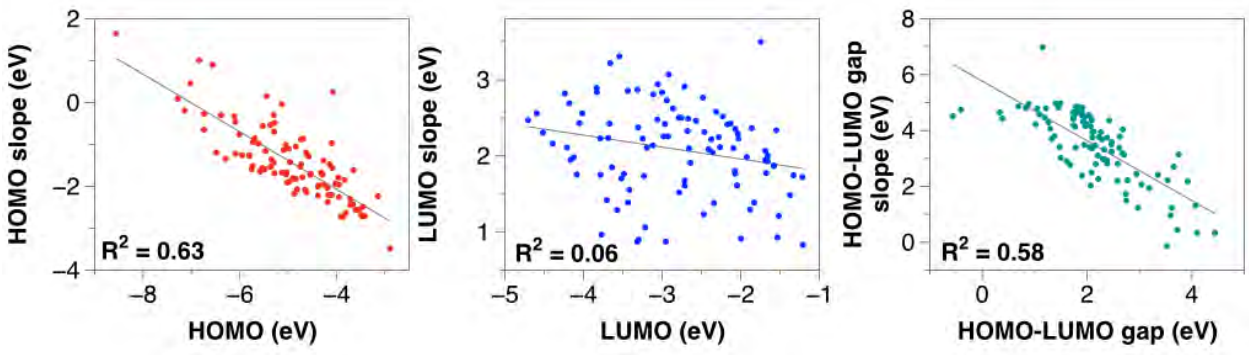
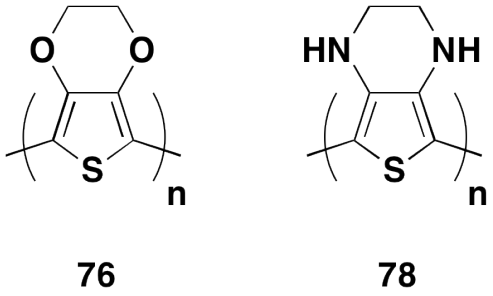


Figure 4.13: Correlations between HOMO, LUMO and HOMO-LUMO gap slopes with the predicted polymer energies.

Finally, we find multiple targets with highly negative LUMO eigenvalues, likely to be good electron acceptors (i.e., high electron affinity). Compounds 28, 29, 55, 56, 57, 58, 65, 66, 93, 94, 96, and 100 (Figure 4.12) all show highly negative LUMO energies. A clear trend is the combination of aromatic substituents with two nitrogens (eg. oxadiazoles, thiadiazoles, diazines, etc.) which strongly stabilizes the LUMO orbital.

4.4 DISCUSSION

An important question from the results, is the surprising success of the one point correlations, for example, accurately predicting the extrapolated polymer HOMO solely from the trimer or tetramer value. Since an infinite number of lines can be fit through one point, (e.g., Figure 4.4), it seems unlikely that only one point is needed. The results indicate, however, a high correlation (R^2 of 0.63) between the slope of the linear regression in HOMO eigenvalues and the extrapolated polymer HOMO (i.e., the y-intercept). A similarly high correlation (R^2 of 0.58) is found for the slope of the HOMO-LUMO gap and the extrapolated gap (Figure 4.13). Such correlations suggest that the degree of delocalization reflected in the slope



76 78

Figure 4.14: PEDOT (76) versus PDAT (78)

is higher in monomers and polymers with electron-rich, less negative HOMO eigenvalues. Indeed, we find that the slope is positive for monomer 29, dinitrothiophene; for this repeat unit, longer oligomers have more negative HOMO eigenvalues than the monomer, suggesting oligo-dinitrothiophene would be harder to oxidize than the monomer. These correlations, for the slope of HOMO eigenvalues and HOMO-LUMO gaps, suggest that electron-rich donor monomers are predicted to have high delocalization and, in general, low HOMO-LUMO gap for the homopolymers.[183]

The results indicate several interesting applications. For example, PEDOT, which is widely used, has very similar properties to PDAT (78), including similar band gap and HOMO energies (Figure 4.14). The trimer HOMO energy for PEDOT (-4.34 eV) and the trimer HOMO energy for PEDAT (-4.33 eV) are very similar, predicting that these will have similarly easy oxidation. Additionally, the extrapolated polymer band gap of these two are similar, suggesting PDAT warrants further investigation.

While most materials for efficient organic photovoltaics use complex donor-acceptor co-polymers, we note that several polythiophenes homopolymers studied here (Figure 4.15) are predicted to have efficiencies above 8%, based on the Scharber criteria.[71] Notably, all four have high (less negative) HOMO energies, suggesting a donor-donor strategy may provide an alternative to the current donor-acceptor design.

Most efforts in the field to improve OPV efficiency have focused on the p-type polymer, but to improve power efficiencies, all properties of the solar cell must be considered and

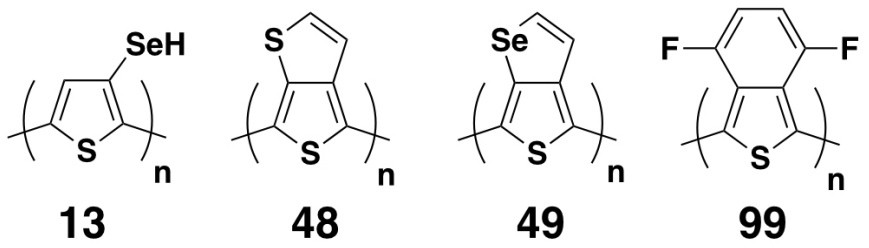


Figure 4.15: Polymers with predicted OPV device efficiencies > 8% by Scharber criteria

improved. Substituted fullerenes such as PCBM has been n-type phases with wide successes in OPV devices.[184] To improve efficiency, changing to a different n-type material with similar LUMO energy might prove worthwhile, by giving stronger optical absorption. Monomers 28, 29, 55-58, 65-66, 93-94, 96 and 100 have LUMO values between -4-5 eV making them potential replacements for PCBM or other fullerene acceptors.

To further test the HOMO, LUMO and gap models, a similar thiophene, outside the initial set of 100 species, with an electron drawing cyano group substitution (Figure 4.16), was considered. For this compound, as expected from our previous discussion, the slope of the LUMO versus 1/N is a positive value (+2.57 eV/[# repeat units]), the slope of the HOMO versus 1/N is a negative value (-0.85 eV/[# repeat units]), and the HOMO-LUMO gap slope is positive (+3.41 eV/[# repeat units]). In addition, the values were tested using

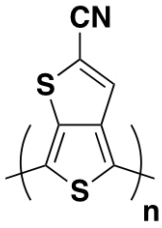


Figure 4.16: Thiophene Test Monomer

the equations of the trimer models shown above to verify that the model is able to accurately predict HOMO and LUMO values from novel monomers. The actual HOMO value of -5.33 eV and the value of the trimer HOMO equation of -5.24 eV and the actual LUMO value of -3.89 eV and the value from the trimer LUMO equation of -3.69 eV show that the calculations performed with the model are accurate within 0.2 eV.

4.5 CONCLUSION

We have shown that across a set of 100 diverse oligothiophene species, the polymer HOMO, LUMO and HOMO-LUMO gaps, computed from DFT, can be accurately estimated from the values of the trimer HOMO, trimer LUMO and trimer HOMO-LUMO gap calculated values. We also show that these approximations can be improved through the models presented above, including simple and easy to calculate properties. In all three cases, the resulting mean unsigned errors are at or below ∼0.2 eV, well within the error of the computational method. Consequently, rather than performing multiple oligomer calculations to extrapolate the polymer electronic structure, most properties can be estimated readily from modest-sized oligomers.

Polymer reorganization energies, related to the hole transport, are also shown to be predicted accurately from small oligomers. The pentamer reorganization energies, as representatives for the polymer, are shown to be highly correlated with the trimer reorganization energy, but this correlation and accuracy increase with other descriptors.

We speculate that the accuracy of the one point model, based on the properties of the trimer, occur because this length approximates the dihedral angles of the polymer based on steric crowding between ligands and other effects. Moreover, while an infinite number of possible lines could fit through one point, we find the slope of the regression lines are correlated. That is, species with high, less negative HOMO energies, show greater shifts as a function of oligomer length. Moreover, strong acceptor monomers, such as 3,4-dinitrothiophene 29, show an unusual oligomer slope, in which dimers and oligomers are predicted to be harder to oxidize (more negative HOMO) than the monomer.

Our overall trends related to individual molecules did not yield surprising results, but show that large, bulky groups reduce mobility as the polymer chain increases in length and a more highly twisted oligomer backbone emerges. The high correlation between the slope of the linear regression in HOMO eigenvalues and the extrapolated polymer HOMO and analogous comparison with the HOMO-LUMO gap suggest that the degree of delocalization reflected in the slope is higher in monomers and polymers with electron-rich, less negative HOMO eigenvalues. Additionally, we have demonstrated that the amine analogue to PEDOT, with its similar chemical electronic structure requires further attention from the community. Finally, we find that four homopolymers in our sample group yield predicted OPV device efficiencies $> 8\%$ by Scharber criteria, have high (less negative) HOMO energies, suggesting a donor-donor strategy may provide an alternative to the current donor-acceptor design. Models, such as those presented here, can be used in larger projects exploring chemical space as reliable ways to predict polymer properties from smaller molecules such as trimers and tetramers. We believe that these statistical models can serve as rapid first screens for a wide range of optoelectronic electronic structure properties.

4.6 ACKNOWLEDGEMENTS

We thank the University of Pittsburgh, including the the Center for Simulation and Modeling for computational resources, Pitt Center for Energy for support, and NSF (CBET-1404591). GRH thanks Dr. Noel OBoyle and Prof. Tara Meyer for discussions.

5.0 GENETIC ALGORITHM OPTIMIZATION OF ORGANIC PHOTOVOLTAIC MATERIALS

The text in this chapter has been adapted from a manuscript which is in preparation for submission for publication.

5.1 INTRODUCTION

Computational molecule selection is of interest to many areas of research ranging from drug discovery to discover of new materials for many industries. Many methods are used for this computational selection, each of which has virtues and faults. One method attractive to many researchers is the development of libraries of chemical structures, which can be searched for molecules of interest for specific applications. Although these libraries take significant computational time to generate, once they are constructed, they can be easily searched for molecules with particular features. One library study successfully developed a library with small organic molecules for blue emission with strong oscillator strengths and low singlet-triplet energy gaps that favor thermally activated delayed fluorescence (TADF) emission.[185] A second study generated a library using the Algorithm for Chemical Space Exploration with Stochastic Search (ACSESS) that allows searching the uncharted areas of the small molecule universe and the mining of chemical libraries that do not yet exist.[186] Beratan continued his work to extend the ACSESS library to generate a range of molecular libraries with a set of compounds that is illustrative of the small molecule universe and shows preference for molecules with favorable physical property values.[187]

While library generation helps with discovery of all possible small molecules, the question

remains as to how to search for particular properties in an efficient way? Machine learning methods provide an alternative strategy to generation of large databases. Aln Aspuru-Guzik et. al. report a generative model allowing for efficient searching and optimization through open-ended spaces of chemical compounds. This generative model works by training deep neural networks on hundreds of thousands of existing chemical structures to construct two coupled functions: an encoder and a decoder with their method being demonstrated for design of drug-like molecules and organic light-emitting diodes.[188] While this method allows continuous optimization for molecules, these are inherently discrete species and to ensure accurate results, the encoder and decoder functions require a lot of training. But is it possible to screen without calculating everything first? One approach is the use of general optimization algorithm based on an interpolation of property values on a hypercube which for larger libraries, electronic structure calculations were performed on less than 0.01% of the compounds in the larger libraries,[189] greatly reducing the computational cost of target molecule discovery. Our work uses another method to screen copious numbers of molecules while performing significantly fewer calculations than production of a library with all possible molecule properties of molecules and mutations of molecules screened within a Genetic Algorithm (GA). The benefit of a GA over other methods is that the GA learns as it runs and selects molecules within the given range of property values. In our current work, we look to determine the degree to which our GA speeds up calculations compared with library generation or brute force calculation of all possible molecules. A GA can be a complementary method to machine learning.

In this work, the key question is whether genetic algorithm screening methods can be made reliable and efficient for performing discrete property-driven optimization of molecules. We seek to grow the search space from 500,000 compounds in our previous work[63], ultimately to 50 million molecules, by enlarging the pool of potential monomers from 129 to 1759, and sampling all possible sequences.[102,115,134,135] While still much smaller than all molecular space, by testing multiple runs and establishing a convergence criteria, we find our methods to be 6000-8000 times faster than brute force search. We outline remaining areas of difficulty in growing to larger search spaces, potential solutions, and filtering criteria for potential organic photovoltaic materials. The promise of efficient genetic algorithm sampling of molecular

properties can be applied to many other molecular search areas in the future.

5.2 COMPUTATIONAL METHODS

5.2.1 Monomer Data Sets

Five data sets, comprising of 129, 442, 611, 908, 1235 and 1759 monomers, were prepared for this study by selecting small monomers that are likely to be used in organic photovoltaics. Monomers were selected from literature reports or obvious synthetic modifications of conjugated monomers and span a range of aromatic and conjugated species. For this study, most of the species studied contain a combination of C, H, N, O, S, F and those containing Si and Se were excluded. In addition, we restricted polymerization sites to those considered most synthetically likely. SMILES for these data sets are shown in Tables B1-B41 and molecule images are shown in Figures B1-B41. Please note, that while there is some overlap in the data sets, they are not identical sets from one sized data set to the next. A range of electron-donating and electron-with-drawing substituents were considered. The monomers span a wide range of electronic properties as shown in Table 5.1, with highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) eigenvalues displaying a several electron-volt span for each data set. For comparison, thiophene is computed to have a HOMO eigenvalue of -8.54 eV and a LUMO at -2.03 eV.

5.2.2 Generation of Optimized 3D Structures

The 3D structure of a homotetramer was generated using a multistep process starting with the SMILES string for the polymer.[72] An initial 3D structure was generated using Open Babel 2.2.3[104] (accessed through its Python interface Pybel)[105] and minimized using the MMFF94 force field[82–86] (500 steps using steepest descent minimization, convergence at 1.0^{-4} kcal/mol). Next a weighted-rotor search (MMFF94, 100 iterations, 20 geometry optimization steps) was carried out to find a low-energy conformer. This was then further optimized using MMFF94 (500 steps). Finally, Gaussian09[154] was used to optimize the structure using the

Number of Monomers	Monomer HOMO (eV)		Monomer LUMO (eV)		Tetramer HOMO (eV)		Tetramer LUMO (eV)	
	Min	Max	Min	Max	Min	Max	Min	Max
129	-8.59	-4.17	-3.93	-1.39	-10.33	-5.71	-3.47	+0.58
442	-8.13	-2.70	-5.56	+1.90	-10.39	-4.19	-4.73	+2.86
611	-8.59	-1.29	-4.41	+2.40	-11.32	-3.31	-5.56	+2.92
908	-8.59	-1.22	-4.17	+2.29	-10.66	-3.92	-4.74	+2.45
1235	-8.59	-1.50	-4.41	+2.02	-11.00	-3.25	-4.51	+2.92
1759	-8.59	-2.74	-4.45	+2.40	-11.32	-3.25	-4.73	+2.86

Table 5.1: Computed ZINDO HOMO and LUMO eigenvalue ranges for the monomers and homotetramers in each data set indicating the increasing diversity in the search pools with increased number of monomers.

PM6 semiempirical method.[68] The entire procedure required ∼8 min per oligomer on one CPU core.

5.2.3 Prediction of Electronic Structure and Optical Excitation Energies

The energies and oscillator strengths of the 15 lowest- energy electronic transitions were calculated from the PM6- optimized geometry[68] using the ZINDO/S method[69] as implemented in Gaussian09[154]. The Python library cclib[190] was used to extract the molecular orbital eigenvalues, energies and oscillator strengths of the electronic transitions. The accuracy of this method was tested in our previous work[63] and was shown to be sufficient.

5.2.4 Synthetic Accessibility

To limit the possible search space and to concentrate on the most synthetically relevant species, we considered copolymers formed by preparing a dimer of two different monomers, followed by polymerization to make tetramers of all possible sequences (as illustrated in Figure 5.1) generating 272,480 tetramers (131 monomer set), 3,118,752 tetramers (442 monomer set), 5,963,360 tetramers (611 monomer set), 13,205,952 tetramers (909 monomer set),

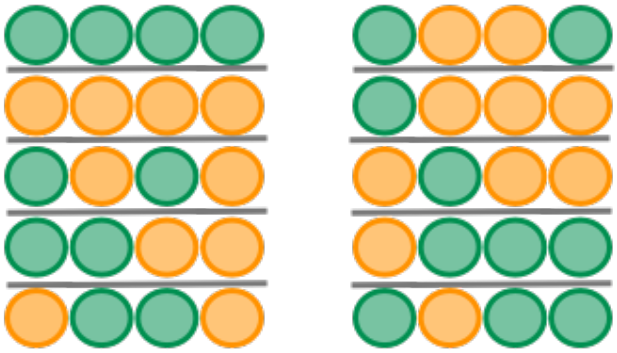


Figure 5.1: Tetramer sequences permitted within the genetic algorithm for the combination of two monomers.

24,383,840 tetramers (1235 monomer set) and 49,477,152 tetramers (1759 monomer set). In addition, hexamers were tested for the data sets containing 131, 442 and 909 monomers, increasing the number of molecules screened to over 52 million compounds.

5.2.5 Calculation of Energy Conversion Efficiency

As with our previous work, the predicted power conversion efficiency was calculated as described by Scharber et al.[71] on the basis of the orbital energy levels and first excitation energy of the oligomer/polymer donor material relative to (PCBM). Our implementation is identical to previous work[63] such that the objective function used was either to maximize the calculated energy conversion efficiency or to minimize the Euclidean distance to a desired HOMO and electronic transition energy (-5.7 eV and 1.39 eV respectively). Better performance was found with the latter and this was what was used for production runs. In both cases, the value of the objective function was penalized by taking into consideration the oscillator strength of the associated electronic transition. For a particular polymer, we chose the lowest energy singlet transition with oscillator strength greater than 1.0. Where none existed (among the 15 calculated), we used the transition with the maximum oscillator strength but scaled the resulting efficiency by the value of the oscillator strength (when the objective function was the efficiency) or added 1.0 minus the oscillator strength to the

distance (when the objective function was the distance). In this way, the objective function tended towards polymers which had both the desired electronic properties and electronic transitions with reasonable oscillator strength.

5.2.6 Genetic Algorithm

A genetic algorithm (GA) is a stochastic method for global optimization based on concepts from evolutionary biology in which a population of chromosomes is optimized in successive generations by applying the evolutionary operators of crossover, mutation and selection. In our study, the chromosomes were the candidate polymers, and successive generations minimized the deviation from the HOMO and electronic transition energy values at the point of maximum efficiency.

Our implementation of the GA using Python, such that each chromosome consisted of a base dimer along with a composition index that indicated how to generate the polymer from the base dimer. An initial population of N polymers was generated consisting of N random dimers with randomly chosen values for the composition indices. A selection operator chose N/5 polymers using tournament selection with tournaments of size 3. Once selected, that polymer was removed from the pool for further selection. Two polymers were randomly selected (with replacement) from this set for crossover. Their base dimers were combined to form two new base dimers which, when combined with random values for the composition index, yielded two new child polymers. Each child was subject to a mutation opera-tor which, for each monomer of the base dimer, had a 3 in 4 chance of replacing it with a monomer randomly chosen from the seven monomers most similar to it (similarity was determined as described in Methods). Children that were duplicates of current population members or of other children were dis-carded. The process of crossover and mutation was repeated until N offspring were created. The next generation was then formed from the N/2 best chromosomes in the original population along with the N/2 best offspring.

During both training and production, the genetic algorithm was run for 100 generations (at which point convergence had occurred). The number of chromosomes was set at 64 after training on the tetramer data showed that this achieved an acceptable compromise

between run-time and performance. A key feature of our genetic algorithm implementation was the mutation operator allowing for mutation between monomers with similar electronic properties. To define similar, for each of the monomers we generated 3D structures (as described above) of the corresponding homopolymer of length 4 and carried out a ZINDO/S single-point calculation. Similar monomers were defined as those whose homopolymers had similar LUMOs and similar HOMO-LUMO gaps (measured by Euclidean distance).

5.2.7 Analysis

Results from the genetic algorithm were analyzed using python with numpy[191] and pandas[192] modules to generate histograms of monomers most often chosen and determine spearman correlations and the percentage of monomers which were identical in different data sets to identify the top monomers and the number of generations needed for the convergence of the set of top monomers.

5.3 RESULTS AND DISCUSSION

5.3.1 Scaling of our GA to massive search spaces

Determining the number of generations required to converge the list of top performing molecules is essential to scaling a genetic algorithm to large search spaces. Ideally, data from several runs, with different initial populations, can provide a set of monomers in which the most frequently chosen monomers are identical in each set. Data set convergence to these sets of top monomers was determined through calculation of the Spearman rank correlation that quantifies how well two lists are described with a monotonic function when ranked in order of energy. Two identically ordered lists produce a perfect Spearman correlation of +1, and an inversely ordered list provides a perfectly inverse Spearman correlation of 1. (Figure 1.7)

For each of the five data sets described in the computational methods section, Spearman correlations (ρ) were calculated at intervals within the range of 1-100 generations. Each set of

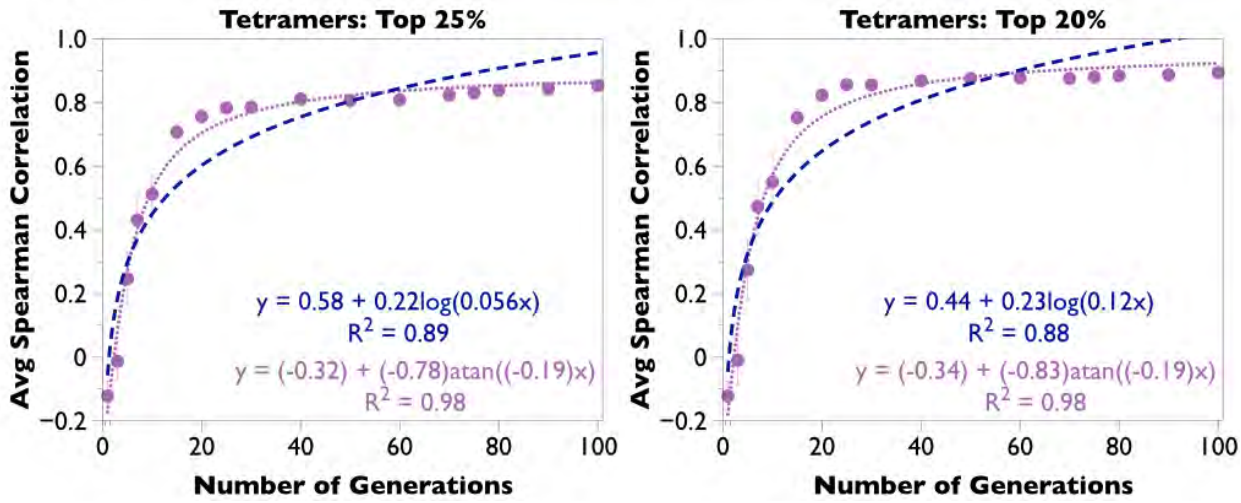


Figure 5.2: Justification of choice of tan^{-1} function to fit logarithmic type data

data was then fit with a line of best fit to determine number of generations to convergence. Convergence of each data set to a set of top monomers was approximated at the value where the average Spearman correlation is equal to 0.50. This cutoff was chosen due to the logarithmic scaling exhibited by the data, and therefore it would be necessary to calculate vastly more generations to achieve greater correlation. An analysis of this threshold value is performed below. Equations of best fit for the top 5, 10, 15, 20 and 25% of the data were determined and used to calculate the number of generations required to achieve data set convergence. (Figure 5.3)

Error in the generations to convergence were calculated from the equation of the tan^{-1} line of best fit. Since we have defined convergence as the point when the Spearman correlation is equal to 0.50, the first step is to calculate the value of x, the number of generations, when y, the average Spearman correlation, is equal to 0.50. Next the values of x are calculated at y + s of the residual (error in the fit) and y - s of the residual. The error is then defined as the value of x at y plus s of the residual minus the the value of x at y minus s of the residual. This calculation gives the total error and therefore divide this number by 2 to determine the error above and below the calculated generations to convergence. For the top 25% of the

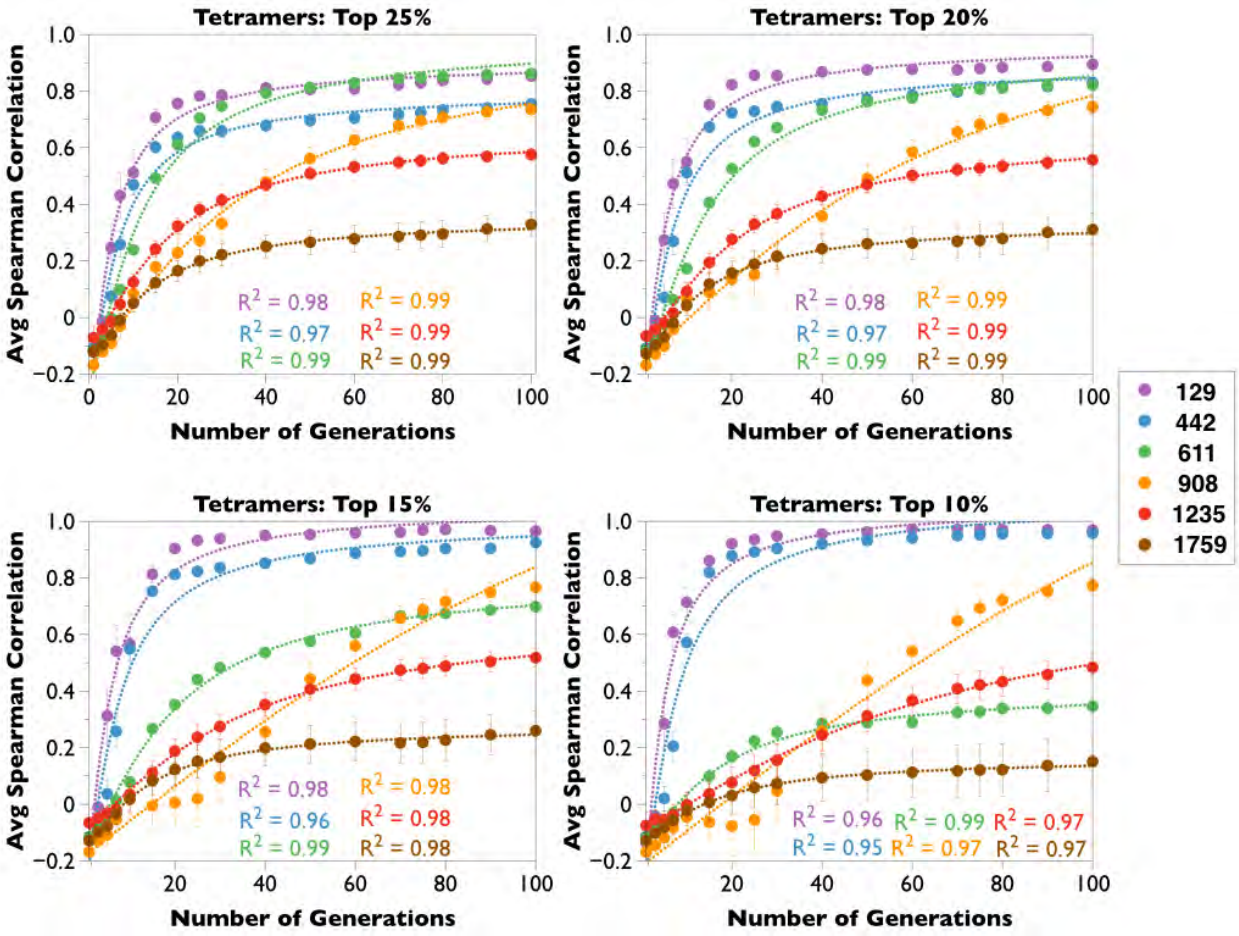


Figure 5.3: Plots of average spearman correlation at each number of generations for the top 25%, 20%, 15% and 10% of the data sampled. The line of best fit shown here are tan^{-1} fits of the average Spearman correlation at each data point.

Top 25%

Number of Monomers	Equation of Fit	x value when y=0.50	s of residual	0.5+s	Value of x with 0.5 + s	0.5-s	Value of x with 0.5 - s	Error range
129	y = (-0.32) + (-0.78)*atan((-0.19)*x)	9.2	0.04	0.54	10.4	0.46	8.2	2.2
442	y = (-0.28) + (-0.69)*atan((-0.15)*x)	14.15	0.048	0.548	17.15	0.452	11.92	5.23
611	y = (-0.26) + (-0.8)*atan((-0.084)*x)	16.65	0.045	0.545	18.8	0.455	14.81	3.99
908	y = (-0.18) + (-0.75)*atan((-0.031)*x)	41.21	0.022	0.522	43.8	0.478	38.81	4.99
1235	y = (-0.13) + (-0.51)*atan((-0.059)*x)	48.61	0.014	0.514	53.27	0.486	44.63	8.64
1759	y = (-0.16) + (-0.33)*atan((-0.074)*x)	no solutions exist						

Table 5.2: Error in the calculation of the number of generations to convergence of data for the top 25% of data. tan^{-1} equations and the lines of best fit are from all calculated data points.

data, the error values are summarized in Table 5.2.

The results for the number of generations to convergence are summarized in Table 5.3 and show that as the data set increases in magnitude, the number of generations needed for the convergence of the top monomer set also increases only modestly. For completeness, the top 25% of the hexamer data was analyzed and results corroborated the tetramer data: the 129 monomers set converges in 5.99 $\pm$ 0.98 generations, the 442 monomer set converges in 12.86 $\pm$ 1.52 generations and the 909 monomer set converges in 57.15 $\pm$ 10.7 generations. These values are within error of the tetramer values showing that the data set is effectively being screened. Using the number of generations required for each of the different sized data sets to achieve to convergence, it is possible predict the number of generations to convergence for data sets of different. (Figure 5.4)

For data sets smaller than 1235 monomers, the model is slightly quadratic or possibly linear, but at larger values (i.e., 1759) the data does not converge to a set of top monomers since the search space is evidently too large. The data sets examining the top 25% and 20% both show R^2 values for the quadratic fit of 0.93 for the prediction of the number of generations required to achieve convergence of a data set and R^2 values for the linear fit of 0.91, excluding the point for 1759 monomers.

Number of Monomers	25%	20%	15%	10%
129	9.2 ± 1.1	8.4 ± 1.1	7.27 ± 0.93	6.76 ± 1.1
442	14.2 ± 2.6	11.7 ± 2.0	10.57 ± 1.94	10.4 ± 2.2
611	16.7 ± 2.0	20.4 ± 2.2	33.16 ± 2.77	Does not converge
908	41.2 ± 2.5	51.3 ± 3.6	59.99 ± 5.72	61.76 ± 6.6
1235	48.6 ± 4.3	60.7 ± 5.2	80.38 ± 5.26	97.42 ± 3.8
1759	Does not converge			

Table 5.3: The number of calculated generations to convergence of the top monomers for the top X% of the tetramer data. Convergence was calculated using the equations of fit to exceed 0.5 correlation.

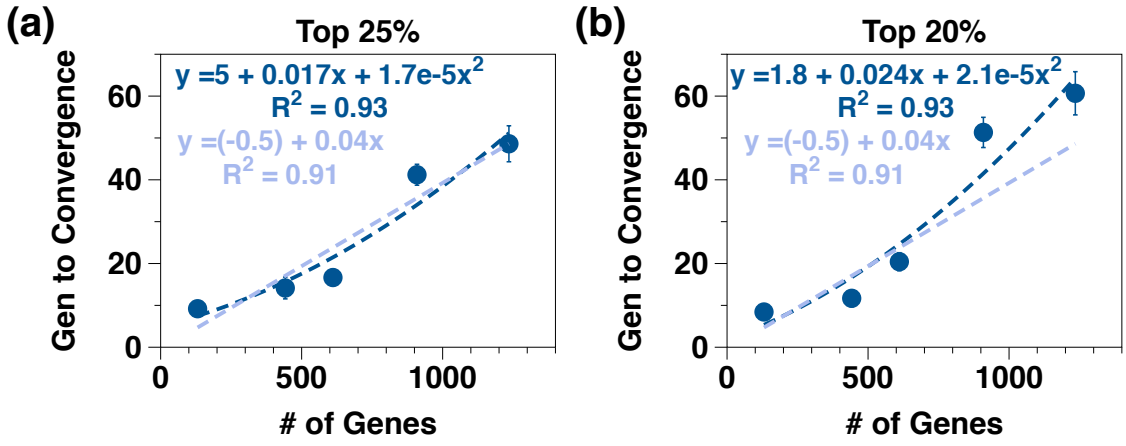


Figure 5.4: Number of generations required for top monomer convergence for different sized data sets. For both (a) the top 20% and (b) the top 25% data sets, the number of generations required for data set convergence is quadratic or linear excluding the 1759 monomer data set which did not achieve convergence

The error in the number of generations to convergence, as reported in Table 5.3 was determined by taking the residual from the line of best fit divided by the square root of the sample size. The number of generations to convergence was then calculated for 0.5 plus and minus this value and the error was determined by subtracting the two new values for the generations to convergence. Unsurprisingly, as the data set increases in size, the error in the number of generations to convergence also increases since more space to sample, with all other parameters remaining constant, and therefore not efficiently sampling the entire data set, as completely as in smaller data sets.

5.3.2 Efficiency of our GA approach

The discrepancy in the number of generations to convergence with 1235 and 1759 monomers demonstrates that a region exists in which the number of monomers in a data set have maximum efficiency. Each generation in our screening consists of 64 calculations (one for each chromosome). The fraction of the calculations which we must perform to achieve convergence for a data set of a particular size is the number of calculations to convergence (i.e., 64 calculations per generation times the number of generations to convergence) divided by the number of calculations in an exhaustive search for that data set. The speedup is therefore the reciprocal of this value or the number of calculations in an exhaustive search divided by the number of calculations to convergence. While the 'easiest' solution to screening millions of molecules would seem to be to screen through an exhaustive search where calculations are performed for each molecule, in reality this is a very slow and therefore costly approach. As illustrated in Figure 5.5, our GA converges to a set of top monomers with a speedup of $\sim$6000x over brute force (for 1235 monomers across the top 20%), and a speedup of $\sim$8000x (for 1235 monomers across the top 25%). Since the largest search space did not converge, for future screening of large data sets, a tournament style approach would likely improve efficiency with increasingly large groups. In a tournament approach, the initial monomer pool would be divided into several optimally sized groups ($\sim$1,000 monomers each). Each group can be screened for top monomers and then the top monomers can compete against one another to determine the overall top monomers.

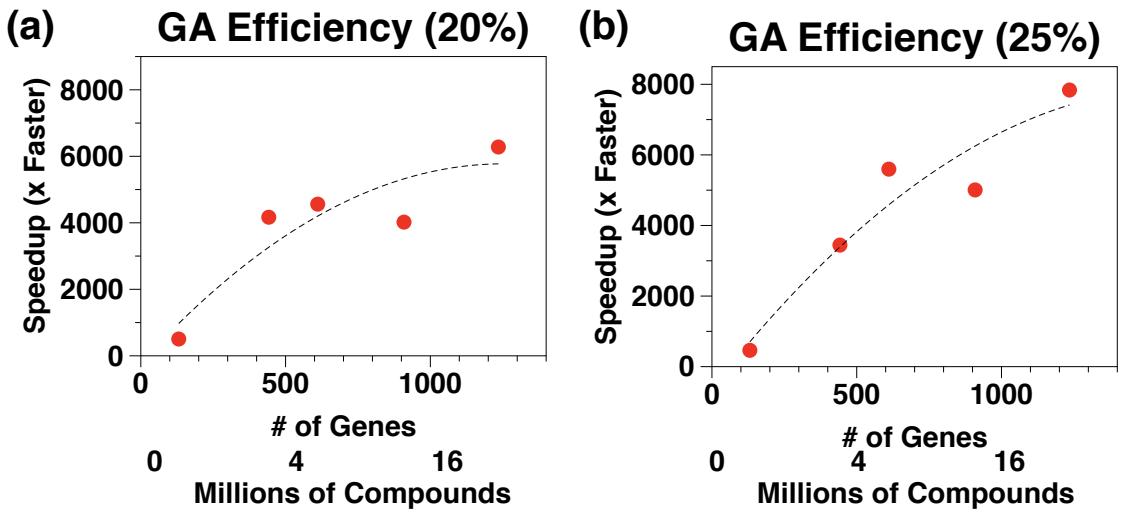


Figure 5.5: The speedup as calculated from the number of calculations performed when running the genetic algorithm compared with the number of calculations required for an exhaustive search. As shown in (a) top 20% and (b) top 25%.

5.3.3 Monomer Hot Spots

Initial random selection of monomers yield a diverse population of HOMO and LUMO energies after the first generation in each set of data, independent of the size of the data set. However, as the GA proceeds through multiple generations, HOMO and LUMO energies which are selected in the remaining population of monomers narrows to a 'hot spot' that emerges within each group of data. (Figure 5.6) These hotspots emerge as the number of generations approaches the number of generations at which the data set converges. In each of the data sets studied, tetramers with HOMO energies ranging from -9 to -5 eV and LUMO energies in the range of -3 to +1 eV are selected with high frequency as candidate materials for solar cells. The existence of an energy 'hot spot' can be used to prescreen candidates in future experiments to eliminate tetramers with energies far outside this range and thus excluding them from the similarity matrix. This would enable more thorough screening of larger data sets in fewer generations as some molecules are eliminated before the GA begins working on the selected data.

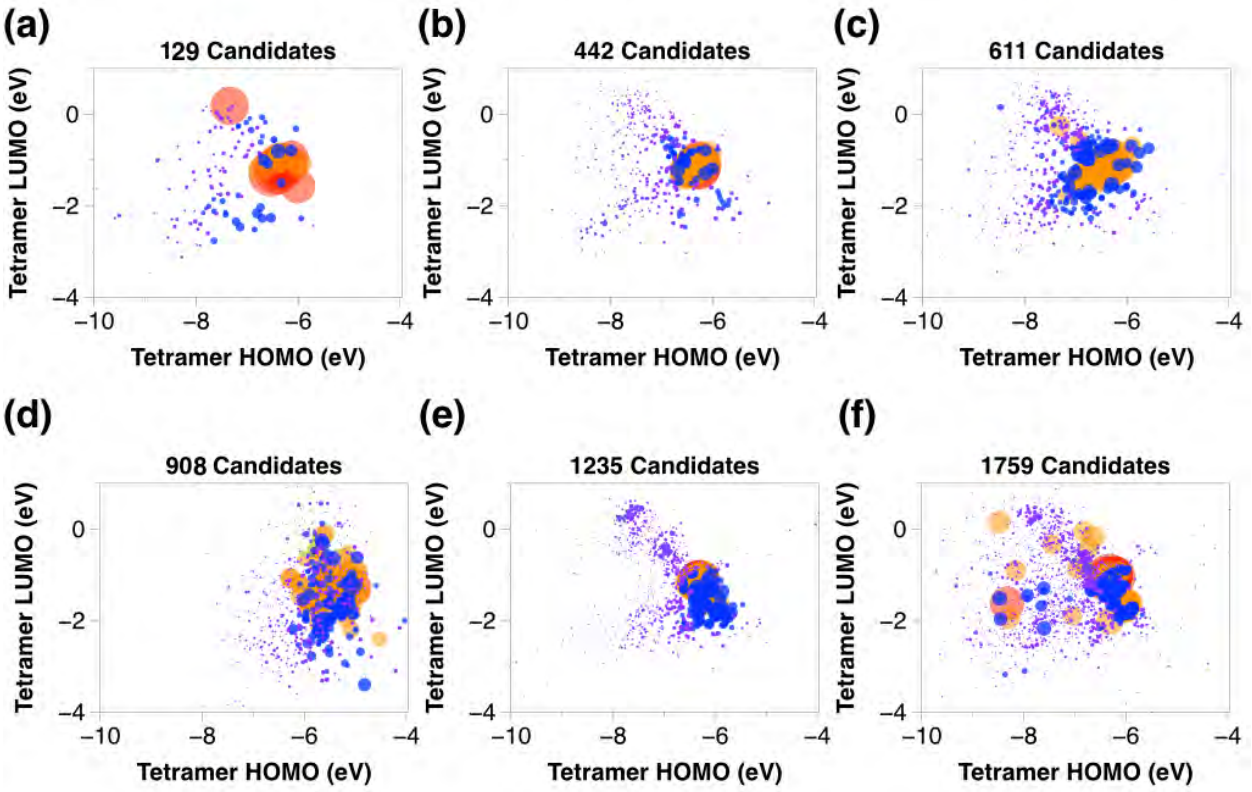


Figure 5.6: "Hotspots" of homotetramer HOMO and LUMO data (after 100 generations), with size and color of the spot normalized based on the frequency of occurrence of each monomer.

5.3.4 Additional Predictive Properties

Homotetramer HOMO and LUMO energies have significant impact on the species selected by the GA. This can allow for eliminating molecules with energies significantly outside these thresholds. To further improve the GA, it would be beneficial to identify other predictive properties that can reliably remove uninteresting molecules and focus toward useful candidate molecules. A stepwise regression was performed to determine properties which effect the frequency with which a molecule is chosen and included properties such as tetramer HOMO, tetramer LUMO, tetramer excitation energy, tetramer oscillation strength, monomer HOMO and monomer LUMO. For each data set, the parameters selected that control the frequency of monomer selection are tetramer HOMO and tetramer LUMO, which are already used in the GA. Interestingly, an additional parameter of the tetramer excitation energy effects the frequency in each of the data sets examined. A random forest analysis corroborated this finding. In future work, generation of the similarity matrix, will include the tetramer HOMO, tetramer LUMO and tetramer excitation energy, leading the similarity matrix to be a better predictor of the ideal materials and provide top monomer convergence in fewer generations.

5.3.5 Top Monomers from the GA

While the determination of the most likely monomer candidates for solar applications is the goal of running this genetic algorithm, deciding how to define this group of top monomers was not straight forward. In our previous work with 129 monomers, we were able to analyze the top 25 monomers. When comparing sets of data with substantially different numbers of monomer, we realized that examining the top X% of the data set was a more meaningful comparison since most of the data comes from a small subset of the monomers as shown by the histograms in Figure 5.7. To determine which percentages to examine, the number of monomers which comprise 25%, 50%, 75%, 80%, 85%, 90% and 95% were examined and converted to percentages of the total data for comparison between different sized data sets. The results from this are reported in Table 5.4 and show that up to 90% of the data comes from a small percentage of the total number of monomers. We therefore analyze the top

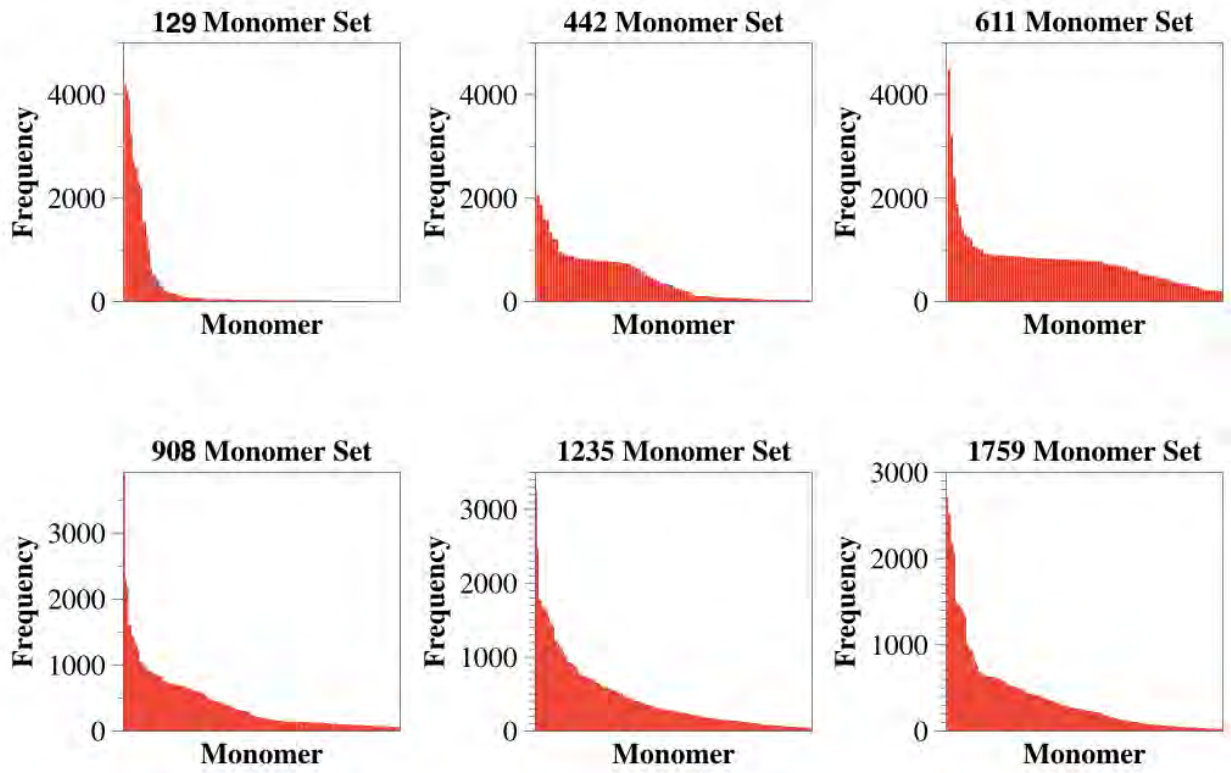


Figure 5.7: Histogram of frequency of the selection of each monomer after 100 generations in different sized monomer sets.

Percentage of Top Candidates	Fraction of monomers that comprise top X% of candidates after 100 generations					
	129	**442**	**611**	**908**	**1235**	**1759**
25%	2%	1%	2%	1%	1%	1%
50%	5%	3%	2%	4%	3%	2%
75%	8%	5%	12%	6%	6%	5%
80%	10%	6%	14%	7%	7%	5%
85%	12%	7%	17%	9%	9%	6%
90%	18%	10%	23%	11%	11%	8%
95%	34%	23%	36%	27%	19%	16%

Table 5.4: Analysis of the percentage of monomers which comprise different percentages of data reveals that up to 90% of the final data comes from a small fraction of the initial data set.

10% and the top 20% of the monomers as these data sets contain most of the data from a given run of the GA.

Analysis to determine the top monomers in the top 10% and top 20% of the data set and whether there is overlap in the top data sets unaffected by the starting pool of monomers was performed on the percentage of the original data set, rounded up to whole numbers. The number of monomers for the top 10% analysis was 14 (129 monomer set), 45 (442 monomer set), 62 (611 monomer set), 91 (909 monomer set), 124 (1235 monomer set) and 176 (1759 monomer set). Likewise, the number of monomers for the top 20% analysis was 27 (129 monomer set), 89 (442 monomer set), 123 (611 monomer set), 182 (909 monomer set), 247 (1235 monomer set) and 352 (1759 monomer set). Analysis was performed on the data after 100 generations and after the number of generations calculated for each size data set to converge to a set of top monomers and these calculations both yield similar results reinforcing our calculations which predict the results of the number of generations to reach a set of top monomers. Table 5.5 presents the number of monomers from each smaller group of data that overlap with the larger data sets, in a pairwise analysis after 100 generations in each case. In both analysis of both the top 10% and 20% of the candidate molecules,

	Top 10% of Candidates						Top 20% of Candidates					
	129	**442**	**611**	**908**	**1235**	**1759**	**129**	**442**	**611**	**908**	**1235**	**1759**
129		85%	64%	79%	64%	71%		59%	52%	48%	52%	59%
442			70%	67%	60%	60%			62%	56%	54%	54%
611				47%	42%	40%				36%	40%	39%
908					33%	48%					31%	59%
1235						75%						65%

Table 5.5: Overlap analysis of the pairwise combination of the data for the top 10% and top 20% of the candidates reveals that most of the candidates that are chosen in the smallest data group (129) are still chosen as important candidates when the monomer pool is increased by more that 10-fold.

an overwhelming number of monomers from the smaller pool of candidates appear in the larger sets of data, even through multiple runs each starting with a different random seed of candidates from the given data set. Building on the finding that there is significant overlap in the pairwise analysis of the different sized data sets, we questioned whether there are a group of candidate molecules which appear in the top 10% and the top 20% of all data sets examined in this study. Figure 5.8 illustrates the seven monomers that are present in all data sets in both the top 10% and the top 20% of the data. While only three of these species have been synthesized, all seven have previously been cited in papers or patents as potential materials for improved OPV materials.

5.3.6 Sequence Analysis

As with the top monomer analysis, we examined the sequences chosen for each of the monomer pairs selected by the GA to determine if the GA can select a method to combine two monomers in a more favorable manner to generate more efficient materials for solar energy. After examining all the data sets, we determine that each of the sequences is chosen about equally except for AAAA which is surprisingly chosen 10 times more often compared with the other possible sequences. Calculations of three data sets with hexamers (129, 442

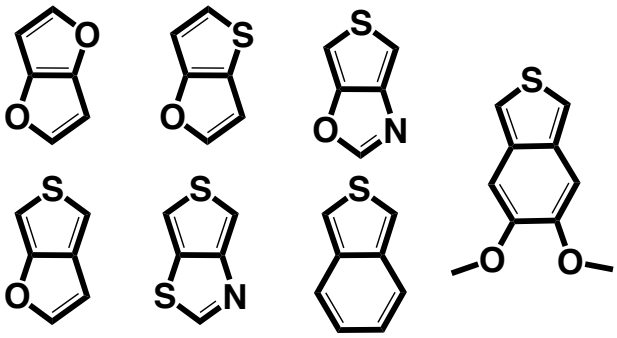


Figure 5.8: The top seven monomers which appear across all data sets for the top 10% and top 20% of the data.

and 908) indicate that the same bias exists with hexamers. In future experiments, adjusting the algorithm to place more importance on the sequence selection may help further screen materials. Interestingly, despite the ubiquitous use of copolymers in experimental investigation of organic electronics, the GA indicates selected homopolymers may still have interesting properties.

5.3.7 Top Monomer Pairs

In addition to analysis of top monomers chosen, selected monomer pairs were also examined. In each step of the GA, two monomers are chosen to form a co-oligomer and therefore, it is logical to look at the pairs with the top performance and frequency. Data was examined after 100 generations for each of the tetramer runs by selecting each of the two monomers, ensuring a unique combination (e.g., AB = BA) and then counting the frequency of each pair. Figure 5 illustrates the eight monomer pairs that were selected most often across the 442, 611, 908, 1235 and 1759 data sets. The set of data with 129 candidates is excluded from this analysis, since some of the monomers were not present in this smaller set. This data is corroborated with the hexamers from the 442 and 908 runs which show the same set of top monomers. Clearly the isobenzothiophene, a known low-band gap system, is an influential motif, particularly when combined with electron-donating substituents such as

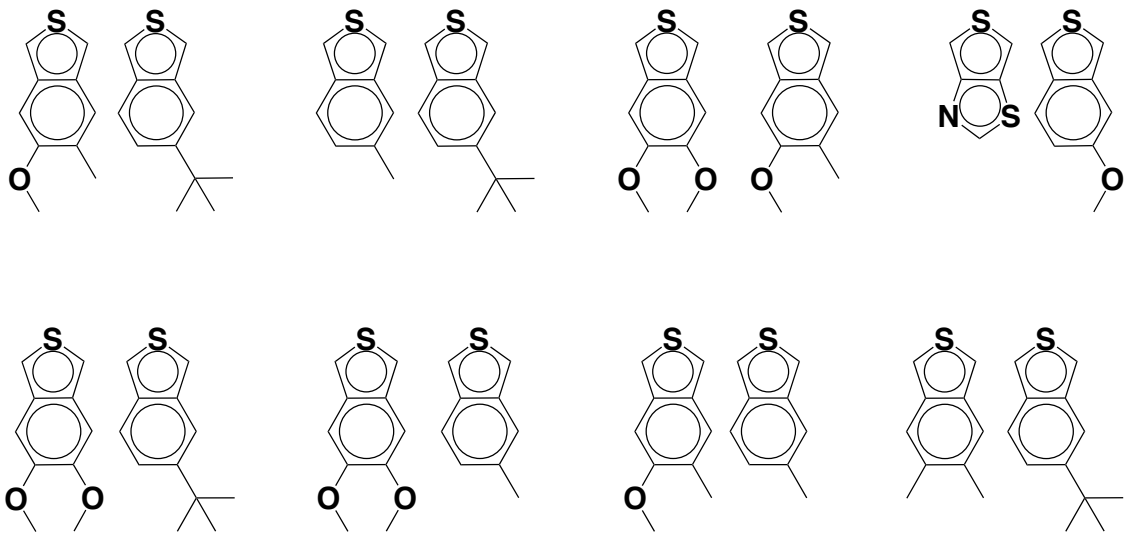


Figure 5.9: Monomer pairs from tetramer GA runs with 442, 611, 909, 1235 and 1759 candidates for the top 20% of the monomers in each set after 100 generations. The set of data with 129 candidates is excluded from this illustration since some of the monomers included these common pairs from other runs were not present in the 129 set and therefore were not chosen as top monomers pairs.

alkyl or alkoxy functional groups.

5.4 CONCLUSION

Genetic algorithm methods are known as efficient techniques for optimizing discrete variables, such as molecular structures. This work shows that, when selecting conjugated oligomers, the GA approach is 6000-800 times faster than brute force. Since the GA is independent of computational methods used, it can provide rapid filtering of lead compounds using first principles, semiempirical, or machine learning methods alike. Moreover, we find only modest scaling of the number of generations required to converge the set of top candidates up to a search space of 25 million compounds. Unfortunately, convergence, as judged by the Spearman rank correlation of the top candidates, is not found for a larger search space. Instead, the GA appears stuck in local regions and cannot explore the entire (vast) search space. In future work, we believe a divide-and-conquer approach that partitions large searches into smaller regions, combined with a competition among top candidates will address this problem. Moreover, we find all searches, regions of ?hot spots? (Figure 3) which comprise monomers frequently incorporated in top candidates. This suggests broader searches can perform some level of initial filtering based on these properties. That is, a new monomer found far from a hotspot is unlikely to be among top candidates and can likely be ignored. Predicting the frequency of monomer genes among top candidates will clearly improve the efficiency of the GA search the presence of such hotspots suggests that for organic electronic materials, statistical and machine learning approaches can rapidly identify interesting new leads. Molecular space is known to be enormous, but the application of efficient GA search techniques offers great promise for finding optimal and near-optimal targets for a wide range of computationally-driven properties. The techniques outlined here for organic electronic materials can easily be adopted for many other electronic structure properties, from redox potentials and activation energies, to polarity, polarizability and dielectric constants.

5.5 ACKNOWLEDGEMENTS

We thank the University of Pittsburgh, including the Center for Energy for support, and NSF (CBET-1404591). GRH thanks Dr. Noel O'Boyle and Prof. Tara Meyer for discussions.

6.0 A SOBERING ASSESSMENT OF CLASSICAL FORCE FIELD METHODS FOR LOW ENERGY CONFORMER PREDICTIONS

The text in this chapter has been adapted from Kanal, I. Y.; Keith, J. A.; Hutchison, G. R., *A Sobering Assessment of Classical Force Field Methods for Low Energy Conformer Predictions*, submitted to the Journal of Chemical Theory and Computation. The author's contribution to the work includes most of the analysis for the project.

6.1 INTRODUCTION

Molecular mechanics (MM) using classical force fields is a highly efficient way to calculate molecular energies and gradients of up to millions of atoms[193]. Given their efficiency, they are also widely used for screening and filtering large numbers of molecular structures for atomic scale properties for solar materials[63], computational drug design[194,195], and/or conformer searching[104,196–198]. In all cases, the quality of the screening naturally depends on the accuracy of the force fields, and a careful assessment is therefore needed to establish their utility in these applications. The present work focuses on assessing the accuracy of classical MM methods used in conformer search applications. Most molecules with four or more atoms have some level of conformational flexibility, and even small molecules possess multiple thermally-accessible conformer geometries.[199] Although classical force fields are widely used to identify low energy conformers, recent studies have questioned the reliability of classical force field methods.[73] Kaminsk and Jensen have also reported detailed benchmarking studies of conformational energies of amino acids, showing limitations of classical force fields with fixed charges for biomolecular applications[198,200]. Consequently, many works consider-

ing the use of conformer generation tools, benchmarks are performed, not by considering a low-energy geometry, but by comparing the geometry of an experimental crystal structure against some ensemble (e.g., 50-100) of conformers.[78],[201] Given a reasonable tool, one might guess that generating enough conformers should produce something close to the experimental geometry, so finding a method, such as force field energies, to score or rank conformers is critical. This creates a need for deeper understanding of the limitations of classical force fields across broader chemical applications. We find several common assumptions are often made to rationalize the use of classical force fields for conformer searches (or other similar applications, such as molecule-protein docking). One assumption is that energy calculations from a classical force field need not be highly accurate to obtain reasonable molecular geometries. A second assumption is that a well-trained force field will be reasonably accurate for molecular structures that fall within the chemical space of the fitted parameterization, even if it performs poorly on species outside of the fitted parameterization. The last assumption is that even though force fields may or may not reliably identify the lowest energy conformer, they can be used to locate low energy conformers in a reliable fashion. In the present work we have carried out a comprehensive investigation to assess the validity of each of these assumptions.

6.2 TEST SET SELECTION

A data set consisting of x-ray crystal structures of 700 small molecules capable of being in multiple conformers was provided to us by Eberjer[78] and were derived from the work of Hawkins et al.[201] along with ligands from the Astex Diverse Set[202].

6.3 COMPUTATIONAL METHODS

We generated geometrically diverse conformers using Open Babel[104] for each molecule in the data set. Up to 250 conformers were generated using a genetic algorithm set to maxi-

mize the root-mean-square displacement (RMSD) between conformers.[104] From the starting geometry of each conformer, conjugate gradient geometry optimizations were performed using Open Babel with the MMFF94[82–86], UFF[80] and GAFF[203] classical force fields or with the PM7[204] semiempirical method using OpenMOPAC.[205] Kohn-Sham Density Functional Theory (DFT) electronic energy calculations were carried out on subsets of these geometries using ORCA[206] with the B3LYP exchange correlation functional,[111,207] the def2-SVP[208,209] basis set, the RI and RIJCOSX[210] approximations, and the D3BJ[211] dispersion correction scheme. To our knowledge this is the most extensive computational validation set to date for studying low energy conformers molecules.

6.4 ANALYSIS

Data analysis was performed using Python scripts with the numpy[191] and scipy.stats libraries incorporating the pandas[192] module. We report Spearman correlations that relate to how well two variables can be described with a monotonic function when ranked in order of energy values. A perfect Spearman correlation is +1, and a perfect inverse correlation is -1. Besides Spearman correlations, we also report R^2 values, x-coefficients, coefficients of the intercepts, and slopes for up to 250 conformers for each of the 700 molecules. The Python scripts used to perform the calculations are available at https://github.com/ghutchis/conformer-scoring.

6.5 RESULTS AND DISCUSSION

Energies of each conformer were analyzed with the OLS (ordinary least squares) regression as integrated with the pandas and numpy libraries in Python to determine R^2 values. Figure 6.1a illustrates how one R^2 value is obtained by calculating the correlation between 250 different conformations optimized using MMFF94 and PM7 for one single molecule ('astex_1l7f'). Note that comparing MMFF94 and PM7 conformer energies consistently results in large scatter and a very low R^2 value. Histograms of all R^2 values across the molecular data set

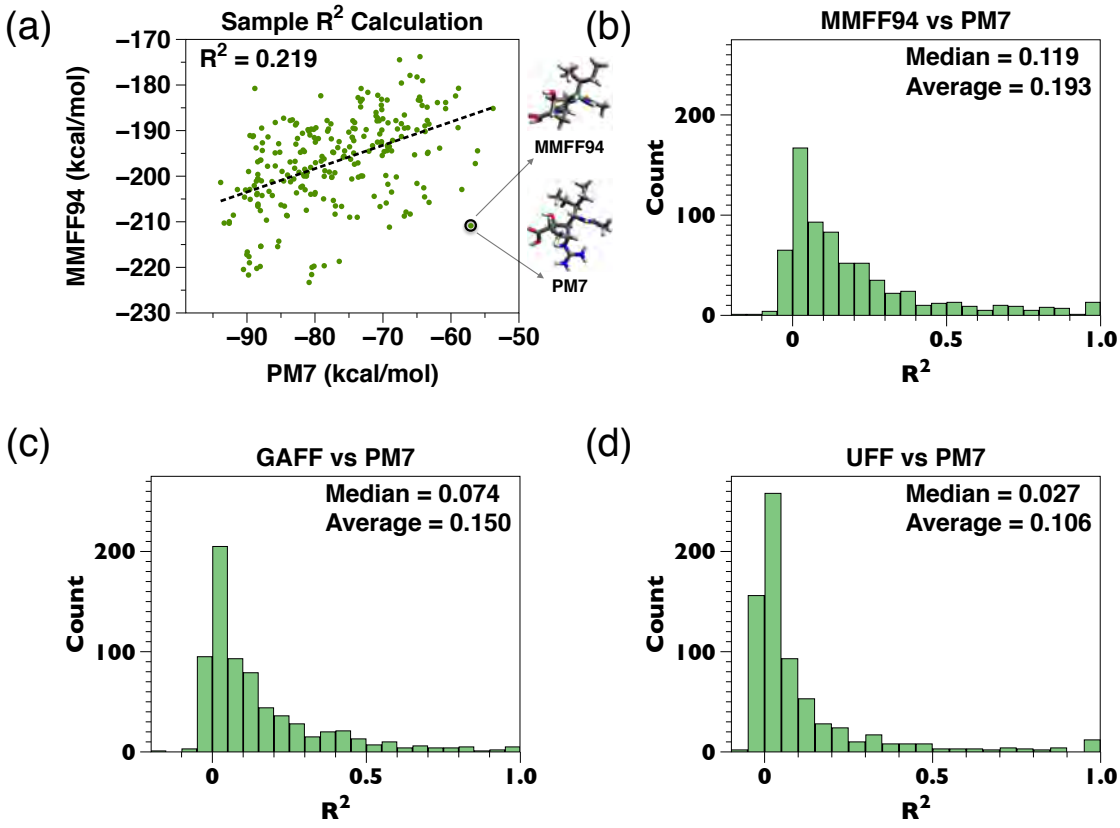


Figure 6.1: (a) 250 data points representing different conformers used to calculate one R^2 value for the 'astex_1l7f' molecule. Similar analyses were performed on up to 250 conformers for each of the 700 molecules in our dataset. The circled point on the plot represents a conformer that contributes to the low R^2 value by having significantly different MMFF94 and PM7 geometries. (b-d) Histograms of 700 R^2 values obtained from the entire data set. MMFF94, GAFF, and UFF all show similarly poor correlation with higher level PM7 methods.

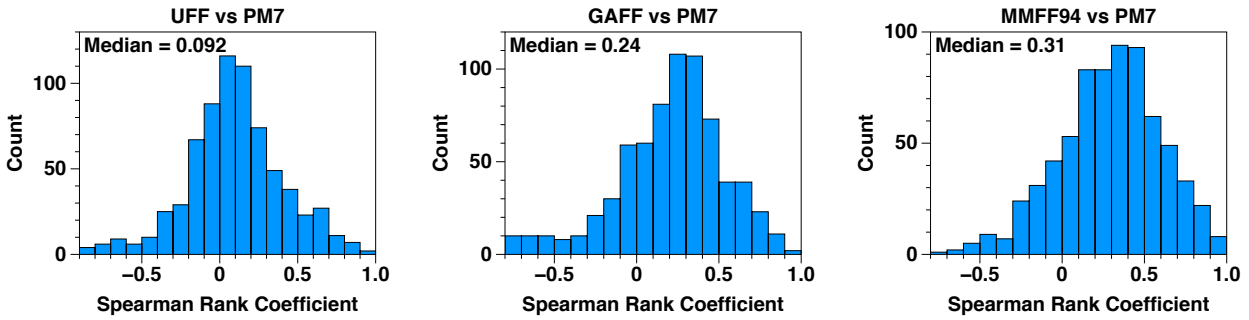


Figure 6.2: Spearman Rank correlations for each of the data set combinations which are discussed in the text with regard to R^2 values. In determining the accuracy of a method, the researcher would expect either the absolute energy values to be accurate, which is explained through the R^2 values or the expectation would be that even if the energy values are not accurate than the order of the ranking of energies from largest to smallest should be accurate. These values show that the ranking as described by the Spearman rank correlation are not accurate.

	R^2		Spearman rank	
	Median	**Average**	**Median**	**Average**
UFF vs. PM7	0.027	0.106	0.092	0.099
GAFF vs. PM7	0.074	0.150	0.240	0.210
MMFF94 vs. PM7	0.119	0.193	0.312	0.291
PM7//MMFF94 vs. PM7	0.276	0.219	0.455	0.434
MMFF94 vs. DFT//PM7	-0.001	0.107	-0.103	-0.086
PM7//MMFF94 vs. DFT//PM7	-0.008	0.090	0.455	0.434
MMFF94//PM7 vs. DFT//PM7	0.318	0.365	-0.455	-0.349
PM7 vs. DFT//PM7	0.263	0.340	0.618	0.488
DFT//MMFF94 vs. DFT//PM7	0.017	0.125	0.200	0.191

Table 6.1: Median and average R^2 values and Spearman correlations for the full data set.

are shown in Figures 6.1(b-d). We find that correlations between classical force fields and semiempirical PM7 are very poor. Spearman rank correlations also demonstrate similarly poor results (Figure 6.2). Table 6.1 shows the median and average R^2 and Spearman correlation values for all data. Note that for all methods, median and average R^2 correlation between force fields and PM7 are a paltry 0.1-0.2.

Although MMFF94 has perceived reliability in generating molecular geometries,[212] these data suggest that all classical force fields have similarly large problems reliably identifying and ranking structurally diverse conformers. *Thus, the assumption that force fields can reliably represent trends in low energy conformers compared to higher level quantum chemistry methods is simply not safe.*

The data above show that MMFF94 demonstrates slightly better correlation with PM7 compared to UFF and GAFF, so we considered PM7 single point energies calculated on MMFF94-optimized geometries (i.e. PM7//MMFF94 calculations). Figures 6.3(a-b) and Table 6.1 show that PM7//MMFF94 data has slightly higher median/average R^2 values (0.276/0.219) compared to MMFF94 data (0.119/0.193) vs. PM7 data. Median and average Spearman rank correlations show a similar trend for MMFF94 (0.312/0.291) and PM7//MMFF94 (0.455/0.434) vs. PM7. Although these results demonstrate slightly improved correlations, the correlation of PM7//MMFF94 with PM7 is still underwhelming,

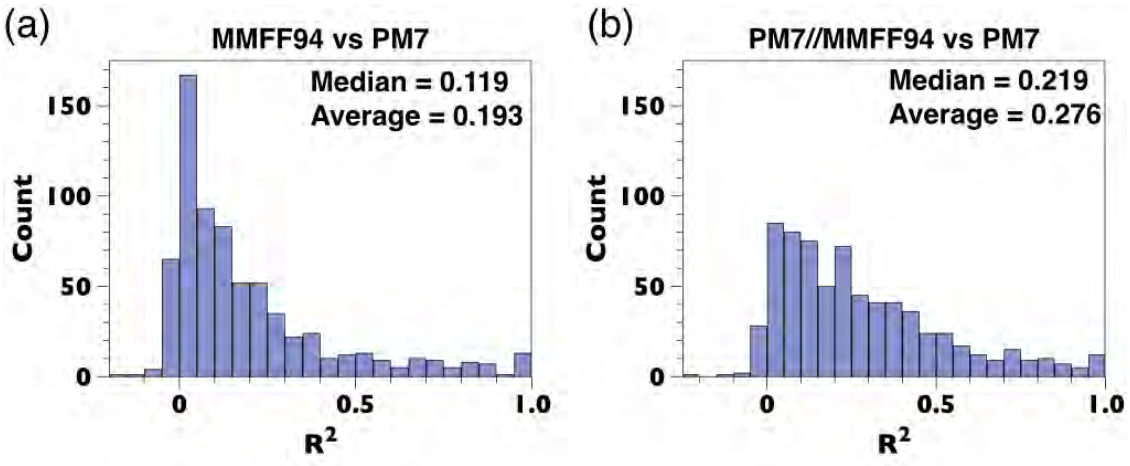


Figure 6.3: Histograms of R^2 values obtained using MMFF94 and PM7//MMFF94 data vs. PM7 data. Higher level single point energy calculations carried out on geometries obtained from classical force fields are only slightly more correlated to higher level theory.

suggesting that MMFF94-optimized geometries are unreliable.

6.6 COMPARISON WITH DFT

We now assess the quality of MMFF94 and PM7 energies and geometries using DFT (B3LYP-D3BJ/def2-SVP calculations). We calculated DFT single point energies (i.e. DFT//PM7 and DFT//MMFF94) for up to ten of the lowest energy conformers from separate PM7 and MMFF94 optimizations on each of the 700 molecules. Although the accuracy of this DFT approach is expected to be deficient compared to more robust electronic structure methods with larger basis sets, it provides a practical representation of a method that should be more reliably accurate than PM7. Figure 6.4 shows histograms of R^2 values for MMFF94, PM7//MMFF94, and PM7, each vs. DFT//PM7 calculations. The data show that standard MM calculations provide wholly unreliable representations of conformers. The median/average R^2 values are (a) -0.001/0.107 for MMFF94, (b) -0.008/0.090 for

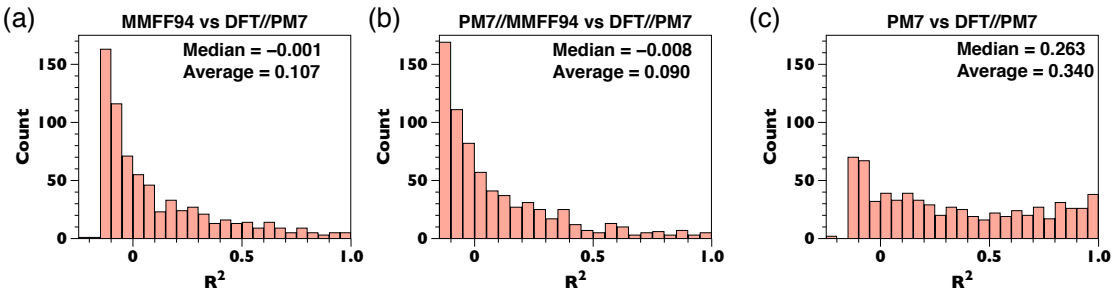


Figure 6.4: Histograms of (a) MMFF94, (b) PM7//MMFF94, and (c) PM7, each vs. DFT//PM7 R^2 data. Calculations utilizing force fields correlate very poorly with DFT//PM7 R^2 data, but PM7 correlates less poorly vs. DFT//PM7.

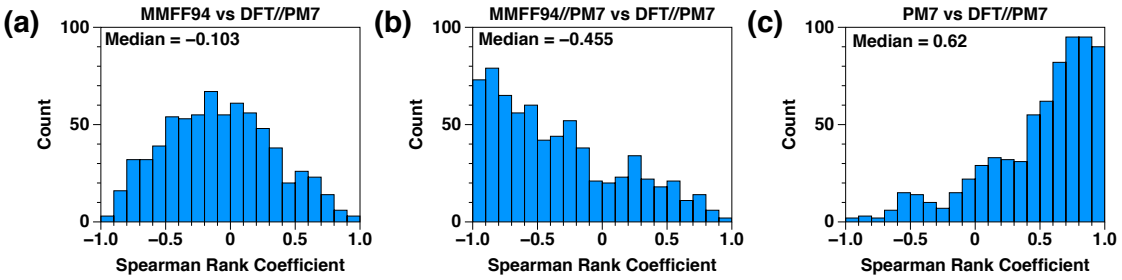


Figure 6.5: Spearman rank correlation data for MMFF94, MMFF94//PM7 and PM7 calculations vs. DFT//PM7 calculations. The results show poor correlation of the ordered ranking of the energies of the molecules between MMFF94 and MMFF94//PM7 with DFT//PM7 and good correlation between the PM7 and DFT//PM7 energies ordered rankings. This suggests that MMFF94 provides poor ordered rankings even when performed on a molecule already optimized with PM7 and that PM7 provides molecule energies in similar orders to the energies of DFT.

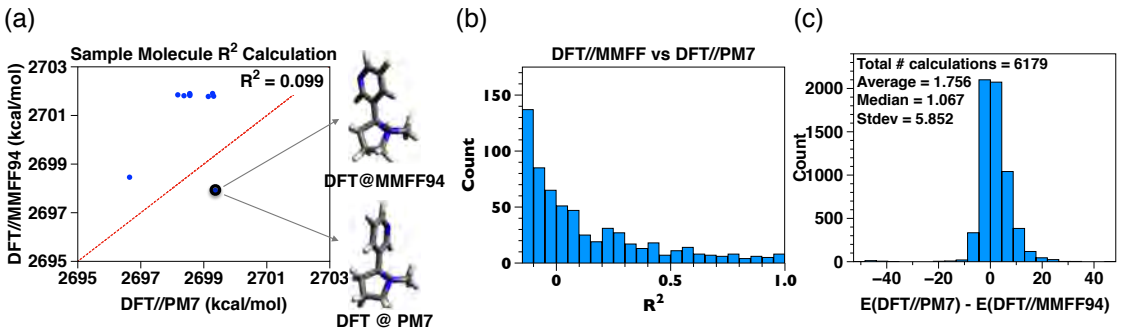


Figure 6.6: (a) 10 data points representing different conformers used to calculate one R^2 value for the 'astex_1p2y' molecule. (b) Histograms of 6179 R^2 values obtained from a subset of the full data set. The results show poor correlation for DFT//MMFF94 vs. DFT//PM7 data. (c) Histogram of the atomization energy differences E(DFT//PM7) - E(DFT//MMFF94), showing that PM7 optimized geometries are on average lower in energy than MMFF94 geometries within the DFT model.

PM7//MMFF94, (c) 0.263/0.340 for PM7 data, each vs. the DFT//PM7 data. Spearman rank correlations show similar results as shown in Figure 6.5 with median/average values of (a) -0.103/-0.086 for MMFF94, (b) -0.455/0.434 for PM7//MMFF94, and (c) 0.618/0.488 for PM7, each vs. DFT//PM7, as shown in Table 6.1. Though R^2 correlations are not particularly good for PM7 vs. DFT//PM7, the Spearman rank correlations are better, showing that PM7 is significantly more reliable for ranking conformers.

Figures 6.6a-b show correlations between DFT//MMFF94 and DFT//PM7 data. Note that Figure 6.6a shows energies as atomization energies, where the larger number represents a more strongly bound state. The median/average values for R^2 values are 0.017/0.125, and Spearman rank correlations are also similarly poor (Average/median of 0.19/0.20). Figure 6.6c shows a histogram of DFT//PM7 - DFT//MMFF94 atomization energies having an average of 1.76 kcal/mol, median of 1.07 kcal/mol and standard deviation of 5.85 kcal/mol. In short, using B3LYP-D3BJ calculations, we find there is frequently a very poor correlation between PM7-optimized and MMFF94-optimized geometries for the same initial conformer.

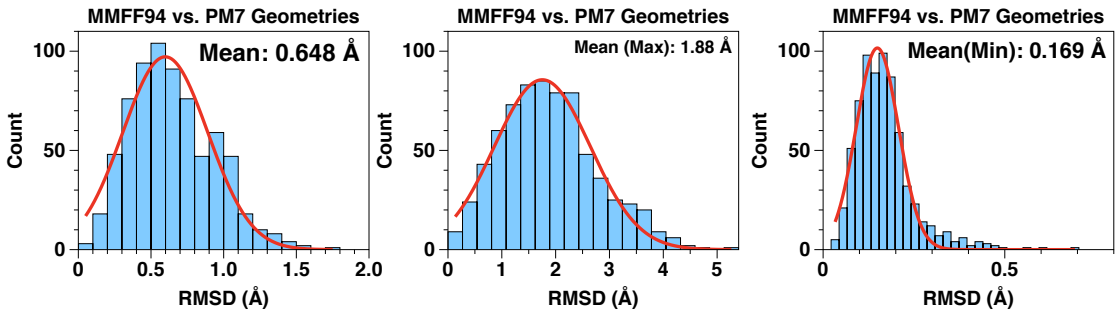


Figure 6.7: RMSD data shows that when beginning with the same starting pose, the optimized MMFF94 and PM7 geometries differ on average by 0.6 Å per heavy atom. The mean of the minimum RMSD values shows that beginning with the same starting pose, the optimized MMFF94 and PM7 geometries minima differ on average by 0.2 Å per heavy atom and the maxima of the average RMSD differ by 1.9 Å per heavy atom.

Moreover, Figure 6.6c indicates that the PM7 geometries are, on average, more stable than corresponding MMFF94 geometries according to the DFT calculations. Therefore, even if higher-level DFT methods are used to evaluate the energies of geometries from MMFF94 optimizations, they are not as reliable as geometries from PM7 optimizations.

In short, since the energies of MMFF94 and PM7 are different, the potential energy surfaces strongly differ. Even when beginning from the same starting conformer geometry, both methods frequently result in quite different optimized geometries. While none of the methods show strong correlation with one another, the worst correlations with DFT//PM7 data are those that involve classical force fields (Figure 6.7).

6.7 ENERGETIC RANGES: HOW MANY CONFORMERS IN AN ENSEMBLE?

Conformer searches aim to identify the most stable conformer or ensemble of conformers. Open source and commercial conformer generation software packages can automate the gen-

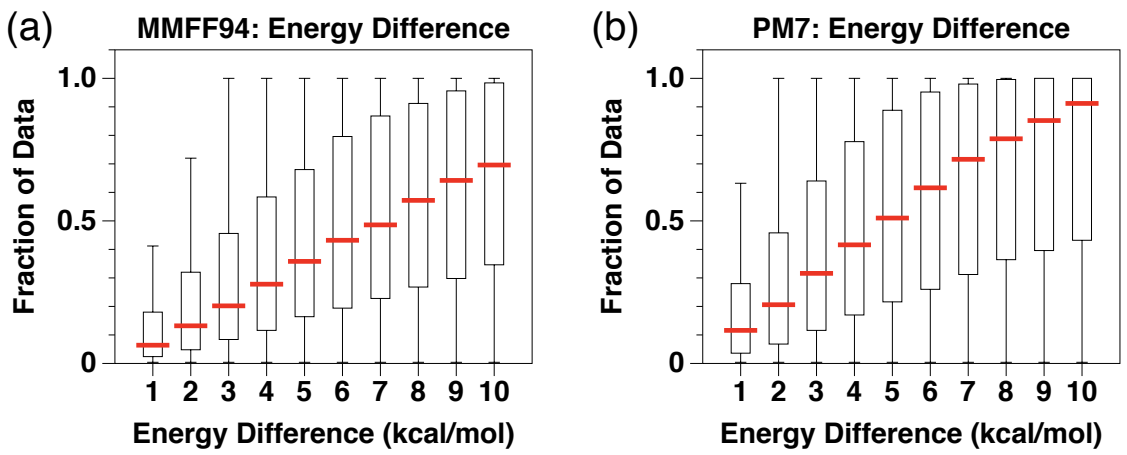


Figure 6.8: Fraction of the data set within energy differences ranging from 1-10 kcal/mol when using (a) MMFF94 and (b) PM7. The red lines represent the median value, the bottom of the square box represents the first quartile, the top of the box represents the third quartile and the endpoints at the top and bottom of the lines represent the maxima and the minima, respectively.

eration of hundreds or potentially thousands of conformers.[74,75,79,213–215] However, as shown above, classical force fields simply do not provide reliable energies or geometries conformer screening. To identify a practical solution, we determined the fraction of the conformers in our data sets that were within a given energy range of the lowest energy geometry, as computed by a particular method. The number of conformers that were within 1-10 kcal/mol at 1 kcal/mol intervals were then counted. Figure 6.8 summarizes these results. Figure 6.8a shows that ∼6% of the conformers generated using MMFF94 are within 1 kcal/mol of the lowest energy conformer, while 70% of the conformers generated are within 10 kcal/mol. In the case of PM7 data, ∼12% of the conformers are within 1 kcal/mol of the lowest energy conformer and 91% are within 10 kcal/mol. This shows that the potential energy surfaces from MMFF94 (and presumably other classical force fields) and PM7 methods are very different.

6.8 USING FORCE FIELDS FOR ROUGH OPTIMIZATION

Computationally efficient methods are often used for fast and rough geometry optimization so that fewer optimization steps are needed for further optimizations with higher level methods. Our data indicate that using force fields for rough optimizations is actually inefficient and likely counter-productive. Figure 6.3 shows that PM7//MMFF94 data poorly correlates with PM7 data. Moreover, the MMFF94 potential energy surface for conformers appears to be very different from that from PM7 (Figure 6.9). The average PM7 gradient norm when starting from an MMFF94 optimized conformers is 140 kcal/Å, and the minimum gradient norm is 50 kcal/Å, showing that MMFF94-optimized geometries are often not close to their corresponding PM7-optimized geometries. The average heavy-atom root mean square displacement (RMSD) between MMFF94 and PM7 optimized geometries starting from the same initial state is 0.6Å (Figure 6.7). It is thus not a surprise then that MMFF94 and PM7 geometry optimizations result in very different final geometries and very different energy rankings.

For this reason, the use of classical MM methods for optimizing molecular structures having multiple torsional degrees of freedom is only advised if the precision and accuracy of the final structures and rankings obtained from the conformer searches is of little or no concern. For example, while the correlation between any given conformer geometry and the experimental crystal structure will depend on the method used to generate the initial conformer ensemble, we find that the lowest RMSD for each molecule is, on average closer to the experimental crystal geometry for PM7 than MMFF94 by 0.02Å (Figure 6.10a) and generally larger (0.03-0.06Å), depending on the number of rotatable bonds (Figure 6.10b,c). In short, the energetic rankings from PM7 better correlate with DFT and the geometries are frequently closer to the experimental crystal structure geometry.

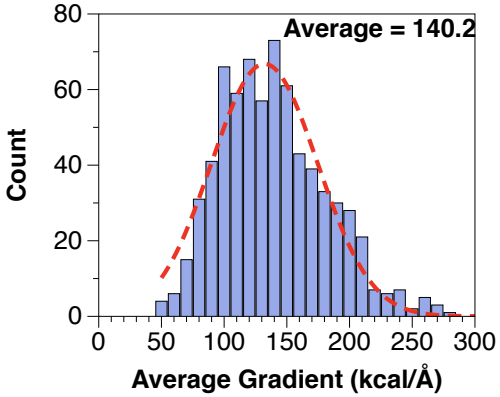


Figure 6.9: Histogram of the average gradient norm across all conformers for a molecule using PM7 on an MMFF94 optimized geometry. The gradients are expected to be close to 0. An analogy to this figure would indicate that if you are on solid ground (i.e., low energy) in MMFF94 world, then when teleported to PM7 world, on average you would be ∼140 feet in the air. Roughly the same gradients are used for the initial non-optimized conformers. PM7 world is not the same as MMFF94 world. The potential energy surfaces do not map, as expected.

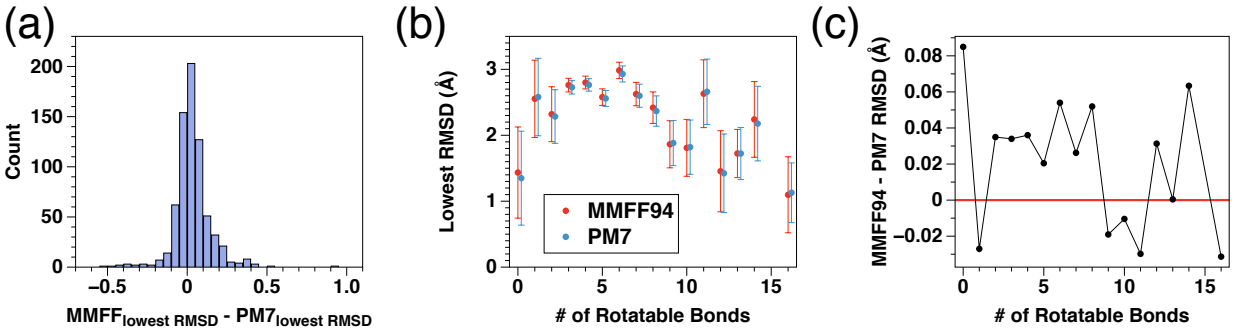


Figure 6.10: (a) Lowest RMSD compared with the number of rotatable bonds show on average closer to the experimental crystal geometry for PM7 than MMFF94 by 0.02 and (b-c) generally larger (0.03-0.06), depending on the number of rotatable bonds

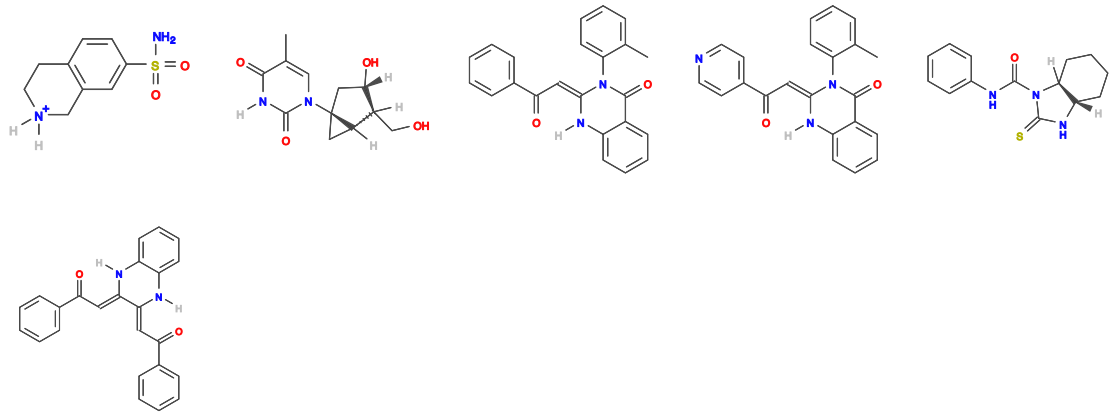


Figure 6.11: Molecules that resulted in R^2 values greater than or equal to 0.80 with MMFF94 vs DFT, MMFF94 vs PM7 and PM7 vs. DFT.

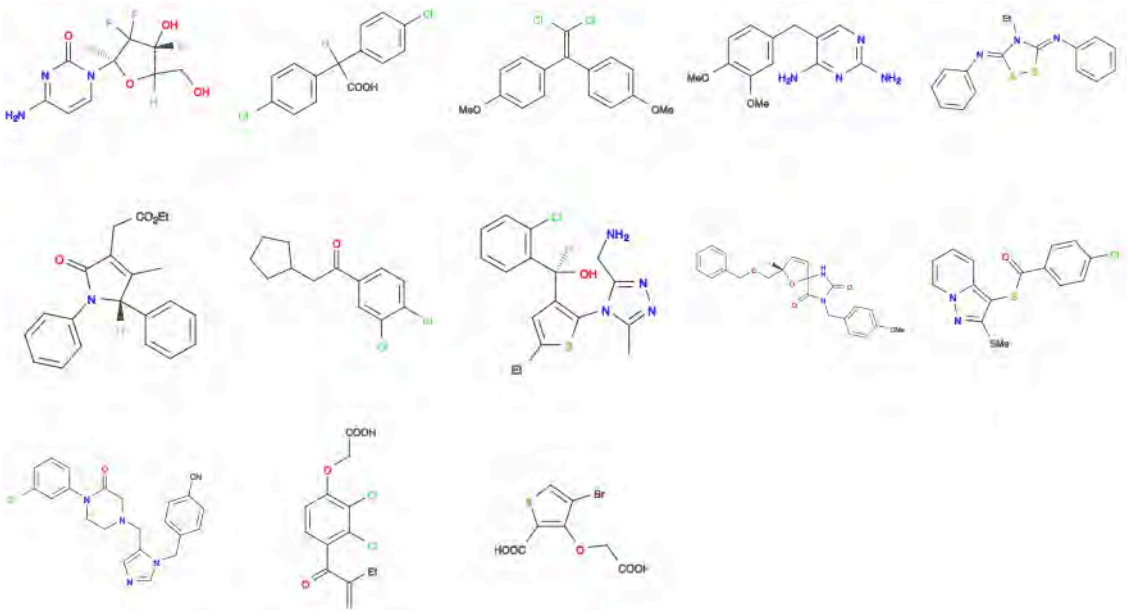


Figure 6.12: Molecules that resulted in R^2 values below zero with MMFF94 vs DFT//PM7, MMFF94 vs PM7 and PM7 vs. DFT//PM7.

6.9 ANALYSIS OF PROBLEM MOLECULES

Classical force fields are parameterized, and thus it is possible that poor performance reflects a need for improved parameterizations. Some of the molecules in our data set had R^2 values uniformly greater than or equal to 0.80 for MMFF94 vs. PM7, MMFF94 vs. DFT//PM7, and PM7 vs DFT//PM7 calculations (Figure 6.11). In these cases, classical force field parameterizations are doing a respectable job identifying and ranking conformers. There were also cases where molecules had R^2 0, between lower-level methods and higher-level methods (Figure 6.12). Visual comparison of molecules in Figures 6.11 and 6.12 suggests that molecules that classical force fields have difficulty with have more rotatable torsions and/or contain halides. However, we screened all of the functional groups with SMILES across our entire data set and actually found no statistical evidence of specific functional groups being more present in problem cases than in the well-performing cases. We also note that there were many molecules with R^2 values near zero. In addition, 45 molecules demonstrated R^2 values below 0.05 as calculated from MMFF94 vs. PM7, MMFF94 vs. DFT//PM7, and PM7 vs. DFT//PM7 calculations.

In short, our statistical analysis indicates that the poor performance of MMFF94 (and presumably other classical force fields) is not simply due to a particular failure in parameterization and that a solution requires more or better fitting. Instead, the issue is systematic, and neither the energies nor the optimized geometries of classical force fields should be trusted for conformational searching or related applications. The energies, and in turn the potential energy surface produced by general-purpose force fields like MMFF94 in general do not correlate with more accurate quantum chemical methods such as PM7 or even more accurate hybrid DFT calculations that account for dispersion. Similar investigations of other generic force fields in these applications is warranted, and we provide all of our dataset free of charge at https://github.com/ghutchis/conformer-scoring for this purpose.

6.10 CONCLUSION

We have quantitatively and statistically assessed the accuracy and reliability of classical force fields used in conformer searching applications. Their performances across a large data set of organic molecules shows severe problems that indicate that they are unreliable for conformer searching and/or filtering of low and high energy geometries. Three widely used force fields for general chemistry were investigated (i.e., MMFF94, UFF, and GAFF), and all were found to perform similarly poorly. We assess that all are wholly unreliable for conformer screening despite conventional wisdom. As noted above, conventional assumptions have suggested that even if energies from classical force fields are not entirely accurate, they can produce reasonably high-quality geometries. We actually find that neither classical force field energies nor their geometries seem relatable to data obtained using higher level PM7 or DFT//PM7 energy calculations. This causes the potential energy surfaces from classical force fields that describe complicated multi-dimensional torsional space to be very different from those that would be obtained from higher level methods. Thus, classical force fields should not be trusted to produce accurate potential energy surfaces for large molecules. Moreover, using classical force fields as an initial screen to optimize geometries and/or rank low and high energy geometries makes intuitive sense, but carrying out this procedure with generic classical force fields is likely actually counterproductive. We find not only large deviation between MMFF94-optimized and PM7-optimized geometries obtained from the same initial structure, but the gradients of the MMFF94 method on a PM7 geometry (and vice versa) are also quite large. In current applications, we prescribe that regardless of the software used to generate conformer ensembles, one should generate a diverse set of geometries (e.g., using RMSD diversity) and perform geometry optimizations and subsequent energy calculations using the highest level quantum chemical methods that are tractable. We note that semiempirical methods such as PM7 can be used quite rapidly on modern computing architectures, and this method correlates less poorly with DFT//PM7 data and appears less likely to omit potentially important conformers within 10 kcal/mol of the minimum energy structure. We do not mean to suggest that all force field methods are unreliable for conformer searching, but we have noted that these problems do not seem to be due to the presence of specific functional

groups in some molecules, and thus a need for better parameters. Careful parameterizations and customized force fields derived from quantum chemical methods are certainly useful for specific applications.[216,217] In the short term, we suggest that future parameterizations should attempt to consider more training with non-equilibrium geometries and multiple conformers to ensure that the potential energy surfaces of the force fields better represent the higher-level quantum chemical methods than they do currently. In the long term we note that our work highlights an urgent need for methods that can rapidly and reliably screen drug-like organic molecules.

6.11 ACKNOWLEDGEMENTS

JAK acknowledges financial support from the Department of Chemical & Petroleum Engineering at the University of Pittsburgh, the Central Research Development Fund, and the R. K. Mellon Foundation. GRH and IYK acknowledge financial support from the NSF (CBET-1404591). We thank Rohith Amruthur, Jeffrey Carr, Yinan Kang, and Yaqun Zhu for help in early stages of this project.

7.0 CONCLUSION

In this chapter, the findings of this thesis are summarized and future directions for the work are discussed.

7.1 SUMMARY

7.1.1 Polymer Predictive Properties

7.1.1.1 Sequences The results from our tetramer studies suggested that the sequence in which the monomers are arranged on average has an effect on the energy. When expanded to hexamers, it is apparent that there is a more complicated effect that depends on block length and placement of that block or sequence within a hexamer. Since sequence also controls polymer conformation and packing motifs, we believe further investigation is needed to predict and control charge transport and other related solid-state properties. In our collaboration with the Meyer group, we find that sequence is important in both solar cell performance and in related properties and through use of calculations we find that it is possible to explore sequence-space to increase our understanding of structure/function correlations and to direct synthesis.

7.1.1.2 Inverse Design of Conjugated Polymers We have shown that across a set of 100 diverse oligothiophene species, the polymer HOMO, LUMO and HOMO-LUMO gaps, computed from DFT, can be accurately estimated from the values of the trimer HOMO, trimer LUMO and trimer HOMO-LUMO gap calculated values. We also show that these

approximations can be improved through the models presented above, including simple and easy to calculate properties. Rather than performing multiple oligomer calculations to extrapolate the polymer electronic structure, most properties can be estimated readily from modest-sized oligomers.

Polymer reorganization energies, related to the hole transport, are also shown to be predicted accurately from small oligomers. The pentamer reorganization energies, as representatives for the polymer, are shown to be highly correlated with the trimer reorganization energy, but this correlation and accuracy increase with other descriptors.

Our overall trends related to individual molecules did not yield surprising results, but show that large, bulky groups reduce mobility as the polymer chain increases in length and a more highly twisted oligomer backbone emerges. The high correlation between the slope of the linear regression in HOMO eigenvalues and the extrapolated polymer HOMO and analogous comparison with the HOMO-LUMO gap suggest that the degree of delocalization reflected in the slope is higher in monomers and polymers with electron-rich, less negative HOMO eigenvalues. Additionally, we have demonstrated that the amine analogue to PEDOT, with its similar chemical electronic structure requires further attention from the community. Finally, we find that four homopolymers in our sample group yield predicted OPV device efficiencies $> 8\%$ by Scharber criteria, have high (less negative) HOMO energies, suggesting a donor-donor strategy may provide an alternative to the current donor-acceptor design.

7.1.2 Genetic Algorithm

Genetic algorithm methods are known as efficient techniques for optimizing discrete variables, such as molecular structures. This work shows that, when selecting conjugated oligomers, the GA approach is 6000-800 times faster than brute force. Since the GA is independent of computational methods used, it can provide rapid filtering of lead compounds using first principles, semiempirical, or machine learning methods alike. Moreover, we find only modest scaling of the number of generations required to converge the set of top candidates up to a search space of 25 million compounds.

7.1.3 Computational Methods

We have quantitatively and statistically assessed the accuracy and reliability of classical force fields used in conformer searching applications. Their performances across a large data set of organic molecules shows severe problems that indicate that they are unreliable for conformer searching and/or filtering of low and high energy geometries. Three widely used force fields for general chemistry were investigated (i.e., MMFF94, UFF, and GAFF), and all were found to perform similarly poorly. We assess that all are wholly unreliable for conformer screening despite conventional wisdom.

Conventional assumptions have suggested that even if energies from classical force fields are not entirely accurate, they can produce reasonably high-quality geometries. We actually find that neither classical force field energies nor their geometries seem relatable to data obtained using higher level PM7 or DFT//PM7 energy calculations. This causes the potential energy surfaces from classical force fields that describe complicated multi-dimensional torsional space to be very different from those that would be obtained from higher level methods. Thus, classical force fields should not be trusted to produce accurate potential energy surfaces for large molecules. Moreover, using classical force fields as an initial screen to optimize geometries and/or rank low and high energy geometries makes intuitive sense, but carrying out this procedure with generic classical force fields is likely actually counterproductive. We find not only large deviation between MMFF94-optimized and PM7-optimized geometries obtained from the same initial structure, but the gradients of the MMFF94 method on a PM7 geometry (and vice versa) are also quite large. In current applications, we prescribe that regardless of the software used to generate conformer ensembles, one should generate a diverse set of geometries (e.g., using RMSD diversity) and perform geometry optimizations and subsequent energy calculations using the highest level quantum chemical methods that are tractable. We note that semiempirical methods such as PM7 can be used quite rapidly on modern computing architectures, and this method correlates less poorly with DFT//PM7 data and appears less likely to omit potentially important conformers within 10 kcal/mol of the minimum energy structure.

7.2 CONCLUSIONS

This thesis has focused on on development and use of polymer predictive properties for discovery of π-conjugated materials. Applying all of these results to the GA could reduce the calculations required for identification of top monomers. The sequence results could be applied, by allowing only likely sequence arrangements as possible mutations. Future sequence studies will aim to correlate intermolecular interactions, packing, film morphology, and interfacial organization with sequence effects. The result of these sequence studies should allow for combined computational and synthetic rational design of materials that can fulfill the complex set of properties necessary for highly efficient organic solar cells and other applications.

The rate at which predictions of likely candidates from the GA are made, can in future stages be improved through models, such as those presented in the inverse design of polymers study to explore chemical space as a reliable way to predict polymer properties from smaller molecules such as trimers and tetramers. We believe that these statistical models can serve as rapid first screens for a wide range of optoelectronic electronic structure properties. From our study of force fields, we suggest careful consideration before using the force fields as a method within the framework of the GA. While we do not mean to suggest that all force field methods are unreliable for conformer searching, we have noted that these problems do not seem to be due to the presence of specific functional groups in some molecules, and thus a need for better parameters. In the short term, we suggest that future parameterizations should attempt to consider more training with non-equilibrium geometries and multiple conformers to ensure that the potential energy surfaces of the force fields better represent the higher-level quantum chemical methods than they do currently. In the long term we note that our work highlights an urgent need for methods that can rapidly and reliably screen drug-like organic molecules.

Unfortunately, convergence, as judged by the Spearman rank correlation of the top candidates, is not found for a larger search space. Instead, the GA appears stuck in local regions and cannot explore the entire (vast) search space. In future work, we believe a divide-and-conquer approach that partitions large searches into smaller regions, combined

with a competition among top candidates will address this problem. Moreover, we find all searches, regions of 'hot spots' which comprise monomers frequently incorporated in top candidates. This suggests broader searches can perform some level of initial filtering based on these properties. That is, a new monomer found far from a hotspot is unlikely to be among top candidates and can likely be ignored. Predicting the frequency of monomer genes among top candidates will clearly improve the efficiency of the GA search the presence of such hotspots suggests that for organic electronic materials, statistical and machine learning approaches can rapidly identify interesting new leads.

7.3 FUTURE WORK

This work has laid the groundwork for further expanding screening for large groups of compounds through use of the genetic algorithm. We have found the optimal size data set for which our GA can efficiently and effectively search. To expand this work, we suggest using the predictive methods discussed here such as models for polymer property prediction from monomer, dimer and trimer properties to predict polymer properties prior to performing calculations with these molecules. In addition, we suggest the use of "hot spots" to remove molecules from the initial pool whose properties fall far outside the desired property range. Finally, the use of a tournament type scheme will allow screening of larger data sets by breaking a large data set into smaller data sets with approximately 1000 monomers per set, performing the GA on each set. After the results from each GA run in the tournament is completed, the top monomers from each run can then be run again to find the final set of top monomers. This process will allow for the GA to still provide increased speedup over brute force screening of billions of compounds while allowing the GA to operate effectively with increased data set size.

APPENDIX A

THIOPHENE SEARCH: ADDITIONAL FIGURES AND TABLES

Code	IUPAC Name	Code	IUPAC Name
1	thiophene	51	thieno[3,4-d][1,3]oxazole
2	3-fluorothiophene	52	thieno[3,4-d][1,3]thiazole
3	3-chlorothiophene	53	thieno[3,4-d][1,3]selenazole
4	3-bromothiophene	54	1H-thieno[3,4-d]imidazole
5	3-(trifluoromethyl)thiophene	55	thieno[3,4-d]oxadiazole
6	thiophene-3-carbonitrile	56	thieno[3,4-d]thiadiazole
7	3-nitrothiopnene	57	2λ4δ2-Thieno[3,4-c][1,2,5]thiadiazole
8	thiophen-3-amine	58	2λ4-Thieno[3,4-c][1,2,5]selenadiazole
9	3-methylthiophene	59	4-methylidenecyclopenta[c]thiophene
10	thiophen-3-ol	60	cyclopenta[c]thiophen-4-one
11	3-methoxythiophene	61	thieno[3,4-b]thiophene 1-oxide
12	thiophene-3-thiol	62	thieno[3,4-b]thiophene 1,1-dioxide
13	thiophene-3-selenol	63	cyclopenta[c]thiophene-4,6-dione
14	thiophene-3-carboxylic acid	64	thieno[3,4-c]pyrrole-4,6-dione
15	3-ethenylthiophene	65	5-(trifluoromethyl)-4H-thieno[3,4-c]pyrrole-4,6(5H)-dione
16	thiophene-3-carbaldehyde	66	thieno[3,4-c]furan-1,3-dione
17	1-thiophen-3-ylethanone	67	thieno[3,4-c]thiophene-4,6-dione
18	2,2,2-trifluoro-1-thiophen-3-ylethanone	68	selenophenol[3,4-c]thiophene-4,6-dione
19	thieno[3,4-c]thiophene-4,6-dione	69	thieno[3,4-d][1,3]dioxol-2-one
20	3-phenylthiophene	70	thieno[3,4-d][1,3]dioxole-2-thione
21	4-thiophen-3-ylaniline	71	1,3-dihydrothieno[3,4-d]imidazol-2-one
22	3-(4-methoxyphenyl)thiophene	72	1,3-dihydrothieno[3,4-d]imidazole-2-thione
23	3-(4-fluorophenyl)thiophene	73	thieno[3,4-d][1,3]dithiol-2-one
24	3-(4-nitrophenyl)thiophene	74	thieno[3,4-d][1,3]dithiole-2-thione
25	3,4-difluorothiophene	75	4,5,6,7-tetrahydro-2-benzothiophene
26	3,4-dichlorothiophene	76	2,3-dihydrothieno[3,4-b][1,4]dioxine
27	3,4-dibromothiophene	77	3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine
28	thiophene-3,4-dicarbonitrile	78	1,2,3,4-tetrahydrothieno[3,4-b]pyrazine
29	3,4-dinitrothiophene	79	2,3-dihydrothieno[3,4-b][1,4]dithiine
30	3,4-dimethylthiophene	80	2,3-dihydro[1,4]diseleno[2,3-c]thiophene
31	3,4-dimethoxythiophene	81	2,3-dihydrothieno[3,4-b][1,4]oxathiine
32	4-methoxythiophen-3-amine	82	4,7-dihydro-2-benzothiophene
33	4-methoxythiophene-3-carbonitrile	83	thieno[3,4-b][1,4]dioxine
34	4-(trifluoromethyl)thiophene-3-carbonitrile	84	1,4-dihydrothieno[3,4-b]pyrazine
35	4-nitrothiophen-3-amine	85	thieno[3,4-b][1,4]dithiine
36	3-methoxy-4-(trifluoromethyl)thiophene	86	[1,4]diselenino[2,3-c]thiophene
37	4-hydroxythiophene-3-carboxylic acid	87	thieno[3,4-b][1,4]dioxine-2,3-dione
38	3-ethenyl-4-methylthiophene	88	1,4-dihydrothieno[3,4-b]pyrazine-2,3-dione
39	4-mercaptothiophen-3-ol	89	thieno[3,4-b][1,4]dithiine-2,3-dione
40	5,6-dihydro-4H-cyclopenta[c]thiophene	90	[1,4]diselenino[2,3-c]thiophene-2,3-dione
41	thieno[3,4-d][1,3]dioxole	91	2-benzothiophene
42	2,3-dihydro-1H-thieno[3,4-d]imidazole	92	thieno[3,4-b]pyrazine
43	thieno[3,4-d][1,3]dithiole	93	thieno[3,4-d]pyridazine
44	[1,3]diselenolo[4,5-c]thiophene	94	thieno[3,4-d]pyrimidine
45	thieno[3,4-b]furan	95	5,6-dimethoxy-2-benzothiophene
46	1H-thieno[3,4-b]pyrrole	96	2-benzothiophene-5,6-dicarbonitrile
47	1-(trifluoromethyl)-1H-thieno[3,4-b]pyrrole	97	5,6-difluoro-2-benzothiophene
48	thieno[3,4-b]thiophene	98	4,5,6,7-tetrafluoro-2-benzothiophene
49	selenopheno[2,3-c]thiophene	99	4,7-difluorobenzo[c]thiophene
50	4H-cyclopenta[c]thiophene	100	5,6-dinitro-2-benzothiophene

Table A1: IUPAC names of the oligothiophenes studied.

Fig. 1	IUPAC Name	References
1	thiophene[1-3]	(1) Willgerodt, C.; Scholtz, T. *Freiberg i/B. J. Prakt. Chem.* **1910**, *81*, 382. (2) Armour, M.; Davies, A. G.; Upadhyay, J.; Wassermann, A. *J. Polym. Sci., Part A-1: Polym. Chem.* **1967**, *5*, 1527. (3) Biehl, E. R. *Prog. Heterocycl. Chem.* **2011**, *22*, 109.
2	3-fluorothiophene[4,5]	(4) Crestoni, M. E.; Fornarini, S. *Gazz. Chim. Ital.* **1989**, *119*, 203. (5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287.
3	3-chlorothiophene[6-10]	(6) Steinkopf, W.; Kohler, W. *Justus Liebigs Ann. Chem.* **1937**, *532*, 250. (7) Coonradt, H. L.; Hartough, H. D.; Johnson, G. C. *J. Am. Chem. Soc.* **1948**, *70*, 2564. (8) Matsushita Electric Industrial Co., Ltd., Japan . 1984, p 3 pp. (9) Lemaire, M.; Buchner, W.; Garreau, R.; Huynh, A. H.; Guy, A.; Roncali, J. *J. Electroanal. Chem. Interfacial Electrochem.* **1990**, *281*, 293. (10) Xu, J.; Shi, G.; Xu, Z.; Chen, F.; Hong, X. *J. Electroanal. Chem.* **2001**, *514*, 16.
4	3-bromothiophene[5,11-13]	(5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287. (11) Steinkopf, W.; Jacob, H.; Penz, H. *Justus Liebigs Ann. Chem.* **1934**, *512*, 136. (12) Gronowitz, S. *Ark. Kemi* **1954**, *7*, 267. (13) Bargon, J.; Mohmand, S.; Waltman, R. J. *IBM J. Res. Dev.* **1983**, *27*, 330.
5	3-(trifluoromethyl)thiophene[14,15]	(14) Leroy, J.; Rubinstein, M.; Wakselman, C. *J. Fluorine Chem.* **1985**, *27*, 291. (15) Ritter, S. K.; Noftle, R. E.; Ward, A. E. *Chem. Mater.* **1993**, *5*, 752.
6	thiophene-3-carbonitrile[5,16,17]	(5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287. (16) Denton, W. I.; Bishop, R. B.; Socony-Vacuum Oil Co., Inc. . 1951. (17) Campaigne, E. E.; Thomas, H. L. *J. Am. Chem. Soc.* **1955**, *77*, 5365.
7	3-nitrothiopnene[5,18-22]	(5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287. (18) Rinkes, I. J. *Recl. Trav. Chim. Pays-Bas Belg.* **1932**, *51*, 1134. (19) Steinkopf, W.; Hopner, T. *Justus Liebigs Ann. Chem.* **1933**, *501*, 174. (20) Kuraki, Y.; Funatsu, E.; Fuji Photo Film Co., Ltd., Japan . 1990, p 29 pp. (21) Aitken, K. M.; Aitken, R. A. *Sci. Synth.* **2007**, *31b*, 1183. (22) Zhang, J.; Chen, M.; Han, Z.; Cao, W.; Beijing University of Chemical Technology, Peop. Rep. China . 2009, p 13pp.
8	thiophen-3-amine[23-26]	(23) Brunett, E. W., 1967. (24) Abramenko, P. I. *Zh. Vses. Khim. Obshchest.* **1972**, *17*, 478. (25) Brunett, E. W.; Altwein, D. M.; McCarthy, W. C. *J. Heterocycl. Chem.* **1973**, *10*, 1067. (26) Paulmier, C. *Sulfur Rep.* **1996**, *19*, 215.
9	3-methylthiophene[5,13,27-31]	(5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287. (13) Bargon, J.; Mohmand, S.; Waltman, R. J. *IBM J. Res. Dev.* **1983**, *27*, 330. (27) Steinkopf, W. *Justus Liebigs Ann. Chem.* **1914**, *403*, 17. (28) Steinkopf, W. *Justus Liebigs Ann. Chem.* **1914**, *403*, 11. (29) Shepard, A. F.; Henne, A. L.; Midgley, T., Jr. *J. Am. Chem. Soc.* **1934**, *56*, 1355. (30) Linstead, R. P.; Noble, E. G.; Wright, J. M. *J. Chem. Soc.* **1937**,

Table A2: Names and references for oligothiophenes studied (1 of 10)

Fig. 1	IUPAC Name	References
		911. (31) Waltman, R. J.; Bargon, J.; Diaz, A. F. *J. Phys. Chem.* **1983**, *87*, 1459.
10	thiophen-3-ol[32]	(32) Hurd, C. D.; Kreuz, K. L. *J. Am. Chem. Soc.* **1950**, *72*, 5543.
11	3-methoxythiophene[5,33,34]	(5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287. (33) Gronowitz, S. *Ark. Kemi* **1958**, *12*, 239. (34) Tanaka, S.; Sato, M.; Kaeriyama, K. *Polym. Commun.* **1985**, *26*, 303.
12	thiophene-3-thiol[5,35,36]	(5) Schatz, J. *Sci. Synth.* **2002**, *9*, 287. (35) Friedmann, W. *J. Inst. Pet.* **1951**, *37*, 239. (36) Caesar, P. D.; Branton, P. D. *J. Ind. Eng. Chem. (Washington, D. C.)* **1952**, *44*, 122.
13	thiophene-3-selenol[37,38]	(37) Litvinov, V. P.; Gol'dfarb, Y. L.; Bogdanov, V. S.; Konyaeva, I. P.; Sukiasyan, A. N. *J. Prakt. Chem.* **1973**, *315*, 850. (38) Mahatsekake, C.; Ebel, M.; Catel, J. M.; Andrieu, C. G.; Mollier, Y.; Tourillon, G. *Sulfur Lett.* **1988**, *7*, 231.
14	thiophene-3-carboxylic acid[39-41]	(39) Rinkes, I. J. *Recl. Trav. Chim. Pays-Bas Belg.* **1936**, *55*, 991. (40) Campaigne, E.; LeSuer, W. M. *Org. Synth.* **1953**, *33*, No pp. given. (41) Englebienne, P.; Weiland, M. *Chem. Commun. (Cambridge)* **1996**, 1651.
15	3-ethenylthiophene[42-48]	(42) Troyanowsky, C. *Compt. rend.* **1951**, *232*, 2236. (43) Troyanowsky, C. *Bull. Soc. Chim. Fr.* **1955**, 424. (44) Adams, C. R. *J. Catal.* **1968**, *11*, 96. (45) Clarke, J. A.; Meth-Cohn, O. *Tetrahedron Lett.* **1975**, 4705. (46) Trumbo, D. L.; Suzuki, T.; Harwood, H. J. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1983**, *24*, 360. (47) Trumbo, D. L.; Lin, F. T.; Lin, F. M.; Harwood, H. J. *Polym. Bull. (Berlin)* **1992**, *28*, 87. (48) Mori, H.; Takano, K.; Endo, T. *Macromolecules (Washington, DC, U. S.)* **2009**, *42*, 7342.
16	thiophene-3-carbaldehyde[49]	(49) Campbell, T. W.; Kaeding, W. W. *J. Am. Chem. Soc.* **1951**, *73*, 4018.
17	1-thiophen-3-ylethanone[42,43,50-52]	(42) Troyanowsky, C. *Compt. rend.* **1951**, *232*, 2236. (43) Troyanowsky, C. *Bull. Soc. Chim. Fr.* **1955**, 424. (50) Jimenez, J. F.; Roncero, A. V. *An. R. Soc. Esp. Fis. Quim., Ser. A* **1949**, *45B*, 1591. (51) Blanchette, J. A.; Brown, E. V. *J. Am. Chem. Soc.* **1951**, *73*, 2779. (52) Aslanoglu, M.; Abbasoglu, S.; Karabulut, S.; Kutluay, A. *Acta Chim. Slov.* **2007**, *54*, 834.
18	2,2,2-trifluoro-1-thiophen-3-ylethanone[53-55]	(53) Cartoni, G.; Liberti, A.; Zoccolillo, L. *J. Chromatogr.* **1970**, *52*, 347. (54) Marino, G.; Linda, P.; Pignataro, S. *J. Chem. Soc. B* **1971**, 1585. (55) DiMenna, W. S. *Tetrahedron Lett.* **1980**, *21*, 2129.
19	thieno[3,4-c]thiophene-4,6-dione[56-60]	(56) Watts, E. A.; Beecham Group Ltd., UK . 1981, p 10 pp. Cont. (57) Lange, G.; Savard, M. E.; Viswanatha, T.; Dmitrienko, G. I. *Tetrahedron Lett.* **1985**, *26*, 1791. (58) Vanderesse, R.; Marchal, J.; Caubere, P. *Synth. Commun.* **1993**, *23*, 1361. (59) Marchal, J.; Bodiguel, J.; Fort, Y.; Caubere, P. *J. Org. Chem.* **1995**, *60*, 8336. (60) Park, K.-H.; Yoon, Y.-S.; Kang, H.; Lee, J.-C. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2004**, *45*, 309.
20	3-phenylthiophene[61-65]	(61) Chrzaszczewska, A. *Rocz. Chem.* **1925**, *5*, 33. (62) Griffin, C. E.; Martin, K. R. *Chem. Commun. (London)* **1965**, 154. (63) Kaeriyama, K.; Tanaka, S.; Sato, M.; Hamada, K. *Synth. Met.*

Table A3: Names and references for oligothiophenes studied (2 of 10)

Fig. 1	IUPAC Name	References
		1989, *28*, C611. (64) Lemaire, M.; Garreau, R.; Delabouglise, D.; Roncali, J.; Youssoufi, H. K.; Garnier, F. *New J. Chem.* **1990**, *14*, 359. (65) Guerrero, D. J.; Ren, X.; Ferraris, J. P. *Chem. Mater.* **1994**, *6*, 1437.
21	4-(3-thienyl)-benzamine[66,67]	(66) Djukic, B.; Seda, T.; Gorelsky, S. I.; Lough, A. J.; Lemaire, M. T. *Inorg. Chem.* **2011**, *50*, 7334. (67) Zamora, P. P.; Camarada, M. B.; Jessop, I. A.; Diaz, F. R.; del, V. M. A.; Cattin, L.; Louarn, G.; Bernede, J. C. *Int. J. Electrochem. Sci.* **2012**, *7*, 8276.
22	3-(4-methoxyphenyl)thiophene[61,65,68-71]	(61) Chrzaszczewska, A. *Rocz. Chem.* **1925**, *5*, 33. (65) Guerrero, D. J.; Ren, X.; Ferraris, J. P. *Chem. Mater.* **1994**, *6*, 1437. (68) Schmitt, J.; Lespagnol, A. *Bull. Soc. Chim. Fr.* **1950**, 459. (69) Kirsch, G.; Cagniant, D.; Cagniant, P. *J. Heterocycl. Chem.* **1982**, *19*, 443. (70) Montheard, J. P.; Delzant, J. F.; Gazard, M. *Synth. Commun.* **1984**, *14*, 289. (71) Robitaille, L.; Leclerc, M.; Callender, C. L. *Chem. Mater.* **1993**, *5*, 1755.
23	3-(4-fluorophenyl)thiophene[65,72-77]	(65) Guerrero, D. J.; Ren, X.; Ferraris, J. P. *Chem. Mater.* **1994**, *6*, 1437. (72) Sastry, C. V. R.; Marwah, A. K.; Marwah, P.; Rao, G. S.; Shridhar, D. R. *Synthesis* **1987**, 1024. (73) Ferraris, J. P.; Dhurjati, M. S. K.; Loveday, D. C.; Barashkov, N. N.; Hmyene, M.; Henderson, C. R. *Proc. - Electrochem. Soc.* **1997**, *96-24*, 14. (74) Sarker, H.; Gofer, Y.; Killian, J. G.; Poehler, T. O.; Searson, P. C. *Synth. Met.* **1997**, *88*, 179. (75) Ferraris, J. P.; Eissa, M. M.; Brotherston, I. D.; Loveday, D. C.; Moxey, A. A. *J. Electroanal. Chem.* **1998**, *459*, 57. (76) Gofer, Y.; Killian, J. G.; Sarker, H.; Poehler, T. O.; Searson, P. C. *J. Electroanal. Chem.* **1998**, *443*, 103. (77) Naudin, E.; Dabo, P.; Guay, D.; Breau, L.; Belanger, D. *Polym. Mater. Sci. Eng.* **1999**, *80*, 629.
24	3-(4-nitrophenyl)thiophene[78-85]	(78) Dabard, R.; Le, B. J. Y. *C. R. Acad. Sci., Ser. C* **1970**, *271*, 311. (79) Tundo, A.; Camaggi, C. M.; Leardini, R.; Tiecco, M. *J. Chem. Soc. B* **1970**, 1683. (80) Dabard, R.; Le, B. J. Y. *Bull. Soc. Chim. Fr.* **1972**, 4280. (81) Wu, X.; Rieke, R. D. *J. Org. Chem.* **1995**, *60*, 6658. (82) Alhalasah, W.; Holze, R. *J. Solid State Electrochem.* **2005**, *9*, 836. (83) Alhalasah, W.; Holze, R. *Microchim. Acta* **2006**, *156*, 133. (84) Alhalasah, W.; Holze, R. *J. Solid State Electrochem.* **2007**, *11*, 1605. (85) Alhalasah, W.; Holze, R. *ECS Trans.* **2007**, *2*, 45.
25	3,4-difluorothiophene[86-90]	(86) Akesson, B.; Gronowitz, S. *Ark. Kemi* **1967**, *28*, 155. (87) Christiansen, H.; Gronowitz, S.; Rodmar, B.; Rodmar, S.; Rosen, U.; Sharma, M. K. *Ark. Kemi* **1969**, *30*, 561. (88) Novak, I. *J. Org. Chem.* **2001**, *66*, 9041. (89) Salzner, U. *J. Phys. Chem. A* **2010**, *114*, 5397. (90) Jameh-Bozorghi, S.; Il, B. H. S. *J. Fluorine Chem.* **2011**, *132*, 190.
26	3,4-dichlorothiophene[6,7]	(6) Steinkopf, W.; Kohler, W. *Justus Liebigs Ann. Chem.* **1937**, *532*, 250. (7) Coonradt, H. L.; Hartough, H. D.; Johnson, G. C. *J. Am. Chem. Soc.* **1948**, *70*, 2564.

Table A4: Names and references for oligothiophenes studied (3 of 10)

Fig. 1	IUPAC Name	References
27	3,4-dibromothiophene[11,13,31,91]	(11) Steinkopf, W.; Jacob, H.; Penz, H. *Justus Liebigs Ann. Chem.* **1934**, *512*, 136. (13) Bargon, J.; Mohmand, S.; Waltman, R. J. *IBM J. Res. Dev.* **1983**, *27*, 330. (31) Waltman, R. J.; Bargon, J.; Diaz, A. F. *J. Phys. Chem.* **1983**, *87*, 1459. (91) Gronowitz, S. *Acta Chem. Scand.* **1959**, *13*, 1045.
28	thiophene-3,4-dicarbonitrile[92-94]	(92) Morel, J.; Paulmier, C.; Pastour, P. *C. R. Acad. Sci., Paris, Ser. C.* **1968**, *266*, 1300. (93) Otto, P.; Ladik, J. *Synth. Met.* **1990**, *36*, 327. (94) Greenwald, Y.; Xu, X.; Fourmigue, M.; Srdanov, G.; Koss, C.; Wudl, F.; Heeger, A. J. *J. Polym. Sci., Part A: Polym. Chem.* **1998**, *36*, 3115.
29	3,4-dinitrothiophene[95-97]	(95) Blatt, A. H.; Gross, N.; Tristram, E. W. *J. Org. Chem.* **1957**, *22*, 1588. (96) Dell'Erba, C.; Spinelli, D. *Boll. Sci. Fac. Chim. Ind. Bologna* **1968**, *26*, 97. (97) Dell'Erba, C.; Spinelli, D.; Leandri, G. *Gazz. Chim. Ital.* **1969**, *99*, 535.
30	3,4-dimethylthiophene[27,98-101]	(27) Steinkopf, W. *Justus Liebigs Ann. Chem.* **1914**, *403*, 17. (98) Rinkes, I. J. *Recl. Trav. Chim. Pays-Bas Belg.* **1933**, *52*, 1052. (99) Marvel, C. S.; Ryder, E. E., Jr. *J. Am. Chem. Soc.* **1955**, *77*, 66. (100) Jen, K. Y.; Miller, G. G.; Elsenbaumer, R. L. *J. Chem. Soc., Chem. Commun.* **1986**, 1346. (101) Yoshino, K.; Manda, Y.; Sawada, K.; Morita, S.; Takahashi, H.; Sugimoto, R.; Onoda, M. *J. Phys. Soc. Jpn.* **1989**, *58*, 1320.
31	3,4-dimethoxythiophene[102-104]	(102) Fager, E. W. *J. Am. Chem. Soc.* **1945**, *67*, 2217. (103) Overberger, C. G.; Lal, J. *J. Am. Chem. Soc.* **1951**, *73*, 2956. (104) Hagiwara, T.; Yamaura, M.; Sato, K.; Hirasaka, M.; Iwata, K. *Synth. Met.* **1989**, *32*, 367.
32	4-methoxythiophen-3-amine[105-107]	(105) Fernandez, F. M. I.; Hotten, T. M.; Tupper, D. E.; Lilly S. A., Spain . 1989, p 11 pp. (106) Hotten, T. M.; Tupper, D. E.; Fernandez, F. M. I.; Lilly S. A., Spain; Lilly Industries Ltd. . 1989, p 6 pp. (107) Ife, R. J.; Brown, T. H.; Leach, C. A.; SmithKline Beecham Intercredit BV, Neth. . 1989, p 41 pp.
33	4-methoxythiophene-3-carbonitrile[108-112]	(108) Zhou, C.; Li, Q.; Huang, Y.; Liu, R. *Wuli Huaxue Xuebao* **1994**, *10*, 825. (109) Miyaji, K.; Iwamoto, S.; Ota, H.; Shigeta, Y.; Hirokawa, Y.; Nakano, S.; Ishiwata, N.; Nissan Chemical Industries, Ltd., Japan . 2006, p 586 pp. (110) Naganuma, K.; Yokoi, H.; Asahi Kasei Pharma Corporation, Japan . 2006, p 214pp. (111) Hergue, N.; Mallet, C.; Frere, P.; Allain, M.; Roncali, J. *Macromolecules (Washington, DC, U. S.)* **2009**, *42*, 5593. (112) Hergue, N.; Mallet, C.; Savitha, G.; Allain, M.; Frere, P.; Roncali, J. *Org. Lett.* **2011**, *13*, 1762.
34	4-(trifluoromethyl)thiophene-3-carbonitrile	No published references at publication time
35	4-nitrothiophen-3-amine[113-115]	(113) Kasina, S.; Neorx Corp., USA . 1998, p 88 pp. (114)Chirakadze, G. G.; Geliashvili, Z. E.; Razmadze, T. O. *Russ. J. Org. Chem.* **2001**, *37*, 1013. (115) Suponitsky, K. Y.; Masunov, A. E.; Antipin, M. Y. *Mendeleev Commun.* **2009**, *19*, 311.
36	3-methoxy-4-(trifluoromethyl)thiophene	No published references at publication time

Table A5: Names and references for oligothiophenes studied (4 of 10)

Fig. 1	IUPAC Name	References
37	4-hydroxythiophene-3-carboxylic acid[116,117]	(116) Oskay, E. *Hacettepe Bull. Natur. Sci. Eng.* **1973**, *2*, 16. (117) Shvedov, V. I.; Kharizomenova, I. A.; Romanova, O. B.; Vasil'eva, V. K.; Grinev, A. N.; Ordzhonikidze, S., All-Union Scientific-Research Chemical-Pharmaceutical Institute . 1973.
38	3-ethenyl-4-methylthiophene	No published references at publication time
39	4-mercaptothiophen-3-ol	No published references at publication time
40	5,6-dihydro-4H-cyclopenta[c]thiophene[118-121]	(118) MacDowell, D. W. H.; Patrick, T. B.; Frame, B. K.; Ellison, D. L. *J. Org. Chem.* **1967**, *32*, 1226. (119) Cagniant, D.; Cagniant, P.; Merle, G. *Bull. Soc. Chim. Fr.* **1968**, 3828. (120) Garreau, R.; Roncali, J.; Garnier, F.; Lemaire, M. *J. Chim. Phys. Phys.-Chim. Biol.* **1989**, *86*, 93. (121) Ruehe, J.; Berlin, A.; Wegner, G. *Macromol. Chem. Phys.* **1995**, *196*, 225.
41	thieno[3,4-d][1,3]dioxole[122-125]	(122) Reist, E. J.; Calkins, D. F.; Goodman, L. *J. Org. Chem.* **1967**, *32*, 169. (123) Jonas, F.; Heywang, G.; Schmidtberg, W.; Bayer A.-G., Fed. Rep. Ger. . 1989, p 6 (124) Jonas, F.; Heywang, G.; Schmidtberg, W.; Heinze, J.; Dietrich, M.; Bayer A.-G., Fed. Rep. Ger. . 1989, p 15 pp. (125) Ahonen, H. J.; Kankare, J.; Lukkari, J.; Pasanen, P. *Synth. Met.* **1997**, *84*, 215.
42	2,3-dihydro-1H-thieno[3,4-d]imidazole	No published references at publication time
43	thieno[3,4-d][1,3]dithiole[126]	(126) Ahmed, M.; Buchshriber, J. M.; McKinnon, D. M. *Can. J. Chem.* **1970**, *48*, 1991.
44	[1,3]diselenolo[4,5-c]thiophene	No published references at publication time
45	thieno[3,4-b]furan[127,128]	(127) Moursounidis, J.; Wege, D. *Tetrahedron Lett.* **1986**, *27*, 3045. (128) Kumar, A.; Sotzing, G. A. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2005**, *46*, 969.
46	1H-thieno[3,4-b]pyrrole[129-131]	(129) Milun, M.; Trinajstic, N. *Croat. Chem. Acta* **1977**, *49*, 107. (130) Garcia, F.; Galvez, C. *Synthesis* **1985**, 143. (131) Buemi, G. *J. Chim. Phys. Phys.-Chim. Biol.* **1987**, *84*, 1147.
47	1-(trifluoromethyl)-1H-thieno[3,4-b]pyrrole	No published references at publication time
48	thieno[3,4-b]thiophene[132-141]	(132) Ghaisas, V. V.; Tilak, B. D. *Curr. Sci.* **1953**, *22*, 184. (133) Wynberg, H.; Zwanenburg, D. J. *Tetrahedron Lett.* **1967**, 761. (134) Litvinov, V. P.; Gol'dfarb, Y. L. *Adv. Heterocycl. Chem.* **1976**, *19*, 123. (135) Lee, K.; Sotzing, G. A. *Macromolecules* **2001**, *34*, 5746. (136) Lee, K.; Sotzing, G. A. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2002**, *43*, 610. (137) Seshadri, V.; Lee, K.; Sotzing, G. A. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2002**, *43*, 584. (138) Sotzing, G. A.; Lee, B.; Reyes, N.; Smith, M. B. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2002**, *43*, 904. (139) Comel, A.; Sommen, G.; Kirsch, G. *Mini-Rev. Org. Chem.* **2004**, *1*, 367. (140) Litvinov, V. P. *Russ. Chem. Rev.* **2005**, *74*, 217. (141) Liang, Y.; Yu, L. *Acc. Chem. Res.* **2010**, *43*, 1227.
49	selenopheno[2,3-c]thiophene[37,142-146]	(37) Litvinov, V. P.; Gol'dfarb, Y. L.; Bogdanov, V. S.; Konyaeva, I. P.; Sukiasyan, A. N. *J. Prakt. Chem.* **1973**, *315*, 850. (142) Litvinov, V. P.; Sukiasyan, A. N.; Gol'dfarb, Y. L.; Bogacheva, L. V. *Izv. Akad. Nauk SSSR, Ser. Khim.* **1971**, 1592. (143) Yasuike, S.; Kurita, J.; Tsuchiya, T. *Heterocycles* **1997**, *45*,

Table A6: Names and references for oligothiophenes studied (5 of 10)

Fig. 1	IUPAC Name	References
		1891. (144) Zahn, S.; Air Products and Chemicals, Inc., USA . 2009, p 21pp. (145) Zahn, S.; Air Products and Chemicals, Inc., USA . 2009, p 67pp. (146) Patra, A.; Wijsboom, Y. H.; Leitus, G.; Bendikov, M. *Chem. Mater.* **2011**, *23*, 896.
50	4H-cyclopenta[c]thiophene[147-149]	(147) Skramstad, J. *Acta Chem. Scand.* **1969**, *23*, 703. (148) Skramstad, J. *Chem. Scr.* **1973**, *4*, 81. (149) Skramstad, J. *Chem. Scr.* **1973**, *4*, 77.
51	thieno[3,4-d][1,3]oxazole[150]	(150) Leach, A. G.; Kidley, N. J. *J. Chem. Inf. Model.* **2011**, *51*, 1048.
52	thieno[3,4-d][1,3]thiazole[151]	(151) Uy, R.; Yang, L.; You, W. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2011**, *52*, 940.
53	thieno[3,4-d][1,3]selenazole	No published references at publication time
54	1H-thieno[3,4-d]imidazole[152-155]	(152) Cheney, L. C.; Parke, Davis & Co. . 1950. (153) Litvak, S.; Boeckx, R. L. O.; Dakshinamurti, K. *Anal. Biochem.* **1969**, *30*, 470. (154) McCormick, D. B. *Proc. Soc. Exp. Biol. Med.* **1969**, *132*, 502. (155) Beil, W.; Staar, U.; Sewing, K. F. *Eur. J. Pharmacol.* **1990**, *187*, 455.
55	thieno[3,4-d]oxadiazole	No published references at publication time
56	thieno[3,4-d]-1,2,3-thiadiazole	No published references at publication time
57	2λ4δ2-Thieno[3,4-c][1,2,5]thiadiazole[156-160]	(156) Behforouz, M.; Benrashid, R. *Tetrahedron Lett.* **1979**, 4493. (157) Tanaka, S.; Tomura, M.; Yamashita, Y. *Heterocycles* **1994**, *37*, 693. (158) Bakhshi, A. K.; Ago, H.; Yoshizawa, K.; Tanaka, K.; Yamabe, T. *J. Chem. Phys.* **1996**, *104*, 5528. (159) Brocks, G. *J. Phys. Chem.* **1996**, *100*, 17327. (160) Shen, W.; Li, M.; He, R.; Zhang, J.; Lei, W. *Polymer* **2007**, *48*, 3912.
58	2λ4-Thieno[3,4-c][1,2,5]selenadiazole	No published references at publication time
59	4-methylidenecyclopenta[c]thiophene	No published references at publication time
60	cyclopenta[c]thiophen-4-one[161,162]	(161) Albrecht, R.; Schroeder, E. *Arch. Pharm. (Weinheim, Ger.)* **1975**, *308*, 588. (162) Dallemagne, P.; Khanh, L. P.; Alsaidi, A.; Renault, O.; Varlet, I.; Collot, V.; Bureau, R.; Rault, S. *Bioorg. Med. Chem.* **2002**, *10*, 2185.
61	thieno[3,4-b]thiophene 1-oxide	No published references at publication time
62	thieno[3,4-b]thiophene 1,1-dioxide	No published references at publication time
63	cyclopenta[c]thiophene-4,6-dione[163,164]	(163) Khanh, L. P.; Dallemagne, P.; Rault, S. *Synlett* **1999**, 1450. (164) Parrain, J. L.; Thibonnet, J. *Sci. Synth.* **2005**, *26*, 745.
64	thieno[3,4-c]pyrrole-4,6-dione[165-167]	(165) Sice, J. *J. Org. Chem.* **1954**, *19*, 70. (166) Laird, D. W.; Sheina, E. E.; Brown, C. T.; Plextronics, Inc., USA . 2008, p 58pp. (167) Moon, J. M.; Choi, H.; Lee, J. M.; LG Chem, Ltd., S. Korea . 2008, p 26pp.
65	5-(trifluoromethyl)-4H-thieno[3,4-c]pyrrole-4,6(5H)-dione	No published references at publication time
66	thieno[3,4-c]furan-1,3-dione[165,168,169]	(165) Sice, J. *J. Org. Chem.* **1954**, *19*, 70. (168) Hopff, H.; von, d. C. J. *Chimia* **1959**, *13*, 107. (169) Takaya, T.; Hijikata, S.; Imoto, E. *Bull. Chem. Soc. Jap.* **1968**, *41*, 2532.
67	thieno[3,4-c]thiophene-4,6-dione	No published references at publication time
68	selenophenol[3,4-c]thiophene-4,6-dione	No published references at publication time
69	thieno[3,4-d][1,3]dioxol-2-one[170]	(170)Zahn, S.; Ford, M. E.; Air Products and Chemicals, Inc., USA .

Table A7: Names and references for oligothiophenes studied (6 of 10)

Fig. 1	IUPAC Name	References
		2007, p 13 pp.
70	thieno[3,4-d][1,3]dioxole-2-thione[170]	(170) Zahn, S.; Ford, M. E.; Air Products and Chemicals, Inc., USA . 2007, p 13 pp.
71	1,3-dihydrothieno[3,4-d]imidazol-2-one[171-174]	(171) Cheney, L. C.; Piening, J. R. *J. Am. Chem. Soc.* **1945**, *67*, 2252. (172) Cheney, L. C.; Piening, J. R. *J. Am. Chem. Soc.* **1945**, *67*, 2213. (173) Mozingo, R.; Harris, S. A.; Wolf, D. E.; Hoffhine, C. E., Jr.; Easton, N. R.; Folkers, K. *J. Am. Chem. Soc.* **1945**, *67*, 2092. (174) Zahn, S.; Ford, M. E.; Air Products and Chemicals, Inc., USA . 2007, p 29
72	1,3-dihydrothieno[3,4-d]imidazole-2-thione[175-177]	(175) Outurquin, F.; Paulmier, C. *Bull. Soc. Chim. Fr.* **1983**, 159. (176) Weidmann, K.; Herling, A. W.; Lang, H. J.; Scheunemann, K. H.; Rippel, R.; Nimmesgern, H.; Scholl, T.; Bickel, M.; Metzger, H. *J. Med. Chem.* **1992**, *35*, 438. (177) Binder, D.; Pyerin, M.; Schnait, H. *J. Heterocycl. Chem.* **1998**, *35*, 923.
73	thieno[3,4-d][1,3]dithiol-2-one[178-180]	(178) Chiang, L. Y.; Shu, P.; Holt, D.; Cowan, D. *J. Org. Chem.* **1983**, *48*, 4713. (179) Yamada, J.-i.; Amano, Y.; Takasaki, S.; Nakanishi, R.; Matsumoto, K.; Satoki, S.; Anzai, H. *J. Am. Chem. Soc.* **1995**, *117*, 1149. (180) Yamada, J.-i.; Satoki, S.; Mishima, S.; Akashi, N.; Takahashi, K.; Masuda, N.; Nishimoto, Y.; Takasaki, S.; Anzai, H. *J. Org. Chem.* **1996**, *61*, 3987.
74	thieno[3,4-d][1,3]dithiole-2-thione[178,181,182]	(178) Chiang, L. Y.; Shu, P.; Holt, D.; Cowan, D. *J. Org. Chem.* **1983**, *48*, 4713. (181) Gronowitz, S.; Moses, P. *Acta Chem. Scand.* **1962**, *16*, 105. (182) Shu, P.; Chiang, L.; Emge, T.; Holt, D.; Kistenmacher, T.; Lee, M.; Stokes, J.; Poehler, T.; Bloch, A.; Cowan, D. *J. Chem. Soc., Chem. Commun.* **1981**, 920.
75	4,5,6,7-tetrahydro-2-benzothiophene[121,183-195]	(121) Ruehe, J.; Berlin, A.; Wegner, G. *Macromol. Chem. Phys.* **1995**, *196*, 225. (183) Mayer, R.; Kleinert, H.; Richter, S.; Gewald, K. *J. Prakt. Chem. (Leipzig)* **1963**, *20*, 224. (184) Tilak, B. D.; Desai, H. S.; Gupte, S. S. *Tetrahedron Lett.* **1966**, 1953. (185) Jahn, R.; Schmidt, U. *Chem. Ber.* **1975**, *108*, 630. (186) Praefcke, K.; Weichsel, C. *Liebigs Ann. Chem.* **1980**, 1604. (187) Rasmussen, C. A. H.; De, G. A. *Synthesis* **1983**, 575. (188) Wegner, G.; Ruehe, J. *Faraday Discuss. Chem. Soc.* **1989**, *88*, 333. (189) Enkelmann, V.; Ruehe, J.; Wegner, G. *Synth. Met.* **1990**, *37*, 79. (190) Mohamadi, F.; Spees, M. M.; Grindey, G. B. *J. Med. Chem.* **1992**, *35*, 3012. (191) Ehrendorfer, C.; Karpfen, A.; Baeuerle, P.; Neugebauer, H.; Neckel, A. *J. Mol. Struct.* **1993**, *298*, 65. (192) Zotti, G.; Zecchin, S.; Schiavon, G.; Vercelli, B.; Berlin, A.; Dalcanale, E.; Groenendaal, L. B. *Chem. Mater.* **2003**, *15*, 4642. (193) Zotti, G.; Zecchin, S.; Schiavon, G.; Vercelli, B.; Berlin, A. *J. Electroanal. Chem.* **2005**, *575*, 169. (194) Fadhel, O.; Benko, Z.; Gras, M.; Deborde, V.; Joly, D.; Lescop, C.; Nyuliszi, L.; Hissler, M.; Reau, R. *Chem.--Eur. J.* **2010**, *16*, 11340. (195) You, W.; Yan, X.; Liao, Q.; Xi, C. *Org. Lett.* **2010**, *12*, 3930.
76	2,3-dihydrothieno[3,4-b][1,4]dioxine[123,196-201]	(123) Jonas, F.; Heywang, G.; Schmidtberg, W.; Bayer A.-G., Fed. Rep. Ger. . 1989, p 6 (196) Dietrich, M.; Heinze, J.; Heywang, G.; Jonas, F. *J. Electroanal. Chem.* **1994**, *369*, 87.

Table A8: Names and references for oligothiophenes studied (7 of 10)

Fig. 1	IUPAC Name	References
		(197) Pei, Q.; Zuccarello, G.; Ahlskog, M.; Inganaes, O. *Polymer* **1994**, *35*, 1347. (198) Coffey, M.; McKellar, B. R.; Reinhardt, B. A.; Nijakowski, T.; Feld, W. A. *Synth. Commun.* **1996**, *26*, 2205. (199) Sotzing, G. A.; Reddinger, J. L.; Reynolds, J. R.; Steel, P. J. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1996**, *37*, 795. (200) Roncali, J.; Blanchard, P.; Frere, P. *J. Mater. Chem.* **2005**, *15*, 1589. (201) Akagi, K.; Yongsoo, J. *Kagaku Kogyo* **2008**, *59*, 34.
77	3,4-dihydro-2H-thieno[3,4-b][1,4]dioxepine[196,198,202,203]	(196) Dietrich, M.; Heinze, J.; Heywang, G.; Jonas, F. *J. Electroanal. Chem.* **1994**, *369*, 87. (198) Coffey, M.; McKellar, B. R.; Reinhardt, B. A.; Nijakowski, T.; Feld, W. A. *Synth. Commun.* **1996**, *26*, 2205. (202) Heywang, G.; Jonas, F. *Adv. Mater. (Weinheim, Ger.)* **1992**, *4*, 116. (203) Kumar, A.; Welsh, D. M.; Morvant, M. C.; Piroux, F.; Abboud, K. A.; Reynolds, J. R. *Chem. Mater.* **1998**, *10*, 896.
78	1,2,3,4-tetrahydrothieno[3,4-b]pyrazine[204]	(204) Kondo, Y.; Nakano, K.; Otake, T.; Fuchigami, K.; Inaki, S.; Kuraray Co., Ltd., Japan; Tokyo Institute of Technology . 2010, p 33
79	2,3-dihydrothieno[3,4-b][1,4]dithiine[205-208]	(205) Wang, C.; Schindler, J. L.; Kannewurf, C. R.; Kanatzidis, M. G. *Chem. Mater.* **1995**, *7*, 58. (206) Goldoni, F.; Langeveld-Voss, B. M. W.; Meijer, E. W. *Synth. Commun.* **1998**, *28*, 2237. (207) Son, Y.; Kang, K.-S.; Shim, C.-Y.; Choi, J. S.; Lee, D.-Y.; Hong, S. Y. *Polymer (Korea)* **2002**, *26*, 589. (208) Hong, S. Y. *Synth. Met.* **2003**, *135-136*, 439.
80	2,3-dihydro[1,4]diseleno[2,3-c]thiophene[209,210]	(209) Pang, H.; Skabara, P. J.; Crouch, D. J.; Duffy, W.; Heeney, M.; McCulloch, I.; Coles, S. J.; Horton, P. N.; Hursthouse, M. B. *Macromolecules (Washington, DC, U. S.)* **2007**, *40*, 6585. (210) Pang, H.; Skabara, P. J.; Gordeyev, S.; McDouall, J. J. W.; Coles, S. J.; Hursthouse, M. B. *Chem. Mater.* **2007**, *19*, 301.
81	2,3-dihydrothieno[3,4-b][1,4]oxathiine[211,212]	(211) Blanchard, P.; Cappon, A.; Levillain, E.; Nicolas, Y.; Frere, P.; Roncali, J. *Org. Lett.* **2002**, *4*, 607. (212) Wijsboom, Y. H.; Sheynin, Y.; Patra, A.; Zamoshchik, N.; Vardimon, R.; Leitus, G.; Bendikov, M. *J. Mater. Chem.* **2011**, *21*, 1368.
82	4,7-dihydro-2-benzothiophene[213]	(213) Pramanik, A.; Kundu, S. K. *Indian J. Chem., Sect. B: Org. Chem. Incl. Med. Chem.* **2002**, *41B*, 1707.
83	thieno[3,4-b][1,4]dioxine[159,214-216]	(159) Brocks, G. *J. Phys. Chem.* **1996**, *100*, 17327. (214) Leriche, P.; Blanchard, P.; Frere, P.; Levillain, E.; Mabon, G.; Roncali, J. *Chem. Commun. (Cambridge, U. K.)* **2006**, 275. (215) Su, K.; Yang, N.-L. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2006**, *47*, 445. (216) Bhattacharyya, D.; Chelawat, H.; Gleason, K. K. *PMSE Prepr.* **2010**, No pp. given.
84	1,4-dihydrothieno[3,4-b]pyrazine[217-221]	(217) Mookherjee, B. D.; Beets, M. G.; Pittet, A. O.; Mason, M. E.; Theimer, E. T.; Tibbetts, M. S.; Evers, W. J.; Katz, I.; Wilson, R. A.; et, a.; International Flavors and Fragrances Inc. . 1971, p 106 pp. (218) Evers, W. J.; Katz, I.; Wilson, R. A.; Theimer, E. T.; International Flavors and Fragrances Inc. . 1972, p 4 pp. (219) Evers, W. J.; Katz, I.; Wilson, R. A.; Theimer, E. T.; International Flavors and Fragrances Inc. . 1973, p 6 pp. Division of U.S. 3.

Table A9: Names and references for oligothiophenes studied (8 of 10)

Fig. 1	IUPAC Name	References
		(220) Evers, W. J.; Katz, I.; Theimer, E. T.; International Flavors and Fragrances Inc., USA . 1976, p 26 pp. Division of Ger. Offen. 2. (221) Folkes, D. J.; Gramshaw, J. W. *Prog. Food Nutr. Sci.* **1981**, *5*, 369.
85	thieno[3,4-b][1,4]dithiine	No published references at publication time
86	[1,4]diselenino[2,3-c]thiophene	No published references at publication time
87	thieno[3,4-b][1,4]dioxine-2,3-dione	No published references at publication time
88	1,4-dihydrothieno[3,4-b]pyrazine-2,3-dione[175,222-228]	(175) Outurquin, F.; Paulmier, C. *Bull. Soc. Chim. Fr.* **1983**, 159. (222) Motoyama, R. *Nippon Kagaku Zasshi* **1957**, *78*. (223) Motoyama, R.; Imoto, E. *Nippon Kagaku Zasshi* **1957**, *78*, 793. (224) Mohwald, H.; Belov, V.; Schrof, W.; BASF A.-G., Germany . 1997, p 74. (225) Nilsson, B.; Tejbrant, J.; Pelcman, B.; Ringberg, E.; Thor, M.; Nilsson, J.; Jonsson, M.; Pharmacia & Upjohn AB, Swed. . 2000, p 151 pp. (226) Nilsson, B.; Tejbrant, J.; Pelcman, B.; Ringberg, E.; Thor, M.; Nilsson, J.; Jonsson, M.; Biovitrum AB, Swed. . 2002, p 45 pp. (227) Wen, L.; Nietfeld, J. P.; Amb, C. M.; Rasmussen, S. C. *J. Org. Chem.* **2008**, *73*, 8529. (228) Wen, L.; Nietfeld, J. P.; Amb, C. M.; Rasmussen, S. C. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2008**, *49*, 633.
89	thieno[3,4-b][1,4]dithiine-2,3-dione	No published references at publication time
90	[1,4]diselenino[2,3-c]thiophene-2,3-dione	No published references at publication time
91	2-benzothiophene[229-235]	(229) Obolentsev, R. D.; Bukharov, V. G.; Baisheva, A. U. *Khim. Seraorgan. Soedin., Soderzhashch. v Neff. i Nefteprod., Akad. Nauk SSSR Bashkirsk. Filial* **1961**, *4*, 20. (230) Hurd, C. D.; Levetan, R. V.; Macon, A. R. *J. Am. Chem. Soc.* **1962**, *84*, 4515. (231) Mayer, R.; Kleinert, H.; Richter, S.; Gewald, K. *Angew. Chem.* **1962**, *74*, 118. (232) Cava, M. P.; Pollack, N. M. *J. Am. Chem. Soc.* **1966**, *88*, 4112. (233) Wudl, F.; Kobayashi, M.; Heeger, A. J. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **1984**, *25*, 257. (234) Wudl, F.; Kobayashi, M.; Heeger, A. J. *J. Org. Chem.* **1984**, *49*, 3382. (235) Nakaya, T. *Konbatekku* **2007**, *35*, 38.
92	thieno[3,4-b]pyrazine[93,159,175,236-242]	(93) Otto, P.; Ladik, J. *Synth. Met.* **1990**, *36*, 327. (159) Brocks, G. *J. Phys. Chem.* **1996**, *100*, 17327. (175) Outurquin, F.; Paulmier, C. *Bull. Soc. Chim. Fr.* **1983**, 159. (236) Schneller, S. W.; Clough, F. W.; Skancke, P. N. *J. Heterocycl. Chem.* **1976**, *13*, 581. (237) Nayak, K.; Marynick, D. S. *Macromolecules* **1990**, *23*, 2237. (238) Armand, J.; Bellec, C.; Boulares, L.; Chaquin, P.; Masure, D.; Pinson, J. *J. Org. Chem.* **1991**, *56*, 4840. (239) Quattrocchi, C.; Lazzaroni, R.; Kiebooms, R.; Vanderzande, D.; Gelan, J.; Bredas, J. L. *Synth. Met.* **1995**, *69*, 691. (240) Kenning, D. D.; Funfar, M. R.; Rasmussen, S. C. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **2001**, *42*, 506. (241) Zhu, Z.; Waller, D.; Brabec, C. J.; Wiley-VCH Verlag GmbH & Co. KGaA: 2008, p 129. (242) Mondal, R.; Ko, S.; Bao, Z. *J. Mater. Chem.* **2010**, *20*, 10568.
93	thieno[3,4-d]pyridazine[159,239,243-249]	(159) Brocks, G. *J. Phys. Chem.* **1996**, *100*, 17327. (239) Quattrocchi, C.; Lazzaroni, R.; Kiebooms, R.; Vanderzande, D.; Gelan, J.; Bredas, J. L. *Synth. Met.* **1995**, *69*, 691.

Table A10: Names and references for oligothiophenes studied (9 of 10)

Fig. 1	IUPAC Name	References
		(243) Robba, M.; Moreau, R. C.; Roques, B. *Compt. rend.* **1964**, *259*, 3783. (244) Robba, M.; Roques, B.; Robba, Max . 1966, p 16 pp. (245) Robba, M.; Roques, B.; Bonhomme, M. *Bull. Soc. Chim. Fr.* **1967**, 2495. (246) Helland, A.; Skancke, P. N. *Acta Chem. Scand.* **1972**, *26*, 2601. (247) Sha, C. K.; Tsou, C. P. *J. Chin. Chem. Soc. (Taipei)* **1991**, *38*, 183. (248) Kimura, S. M. H.; Konica Co., Japan . 2002, p 34 pp. (249) El-Dean, A. M. K.; Gaber, A. E.-A. M.; El-Gaby, M. S. A.; Eyada, H. A.; Al-Kamali, A. S. N. *Phosphorus, Sulfur Silicon Relat. Elem.* **2004**, *179*, 321.
94	thieno[3,4-d]pyrimidine[93,159,239,250-253]	(93) Otto, P.; Ladik, J. *Synth. Met.* **1990**, *36*, 327. (159) Brocks, G. *J. Phys. Chem.* **1996**, *100*, 17327. (239) Quattrocchi, C.; Lazzaroni, R.; Kiebooms, R.; Vanderzande, D.; Gelan, J.; Bredas, J. L. *Synth. Met.* **1995**, *69*, 691. (250) Robba, M.; Lecomte, J. M.; Cugnon, d. S. M. *C. R. Acad. Sci., Paris, Ser. C* **1968**, *267*, 697. (251) Ibrahim, Y. A.; Elwahy, A. H. M.; Kadry, A. M. *Adv. Heterocycl. Chem.* **1996**, *65*, 235. (252) Varvounis, G.; Giannopoulos, T. *Adv. Heterocycl. Chem.* **1996**, *66*, 193. (253) Juric, A.; Nikolic, S.; Trinajstic, N. *Croat. Chem. Acta* **1997**, *70*, 841.
95	5,6-dimethoxy-2-benzothiophene[254,255]	(254) Wudl, F.; Heeger, A.; Yoshiaki, Y.; Kobayashi, M.; University of California, Berkeley, USA . 1988, p 16 pp. (255) Defieuw, G.; Samijn, R.; Vandezande, D.; Gelan, J. *Res. Discl.* **1992**, *339*, 568.
96	2-benzothiophene-5,6-dicarbonitrile[256]	(256) Hsu, D.-T.; Lin, C.-H. *J. Org. Chem.* **2009**, *74*, 9180.
97	5,6-difluoro-2-benzothiophene	No published references at publication time
98	4,5,6,7-tetrafluoro-2-benzothiophene[257-263]	(257) Brooke, G. M.; Mawson, S. D. *J. Chem. Soc., Perkin Trans. 1* **1990**, 1919. (258) Swann, M. J.; Brooke, G.; Bloor, D.; Maher, J. *Synth. Met.* **1993**, *55*, 281. (259) Brooke, G. M.; Drury, C. J.; Bloor, D.; Swann, M. J. *J. Mater. Chem.* **1995**, *5*, 1317. (260) Kiebooms, R.; Adriaensens, P.; Vanderzande, D.; Gelan, J.; Swann, M. J.; Bloor, D.; Drury, C. J.; Brooke, G. M. *Macromolecules* **1996**, *29*, 5981. (261) Kiebooms, R.; Adriaensens, P.; Vanderzande, D.; Gelan, J.; Swann, M. J.; Bloor, D.; Drury, C. J.; Brooke, G. M. *Synth. Met.* **1997**, *84*, 189. (262) Cornil, J.; Vanderdonckt, S.; Lazzaroni, R.; Dos, S. D. A.; Thys, G.; Geise, H. J.; Yu, L. M.; Szablewski, M.; Bloor, D.; Loegdlund, M.; Salaneck, W. R.; Gruhn, N. E.; Lichtenberger, D. L.; Lee, P. A.; Armstrong, N. R.; Bredas, J. L. *Chem. Mater.* **1999**, *11*, 2436. (263) Uoyama, H.; Nakamura, K.; Tukiji, M.; Furukawa, M.; Uno, H. *Heterocycles* **2007**, *73*, 673.
99	4,7-difluorobenzo[c]thiophene	No published references at publication time
100	5,6-dinitro-2-benzothiophene	No published references at publication time

Table A11: Names and references for oligothiophenes studied (10 of 10)

Code	SMILES
01	c(s1)ccc1
02	c(s1)cc(F)c1
03	c(s1)cc(Cl)c1
04	c(s1)cc(Br)c1
05	c(s1)cc(C(F)(F)(F))c1
06	c(s1)cc(cn)c1
07	c(s1)cc(N(O)(O))c1
08	c(s1)cc(N)c1
09	c(s1)cc(C)c1
10	c(s1)cc(O)c1
11	c(s1)cc(OC)c1
12	c(s1)cc(S)c1
13	c(s1)cc([Se])c1
14	c(s1)cc(COO)c1
15	c(s1)cc(C=O)c1
16	c(s1)cc(C(C)=O)c1
17	c(s1)cc(C(C(F)(F)F)=O)c1
18	c(s1)cc(c2ccccc2)c1
19	c(s1)c(F)c(F)c1
20	c(s1)c(Cl)c(Cl)c1
21	c(s1)c(Br)c(Br)c1
22	c(s1)c(cn)c(cn)c1
23	c(s1)c(N(O)O)c(N(O)O)c1
24	c(s1)c(C)c(C)c1
25	c(s1)c(OC)c(OC)c1
26	c(s1)c(N)c(N)c1
27	c(s1)c(OC)c(cn)c1
28	c(s1)c(N)c(N(O)O)c1
29	c(s1)c(cn)c(C(F)(F)F)c1
30	c(s1)c(O)c(C(O)=O)c1
31	c(s1)c2CCCc2c1
32	c(s1)c2OCOc2c1
33	c(s1)c2NCNc2c1
34	c(s1)c2SCSc2c1
35	c(s1)c2[Se]C[Se]c2c1
36	c(s1)c2occc2c1
37	c(s1)c2Nccc2c1
38	c(s1)c2N(C(F)(F)F)ccc2c1
39	c(s1)c2sccc2c1
40	c(s1)c2[Se]ccc2c1
41	c(s1)c2Cccc2c1
42	c(s1)c2ocnc2c1
43	c(s1)c2scnc2c1
44	c(s1)c2Ncnc2c1
45	c(s1)c2onnc2c1
46	c(s1)c2snnc2c1
47	c(s1)c2N=S=Nc2c1
48	c(s1)c2N=[Se]=Nc2c1
49	c(s1)c2C(=C)ccc2c1
50	c(s1)c2C(=O)ccc2c1
51	c(s1)c2C(S=O)ccc2c1
52	c(s1)c2C(S(=O)=O)ccc2c1
53	c(s1)c2C(=O)CC(=O)c2c1
54	c(s1)c2C(=O)NC(=O)c2c1
55	c(s1)c2C(=O)N(C(F)(F)F)C(=O)c2c1
56	c(s1)c2C(=O)OC(=O)c2c1
57	c(s1)c2C(=O)SC(=O)c2c1
58	c(s1)c2OC(=O)Oc2c1
59	c(s1)c2OC(=S)Oc2c1
60	c(s1)c2NC(=O)Nc2c1
61	c(s1)c2NC(=S)Nc2c1
62	c(s1)c2SC(=O)Sc2c1
63	c(s1)c2SC(=S)Sc2c1
64	c(s1)c2OCCOc2c1
65	c(s1)c2OCCCOc2c1
66	c(s1)c2NCCNc2c1
67	c(s1)c2SCCSc2c1
68	c(s1)c2[Se]CC[Se]c2c1
69	c(s1)c2ccccc2c1
70	c(s1)c2occoc2c1
71	c(s1)c2nccnc2c1
72	c(s1)c2sccsc2c1
73	c(s1)c2[Se]cc[Se]c2c1
74	c(s1)c2oc(o)c(o)oc2c1
75	c(s1)c2Nc(o)c(o)Nc2c1
76	c(s1)c2sc(o)c(o)sc2c1
77	c(s1)c2ccccc2c1
78	c(s1)c2nccnc2c1
79	c(s1)c2cnncc2c1
80	c(s1)c2ncncc2c1
81	c(s1)c2cc(OC)c(OC)cc2c1
82	c(s1)c2cc(cn)c(cn)cc2c1
83	c(s1)c2cc(F)c(F)cc2c1
84	c(s1)c2c(F)c(F)c(F)c(F)c2c1
85	c(s1)c2c(F)ccc(F)c2c1
86	c(s1)c2cc(N(O)O)c(N(O)O)cc2c1
87	c(s1)c2[Se]c(=O)c(=O)[Se]c2c1
88	c(s1)c2[Se]cnc2c1
89	c(s1)c2CCCCc2c1
90	c(s1)c2c(=O)[Se]c(=O)c2c1
91	c(s1)cc(cc)c1
92	c(s1)c(C)c(cc)c1
93	c(s1)cc(c(=O)OC)c1
94	c(s1)c(S)c(O)c1
95	c(s1)c(OC)c(C(F)(F)F)c1
96	c(s1)cc(c2ccc(N)cc2)c1
97	c(s1)cc(c2ccc(OC)cc2)c1
98	c(s1)cc(c2ccc(F)cc2)c1
99	c(s1)cc(c2ccc(N(=O)=O)cc2)c1
100	c(s1)c2OCCSc2c1

Table A12: SMILES of the oligothiophenes studied.

Code	HOMO			LUMO			HOMO-LUMO Gap		
	Intercept	Slope	R^2	Intercept	Slope	R^2	Intercept	Slope	R^2
1	-4.543	-1.722	0.974	-2.480	2.764	0.990	2.063	4.485	0.996
2	-4.704	-1.780	0.978	-2.917	3.066	0.991	1.787	4.847	1.000
3	-4.901	-1.804	1.000	-3.000	2.834	0.999	1.902	4.638	0.999
4	-4.845	-1.824	0.999	-2.940	2.728	0.999	1.905	4.552	1.000
5	-5.537	-1.559	0.989	-3.113	2.806	0.994	2.424	4.365	0.993
6	-5.800	-1.267	0.975	-3.833	2.894	0.998	1.966	4.161	0.995
7	-6.086	-1.221	0.888	-3.784	1.754	0.988	2.302	2.974	0.959
8	-3.836	-1.848	0.994	-1.858	2.364	0.998	1.978	4.212	0.996
9	-4.259	-1.952	0.982	-2.170	2.414	0.987	2.089	4.365	0.985
10	-3.722	-2.447	0.998	-2.228	2.519	0.999	1.494	4.965	0.999
11	-3.534	-2.547	0.998	-2.129	2.423	0.998	1.404	4.971	0.998
12	-4.624	-2.142	0.994	-2.677	2.328	0.979	1.947	4.470	0.998
13	-4.651	-2.079	0.999	-2.756	2.502	1.000	1.895	4.580	1.000
14	-5.039	-1.710	0.981	-2.712	1.608	0.987	2.326	3.318	0.984
15	-4.733	-1.475	0.992	-2.262	2.115	0.998	2.471	3.590	1.000
16	-6.012	-0.762	0.970	-3.305	2.372	0.998	2.708	3.134	0.996
17	-4.816	-2.069	0.991	-2.713	1.678	0.974	2.102	3.747	0.984
18	-5.802	-1.258	0.995	-3.418	1.556	0.983	2.384	2.814	0.990
19	-5.285	-1.686	0.922	-2.658	2.099	0.978	2.626	3.786	0.957
20	-4.536	-1.696	0.998	-2.074	1.802	0.999	2.462	3.498	0.999
21	-4.145	-1.473	0.959	-1.677	1.774	0.998	2.468	3.248	0.992
22	-4.293	-1.720	0.982	-2.029	2.226	0.998	2.264	3.946	0.995
23	-4.753	-1.528	0.996	-2.373	2.218	0.999	2.379	3.745	0.999
24	-5.691	-0.852	0.943	-2.952	0.876	0.940	2.740	1.727	0.942
25	-4.969	-1.881	1.000	-3.054	2.940	0.999	1.915	4.821	0.999
26	-6.567	0.900	0.691	-2.469	1.233	0.900	4.098	0.333	0.099
27	-5.127	-1.672	1.000	-3.059	2.477	0.999	2.068	4.148	1.000
28	-7.287	0.087	0.293	-4.517	2.306	0.985	2.770	2.219	0.968
29	-8.563	1.646	0.866	-4.124	1.979	0.981	4.438	0.333	0.167
30	-4.963	-1.423	0.996	-1.215	1.722	0.997	3.748	3.144	0.997
31	-4.272	-1.381	0.800	-1.657	2.002	0.949	2.614	3.383	0.922
32	-3.356	-3.201	0.969	-1.368	1.027	0.685	1.989	4.228	0.999
33	-5.018	-1.909	0.988	-3.248	2.228	0.992	1.770	4.137	0.999
34	-7.140	-0.189	0.634	-3.787	2.229	0.999	3.353	2.418	0.997
35	-5.097	-1.132	0.995	-3.484	1.765	0.989	1.613	2.897	0.991
36	-5.726	-0.943	0.637	-2.071	1.765	0.971	3.655	2.708	0.934
37	-5.528	-0.555	0.998	-2.340	1.382	1.000	3.187	1.937	0.999
38	-5.053	-1.006	0.970	-1.514	1.214	0.996	3.539	2.219	0.989
39	-3.863	-2.734	0.956	-2.410	2.086	0.989	1.454	4.820	0.993
40	-3.952	-2.257	1.000	-2.037	2.290	1.000	1.915	4.547	1.000
41	-4.100	-2.206	0.999	-2.070	2.352	0.999	2.030	4.558	1.000
42	-3.648	-1.612	0.995	-1.547	2.336	0.976	2.101	3.948	0.990
43	-4.387	-2.003	0.979	-2.463	2.314	0.998	1.924	4.316	0.992
44	-4.735	-0.878	0.645	-2.357	2.049	0.998	2.378	2.927	0.940
45	-3.717	-2.301	0.998	-2.668	2.487	0.998	1.049	4.789	0.998
46	-3.144	-2.234	0.996	-2.312	2.579	0.994	0.832	4.813	0.995
47	-4.629	-1.194	0.984	-2.047	1.615	0.993	2.582	2.808	0.990
48	-3.951	-1.959	0.994	-2.862	2.622	0.996	1.089	4.582	0.995
49	-4.087	-1.563	0.967	-2.788	2.500	0.997	1.298	4.063	0.991
50	-4.025	-1.937	0.999	-2.278	2.247	0.998	1.747	4.184	0.999
51	-4.178	-2.376	0.998	-2.963	2.418	0.998	1.215	4.794	0.998
52	-4.169	-2.207	0.998	-2.981	2.257	0.998	1.188	4.464	0.998
53	-4.189	-2.037	0.998	-2.925	2.256	0.997	1.264	4.293	0.999
54	-3.453	-2.489	0.997	-2.569	2.461	0.997	0.884	4.950	0.997
55	-4.872	-2.174	0.987	-4.181	2.690	0.992	0.691	4.865	0.991

Table A13: HOMO, LUMO and HOMO-LUMO Gap intercepts, slopes and R^2 values for all data. (1 of 2)

Code	HOMO			LUMO			HOMO-LUMO Gap		
	Intercept	Slope	R^2	Intercept	Slope	R^2	Intercept	Slope	R^2
56	-5.038	-1.663	0.998	-4.017	2.558	0.996	1.021	4.222	0.997
57	-3.805	-2.634	0.995	-4.211	2.109	0.987	-0.406	4.744	0.992
58	-3.604	-2.564	0.997	-4.165	1.949	0.976	-0.561	4.512	0.995
59	-4.575	-1.284	1.000	-1.995	0.914	0.996	2.580	2.198	0.999
60	-5.301	-1.434	1.000	-3.215	1.061	0.995	2.086	2.495	0.998
61	-5.349	-1.358	0.993	-3.193	2.109	0.991	2.156	3.466	0.992
62	-5.953	-1.255	0.978	-3.668	2.423	0.995	2.285	3.678	0.991
63	-5.738	-1.514	0.994	-3.573	1.289	0.999	2.165	2.803	0.997
64	-5.757	-1.592	0.995	-3.434	1.389	0.999	2.323	2.982	0.998
65	-6.293	-1.345	0.994	-4.086	1.756	1.000	2.208	3.101	1.000
66	-6.485	-1.194	0.987	-4.393	2.162	0.998	2.092	3.355	0.999
67	-6.742	-0.272	0.134	-3.770	0.966	0.977	2.971	1.238	0.690
68	-6.745	-0.649	0.853	-3.704	1.420	0.996	3.041	2.069	0.982
69	-5.485	-1.589	1.000	-3.660	3.217	0.999	1.825	4.805	1.000
70	-5.574	-1.643	1.000	-3.830	2.838	0.998	1.744	4.481	0.999
71	-4.844	-1.050	0.994	-2.717	2.912	0.998	2.127	3.962	1.000
72	-5.368	-0.535	0.932	-3.313	2.868	0.999	2.054	3.403	0.997
73	-5.364	-1.536	0.977	-3.544	3.308	0.995	1.820	4.844	0.991
74	-5.468	-1.156	0.986	-3.686	2.235	0.998	1.783	3.391	1.000
75	-4.369	-1.973	0.784	-1.685	2.066	0.908	2.684	4.040	0.850
76	-3.421	-2.707	0.999	-1.687	1.943	0.999	1.734	4.649	0.999
77	-3.461	-2.711	0.998	-1.658	1.929	1.000	1.803	4.640	0.999
78	-2.896	-3.483	0.861	-1.743	3.499	0.901	1.153	6.982	0.882
79	-5.281	-0.484	0.726	-1.208	0.831	0.835	4.073	1.315	0.796
80	-5.130	-0.038	0.002	-1.534	0.929	0.319	3.596	0.967	0.187
81	-3.660	-2.427	0.999	-1.923	1.983	0.999	1.737	4.410	0.999
82	-5.292	-0.693	0.797	-1.372	1.485	0.966	3.921	2.178	0.925
83	-4.318	-1.667	0.985	-2.092	2.101	1.000	2.226	3.768	0.997
84	-4.075	0.256	0.359	-1.327	1.747	0.989	2.749	1.491	0.925
85	-5.450	0.155	0.840	-1.829	1.387	0.877	3.620	1.232	0.847
86	-4.988	-1.102	0.333	-2.047	1.340	0.354	2.942	2.442	0.345
87	-5.581	-1.629	0.999	-3.637	1.850	0.993	1.944	3.479	0.999
88	-6.112	-0.328	0.386	-3.436	2.855	0.999	2.675	3.183	0.984
89	-7.021	0.458	0.595	-3.299	0.901	0.983	3.722	0.443	0.645
90	-6.838	1.008	0.408	-3.316	0.873	0.975	3.522	-0.135	0.016
91	-3.963	-1.787	0.994	-2.638	1.909	0.991	1.325	3.696	0.993
92	-3.903	-2.720	0.997	-3.519	1.704	0.995	0.384	4.424	0.996
93	-5.614	-0.980	1.000	-4.238	2.819	0.994	1.376	3.800	0.996
94	-4.386	-2.220	0.996	-4.052	2.429	0.995	0.333	4.649	0.995
95	-4.095	-0.970	0.996	-1.880	1.296	0.997	2.215	2.266	0.997
96	-6.093	-0.459	0.979	-4.596	2.559	0.995	1.498	3.018	0.998
97	-4.595	-1.266	0.995	-3.118	2.138	0.991	1.477	3.404	0.993
98	-5.301	-0.294	0.859	-3.230	1.736	1.000	2.071	2.030	0.998
99	-4.687	-0.866	0.938	-2.783	1.522	0.960	1.903	2.387	0.953
100	-6.386	-0.301	0.465	-4.707	2.469	0.974	1.679	2.769	0.936

Table A14: HOMO, LUMO and HOMO-LUMO Gap intercepts, slopes and R^2 values for all data. (2 of 2)

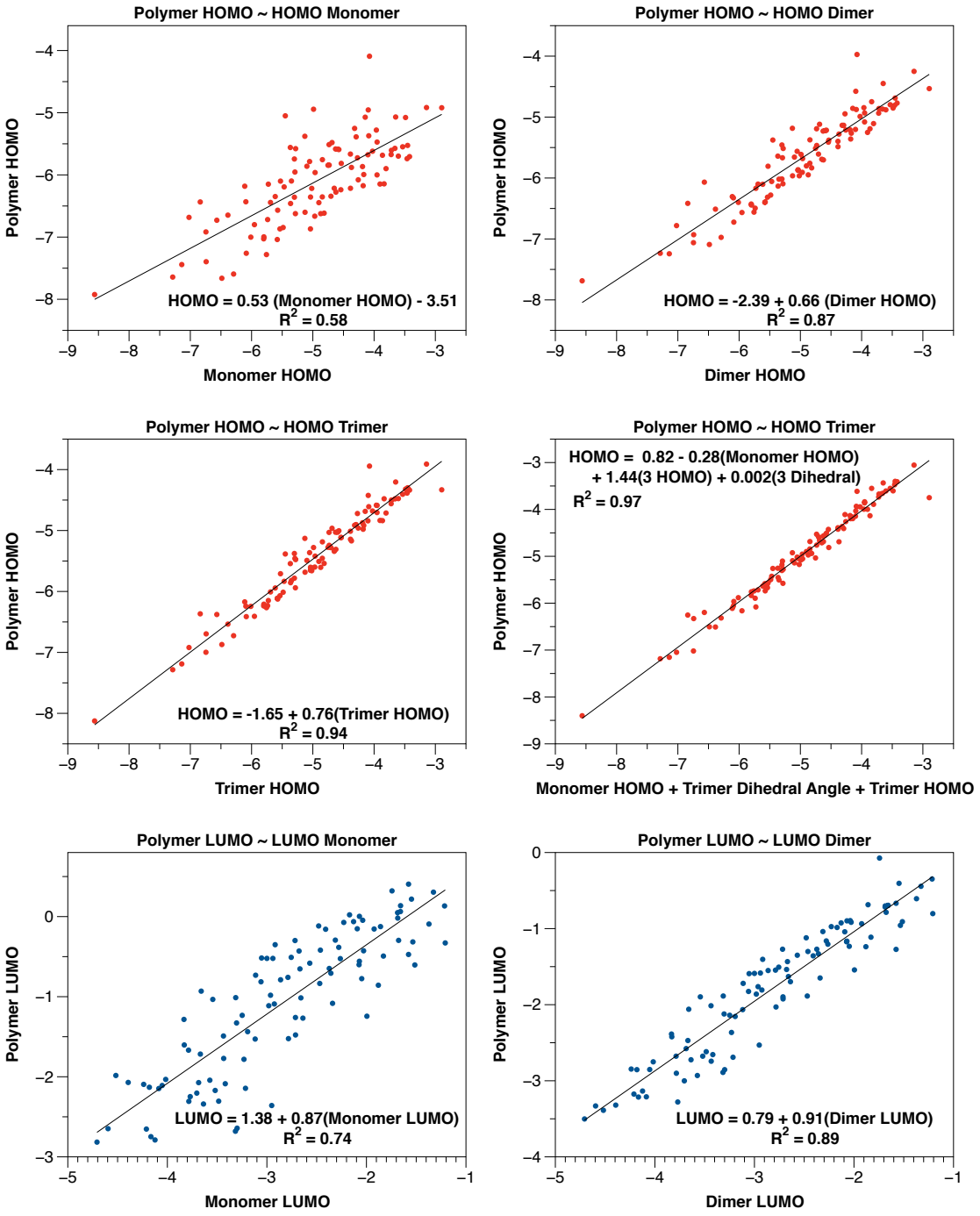


Figure A1: Additional plots for the linear regression models for predicting polymer HOMO, LUMO, and HOMO-LUMO gap from monomers, dimers, trimers and multivariate fits (1 of 2)

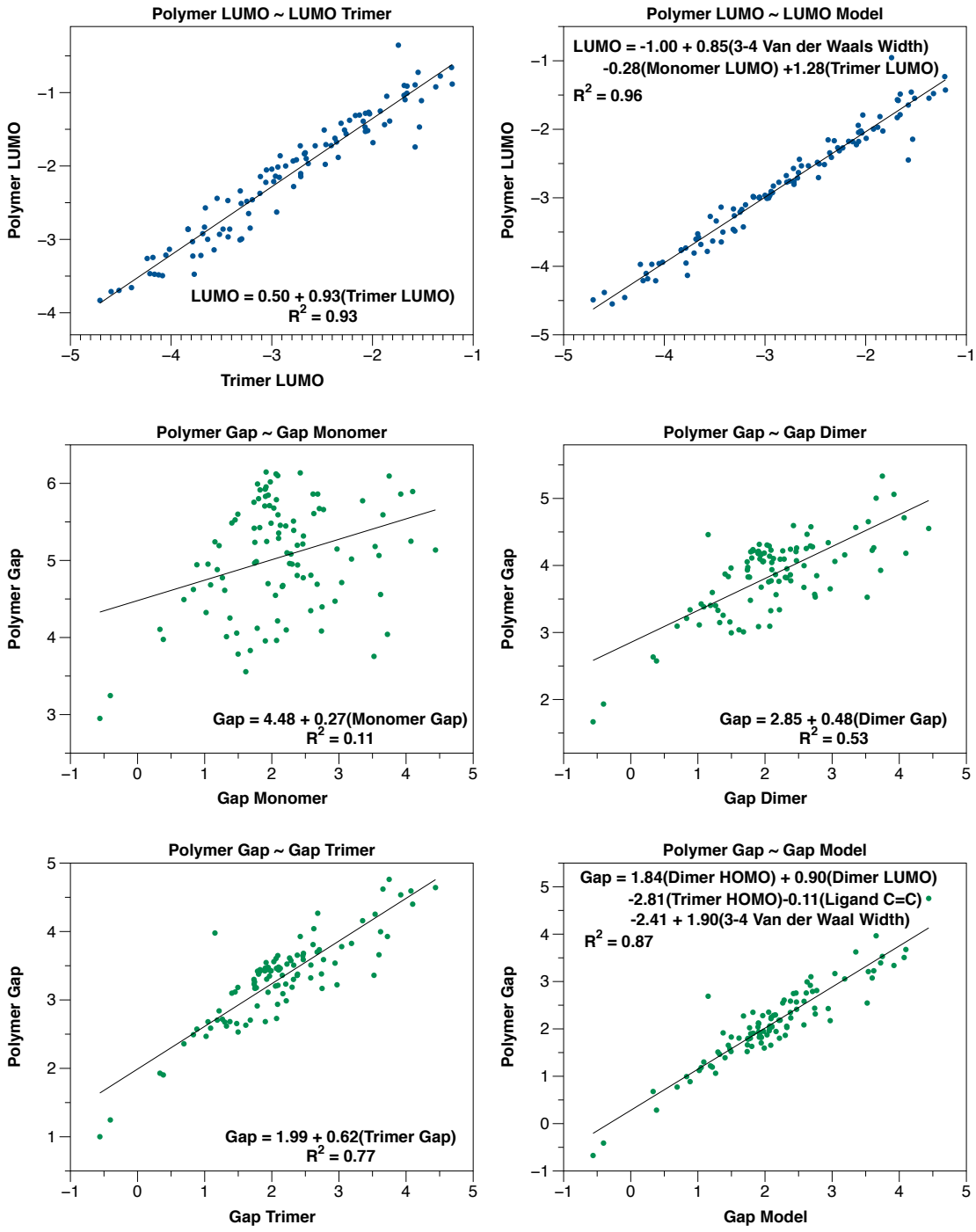


Figure A2: Additional plots for the linear regression models for predicting polymer HOMO, LUMO, and HOMO-LUMO gap from monomers, dimers, trimers and multivariate fits (2 of 2)

Code	Dimer	Trimer	Tetramer	Pentamer
1	0.403	0.426	0.374	0.323
2	0.515	0.407	0.356	0.326
3	0.448	0.433	0.379	0.337
4	0.39	0.353	0.319	0.294
5	0.518	0.709	0.646	0.632
6	0.342	0.366	0.322	0.297
7	0.45	0.486	0.486	0.449
8	1.657	0.658	0.562	0.508
9	0.41	0.488	0.442	0.4
10	0.53	0.43	0.382	0.355
11	0.532	0.431	0.382	0.355
12	0.869	0.523	0.529	0.444
13	0.95	0.58	0.434	0.309
14	0.446	0.521	0.502	0.464
15	0.358	0.391	0.486	0.411
16	0.001	0.613	0.652	0.658
17	0.602	0.478	0.476	0.421
18	0.505	0.532	0.526	0.497
19	0.466	0.742	0.805	0.657
20	0.409	0.381	0.391	0.395
21	0.57	0.488	0.473	0.439
22	0.483	0.396	0.439	0.499
23	0.461	0.389	0.394	0.396
24	0.379	0.501	0.597	0.64
25	0.572	0.422	0.369	0.335
26	0.283	0.364	0.935	1.036
27	0.293	0.325	0.293	0.71
28	0.289	0.283	0.48	0.5
29	0.475	0.641	0.952	1.019
30	0.238	0.989	0.996	0.944
31	0.87	0.892	0.898	0.924
32	0.701	0.68	0.512	0.431
33	1.015	0.748	0.729	0.729
34	0.414	0.633	0.644	0.588
35	0.544	0.609	0.598	0.573
36	0.62	0.64	0.75	1.116
37	0.597	0.584	0.6	0.549
38	0.424	0.755	0.792	0.795
39	0.298	0.417	0.367	0.336
40	0.383	0.34	0.298	0.288
41	0.34	0.476	0.411	0.373
42	0.94	0.873	0.546	0.474
43	0.337	0.45	0.332	0.338
44	0.302	0.222	0.317	0.35
45	0.264	0.243	0.229	0.215
46	0.181	0.243	0.246	0.231
47	0.422	0.634	0.706	0.672
48	0.173	0.261	0.256	0.244
49	0.137	0.219	0.224	0.209
50	0.364	0.255	0.231	0.213

Code	Dimer	Trimer	Tetramer	Pentamer
51	0.304	0.298	0.269	0.247
52	0.198	0.255	0.236	0.219
53	0.145	0.236	0.223	0.209
54	0.222	0.257	0.234	0.216
55	1.752	1.688	1.241	1.286
56	0.843	0.294	0.258	0.242
57	0.157	0.199	0.164	0.121
58	0.137	0.176	0.134	0.084
59	-	0.336	0.336	-
60	0.37	0.25	0.222	0.2
61	-	0.395	0.369	-
62	0.397	0.403	0.402	0.36
63	0.425	0.404	0.369	0.388
64	0.443	0.402	0.348	0.373
65	0.513	0.464	0.405	0.439
66	0.449	0.402	0.347	0.338
67	0.587	0.402	0.569	0.571
68	0.639	0.548	0.732	0.618
69	0.354	0.474	0.405	0.358
70	0.119	0.499	0.43	0.38
71	0.301	0.496	0.475	0.447
72	0.104	0.069	0.451	0.443
73	0.224	0.45	0.514	0.545
74	0.12	0.3	0.476	0.422
75	0.38	0.716	0.769	0.835
76	0.487	0.441	0.387	0.349
77	0.502	0.53	0.454	0.403
78	0.767	1.053	0.907	0.358
79	0.228	0.197	0.816	0.725
80	0.168	0.199	0.241	0.495
81	0.359	0.38	0.346	0.321
82	0.377	0.788	1.008	1.201
83	0.283	0.254	0.092	0.329
84	0.601	0.505	0.332	0.473
85	0.218	0.588	0.523	0.439
86	0.175	0.873	0.476	0.337
87	0.474	0.512	0.441	0.392
88	0.302	0.666	0.603	0.589
89	0.29	0.832	0.641	0.654
90	0.219	0.626	0.785	0.938
91	0.18	0.37	0.405	0.425
92	0.276	0.283	0.249	0.224
93	1.139	0.418	0.43	0.435
94	0.315	0.291	0.25	0.222
95	0.258	0.426	0.441	0.454
96	0.169	0.342	0.396	0.421
97	0.231	0.427	0.455	0.461
98	0.323	0.375	0.382	0.386
99	0.29	0.331	0.342	0.353
100	0.304	0.433	0.478	0.488

Table A15: Reorganization energies for all data

APPENDIX B

GENETIC ALGORITHM OPTIMIZATION OF ORGANIC PHOTOVOLTAIC MATERIALS: ADDITIONAL FIGURES

129 Monomer SMILES		
c(s1)c(S(=O)(=O)C=C2)c2c1	c(s1)c(OCCCO2)c2c1	c(o1)nnc1
c(c(nsn1)c12)ccc2	c(s1)c(SC=CS2)c2c1	c(s1)c(N(=O)=O)c(N)c1
c(s1)cc(N(=O)=O)c1	c(s1)cc(c2ccc(F)cc2)c1	C(C1=C)=CC=C1
c(s1)c(NC(=S)N2)c2c1	c(s1)c(SCS2)c2c1	c(nc1)ccc1
c(o1)ccc1	c([nH]1)ccc1	c(s1)c(cccc2)c2c1
c(s1)c(C)c(C)c1	c(s1)cc(c12)[nH]c(c2s3)cc3	c(cc1)ccc1
c(s1)cc(c12)Cc(c2s3)cc3	c(s1)c(CCC2)c2c1	c(s1)cc(c12)c(=O)c(c2s3)cc3
C(C1=O)=CC=C1	C(=C1C(=O)NC(=O)C12)C3C(=O)NC(=O)C=3C=2	c(s1)cc(C)c1
c(s1)c(C(=O)CC2(=O))c2c1	C1NC(=O)C2C=1C(=O)NC=2	c(s1)cc(C#N)c1
c(s1)c(SC(=O)C(=O)S2)c2c1	C=CC#C	C(S1)=CC2=C1C=C(C2=O)
c1[nH]c(c2)c(c1)[nH]c2	c1sc(cc2c3)c(c1)cc2sc3	c(s1)cc(C(=O)C(F)(F)F)c1
c(s1)c(SN=N2)c2c1	c(s1)c(OC(=O)C(=O)O2)c2c1	c(s1)c(OCCS2)c2c1
c(s1)c(CC=CC2)c2c1	c(s1)c(OC(=O)O2)c2c1	c(s1)c(C#N)c(C(F)(F)F)c1
c(cc1)cc(c12)Cc(c2c3)ccc3	c(cc1)cc(c12)c(=O)c(c2cc3)cc3	c(s1)cc(N)c1
c(cn1)ccc1C=C	c(s1)c(C(=O)C=C2)c2c1	c1oc(c2)c(c1)oc2
c(s1)c(CCCC2)c2c1	C#CC#C	c(s1)cc(OC)c1
c(s1)c(SC=C2)c2c1	c(s1)cc(c12)scc2	c([nH]1)ccc1C#C
c(s1)c(NC(=O)N2)c2c1	C(C1=C)=C(C=CC=C2)C2=C1	c(s1)c(C(=C)C=C2)c2c1
c(s1)c(OC)c(C#N)c1	c(s1)c(cc(F)c(F)c2)c2c1	c(s1)c(O)c(C(=O)O)c1
c(cc1)ccc1N	c(c(non1)c12)ccc2	c(s1)c(OC=CO2)c2c1
c(s1)c(F)c(F)c1	c(s1)c(C(=O)OC2(=O))c2c1	c(s1)c(OC=C2)c2c1
c(s1)c(SC(N(=O)=O)=C2)c2c1	c(s1)cnc1	c(s1)c(NC=N2)c2c1
C=CN=N	c(s1)c(NCCN2)c2c1	c(s1)c(NC(=O)C(=O)N2)c2c1
c(s1)c(SC=N2)c2c1	C(=C1C(=O)OC(=O)C12)C3C(=O)OC(=O)C=3C=2	c(o1)cnc1
c([nH]1)cnc1	c(s1)c(C)c(C=C)c1	c(s1)cc(O)c1
c(s1)c(OC)c(C(F)(F)F)c1	c(s1)c(OC=N2)c2c1	c(s1)c(nccn2)c2c1
C=CC=C	c(s1)cc(C=O)c1	c(s1)c(SC(C#N)=C2)c2c1
c(s1)cc(c12)C(=C(C#N)C#N)c(c2s3)cc3	c(cc1)cc(c12)C(=O)c(c2c3)ccc3	c1oc(c2)c(c1)sc2
c(s1)c(SC(=S)S2)c2c1	c(s1)c(NCN2)c2c1	c(s1)cc(c12)[nH]cc2
c(s1)c(cc(C#N)c(C#N)c2)c2c1	c(s1)c(cnnc2)c2c1	C1=CC(C2)=C(C1)C=C2
c(cc1)ccc1C=C	c(s1)c(cc(OC)c(OC)c2)c2c1	c(s1)cc(C(=O)O)c1
c(s1)c(OC)c(N)c1	c(s1)c(OCCO2)c2c1	c(s1)c(N(=O)=O)c(N(=O)=O)c1
c(s1)c(NC=C2)c2c1	c(s1)c(S(=O)C=C2)c2c1	c(s1)c(C=C2)c2c1
c(cc1)cc(c12)[nH]c(c2c3)ccc3	c(s1)c(SCCS2)c2c1	c(s1)c(ON=N2)c2c1
c([nH]1)ccc1C=C	c(cc1)cc(c12)oc(c2c3)ccc3	c(s1)c(O)c(S)c1
c(s1)nnc1	c(s1)cc(c2ccc(OC)cc2)c1	c(s1)c(OC(=S)O2)c2c1
c1sc(c2)c(c1)sc2	c(s1)c(C(=O)NC2(=O))c2c1	c(s1)c(OCO2)c2c1
c(s1)cc(c2ccc(N(=O)=O)cc2)c1	c(s1)cc(C(=O)C)c1	c(s1)c(c(F)c(F)c(F)c(F)2)c2c1
c(s1)c(NC=CN2)c2c1	c(s1)c(CC=C2)c2c1	c(s1)c(OC)c(OC)c1
c(s1)cc(S)c1	c(s1)c(cc(N(=O)=O)c(N(=O)=O)c2)c2c1	c(s1)c(C#N)c(C#N)c1
C(C1=O)=C(C=CC=C2)C2=C1	c(s1)c(ncnc2)c2c1	c(s1)c(C(=O)SC2(=O))c2c1
c(s1)cc(C(=O)OC)c1	C=C(C1=O)C(=O)C1	c(s1)cc(c2ccccc2)c1
c(s1)ccc1C=C	c(s1)cc(c2ccc(N)cc2)c1	c(s1)c(SC(=O)S2)c2c1

Table B1: List of SMILES for the 129 monomer data set.

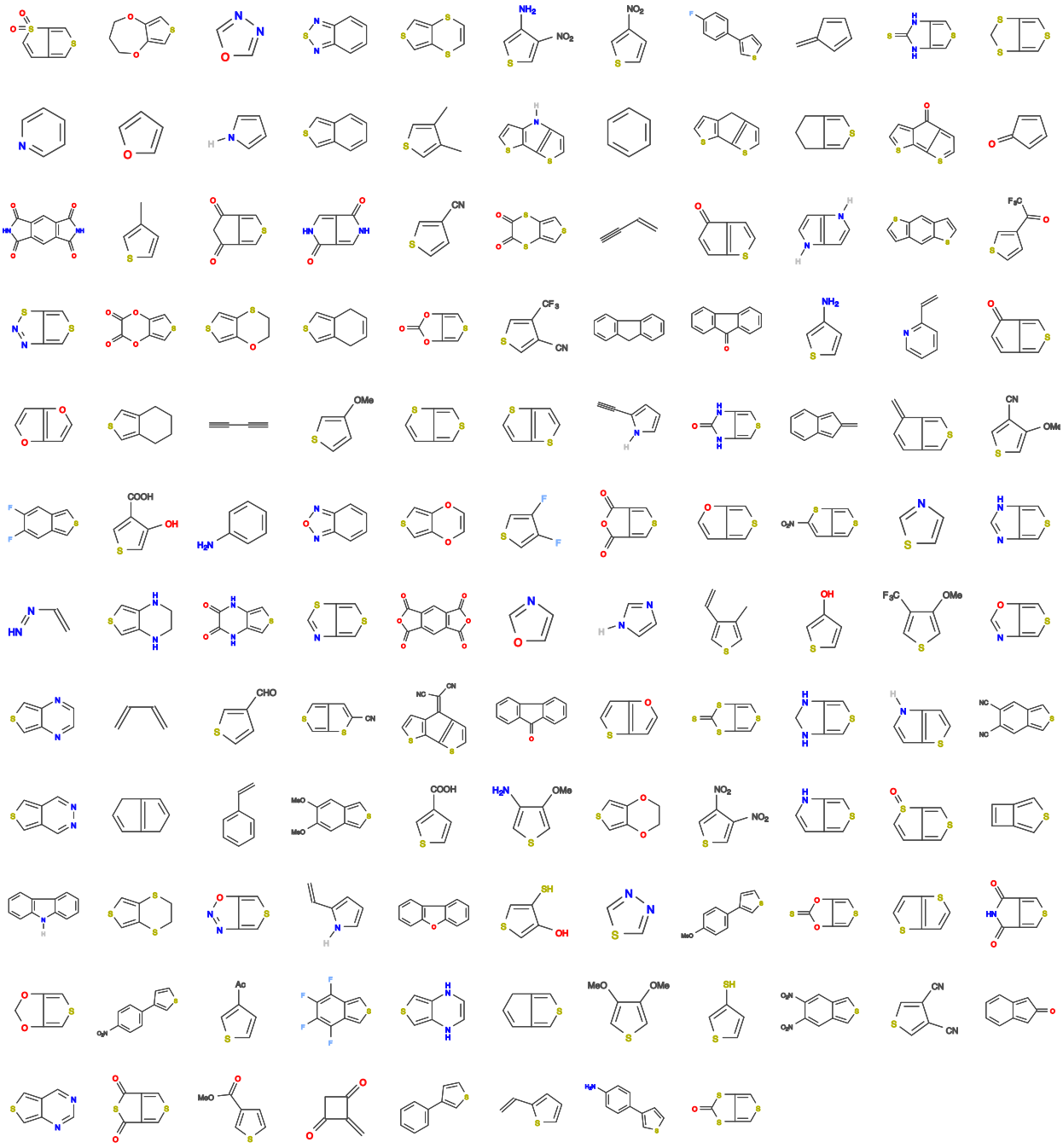


Figure B1: Molecules in the 129 monomer dataset.

C1S[C]2[C]3SCC[C@@H]3C(=O)[C@H]2C1	c1sc(c(c1C(F)(F)F)C(F)(F)F)	c1c(OCC)c(OCC)c(OCC)cc1
c1scc2c1[C@@H](CC[C@H]2N(=O)=O)N	c1ccc(c2c1nn(CC)n2)	c1sc(c(c1CC)CC)
N1[C@H]2[C@H](N[C@@H]3[C@@H](CSC3)N2)N(S1)	c1c(C)c(c2c1ccco2)	c1cc2c(s1)c1c(C2)cc(s1)
c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)	N1[C@H]2CSC[C@H]2N(S1)	c1sc(CC)c(c1CC)
c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)	c1c(c2c(s1)nc1c(c2CC)CCS1)	c1sc(c2c1cc(F)c(F)c2)
c1sc2c(n1)Cc1c2sc(c1)	c1cc(N(=O)=O)c(cc1)	c1c2c(SCC2)c2c(CCS2)c1
c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)	c1sc(C(F)(F)F)cc1	c1sc(c2c1SCCCS2)
c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)	c1sc(c2c1ocn2)	c1sc(c2c1cc(C(=O)C)cc2)
c1sc(c2c1c(C#N)ccc2C#N)	c1sc(c(c1N(=O)=O)C#N)	c1oc(c(c1C)C)
c1sc(c2c1S(=O)(=O)CCC2)	c1sc(c2c1CCC[C@H]2O)	C1=CC=C(C1=O)
C1SC[C@@H]2[C@H]1[C](O[C@@H]2[O])NCC	c1sc2c(c3ccoc3cc2c1)	C=NNc1ccc(cc1)
c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)	c1cc(CN)ccc1CN
C\1=C/O/C=C(/O/C=C/O/C=C/Oc2c(cccc2)O1)	c1sc(c2c1sc(C(=O)CC)c2F)	c1sc(c2c1C(=O)c1ccccc1C2=O)
C1=C(C=C/C/1=C\1/C=CC(=C1)N)N	c1sc(S[CH2])c(c1)	c1sc2cc(C(=O)O)sc2c1
c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C(F)(F)F)C2)	c1sc(C=O)c(c1)	c1c2c(ncc(CC)n2)c(s1)
c1ccc(OCC)c2c1c(OCC)ccc2	c1sc2c(c1)sc1cc(sc21)	c1scc(C#N)c1N
c1cc2c(cc1)c1c(C2(C)C)cc(cc1)	c1sc(c2c1cc(S)c(O)c2)	c1c(CC)sc(c1)
n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)	c1sc2c(c1)oc1c2sc(c1)	C=Cc1oc(c(c1O)O)C=C
C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1	c1c2CCc2sc1	c1sc(c2c1c(N(=O)=O)ccc2N)
c1c2c(ccs1)nc1c2sc(c1)	c1scc(N(O)O)c1N(O)O	c1oc(nn1)
c1scc(C)c1C=C	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2	c1c(C)cc(s1)
c1c2csc2c(nn1)	[C]1C(=O)Oc2c1cc1c(c2)[C](C(=O)O1)	c1sc(c(c1C#N)C#N)
c1sc(c2c1ncs2)	c1sc(c2c1OCCS2)	c1scc2c1sc1c2csc1
c1c2n(CCC)c3c(c2ccc1)cccc3	c1c(O)sc2c1[nH]c1c2sc(c1)	c1sc(c2c1C[C@H](S)CC2)
c1occ(N(=O)=O)c1N	c1scc2c1C(=[S@@](OC)C(=C2)OC)	c1c2c3c(s1)ccc1c3c(cc2)sc1
c1sc(c2c1C(=O)NC2=O)	c1sc(c2c1sc(C(=O)OCC)c2)	c1[nH]c(cc1)C#C
c1oc2c(C(=O)N)c(oc2c1)	c1sc(c2c1OCCCO2)	c1sccc1/C=C(/Cc1ccc(CC)cc1)
C1=[S]C(=S)C2=C1C=C(S2(=O)=O)	c1sc(cc1)C#C	c1[nH]cc2c1[nH]cc2
C1=CC(OC)=C(/C/1=C\1/C=C(OC)C=C1)	c1sc(c2c1c(C(=O)OC)ccc2)	c1sc(c(c1CC)CC)C=C
C=Cc1ccccc1	c1ccc(cc1)	c1[nH]cnc1
N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)	c1sc(c2c1[C@H](S)CC[C@H]2O)	c1sc(c2c1C(=O)CC2=O)
c1sc(c2c1SCS2)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)	c1cc2c(s1)cc(s2)
c1scc(OC)c1N	c1sc(c2c1CC[C@@H](N(=O)=O)C2)	C#Cc1sc(cc1)C#C
c1c(cc2c(c1)nccn2)	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)	c1sc(c(c1C(=O)O)OC)
c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)	c1sc(c2c1ncnc2)	C1=CC=C(C1=S)
c1sc(c2c1SCC(=O)CO2)	N(CC)c1ccc(c(c1)C=C)N(CC)	c1c2C(=O)OCc2ccc1
c1oc2c(c1)C(=O)c1c2oc(c1)	c1sc(c2c1c(C)ccc2)	c1c(CC)cc(s1)
C1SC[C@@H]2[C@@]1(S)C[C@H](OC)[C@@H](OC)C2	c1sc(c2c1[C@H](C)CCC2)	c1sc(c(c1F)F)
C1=C[C@@H]2[C@H](C1)[C@@H]1[C@@H](C=CC1)C2=C	c1sc(c2c1c(C=C)ccc2C)	C1=C2C(C(=O)S1)=C(OC2=O)
c1sc(c2c1[C@H](C=C)CCC2)	c1sc(c2c1[nH]c(=S)[nH]2)	c1cc(c2c(c1)c1c(s2)cccc1)
c1sc(c(c1O)C(=O)O)	c1oc(c(c1C#N)C(F)(F)F)	c1sc(c2c1nc(OCC)c(CN)n2)
C1SCN2[C@H]1NCC2	c1oc2c(c1)C(=S)c1c2oc(c1)	c1sc(c2c1sc(=S)o2)
c1sc(c2c1cc(C(F)(F)F)cc2)	c1sc(c(c1)N(=O)=O)	c1scc2c1c(ccc2O)
C1=[S]C(=O)C2=C1C(=O)[S]=C2	c1sc(c2c1cc(C)c(C)c2)	c1sc(c2c1c(F)ccc2F)
c1sc(c2c1C[C@H](OC)[C@@H](OC)C2)	C1=C(c2cscc12)	c1sc(C)c(c1)
c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)	n1c2cscc2[nH]cc1
c1sc(c2c1cc(S)cc2)	c1c2c3CCSc3c3SCCc3c2c(cc1)	c1c(F)cccc1
c1sc(c2c1cc(C(=O)O)c(O)c2)	c1c2c(OCN2)c(s1)	c1cc2c(cc1)c1c(C2)cc(cc1)
C1=CC2=CC(=NC2=C1)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)	c1c(CC)c(ccc1)
c1c(OC)cc(c(c1)OCC(CC)CC)	c1c2[CH][S]([CH2])[CH]c2c(cc1)	c1sc(c(c1OC)C(F)(F)F)
c1scc2c1CCC[C@H]2C(=O)O	c1oc2cc(sc2c1)	c1sc(c2c1cc(C=C)cc2)
C#Cc1[nH]c(cc1)C#C	C(=S)[CH]C1=[S]C=C(O1)	c1sc2nc(sc2c1)
C1=C2C(C(=O)N1)=C(OC2=O)	c1sc(c2c1cc(C(C)(C)C)cc2)	c1c2nsnc2c(cc1)
c1sc(c2c1c(C(F)(F)F)ccc2C#N)	c1c(C(=O)C)scc1	c1c2nccnc2c(s1)
N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1	c1scc(C(=O)O)c1S	c1c2nc(C)c(C)nc2c(s1)
N1c2cscc2N(CC(=O)C1)	c1c(OC)sc(OC)c1	c1[nH]c(C=C)c(c1)
c1ccc(c2c1cccn2)	c1sc(c2c1nc(OCC)c(OCC)n2)	c1oc(cc1C=O)
c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2	c1c2c(nccc2)c(s1)	[CH]/C=C\1/OCC[C](OCC1)
c1sc(c(c1OC)C#N)	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)	c1sc(c2c1sc(=O)s2)
c1c(C(=O)C(F)(F)F)scc1	C1=Cc2[nH]c(cc2C1)	c1c(C)sc(C)c1
c1sc(c(c1)C(=O)C(F)(F)F)	c1csc2c1c(CC)c(CC)c1c(csc21)	c1sc(c2c1c(C#N)ccc2OC)
c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)	c1sc2c(c1)C(=O)C(=C2)	C1=c2ccccc2=C(C1=C)
c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)	c1sc(c2c1C[C@H](C(=O)OC)CC2)	c1occc1S
N1[C@]2(SC)CSC[C@H]2NC(=C1)	c1sc(c2c1[C@H](C)CC[C@H]2C)	c1occc1N
c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)	c1sc(c2c1CCS(=O)(=O)C2)	c1ccc2c(c1)oc1c2ccc(c1)
c1sc(c2c1CC[C@@H](O)C2)	c1scc(N(=O)=O)c1N	c1sc(c2c1OSCCO2)
C1S[C]2[C]3SCC[C@@H]3C(=C(CN)CN)[C@H]2C1	c1sccc1O	c1sc(c2c1sc(N(=O)=O)c2)

Table B2: List of SMILES for the 442 monomer data set. (Part 1 of 3)

c1sc(c2c1c(C=C)ccc2)	c1sc(c2c1ccc(F)c2)	c1c2c(ncc(C)n2)c(s1)
c1n(CC)c(cc1)	c1sc(c2c1scc2)	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)
C1C[C@@H]2[C@@H](CC1)N(CC)[C@@H]1[C@@H](S2)CCCC1	[CH]C1=C[C](C[C@H]2CSC[C@]12C=O)	c1oc(c(c1OC)C#N)
c1c(OC)c(cc(c1)OC)	c1sccc1N	C1=C2CSC[C@H]2C(=C1)
c1cccc2c1N(C)C(=O)[C]2[C]1C(=O)N(C)c2cc(ccc12)	c1c2c(cccc2)c(s1)	c1c(F)sc(F)c1
c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)	c1c(OCC)cc(OCC)c(c1)	c1[nH]cc(c1C)C=C
C1[C@H]([O])C[C@@H]2[C@@H]1CCCC2	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)	C1=CC=C(C1)
c1c(F)c(F)c(c(c1F)F)	c1sc(c2c1scn2)	c1c2[nH]cnc2c(c2c1[nH]cn2)
c1sc(c2c1ncnc2C#N)	c1sc(c(c1OC)OC)	c1sc(c(c1C#N)C(F)(F)F)
n1c2C=[S][CH]c2n(c2c1c1c3c(cc1)CC)c(CC)ccc23)	c1[nH]c2nc3[nH]c(nc3nc2n1)	c1c(OCC)c(ccc1)
c1scc2c1C(=C[S@@]2[O])	c1sc(c(c1N(=O)=O)N(=O)=O)	c1sc(c2c1oc(=O)o2)
C1=C(C)C(C)=C(C2=N[C@H]3N4[C@@H](N[C@@H]5C[C@H](C)[C@H](C[C@@H]35)C)[C@H]3[C@H](N[C@@H]4[C@H]12)C[C@H]([C@H](C3)C)C)	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)	c1cc(C)cc(c1C)
c1oc(cc1OC)	c1sc(c(c1O)OC)	c1sc(c2c1nc(CCO)c(CN)n2)
N1c2cscc2N(CC1)	c1sc(c2c1[nH]c(=O)[nH]2)	c1[n]c2c(c(=O)c3c2[n]cc3)c1
c1c(C)cc(c(c1)C)C=C	c1n(CC)c(=O)c2c1c(=O)n(CC)c2	c1sc(NC)c(c1C(=O)OC)
c1sc(c2c1CCC[C@H]2N(=O)=O)	c1scc2c1C=C(C2=O)	c1oc(cc1C(=O)C)
c1sc(c2c1[C@H](N)CC[C@H]2N)	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)	C1=C(CC)C(CC)=C(S1(=O)=O)
c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)	c1scc(C(=O)OC)c1	c1oc(cc1C)
c1sc(c2c1C[C@H](C(=O)C)CC2)	c1sc(c2c1C=[S]C=C2)	C1=CC2=C(C1)C=C(C2)
c1scc2c1C[C@@H]([C@@H](OC)C2)N	c1oc(cc1)	c1csc2c1n(CC)c1c2scc1
c1sc(c2c1[C@H](OC)CC[C@H]2OC)	c1[nH]c(c2c1nccn2)	c1sc(c2c1cc(C#N)cc2)
N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)	c1sc(c2c1oc(C#N)c2)	c1nnc(nn1)
c1cc(c(cc1)N(c1ccccc1)c1ccccc1)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)	C(=C)C#C
n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)	c1scce1N(CC)	c1sc(nn1)
c1c2c(sc(C(=O)OCC)c2F)c(s1)	c1sc(c2c1CCC2)	C1=CC(=C2[C@H]1CCS2)
c1scce1/C=C(/Cc1sc(CC)cc1)	c1coc(N(=O)=O)c1	C1=C([CH]S1)C(=C)
c1cc2c(=O)n(C)c(=O)c3c2c2c1C1=C4[C@@H](c2cc3)C=CC2=C4[C@@H](C(=O)N(C2=O)C)C(=C1)	c1oc(cc1C(F)(F)F)	c1c(C(F)(F)F)sc(C#N)c1
N1C(=O)[C](c2c1cc(cc2)C)[C]1C(=O)N(c2c1ccc(C)c2)	c1c(N)ccc(c1)	c1sc(c2c1OCCO2)
C1=C2OCCSC2=C([S@@]1NC)	C1=C(c2c3c(cccc13)ccc2)	c1sc2c(c1)C(=S)c1c2oc(c1)
c1n(C)c(=O)c2c1c(=O)n(C)c2	c1sc(c2c1[CH][N]2)	c1sc(c2c1C(=O)CCC2=O)
c1scc2c1[C@@H](CCC2)N	c1sc(c(n1)C(=O)OC)	c1sc(c2c1cc(C=C)c(C)c2)
c1c2c(nccn2)c(c2c1nccn2)	c1c2cscc2c(N)c(c1)	c1oc(cc1C#N)
C1=C(F)C=C/C/1=C\1/C(=CC(=C1)F)	c1sc(c(c1)OC)	c1sc2c(c1)C(C)(C)c1c2sc(c1)
c1c2n(CC)c3c(cc4c(c3)c3c(C4)cccc3)c2cc2c1c1c(C2)cc2c(c1)C(CC)(CC)c1c2cccc1	c1sc(c2c1sc(=O)c(=O)s2)	c1sc2c(c1)[nH]c(c2)
n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)	c1oc(c(c1C#N)C#N)	C#CC#C
c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)	c1scc2c1c(c(cc2)O)	c1c(ncc2c1non2)
c1c2c(ccs2)c(c2c1cc1c(c2)scc1)	N(CC)c1cscc1N(CC)	c1sc(c2c1cc(C#N)c(C#N)c2)
c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)	C1=CC=C(S1(=O)=O)	c1[nH]c(c(c1)C(=O)C)
c1c2ccsc2cc2c1c(c1c(c2)scc1)	c1sc(c2c1CCCC2)	c1occ(OC)c1N
c1sc(c2c1[C@H](C(F)(F)F)CCC2)	c1sc(c2c1cccc2F)	c1scc2c1C=C(C2)
c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)	c1c(OCC)c(cc(c1)OC)	c1sc(c(c1SCC)SCC)C=C
c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)	c1sc(C(=O)O)cc1O	c1sc(c2c1CC(=O)C(=O)C2)
C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)	c1c(C#N)sc(OC)c1	c1occ(C)c1C=C
c1c2c3c(s1)ccc1c3c(cc2)c(s1)	c1[nH]c(cc1)	c1sc(c2c1sc(=S)s2)
c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](OC)C2)	c1sc(c2c1[nH]cn2)	c1oc(cc1O)
c1sc(c2c1C[C@H](S)[C@@H](O)C2)	c1sc(c2c1nc(CN)c(CN)n2)	c1cnc(c2c1nsn2)
c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)	c1oc(cc1C(=O)O)	c1sc(c2c1oc(N(=O)=O)c2)
c1sc(c2c1[C@H](C(=O)O)CC[C@H]2O)	c1sc(c2c1c(C=O)ccc2)	c1oc(c(c1)N(=O)=O)
c1sc(c2c1C[C@H](C)[C@@H](C)C2)	C=C	c1sc(c2c1occ2)
c1scc2c1C[C@H](C)[C@H](C2)C(=C)	c1sc(c2c1cc[nH]2)	c1c2nonc2c(cc1)
c1ccc2c(c1)C(=O)c1c2ccc(c1)	c1sc(c2c1c(C#N)ccc2)	c1c(CC)c(CC)c(CC)cc1
C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)	c1sc(c(c1)C=O)	c1cc2c(c3c1nsn3)cc(c1c2nsn1)
c1sc2c(ccc3c2ccc2c3ccc3c2scc3)c1	Nc1sc(N(=O)=O)cc1	C1=C2C(C(=O)S1)=CN(C2=O)
c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1	c1c(C)cc(CC)c(C)c1	c1csc2c1c(CC)c1c(c2CC)scc1
c1sc(c2c1C[C@H](N)[C@@H](N(=O)=O)C2)	c1sc(c2c1cc(OC)cc2)	c1sc(c2c1OCSS2)
C1=CC=C([C]1[C]1C=CC=C1N)N	c1csc2c1C(S)=c1c2scc1	C1S[C@@H]2[C@@H]1NCC2
c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)	Nc1sc(c2c1nccn2)	c1sc(c2c1c(S)ccc2)
c1[nH]cc2c1S(=O)(=O)C(=C2)	C=C(C)CN	c1nc2CSCc2nc1
c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)	c1oc(c(c1O)S)	c1sc(c2c1c(S)ccc2O)
c1scc2c1C[C@@H](CC2)C(=C)	c1cc(cnc1)	N1C(=S)C=C(C1=S)
c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)	c1cc2c(cs1)c1c(ccs1)c2	c1oc(c(c1OC)OC)
c1c(c2c(s1)c(OC)c1c(c2OC)scc1)	Sc1cscc1S	c1sc(c2c1OSCS2)
N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)	[C]1c2ccccc2[C](C1=O)	c1sc(c2c1[nH]c(N(=O)=O)c2)
c1scc2c1nc1c3cc(sc3c3sccc3c1n2)	N1CN[C@@H]2[C@H]1N(CN2)	c1c(C(F)(F)F)sc(OC)c1

Table B3: List of SMILES for the 442 monomer data set. (Part 2 of 3)

c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)	c1c2c(cccc2)cc2c1c(ccc2)	c1cscc1C(=O)O
C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1	c1sc2c(c1)C([C](C#N)C#N)=c1c2sc(c1)	c1sc(c(c1O)C#N)
c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)	c1sc(c2c1cc(C=O)cc2)	c1c([CH2])c(c2c1ccsc2)
c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)	c1sc(c2c1[C@H](C(=O)C)CCC2)	c1scc2c1nc(c(NC)n2)N(C)
c1sc(c2c1CC[C@@H](N)C2)	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)	C1=CC2=C(C1)N=C(C2)
c1sc(c2c1cc(C)c(OC)c2)	c1sc(c2c1c(C(F)(F)F)ccc2)	c1oc2CCc2c1
c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)	c1oc2c(c1)C(=O)c1c2sc(c1)	c1sccc1S
c1sc(c(c1OCC)OCC)C=C	c1sc(c2c1[C@H](F)CC[C@H]2F)	Nc1cscc1N
C(=C)c1sc(C)c(c1)	c1sc(c2c1oc(=S)o2)	c1sc(c2c1sc(C(=O)CC)c2)
c1sc(c2c1c(C)ccc2C)	c1sc(N(=O)=O)cc1	c1sc(c2c1cc(C)cc2)
c1sc(c2c1OCO2)	C1Oc2cscc2OC(C1=O)	c1sc(c2c1SCCS2)
C1=[S]c2cscc2[S]=C1	c1sc(c2c1cc(OC)c(OC)c2)	c1csc2c1cc(c1c2scc1)
c1sc(c2c1CCC[C@H]2C=O)	c1sc(c2c1c(C(F)(F)F)ccc2OC)	c1sc(c2c1nccn2)
c1sc(c2c1oc(=O)c(=O)s2)	c1oc2cc(oc2c1)	c1scc2c1N(CCN2C(=O)O)
c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)	c1c([n]c2[CH]Sc12)	c1sc(cc1)N
C1=C(CC2=C1C[CH]2)	c1scc2c1C=C(S2(=O)=O)	c1sc(c2c1cc(C(=O)OC)s2)
c1c(C)cc(c(c1)C)		

Table B4: List of SMILES for the 442 monomer data set. (Part 3 of 3)

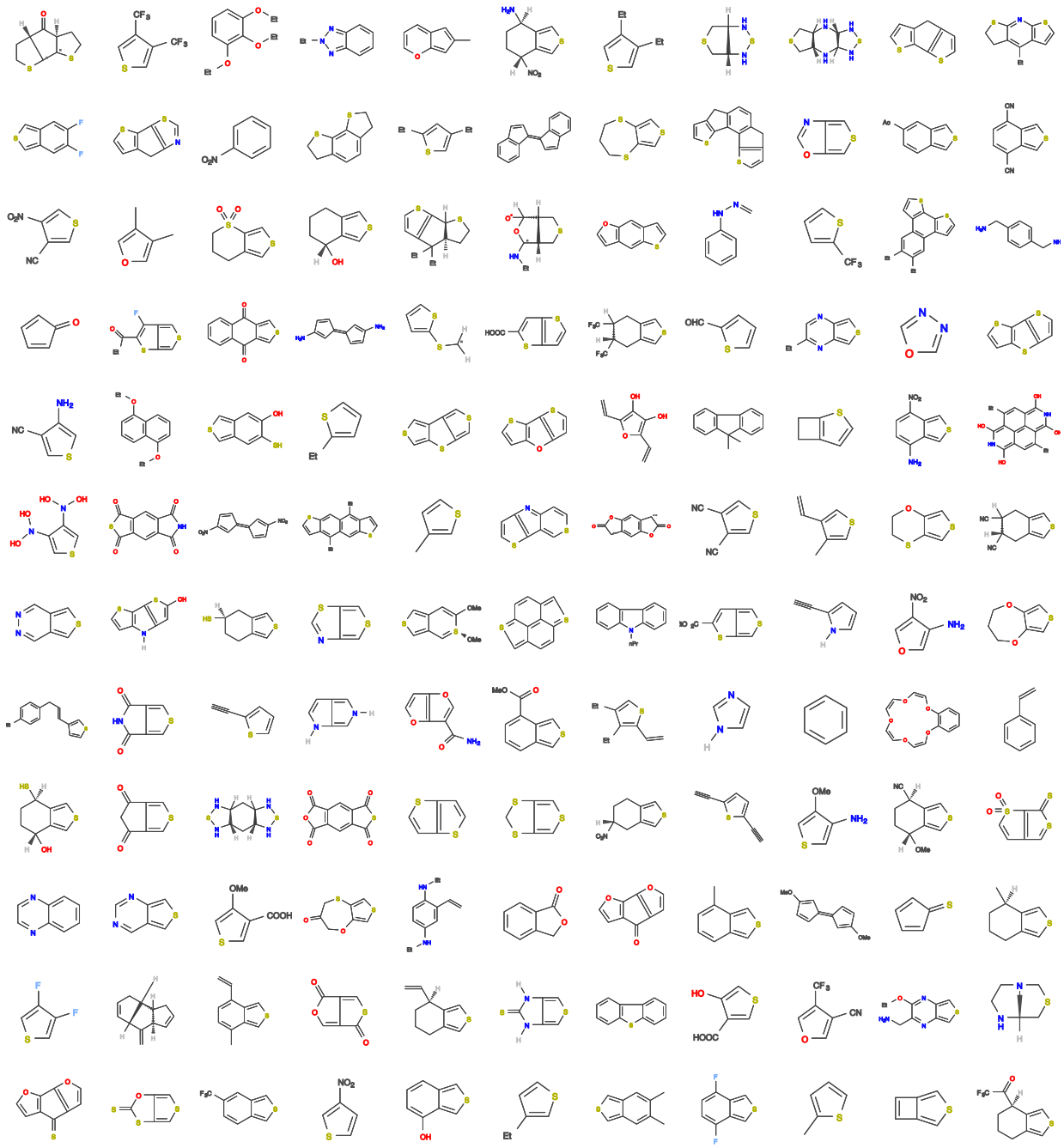


Figure B2: Molecules in the 442 monomer dataset (1 of 4).

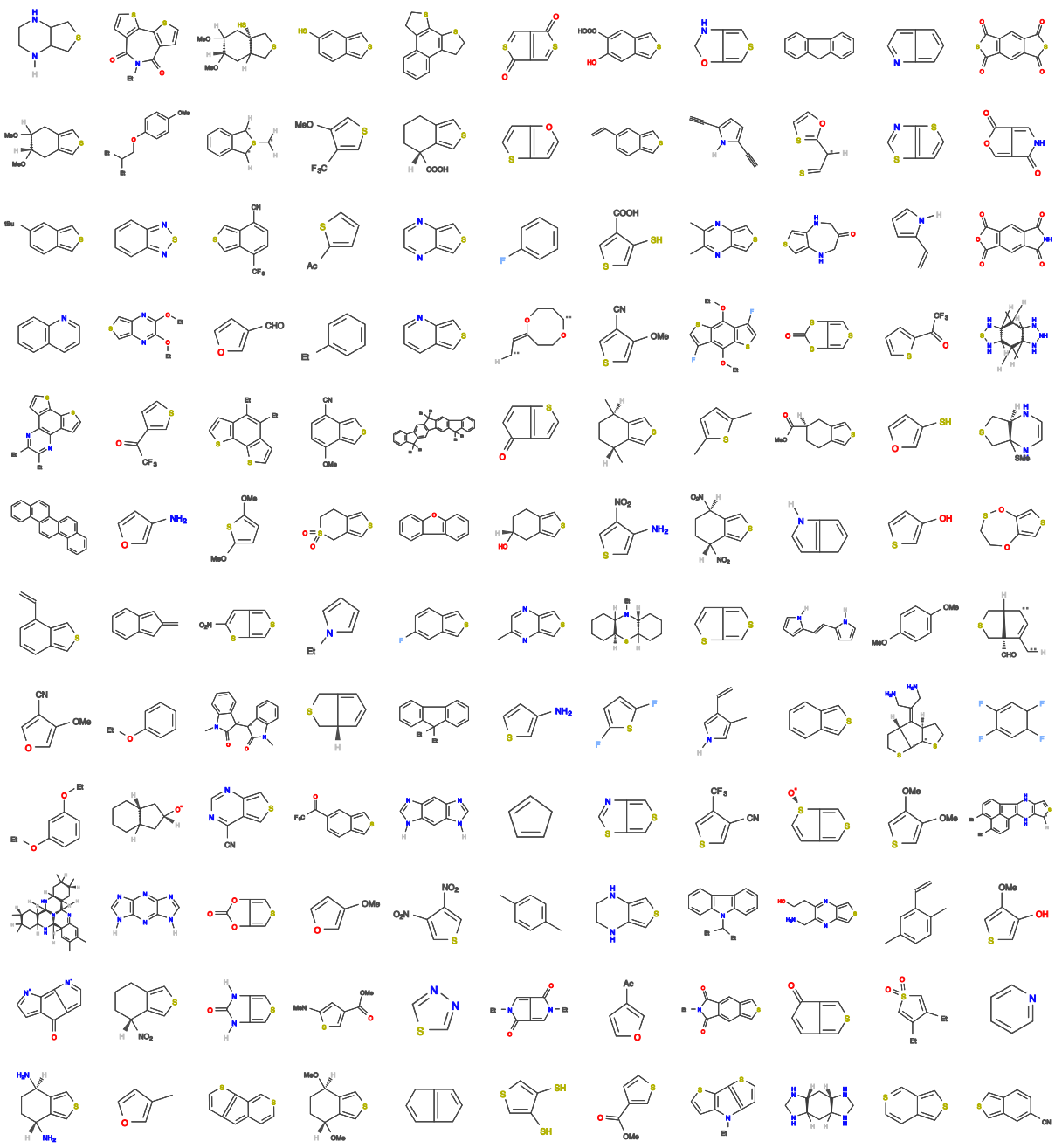


Figure B3: Molecules in the 442 monomer dataset (2 of 4).

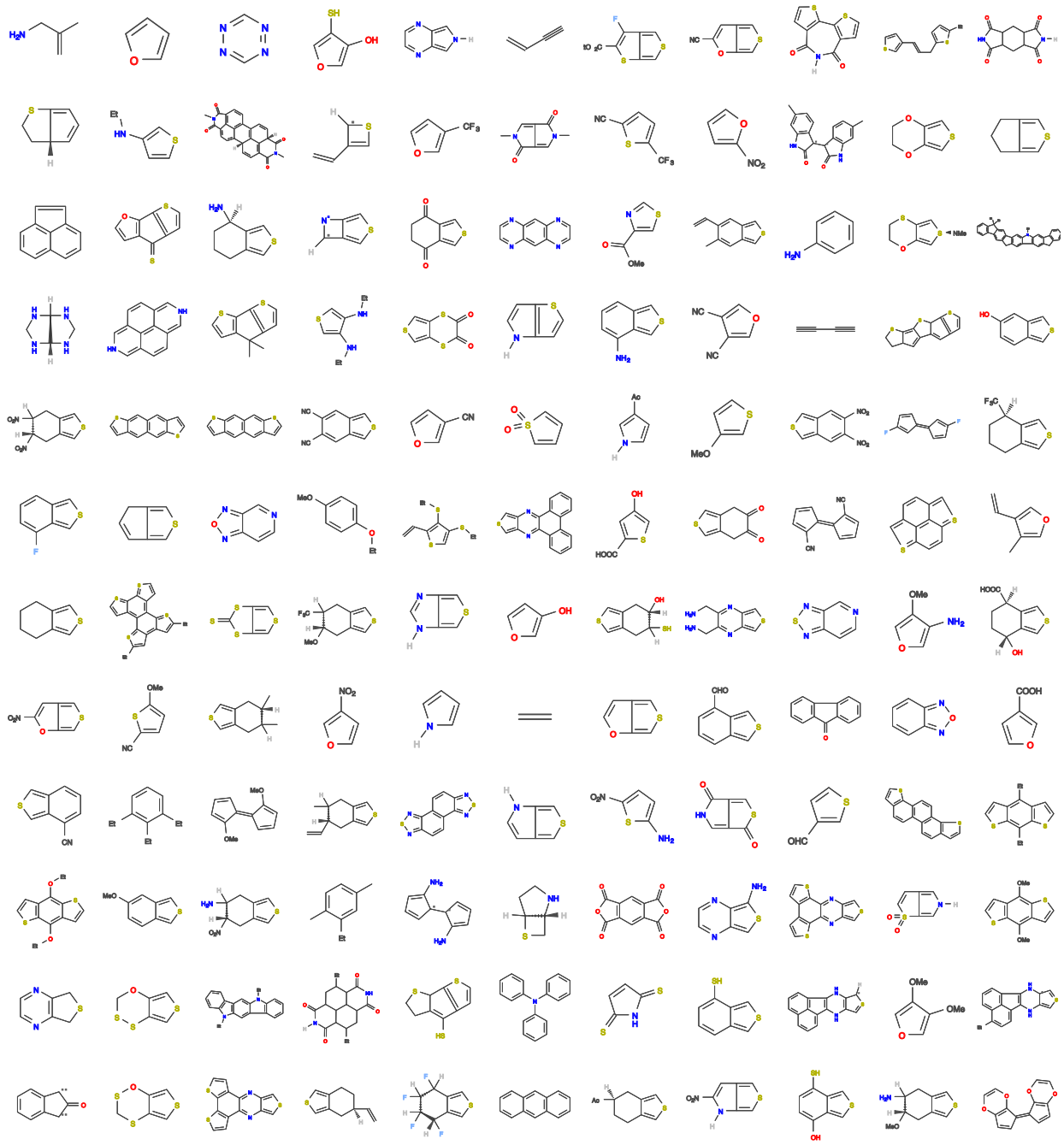


Figure B4: Molecules in the 442 monomer dataset (3 of 4).

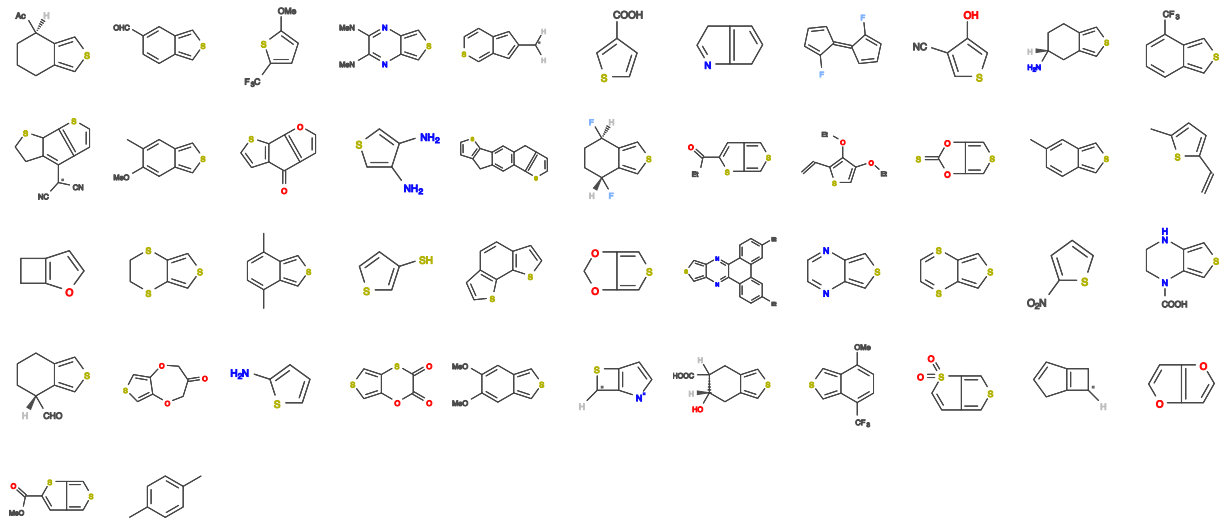


Figure B5: Molecules in the 442 monomer dataset (4 of 4).

C=C	c1c(C)pc2c1pc(C)c2	c1oc2c(n1)cc1c(c2)oc(n1)
C#CC#C	c1c(cc2c(c1)nccn2)	c1sc(c2c1c(N(=O)=O)ccc2)
C1OCC1	c1sc(c2c1c(C)ccc2)	c1sc(c2c1c(C(=O)O)ccc2O)
C=CC=C	c1sc(c(c1O)C(=O)O)	C1=Cc2cc3=CC(=Cc3cc2=C1)
c1cCcc1	c1sc(c2c1sc(=S)o2)	c1sc(c2c1cc(C#N)c(OC)c2)
C(=C)N=N	c1sc(c(c1)N(=O)=O)	c1sc(c2c1cc(C(=O)OC)cc2)
c1occc1S	c1sc(c2c1cc(S)cc2)	c1sc(c2c1SC(=O)CC(=O)O2)
c1occc1N	C1=CC2=CC(=NC2=C1)	c1c(O)sc2c1[nH]c1c2sc(c1)
c1sccc1O	c(cc1)c2cC(C)cc2c1	c1sc(c2c1c(N(=O)=O)ccc2N)
c1sccc1N	c1sc(c2c1sc(=O)s2)	N(CC)c1ccc(c(c1)C=C)N(CC)
c1sccc1S	C1=Cc2[nH]c(cc2C1)	c1oc2c(c1)C(=O)c1c2oc(c1)
C=C(C)CN	c1c(OC)c(cc(c1)OC)	c1cc(c2c(c1)c1c(s2)cccc1)
c(B1)ccc1	c1sc(c2c1ncnc2C#N)	c1sc(c2c1cc(C(F)(F)F)cc2)
c1SCSc1cc	c(cc1)c2cs(c)cc2c1	c1cc2c(cc1)c1c(C2)cc(cc1)
c1sc(nn1)	C1=C(CC2=C1C[CH]2)	c1sc(c2c1cc(C(C)(C)C)cc2)
Sc1cscc1S	c(oc1c(F)s2)csc1c2	c1sc(c2c1nc(OCC)c(OCC)n2)
c1oc(cc1)	C/C=C\1/OCCC(OCC1)	c1sc(c2c1cc(C#N)c(C#N)c2)
Nc1cscc1N	c1[nH]c(c2c1nccn2)	C1=C(CC)C(CC)=C(S1(=O)=O)
c1oc(nc1)	c1sc(c(n1)C(=O)OC)	c1sc2c(c1)C(=S)c1c2oc(c1)
c1scc(n1)	c1c2cscc2c(N)c(c1)	c1sc(c(c1N(=O)=O)N(=O)=O)
c1ccc(s1)	c1scc2c1c(c(cc2)O)	c1sc(c2c1OCC[C@@H](SC)O2)
c1ccc(cc1)	c1sc(c2c1oc(=O)o2)	c1scc2c1C[C@@H](CC2)C(=C)
c1occc1C=C	c1sc(c2c1sc(=S)s2)	c1sc(c2c1c(C(F)(F)F)ccc2)
c1cc(cnc1)	c1sc(c2c1ccc(F)c2)	c1oc2c(c1)C(=O)c1c2sc(c1)
c1oc(cc1C)	c1oc(c(c1)N(=O)=O)	c(s1)c2c(=O)n(C)c(=O)c2c1
c1nnc(nn1)	c1c(C)cc(CC)c(C)c1	c1sc(c2c1CC[C@@H](C=O)C2)
c1oc(cc1O)	c(s(c)1)c2ccccc2c1	c1sc(c2c1cc(SCC)c(SCC)c2)
c1sc(cc1)S	c1sc(c2c1oc(=S)o2)	c1sc(c2c1c(F)c(F)c(F)c2F)
c1sc(cc1)N	c1sc(c2c1cc(C)cc2)	c1cc2c(C(=O)N(C2=O)CC)cc1
c1c2CCc2sc1	c1sc(c(c1C)S[CH2])	c1oc(c(c1N(=O)=O)N(=O)=O)
c1c(F)cccc1	c1sc(CC)c2c1nc(s2)	c1c(c2c(s1)c1c(n2CC)CCS1)
c1oc(cc1OC)	c1scc2c1c(c(cc2)N)	c1c(F)c(F)c(c2c1nn(CC)n2)
c1ccc(=O)c1	c1cc(c2c1[nH]cnc2)	c1sc2c(c1)C(=O)c1c2sc(c1)
c1cc(SC)sc1	C1=C(Oc2c(csc2)O1)	c1c2n(CC)c3c(c2ccc1)cccc3
c1oc2CCc2c1	C1=C(Cc2c(C1)nsn2)	c1oc2c(c1)C(=S)c1c2oc(c1)
c1c(C)cc(s1)	c1oc(c(c1O)C(=O)O)	c1c(c2c(s1)nc1c(c2CC)CCS1)
c(c(C)1)ccc1	c1cc(N(=O)=O)c(cc1)	c1cc2c(s1)c1c([nH]2)cc(s1)
c1sc(cc1)C#C	c1scc(N(O)O)c1N(O)O	c1c2c3CCSc3c3SCCc3c2c(cc1)
c1scc(OC)c1N	C1=c2ccccc2=C(C1=C)	c1c2[nH]cnc2c(c2c1[nH]cn2)
c1sc(C)c(c1)	c1sc(c2c1c(F)ccc2F)	c1[nH]c2nc3[nH]c(nc3nc2n1)
c1oc(cc1C=O)	c1sc(c(c1)C(F)(F)F)	c1sc(c2c1CCC[C@H]2N(=O)=O)
c1sccc1N(CC)	c(s1)c(CCN)c(CCN)c1	c(s1)c2c(=O)N(CC)c(=O)c2c1
c1oc(cc1C#N)	c1sc(c(c1C(=O)O)OC)	c1sccc1/C=C(/Cc1sc(CC)cc1)
c1occ(OC)c1N	c1c(C)cc(c(c1)C)C=C	c1sc(c2c1cc(C(=O)O)c(O)c2)

Table B5: List of SMILES for the 611 monomer data set. (Part 1 of 5)

c1[nH]c(cc1)	C1=[S]c2cscc2[S]=C1	c1n(C)c(=O)c2c1c(=O)n(C)c2
c1nc2CCNS2c1	C1=CC2=C(C1)C=C(C2)	c1csc2c1c(CC)c1c(c2CC)scc1
C=Cc1csc(c1)	c1sc(c2c1oc(C#N)c2)	c1[nH]cc2c1S(=O)(=O)C(=C2)
c1sc(O)c(c1)	c1sc2c(c1)[nH]c(c2)	c1sc(c2c1[nH]c(N(=O)=O)c2)
c1sc(cc1)C=C	c1c(OCC)c(cc(c1)OC)	n1c(C)c2c(c1=O)c(C)n(c2=O)
c1oc(c(c1C)C)	C=Cc1oc(c(c1O)O)C=C	c(s1)c2c(=O)n(CC)c(=O)c2c1
C1=CC=C(C1=O)	c1sc(c2c1cc(OC)cc2)	c1cc(CC)c(c(c1)CC)/C=C(/C)
c1scc(C#N)c1N	c1sc(c2c1c(S)ccc2O)	c1[nH]c(=O)c2c1c(=O)[nH]c2
c1c(CC)sc(c1)	C1=CC2=C(C1)N=C(C2)	c1sc(c(c1C(F)(F)F)C(F)(F)F)
c1scc(C)c1C=C	c1sc(c2c1c(C)ccc2C)	c1sc(c2c1C(=O)c1ccccc1C2=O)
C1=CC=C(C1=S)	C1=C(C(=C[S@@]1O)O)	c(c(CC)c1)c(c(=C)cn)c(CC)c1
c1sc(c(c1F)F)	c1sc(OC)c(C(=O)C)c1	c1sc(c2c1[C@H](S)CC[C@H]2O)
c1c(C)Nc(C)c1	c1sc(c2c1sc(C#N)c2)	C1=[S]C(=O)C2=C1C(=O)[S]=C2
c1c(CC)cc(s1)	c1sc(c2c1CCC[C@H]2O)	c(s1)cc(c(=O)2)c1c(s3)c2cc3
c1c(nc2cSc12)	c1sc2c(c3ccoc3cc2c1)	c1sc(c2c1[C@H](C)CC[C@H]2C)
c1c(C)sc(C)c1	c1sc2cc(C(=O)O)sc2c1	c1c(=O)oc2c1c(cc1cc(o)oc21)
c1n(CC)c(cc1)	c1sc(c2c1C(=O)CC2=O)	c1sc(c2c1c(C(F)(F)F)ccc2OC)
c1c(F)sc(F)c1	c1sc(c2c1SCC(=O)CO2)	c1c2c(sc(C(=O)OCC)c2F)c(s1)
c1cnc(cc1)C=C	c1sc(c2c1c(C=C)ccc2)	c1sc2c(c1)C(C)(C)c1c2sc(c1)
c1c(S)sc(O)c1	c1sc(c2c1cc(C=C)cc2)	c1ccc2c(c1)C(=O)c1c2ccc(c1)
c1oc(c(c1O)S)	N1c2cscc2N(CC(=O)C1)	c1sc(c2c1C[C@H](C(=O)C)CC2)
c1c(N)ccc(c1)	c1c2c(ncc(C)n2)c(s1)	c1sc(c2c1[C@H](C(=O)C)CCC2)
c1sc(c(c1)OC)	C1=C2CSC[C@H]2C(=C1)	c1sc(c2c1[C@H](F)CC[C@H]2F)
c1occ(C)c1C=C	c1c(F)c(F)c(c(c1F)F)	c(s1)cc(c(=S)2)c1c(s3)c2cc3
c1nc2cscc2nc1	c1c(OCC)cc(OCC)c(c1)	c1sc(c2c1[C@H](N)CC[C@H]2N)
C1SC=C1C(=O)C	c1sc(NC)c(c1C(=O)OC)	c1sc(cc1)c1ccc(s1)c1sc(cc1)
c1cscc1C(=O)O	C1=CC(=C2[C@H]1CCS2)	c1sc(c2c1cc(N(=O)=O)c(N)c2)
c1sc(c2c1CN2)	c1csc2c1cc(c1c2scc1)	c1sc(c2c1c(C(F)(F)F)ccc2C#N)
c1sc(c(c1C)C)	c1[nH]c2cc([nH]c2c1)	c1csc2c1c(CC)c(CC)c1c(csc21)
c1sc(c(c1O)S)	c1sc(c2c1C(=O)NC2=O)	c1sc(c2c1C[C@H](C(=O)OC)CC2)
c1sc(c2c1ocn2)	c1[nH]c(c(c1)C(=O)C)	c1n(CC)c(=O)c2c1c(=O)n(CC)c2
C1SC=C1C(=O)OC	c1sc(c(c1SCC)SCC)C=C	c1c2ccsc2cc2c1c(c1c(c2)scc1)
C=NNc1ccc(cc1)	c1sc(c2c1c(C=O)ccc2)	c1c2c3c(s1)ccc1c3c(cc2)c(s1)
c(s1)c2NCOc2c1	c1sc(c2c1c(C#N)ccc2)	c1cc2c(c3c1nsn3)cc(c1c2nsn1)
c1cc(CN)ccc1CN	c1c(CC)c(CC)c(CC)cc1	c1sc(c2c1CC[C@@H](C(=O)O)C2)
c1sc(C=O)c(c1)	c1sc(c2c1cc(C=O)cc2)	c1ccc(c2c1c1c(c3c2ccs3)scc1)
c(s1)c2scnc2c1	c1sc(c(c1OCC)OCC)C=C	c1c2c(ccc3c2ccs3)c2c(ccs2)c1
c1sc(c2c1ncs2)	c1sc(c2c1C(=O)OC2=O)	c1sc(c2c1[C@H](N)CC[C@H]2OC)
c(s1)c2ncsc2c1	C1S=C(/C(=N/N)/C1=S)	N1[C@@H]2CSC(=S)[C@H]2N(CC1)
C1=C(c2cscc12)	c1sc(c2c1cc(C#N)cc2)	C1C2CC(=C)O[C@H](C1C=CC2C)CCC
c1oc2cc(sc2c1)	c1[nH]c2[nH]c(nc2n1)	C1=C(C=C/C/1=C\1/C=CC(=C1)N)N
c1sc2nc(sc2c1)	c1sc(c2c1C(=O)SC2=O)	c1scc2c1C(=[S@@](OC)C(=C2)OC)
c1c(CC)c(ccc1)	c1sc2c(c1CC)sc(c2CC)	c1sc(c2c1CC[C@@H](N(=O)=O)C2)
c1sc(c(c1)C#N)	c1sc(c2c1SCC(=O)CS2)	C1=[S]C(=S)C2=C1C=C(S2(=O)=O)

Table B6: List of SMILES for the 611 monomer data set. (Part 2 of 5)

c(c1)cc(cc)cc1	c1csc2c1cc1c(ccs1)c2	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)
c1sc(c2c1scc2)	c1sc2c(n1)Cc1c2sc(c1)	c1sc(c2c1[C@H](C(F)(F)F)CCC2)
c1sc(c2c1scn2)	c1sc(c(c1N(=O)=O)C#N)	c1sc(c2c1[C@H](C)CC[C@H]2C=C)
c1cc(C)cc(c1C)	c1sc2c(c1)sc1cc(sc21)	c1cc2c(s1)c1c(C32OCCO3)cc(s1)
c1sc(c(c1O)OC)	c1sc(c2c1cc(S)c(O)c2)	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)
c1sc(c2c1CCC2)	c1sc(c2c1c(C=C)ccc2C)	c1sc(c2c1[C@H](OC)CC[C@H]2OC)
c1sc(c2c1SCS2)	c1sc(c2c1cc(C)c(C)c2)	c1sc(c2c1C[C@H](C(F)(F)F)CC2)
c1sc(c2c1occ2)	c1sc(c(c1OC)C(F)(F)F)	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)
c1sc(c(c1)C=O)	c1c2nc(C)c(C)nc2c(s1)	c1cc2c(s1)c1c(C2(CC)CC)cc(s1)
c1sc(c2c1OCO2)	c1sc2c(c1)C(=O)C(=C2)	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)
C1C([CH2])CCC1	c1cnc2c1c(=S)c1c2ncc1	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)
c1c(sc(OC)c1)N	c1csc2c1n(CC)c1c2scc1	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1
c1oc2cc(oc2c1)	c1scc2c1[C@@H](CCC2)N	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)
N1c2cscc2N(C1)	c1sc(c2c1C(=O)CCC2=O)	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)
c(nc1c2)cs1nc2	C1S[C@@H]2[C@@H]1NCC2	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)
c1c(OC)c(ccc1)	c1csc2c1C(S)=c1c2scc1	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)
c1sc(c2c1nco2)	c1cccc2c1n(c1c2cccc1)	c1c2c(ccs2)c(c2c1cc1c(c2)scc1)
c1sc(c(c1CC)CC)	c1oc(c(c1C(F)(F)F)OC)	c1sc(c2c1C[C@H](S)[C@@H](O)C2)
c1sc(CC)c(c1CC)	C1=[S]C(=C(O1)[N]C=O)	c1sc(c2c1C[C@H](C)[C@@H](C)C2)
c1sc(c2c1OCCS2)	C1SC[C@](C1)(C(=O)N)N	c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1
c1c2cscc2c(nn1)	c1scc2c1N(CCN2C(=O)O)	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)
c1[nH]c(cc1)C#C	c1scc2c1C=C(S2(=O)=O)	c1scc2c1C[C@@H]([C@@H](OC)C2)N
C#Cc1sc(cc1)C#C	c1sc(CC)c2c1c(c[nH]2)	c(c1)c2nccnc2c1c(c1)c2nccnc2c1
c1sc(c2c1ncnc2)	c1scc2c1c1c(s2)sc(c1)	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)
c1c2nsnc2c(cc1)	C1=[S]C(=C(S1)C(=O)C)	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)
c1c(C(=O)C)scc1	c1sc(c2c1nc(N)c(N)n2)	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)
c1c2nccnc2c(s1)	c1c(F)c(F)c(c2c1nsn2)	c1sc(c2c1C[C@H](F)[C@@H](F)C2)
c1c(OC)sc(OC)c1	n1c2cscc2n(c(=O)c1=O)	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)
c1c(OCC)c(ccc1)	N1[C@H]2CSC[C@H]2N(S1)	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)
c1sc(c(c1OC)OC)	c1c2c(ccs1)nc1c2sc(c1)	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)
c1cnc(c2c1nsn2)	c1sc(c2c1C[C@H](S)CC2)	C1=C(F)C=C/C/1=C\1/C(=CC(=C1)F)
N1c2cscc2N(CC1)	c1oc2c(C(=O)N)c(oc2c1)	c1scc2c1C[C@H](C)[C@H](C2)C(=C)
c1oc(cc1C(=O)C)	c(s1)c2ccs(=N)(=O)c2c1	c1sc(c2c1[C@H](C#N)CC[C@H]2C#N)
c1c(oc2cscc2s1)	c1sc(c2c1[C@H](C)CCC2)	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)
c1sc(c2c1OCCO2)	c1oc(c(c1C#N)C(F)(F)F)	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)
c1scc2c1C=C(C2)	C(=S)[CH]C1=[S]C=C(O1)	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)
c1c(ncc2c1non2)	c(o1)c2c(=O)Nc(=O)c2c1	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)
c1sc(c2c1CCCC2)	c1c(C(=O)C(F)(F)F)scc1	c1scc2c1[C@@H](CC[C@H]2N(=O)=O)N
c1c2nonc2c(cc1)	c1sc(c2c1c(C#N)ccc2OC)	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3
c1oc(cc1C(=O)O)	c1c2c(ncs2)c(c2c1ncs2)	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1
c1sc(c2c1OCSS2)	c1sc(c(c1C#N)C(F)(F)F)	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)
c1oc(c(c1OC)OC)	c1scc2c1C(=C[S@@]2[O])	c1sc(c2c1[C@H](C(=O)O)CC[C@H]2O)
c1sc(c2c1OSCS2)	c1sc(c2c1CCC[C@H]2C=O)	c(c(=[BH])1)ccc1c(c(=[BH])1)ccc1
c1sc(c2c1nccn2)	c1cc2c(cs1)c1c(ccs1)c2	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)

Table B7: List of SMILES for the 611 monomer data set. (Part 3 of 5)

c1sc(c(c1O)C#N)	c1sc(c(c1OCC)ON(=O)=O)	c1cc2c(s1)c1c(cc2)c2c(cc1)cc(s2)
c1sc(c2c1SCCS2)	c1c(C(F)(F)F)sc(C#N)c1	C1=CC(OC)=C(/C/1=C\1/C=C(OC)C=C1)
n1c(C)cc(c1C=C)	c1c(OCC)cc(CC)c(OCC)c1	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)
c1c2cccc(c2cs1)	c1sc2C3=C([CH]S3)Cc2c1	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)
c1sc(c2c1SCCCS2)	c1c([CH2])c(c2c1ccsc2)	C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)
c1cc2cc(o)cc2cc1	c1sc(c2c1cc(C)c(OC)c2)	c1sc(c2c1C[C@H](C#N)[C@@H](OC)C2)
c1sc(c2c1OCCCO2)	c(c(CC)1)ccc1c(c1)ccc1	c1sc(c2c1nc1c3cccnc3c3ncccc3c1n2)
c1cnc(c2c1nccn2)	c1c(OC)cc(c(c1)OCC)C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)
c1cc2c(s1)cc(s2)	C=C=C1CC2=C(C1)C(=CC2)	c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)
c1c2cS(F)cc2ccc1	c1sc(c2c1[nH]c(C#N)c2)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)
C1SCN2[C@H]1NCC2	c1oc(cc1C(=O)C(F)(F)F)	c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)
c1scc2c1c(ccc2O)	c(s1)c2C(NCC)OC(O)c2c1	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)
n1c2cscc2[nH]cc1	c1sc(c2c1[C@H](S)CCC2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)
c1c2c(OCN2)c(s1)	c1c(OCCC)c(cc(c1)OCCC)	c1sc(c2c1C[C@H](C(=O)C(F)(F)F)CC2)
N1C(=O)C=C(C1=O)	c1cc(cc2c1ccc1c2cccc1)	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)
c1scc(C(=O)O)c1S	NCc1cc(OCC)c(cc1OCC)CN	c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)
c1ccc(c2c1cccn2)	c1sc(c2c1cc(C(=O)C)cc2)	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)
c1sc(c(c1OC)C#N)	c1sc(c2c1c(C#N)ccc2C#N)	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)
c1sc(c2c1OSCCO2)	c1sc(c2c1S(=O)(=O)CCC2)	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)
c(s1)c(O)c(NN)c1	c1scc2c1CCC[C@H]2C(=O)O	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)
c1oc(c(c1OC)C#N)	c1sc(c2c1CCS(=O)(=O)C2)	c(s1)cc(c(=C(CN)(CN))2)c1c(s3)c2cc3
c1[nH]cc(c1C)C=C	c1ccc2c(c1)oc1c2ccc(c1)	C1=CC(C#N)=C/C/1=C/1\C(=CC(=C1)C#N)
c1coc(N(=O)=O)c1	c1sc(c2c1CC[C@@H](O)C2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2OC)
c1sc(c2c1cccc2F)	c1sc(c2c1sc(N(=O)=O)c2)	c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)
c1sc(C(=O)O)cc1O	c1cc2occoccoccoc2cc1	c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)
c1c(C#N)sc(OC)c1	c1sc(c2c1oc(=O)c(=O)s2)	C1=CC(N(=O)=O)C(C2C=CC=C2N(=O)=O)=C1
C(C(=O)1)C(=O)C1	C1=C(c2c3c(cccc13)ccc2)	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)
Nc1sc(c2c1nccn2)	N1CN[C@@H]2[C@H]1N(CN2)	c1sc(c2c1C[C@H](N)[C@@H](N(=O)=O)C2)
N1C(=S)C=C(C1=S)	c1sc(c2c1sc(=O)c(=O)s2)	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)
c1sc(C(=O)OC)cc1	c1sc(c2c1CC(=O)C(=O)C2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C#N)
C1=C(Cc2cscc2C1)	c1sc(c2c1nc(CN)c(CN)n2)	c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)
c1oc(cc1C(=O)OC)	c1sc(c2c1oc(N(=O)=O)c2)	c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)
c1sc(N(=O)=O)cc1	c1scc2c1nc(c(NC)n2)N(C)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)
c1sc(c(c1C=C)OC)	c1sc(c2c1CC[C@@H](N)C2)	c1c(C(=O)c2cscc2C(=O)c2ccccc2)ccc(c1)
c1scc(C(=O)OC)c1	c1sc(c2c1sc(C(=O)CC)c2)	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3
c1sc(c2c1OCCCC2)	c1c2nc(CC)c(CC)nc2c(s1)	c1nc2c(s1)[C@H]1[C@H](C2(CC)CC)N(CS1)
c1c(C)cc(c(c1)C)	c1c(OCC)cc(c(c1)OCC)C=C	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)
c1sc(c2c1OCCCS2)	c1[nH]c(cc1)C(C(=O)N)=C	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)
c1c2ccccc2c(c1=O)	c(s1)c2cc(OC)c(OC)cc2c1	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](OC)C2)
C1=C([CH]S1)C(=C)	c1sc(c2c1CC[C@@H](C)C2)	N1CN[C@@H]2N[C@H]3[C@@H](N[C@H]12)NCN3
c1sc(c(c1C#N)C#N)	c1sc(c2c1cc(OC)c(OC)c2)	c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)
c1occ(N(=O)=O)c1N	c1nc2c(c(=O)c3c2ncc3)c1	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)
c1sc(c2c1cc(C)s2)	c1sc(c2c1[C@H](OC)CCC2)	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C#N)C2)
c1c2C(=O)OCc2ccc1	c1sc(c2c1cc(C(=O)OC)s2)	N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)

Table B8: List of SMILES for the 611 monomer data set. (Part 4 of 5)

c1scc2c1sc1c2csc1	c1sc(c2c1oc(=O)c(=O)o2)	c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2
c1[nH]c(C=C)c(c1)	c1sc(c2c1c(C(=O)C)ccc2)	N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)
c1scc(N(=O)=O)c1N	C1=C2C(C(=O)O1)=C(OC2=O)	C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1
c1sc(C(F)(F)F)cc1	c1sc(c2c1sc(C(=O)CC)c2F)	C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1
c1c2c(cccc2)c(s1)	c1sc(c2c1sc(C(=O)OCC)c2)	c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)
c1oc(cc1C(F)(F)F)	C1=C2C(C(=O)S1)=C(OC2=O)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C(F)(F)F)
N(CC)c1cscc1N(CC)	c1sc(c2c1[C@H](C=C)CCC2)	N1[C@H]2[C@H](N[C@@H]3[C@@H](CSC3)N2)N(S1)
c1oc(c(c1C#N)C#N)	c1sc(c2c1[nH]c(=S)[nH]2)	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)
c1c2c(nccc2)c(s1)	c1sc(c2c1nc(OCC)c(CN)n2)	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)
C1=CC=C(S1(=O)=O)	c1sc(c(c1CC)OCCS(O)(O)O)	c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)
c1sc(c2c1[nH]cn2)	C1=C2C(C(=O)N1)=C(OC2=O)	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)
c1sc(c2c1cc[nH]2)	c1sc(c(c1)C(=O)C(F)(F)F)	c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)
Nc1sc(N(=O)=O)cc1	c1sc(c2c1nc(CCO)c(CN)n2)	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)
C(=C)c1sc(C)c(c1)	c1sc(c2c1[nH]c(=O)[nH]2)	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)
c1sc(c(c1OCC)OCC)	c1c2c(nccn2)c(c2c1nccn2)	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C(F)(F)F)C2)
c1sc(c2c1[nH]nc2)	c1ccc(OCC)c2c1c(OCC)ccc2	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)
c1scc2c1c(ccc2OC)	C1=C2C(C(=O)S1)=CN(C2=O)	C1=C[C@@H]2[C@H](C1)[C@@H]1[C@@H](C=CC1)C2=C
c1sc(c(c1)C(=O)C)	c1c2c(cccc2)cc2c1c(ccc2)	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)
c1sc(C(=O)O)c(c1)	c1sc(c2c1cc(N(=O)=O)cc2)	C1C(=O)O[C@@H]2[C@H]1CC[C@@H]1[C@H]2OC(=O)C1
c1c(C)c(c2c1ccco2)	c1sc2c(c1)c1c(scc1)c(c2)	C1C(=O)O[C@H]2[C@@H]1CC[C@H]1[C@@H]2OC(=O)C1
C1Oc2cscc2OC(C1=O)	c1csc2c1c(C)c(C)c1c2scc1	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)
c1sc(c2c1c(S)ccc2)	c1sc2c(c1)n(C)c1c2sc(c1)	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)
c1[nH]cc2c1[nH]cc2	c1sc(c2c1CC[C@@H](OC)C2)	C1C[C@@H]2[C@@H](CC1)N(CC)[C@@H]1[C@@H](S2)CCCC1
c1sc(c(c1CC)CC)C=C	c1ccc(c2c1nc(CC)c(CC)n2)	c(s1)cc(c(OC(CCC)CCC)2)c1c(OC(CCC)CCC)c(c3)c2sc3
c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)		
c1ccc2c3cc4n(CC)c5cc6c(cc5c4cc3Cc2c1)Cc1c6ccc(c1)		
c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)		
c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)		
N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1		
c1cc2c(cc1)c1c(=C2C(CN)CN)cc2c(c1)=C(C(CN)CN)c1c2ccc(c1)		
c(s1)c(CCCCCC)cc1c(s1)c2c(=O)n(C)c(=O)c2c1c(s1)c(CCCCCC)cc1		
c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)		
c1c(SC)c(SC)c(c2c1nc1c(n2)c2c(nc3cc(SC)c(SC)cc3n2)c2c1nc1cc(SC)c(SC)cc1n2)		
c1cc2c(=O)n(C)c(=O)c3c2c2c1C1=C4[C@@H](c2cc3)C=CC2=C4[C@@H](C(=O)N(C2=O)C)C(=C1)		
C1=C(C)C(C)=C(C2=N[C@H]3N4[C@@H](N[C@@H]5C[C@H](C)[C@H](C[C@@H]35)C)[C@H]3[C@H](N[C@@H]4[C@H]12)C[C@H]([C@H](C3)C)C)		

Table B9: List of SMILES for the 611 monomer data set. (Part 5 of 5)

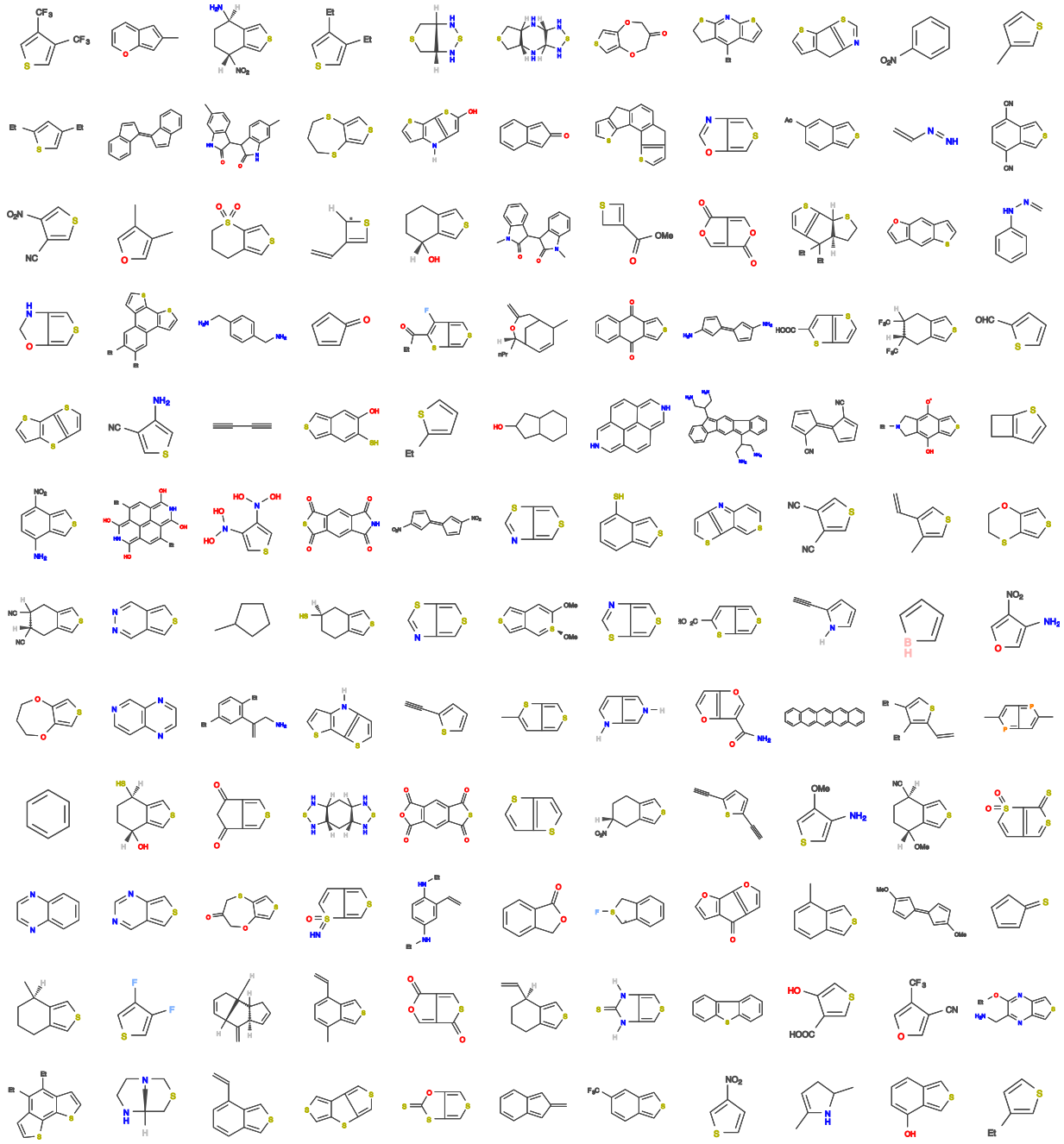


Figure B6: Molecules in the 611 monomer dataset (1 of 5).

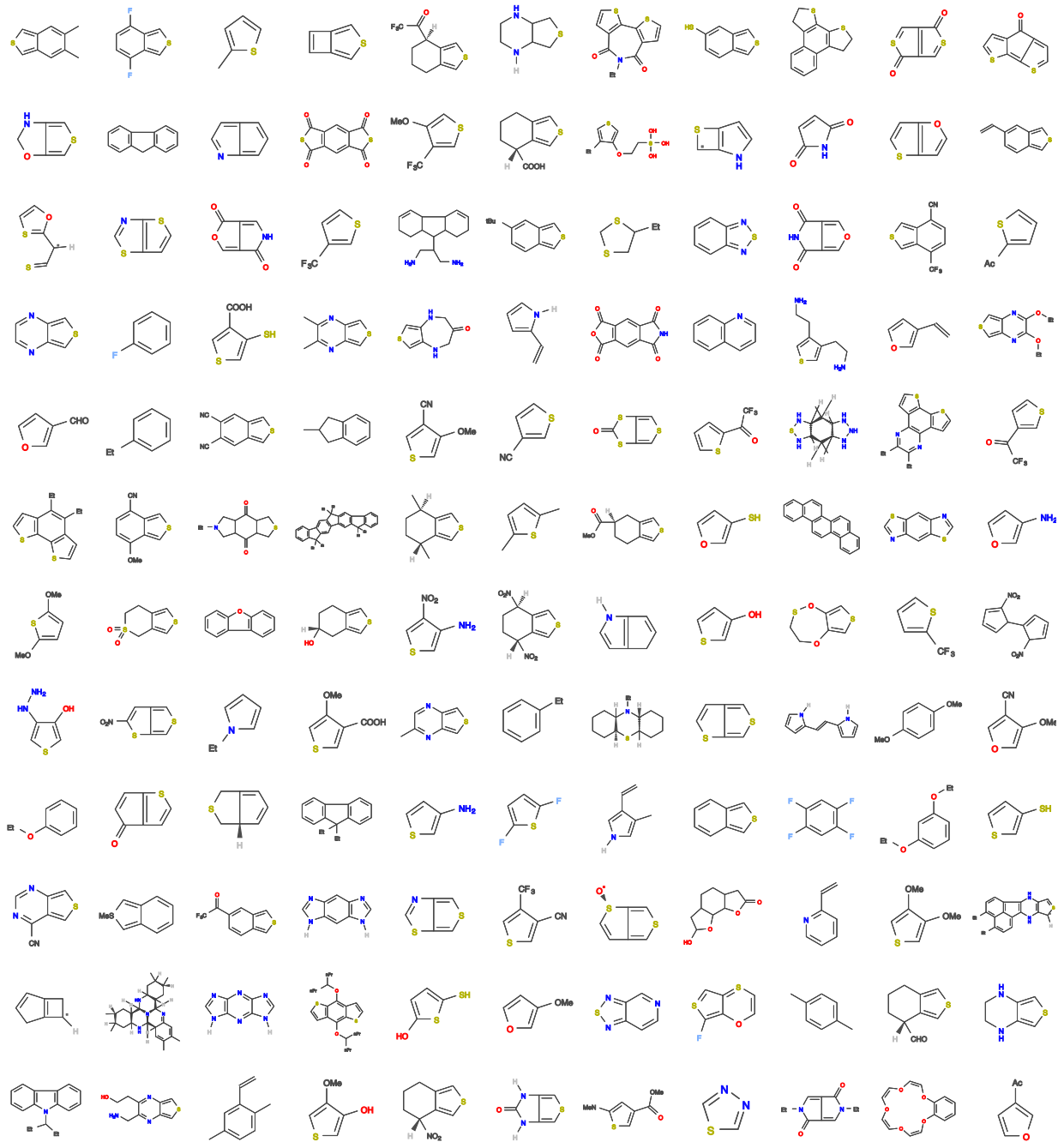


Figure B7: Molecules in the 611 monomer dataset (2 of 5).

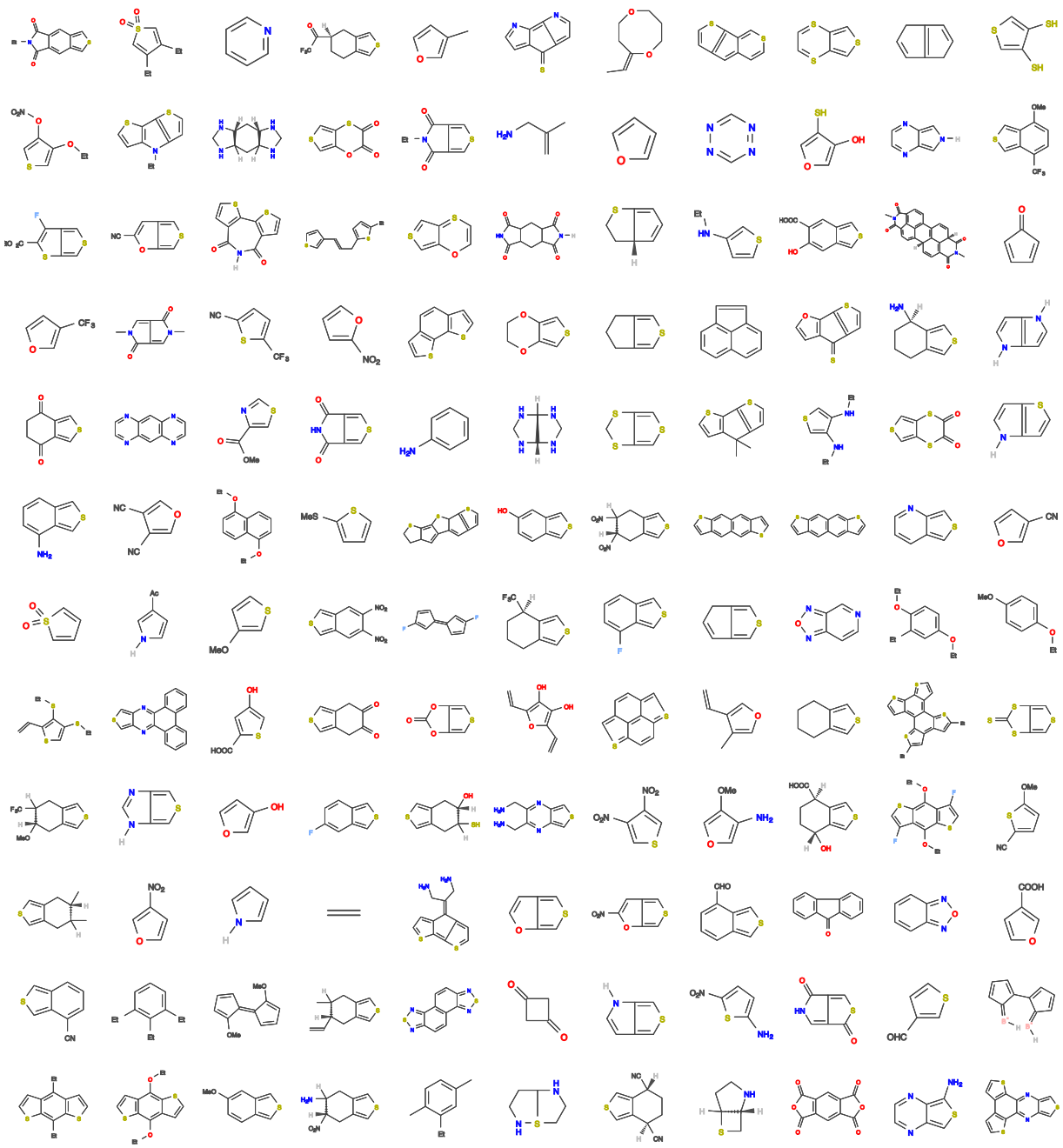


Figure B8: Molecules in the 611 monomer dataset (3 of 5).

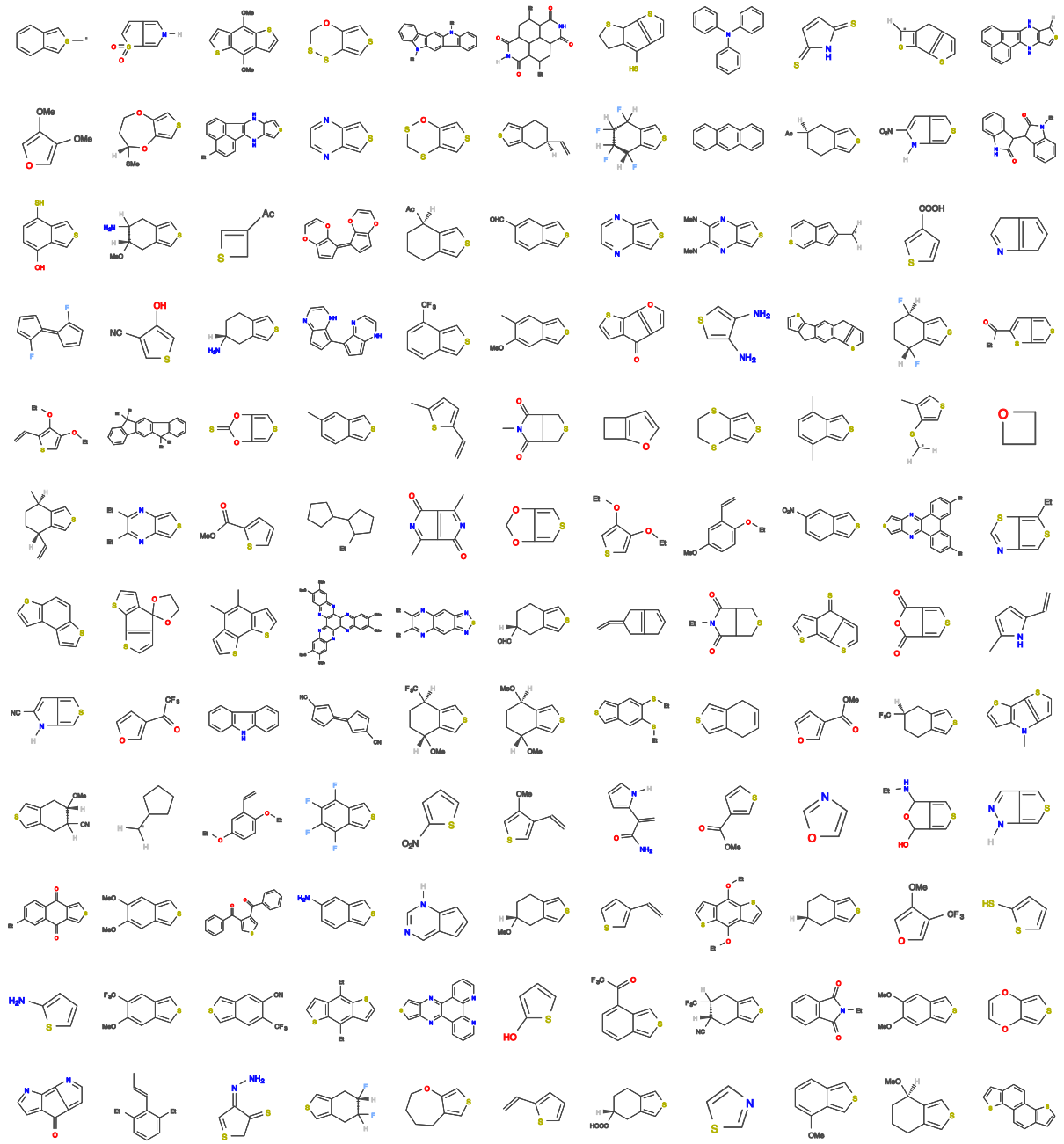


Figure B9: Molecules in the 611 monomer dataset (4 of 5).

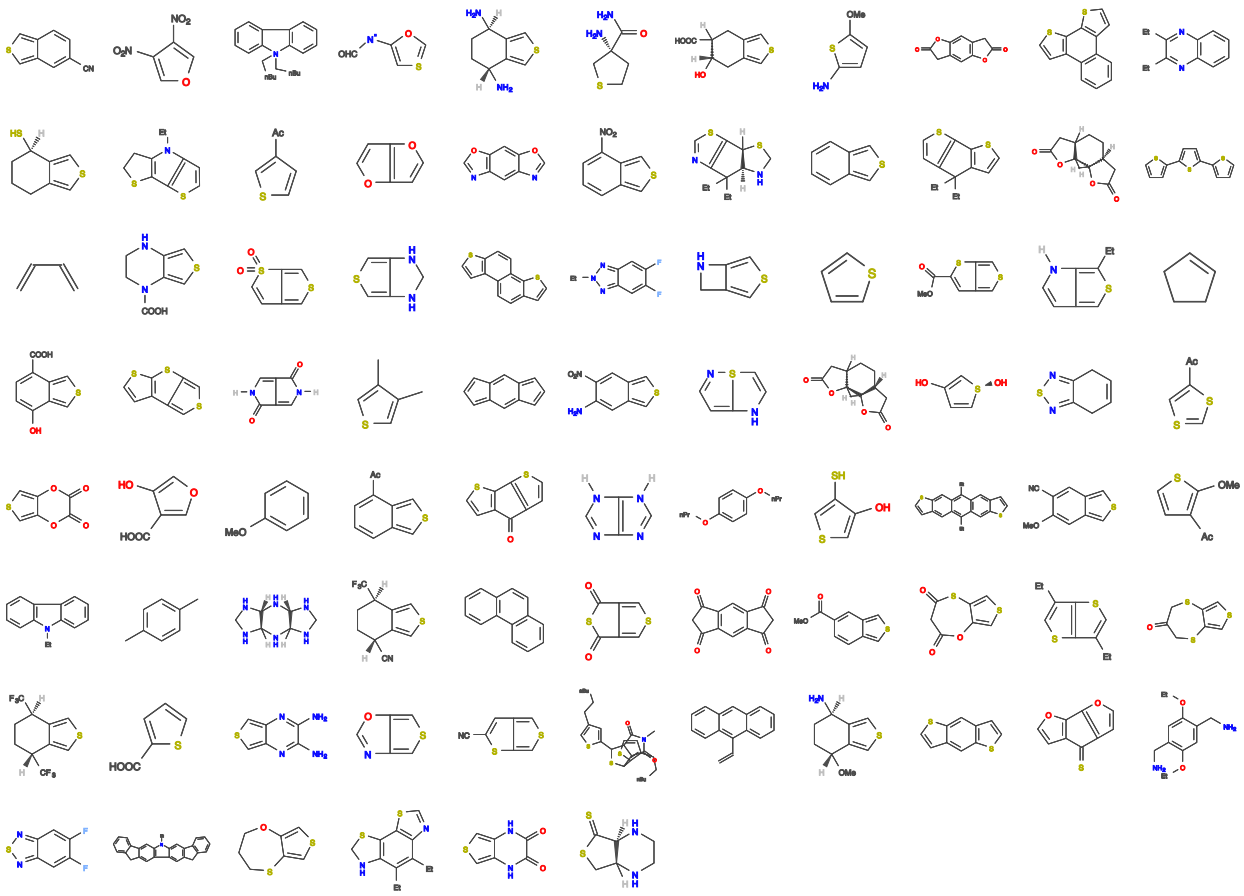


Figure B10: Molecules in the 611 monomer dataset (5 of 5).

c1cCcc1	C1=C(Cc2c(C1)nsn2)	c1c2c(ncs2)c(c2c1ncs2)C#C
c1occc1S	c1sc(c2c1c(S)ccc2)	c1sc(c2c1c(N(=O)=O)ccc2N)
c1occc1N	c1oc(c(c1O)C(=O)O)	c1c(OC)cc(c(c1)OCC)C=CC#C
c1sccc1O	c1sc(c2c1nccn2)C#C	N(CC)c1ccc(c(c1)C=C)N(CC)
c1sccc1N	c1ccc(c2c1nn(CC)n2)	c1oc2c(c1)C(=O)c1c2oc(c1)
c1sccc1S	c1cc(N(=O)=O)c(cc1)	c1cc(c2c(c1)c1c(s2)cccc1)
c1oc(nn1)	c1sc(c(c1OC)C#N)C#C	c1oc2c(c1)C(=S)c1c2oc(c1)
c1SCSc1cc	c(s1)c(O)c(NN)c1C#C	c1sc(c2c1cc(C(F)(F)F)cc2)
Nc1cscc1N	c1sc(c2c1cccc2F)C#C	c1cc2c(cc1)c1c(C2)cc(cc1)
Sc1cscc1S	c1oc(cc1C(=O)OC)C#C	c1c(OC)cc(c(c1)OCC(CC)CC)
c1sc(nn1)	C1SCN2[C@H]1NCC2C#C	c1sc(c2c1cc(C(C)(C)C)cc2)
c1oc(cc1)	c1scc(N(O)O)c1N(O)O	c1sc(c2c1nc(OCC)c(OCC)n2)
c1ccc(s1)	C1=C(Cc2cscc2C1)C#C	c1sc(c2c1c(C(F)(F)F)ccc2)
c1oc(nc1)	c1sc(c2c1c(C)ccc2C)	c1oc2c(C(=O)N)c(oc2c1)C#C
c1scc(n1)	c1sc(c2c1c(S)ccc2O)	c(s1)c2c(=O)n(C)c(=O)c2c1
c1ccc(cc1)	C1=c2ccccc2=C(C1=C)	c(c(CC)1)ccc1c(c1)ccc1C#C
c1cCcc1C#C	c1sc(c2c1c(F)ccc2F)	c1c(C(=O)C(F)(F)F)scc1C#C
c1occc1C=C	c1sc(c2c1SCCCS2)C#C	N1[C@H]2CSC[C@H]2N(S1)C#C
c1oc(cc1C)	N1C(=O)C=C(C1=O)C#C	C1=C(CC)C(CC)=C(S1(=O)=O)
c1nnc(nn1)	c1sc(c(c1C(=O)O)OC)	c1c2c(SCC2)c2c(CCS2)c1C#C
c1oc(cc1O)	c1coc(N(=O)=O)c1C#C	c1oc(c(c1C#N)C(F)(F)F)C#C
c1cc(cnc1)	c1sc(c(c1C=C)OC)C#C	c1sc2c(c1)C(=S)c1c2oc(c1)
c1sccc1SC#C	C=Cc1oc(c(c1O)O)C=C	c1oc(cc1C(=O)C(F)(F)F)C#C
c1ccc(=O)c1	cc(sc1n2)nc1sc2cC#C	c1sc(c2c1cc(C#N)c(C#N)c2)
c1c2CCc2sc1	c1cc2c(s1)cc(s2)C#C	c1sc(c(c1N(=O)=O)N(=O)=O)
c1sccc1NC#C	C1=[S]c2cscc2[S]=C1	c1cc2c(cs1)c1c(ccs1)c2C#C
c1c(F)cccc1	C1=CC2=C(C1)C=C(C2)	c(o1)c2c(=O)Nc(=O)c2c1C#C
c1occc1NC#C	c1sc(c2c1OSCCO2)C#C	c1oc2c(c1)C(=O)c1c2sc(c1)
c1oc(cc1OC)	c1ccc(c2c1cccn2)C#C	c1c(O)sc2c1[nH]c1c2sc(c1)
c1sccc1OC#C	c1sc(c2c1oc(C#N)c2)	c1sc(c2c1cc(SCC)c(SCC)c2)
c1oc2CCc2c1	c1sc2c(c1)[nH]c(c2)	c1sc(c2c1c(F)c(F)c(F)c2F)
c1c(C)cc(s1)	c1sc(c2c1OCCCO2)C#C	c(s1)c2ccs(=N)(=O)c2c1C#C
c(c(C)1)ccc1	c1scc(C(=O)OC)c1C#C	c1sc2c(c1)C(=O)c1c2sc(c1)
c1scc(OC)c1N	c1c(OCC)c(cc(c1)OC)	c1c2n(CC)c3c(c2ccc1)cccc3
c1SCSc1ccC#C	Nc1sc(c2c1nccn2)C#C	c1cc(cc2c1ccc1c2cccc1)C#C
c1oc(cc1C=O)	C1=CC2=C(C1)N=C(C2)	c1cc2c(C(=O)N(C2=O)CC)cc1
c1sccc1N(CC)	c1c(C#N)sc(OC)c1C#C	c1c(C(F)(F)F)sc(C#N)c1C#C
c1oc(cc1C#N)	C(C(=O)1)C(=O)C1C#C	c1oc(c(c1N(=O)=O)N(=O)=O)
c1occ(OC)c1N	c1sc(c2c1OCCCC2)C#C	c1c(c2c(s1)c1c(n2CC)CCS1)
c1ccc(s1)C#C	c1sc(c2c1cc(OC)cc2)	c1c(F)c(F)c(c2c1nn(CC)n2)

Table B10: List of SMILES for the 908 monomer data set. (Part 1 of 8)

c1oc(cc1)C#C	c1sc(c(c1)C(F)(F)F)	c1sc(c2c1c(C#N)ccc2OC)C#C
c1oc(nc1)C#C	c1sc(c2c1sc(C#N)c2)	c1sc(c2c1[nH]c(C#N)c2)C#C
c1sc(nn1)C#C	c1sc(c2c1OCCCS2)C#C	c1c(c2c(s1)nc1c(c2CC)CCS1)
C=Cc1csc(c1)	c1sc(C(=O)OC)cc1C#C	c1sc(c2c1oc(N(=O)=O)c2)C#C
c1oc(nn1)C#C	C1=C(C(=C[S@@]1O)O)	C1=C2OCCSC2=C([S@@]1NC)C#C
c1oc(c(c1C)C)	c1scc2c1c(ccc2O)C#C	c1scc2c1nc(c(NC)n2)N(C)C#C
C1=CC=C(C1=O)	c1cc2cc(o)cc2cc1C#C	c1sc(c2c1oc(=O)c(=O)o2)C#C
c1cc(cnc1)C#C	c1c2c(OCN2)c(s1)C#C	c1sc(c2c1sc(N(=O)=O)c2)C#C
c1scc(C#N)c1N	c1sc(C(F)(F)F)cc1C#C	c1cc2occoccoccoccoc2cc1C#C
c1c(CC)sc(c1)	c1c2c(ncc(C)n2)c(s1)	c1c2n(CCC)c3c(c2ccc1)cccc3
C1=CC=C(C1=S)	c1sc2c(c3ccoc3cc2c1)	c1cc2c(s1)c1c([nH]2)cc(s1)
c1sc(c(c1F)F)	c1sc2cc(C(=O)O)sc2c1	c1sc(c2c1cc(C(=O)OC)s2)C#C
c1c(C)Nc(C)c1	c1sc(c2c1C(=O)NC2=O)	c1sc(c2c1nc(CN)c(CN)n2)C#C
c1c(CC)cc(s1)	C1=CC=C(S1(=O)=O)C#C	c1sc(c2c1CCS(=O)(=O)C2)C#C
c1ccc(cc1)C#C	c1sc(c2c1C(=O)CC2=O)	c1c2c3CCSc3c3SCCc3c2c(cc1)
c1cscc1C(=O)O	c1sc(c2c1SCC(=O)CO2)	c1sc(c2c1oc(=O)c(=O)s2)C#C
c1c(nc2cSc12)	c1c2c(nccc2)c(s1)C#C	c1sc(c2c1c(C(=O)C)ccc2)C#C
c1nnc(nn1)C#C	C(=C)c1sc(C)c(c1)C#C	c(s1)c2cc(OC)c(OC)cc2c1C#C
c1oc(cc1O)C#C	c1sc(c2c1cc(C#N)cc2)	c1c2[nH]cnc2c(c2c1[nH]cn2)
c1oc(cc1C)C#C	c1scc(N(=O)=O)c1NC#C	c1[nH]c2nc3[nH]c(nc3nc2n1)
c1c(C)sc(C)c1	c1sc(c2c1cc[nH]2)C#C	c1sc(c2c1S(=O)(=O)CCC2)C#C
c1n(CC)c(cc1)	c1sc(c2c1cc(C=C)cc2)	c1sc(c2c1cc(C=C)c(C)c2)C#C
c1c(F)sc(F)c1	N1c2cscc2N(CC(=O)C1)	n1c(C)c2c(c1=O)c(C)n(c2=O)
c1cnc(cc1)C=C	c1sc(C(=O)O)c(c1)C#C	c1sc(c2c1sc(C(=O)CC)c2)C#C
c1c(S)sc(O)c1	c1scc2c1c(ccc2OC)C#C	c1sc(c2c1cc(C(=O)O)c(O)c2)
c1oc(c(c1O)S)	c1sc(c2c1c(C=C)ccc2)	c1n(C)c(=O)c2c1c(=O)n(C)c2
c1c(N)ccc(c1)	c1oc(c(c1C#N)C#N)C#C	c1sc(c2c1cc(OC)c(OC)c2)C#C
c1sc(c(c1)OC)	C1=C([CH]S1)C(=C)C#C	c1ccc2c(c1)oc1c2ccc(c1)C#C
c1occ(C)c1C=C	C1=C2CSC[C@H]2C(=C1)	c1sc(c2c1cc(C(=O)C)cc2)C#C
c1nc2cscc2nc1	c1c(F)c(F)c(c(c1F)F)	c1sc(c2c1[nH]c(N(=O)=O)c2)
C1SC=C1C(=O)C	c1c(OCC)cc(OCC)c(c1)	c1cc2c(s1)c1c(C2)cc(s1)C#C
c1scc(C)c1C=C	c1sc(c2c1cc(C=O)cc2)	c1sc(c2c1sc(=O)c(=O)s2)C#C
c1oc(c(c1F)F)	c1scc2c1C=C(C2=O)C#C	c(s1)c2c(=O)n(CC)c(=O)c2c1
c1sc(c2c1CN2)	c1occ(N(=O)=O)c1NC#C	c1c2c3c(s1)ccc1c3c(cc2)sc1
c1sc(c(c1C)C)	c1sc(c2c1[nH]cn2)C#C	c1cc(CC)c(c(c1)CC)/C=C(/C)
c1occc1C=CC#C	c1n(CC)nc(c1CC)CCC#C	c(s1)c2c(=O)N(CC)c(=O)c2c1
c1sc(c(c1O)S)	c1sc(c(c1SCC)SCC)C=C	c1sc(c2c1c(C#N)ccc2C#N)C#C
c1sc(c2c1ocn2)	c1sc(c2c1c(C=O)ccc2)	c1sccc1/C=C(/Cc1sc(CC)cc1)
C1SC=C1C(=O)OC	c1c(CC)c(CC)c(CC)cc1	c1sc(c(c1C(F)(F)F)C(F)(F)F)
C=NNc1ccc(cc1)	c1sc(NC)c(c1C(=O)OC)	c1sc(c2c1C(=O)c1ccccc1C2=O)

Table B11: List of SMILES for the 908 monomer data set. (Part 2 of 8)

c(s1)c2NCOc2c1	c1oc(cc1C(F)(F)F)C#C	C1=C2C(C(=O)S1)=CN(C2=O)C#C
c1cc(CN)ccc1CN	c1sc(c2c1c(C#N)ccc2)	c1sccc1/C=C(/Cc1ccc(CC)cc1)
c1oc(cc1OC)C#C	c1sc(c(c1OCC)OCC)C=C	c(c(CC)c1)c(c(=C)cn)c(CC)c1
c1sc(c2c1ncs2)	c1csc2c1cc(c1c2scc1)	c1sc(c2c1cc(N(=O)=O)cc2)C#C
c(s1)c2ncsc2c1	c1sc(c2c1[nH]nc2)C#C	c1c2c(nccn2)c(c2c1nccn2)C#C
c1c2CCc2sc1C#C	c1sc(c2c1SCC(=O)CS2)	c1sc(c2c1nc(CCO)c(CN)n2)C#C
c1oc2cc(sc2c1)	c1sc(c2c1C(=O)OC2=O)	C1=[S]C(=O)C2=C1C(=O)[S]=C2
c1c(CC)c(ccc1)	c1c(OCC)c(cc(c1)OCC)	c(s1)cc(c(=O)2)c1c(s3)c2cc3
c1sc(c(c1)C#N)	Nc1sc(N(=O)=O)cc1C#C	c1sc(c2c1nc(OCC)c(CN)n2)C#C
c(c1)cc(cc)cc1	C1S=C(/C(=N/N)/C1=S)	c1sc(c2c1c(C(F)(F)F)ccc2OC)
c1sc(c2c1scc2)	c1c2c(cccc2)c(s1)C#C	c1c(=O)oc2c1c(cc1cc(o)oc21)
c1oc2CCc2c1C#C	c1sc(c(c1O)C(F)(F)F)	c1ccc(c2c1nc(CC)c(CC)n2)C#C
c1cc(C)cc(c1C)	c1csc2c1cc1c(ccs1)c2	c1sc2c(c1)c1c(scc1)c(c2)C#C
c1sc(c(c1O)OC)	c1sc(c2c1cc(C)s2)C#C	c1sc(c2c1cc(C(=O)OC)cc2)C#C
c1sc(c2c1scn2)	c1sc(c2c1C(=O)SC2=O)	C1=C2C(C(=O)N1)=C(OC2=O)C#C
c1sc(c2c1CCC2)	c1sc2c(c1CC)sc(c2CC)	c1c2c(sc(C(=O)OCC)c2F)c(s1)
c1sc(c2c1SCS2)	c1sc(c(c1)C(=O)C)C#C	c1sc(c(c1)C(=O)C(F)(F)F)C#C
c(nc1c2)cs1nc2	c1sc(c2c1cc(F)c(F)c2)	c1sc2c(c1)n(C)c1c2sc(c1)C#C
c1sc(c2c1occ2)	c1sc2c(n1)Cc1c2sc(c1)	c1sc2c(c1)C(C)(C)c1c2sc(c1)
c1c(F)cccc1C#C	c1c2cscc2c(N)c(c1)C#C	c1sc(c2c1c(N(=O)=O)ccc2)C#C
c1sc(c(c1)C=O)	c1oc(c(c1)N(=O)=O)C#C	c1ccc2c(c1)C(=O)c1c2ccc(c1)
c1sc2nc(sc2c1)	c1sc(c(c1N(=O)=O)C#N)	c1sc(c2c1SC(=O)CC(=O)O2)C#C
c1sc(c2c1OCO2)	c1oc(c(c1O)C(=O)O)C#C	c1sc(c2c1[nH]c(=O)[nH]2)C#C
c1c(sc(OC)c1)N	c1c2c(ncc(CC)n2)c(s1)	c(c(N)1)ccc1c(c(N)1)ccc1C#C
c1oc2cc(oc2c1)	c1c(C)c(c2c1ccco2)C#C	c1c2c(cccc2)cc2c1c(ccc2)C#C
N1c2cscc2N(C1)	c1sc(c2c1cc(C)c(C)c2)	c1sc(c2c1c(C(=O)O)ccc2O)C#C
c1c(OC)c(ccc1)	c1sc2c(c1)sc1cc(sc21)	c1sc(c2c1sc(C(=O)CC)c2F)C#C
c1sc(c2c1nco2)	c1sc(c2c1cc(S)c(O)c2)	c1sc(c2c1[nH]c(=S)[nH]2)C#C
c1sc(c(c1CC)CC)	c1sc2c(c1)oc1c2sc(c1)	c1sc(c2c1sc(C(=O)OCC)c2)C#C
c1sc(CC)c(c1CC)	c1sc(c2c1c(C=C)ccc2C)	c1sc(cc1)c1ccc(s1)c1sc(cc1)
c1sc(c2c1OCCS2)	c1c(cc2c(c1)nccn2)C#C	c1csc2c1c(CC)c1c(n2)scc1C#C
c1c2cscc2c(nn1)	c1sc(c(c1OC)C(F)(F)F)	c1sc(c2c1c(C(=O)OC)ccc2)C#C
C=Cc1csc(c1)C#C	c1sc(c2c1oc(=O)o2)C#C	c1sc(c2c1cc(N(=O)=O)c(N)c2)
C#Cc1sc(cc1)C#C	C1=C(Cc2c(C1)nsn2)C#C	c1csc2c1c(C)c(C)c1c2scc1C#C
c1sc(c2c1ncnc2)	c1c2nc(C)c(C)nc2c(s1)	c1sc(c2c1cc(C#N)c(OC)c2)C#C
c1sccc1N(CC)C#C	c1sc(c2c1oc(=S)o2)C#C	c1sc(c2c1c(N(=O)=O)ccc2N)C#C
c1c(C(=O)C)scc1	c1sc(c(c1CC)CC)C=CC#C	c1sc(c2c1cc(C(F)(F)F)cc2)C#C
c1c2nccnc2c(s1)	c1sc(c(c1O)C(=O)O)C#C	c(s1)c2c(=O)n(C)c(=O)c2c1C#C
c1oc(cc1C#N)C#C	C1=C(Oc2c(csc2)O1)C#C	c1c(O)sc2c1[nH]c1c2sc(c1)C#C
c1c(OC)sc(OC)c1	c1sc2c(c1)C(=O)C(=C2)	c1sc(c2c1c(C(F)(F)F)ccc2C#N)

Table B12: List of SMILES for the 908 monomer data set. (Part 3 of 8)

c1sc(c(c1OC)OC)	c1sc(c2c1ncnc2C#N)C#C	c1csc2c1c(CC)c(CC)c1c(csc21)
c1cnc(c2c1nsn2)	c1sc(c2c1c(C)ccc2)C#C	c1sc(c2c1cc(SCC)c(SCC)c2)C#C
N1c2cscc2N(CC1)	C1Oc2cscc2OC(C1=O)C#C	c1n(CC)c(=O)c2c1c(=O)n(CC)c2
c1oc(cc1C(=O)C)	c1cnc2c1c(=S)c1c2ncc1	c1sc(c2c1cc(C#N)c(C#N)c2)C#C
c1oc(cc1C=O)C#C	c1csc2c1n(CC)c1c2scc1	c1oc2c(c1)C(=O)c1c2oc(c1)C#C
c1sc(c2c1OCCO2)	c1sc(c2c1sc(=O)s2)C#C	c1c2c3c(s1)ccc1c3c(cc2)c(s1)
c(c(C)1)ccc1C#C	c1sc(c2c1c(S)ccc2)C#C	c1sc(c2c1c(C(F)(F)F)ccc2)C#C
c1scc2c1C=C(C2)	c(cc1)c2cC(C)cc2c1C#C	c1oc2c(c1)C(=S)c1c2oc(c1)C#C
c1c(ncc2c1non2)	c1sc(c2c1sc(=S)o2)C#C	c1sc(c2c1cc(C(C)(C)C)cc2)C#C
c1sc(c2c1CCCC2)	c1sc(c2c1C(=O)CCC2=O)	c1cc2c(c3c1nsn3)cc(c1c2nsn1)
c1c(C)cc(s1)C#C	c1scc2c1C=C(S2(=O)=O)	c1sc(c2c1c(F)c(F)c(F)c2F)C#C
c1sc(c2c1OCSS2)	c1sc(c2c1cc(C)cc2)C#C	c1cc(c2c(c1)c1c(s2)cccc1)C#C
c1oc(c(c1OC)OC)	c1sc(c(c1)N(=O)=O)C#C	c1cc2c(cc1)c1c(C2)cc(cc1)C#C
c1sc(c2c1OSCS2)	C#Cc1[nH]c(cc1)C#CC#C	N(CC)c1ccc(c(c1)C=C)N(CC)C#C
c1sc(c2c1nccn2)	c1c(OC)c(cc(c1)OC)C#C	c1ccc(c2c1c1c(c3c2ccs3)scc1)
c1oc(cc1C(=O)O)	c1c(C(F)(F)F)sc(OC)c1	C1=C(CC)C(CC)=C(S1(=O)=O)C#C
c1c(OCC)c(ccc1)	c1sc(c2c1ccc(F)c2)C#C	c1c2c(ccc3c2ccs3)c2c(ccs2)c1
c1sc(c2c1SCCS2)	c1cccc2c1n(c1c2cccc1)	c1c(OC)cc(c(c1)OCC(CC)CC)C#C
n1c(C)cc(c1C=C)	c(cc1)c2cs(c)cc2c1C#C	c1sc(c(c1N(=O)=O)N(=O)=O)C#C
c1c2nonc2c(cc1)	c(oc1c(F)s2)csc1c2C#C	c1csc2c1c(CC)c1c(c2CC)scc1C#C
c1c2cccc(c2cs1)	c1sc(c2c1cc(S)cc2)C#C	c1c2[nH]cnc2c(c2c1[nH]cn2)C#C
c1sc(c(c1O)C#N)	c1scc2c1c(c(cc2)O)C#C	C1C2CC(=C)O[C@H](C1C=CC2C)CCC
c1c2nsnc2c(cc1)	c1scc2c1N(CCN2C(=O)O)	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)
c1c(oc2cscc2s1)	c1oc(c(c1C(F)(F)F)OC)	C1=[S]C(=S)C2=C1C=C(S2(=O)=O)
c1c(CC)cc(s1)C#C	c(s(c)1)c2ccccc2c1C#C	c1cc2c(s1)c1c([nH]2)cc(s1)C#C
c1cc2c(s1)cc(s2)	c1scc2c1c(c(cc2)N)C#C	c1c2c3c(s1)ccc1c3c(cc2)sc1C#C
c1cscc1C(=O)OC#C	c1c(C)cc(CC)c(C)c1C#C	n1c(C)c2c(c1=O)c(C)n(c2=O)C#C
c1sc(c2c1CN2)C#C	c1sc(CC)c2c1c(c[nH]2)	c1sc(c2c1cc(C(=O)O)c(O)c2)C#C
c1c(S)sc(O)c1C#C	c1scc2c1c1c(s2)sc(c1)	c1cc2c(s1)c1c(C2(CC)CC)cc(s1)
c1occ(C)c1C=CC#C	C1=[S]C(=C(S1)C(=O)C)	c(s1)c2c(=O)N(CC)c(=O)c2c1C#C
c1sc(c2c1OCCCO2)	c1sc(c(n1)C(=O)OC)C#C	c1sc(c2c1[nH]c(N(=O)=O)c2)C#C
c1scc(C#N)c1NC#C	c1sc(c2c1nc(N)c(N)n2)	c1cc2c(s1)c1c(C32OCCO3)cc(s1)
c1c(CC)sc(c1)C#C	c1sc(c(c1C)S[CH2])C#C	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)
c1c2cS(F)cc2ccc1	c1c(F)c(F)c(c2c1nsn2)	c1c2c3CCSc3c3SCCc3c2c(cc1)C#C
c1sc(c2c1OSCCO2)	c1sc(c2c1sc(=S)s2)C#C	c1n(C)c(=O)c2c1c(=O)n(C)c2C#C
C1SCN2[C@H]1NCC2	n1c2cscc2n(c(=O)c1=O)	c1c(c2c(s1)nc1c(c2CC)CCS1)C#C
c1scc2c1c(ccc2O)	N1[C@H]2CSC[C@H]2N(S1)	c(s1)c2c(=O)n(CC)c(=O)c2c1C#C
n1c2cscc2[nH]cc1	c1sc(c2c1cc(OC)cc2)C#C	c1[nH]c2nc3[nH]c(nc3nc2n1)C#C
c1c2c(OCN2)c(s1)	c1sc(c(c1)C(F)(F)F)C#C	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)
c1scc(C)c1C=CC#C	c1c2c(ccs1)nc1c2sc(c1)	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)

Table B13: List of SMILES for the 908 monomer data set. (Part 4 of 8)

c1scc(C(=O)O)c1S	c1sc(c2c1c(F)ccc2F)C#C	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)
c1oc(c(c1F)F)C#C	c1oc2c(C(=O)N)c(oc2c1)	c(s1)cc(c(=O)2)c1c(s3)c2cc3C#C
c1ccc(c2c1cccn2)	c1ccc(c2c1nn(CC)n2)C#C	c(s1)cc(c(=S)2)c1c(s3)c2cc3C#C
c1sc(c(c1OC)C#N)	c(s1)c2ccs(=N)(=O)c2c1	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1
c1sc(c(c1)OC)C#C	c1oc(c(c1C#N)C(F)(F)F)	c1sc(c2c1cc(N(=O)=O)c(N)c2)C#C
c1n(CC)c(cc1)C#C	C=Cc1oc(c(c1O)O)C=CC#C	c(c(CC)c1)c(c(=C)cn)c(CC)c1C#C
c1oc(c(c1O)S)C#C	c1c2c(SCC2)c2c(CCS2)c1	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)
c1sc(c2c1SCCCS2)	c(o1)c2c(=O)Nc(=O)c2c1	c1c(=O)oc2c1c(cc1cc(o)oc21)C#C
c1sc(C(=O)O)cc1O	c1c(C(=O)C(F)(F)F)scc1	c1c2c(ccs2)c(c2c1cc1c(c2)scc1)
c1c(N)ccc(c1)C#C	c1sc(c2c1c(C#N)ccc2OC)	c(c1)c2nccnc2c1c(c1)c2nccnc2c1
c1c(F)sc(F)c1C#C	c1cc2c(cs1)c1c(ccs1)c2	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)
c1coc(N(=O)=O)c1	c1sc(c(c1C#N)C(F)(F)F)	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)
c1nc2cscc2nc1C#C	c1sc2c(c1)[nH]c(c2)C#C	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)
c1sc(C(=O)OC)cc1	c1sc(c(c1OCC)ON(=O)=O)	c1sc(c2c1C(=O)c1ccccc1C2=O)C#C
c1sc(c2c1cccc2F)	c1cc(N(=O)=O)c(cc1)C#C	c1sc(c(c1C(F)(F)F)C(F)(F)F)C#C
c1c(C#N)sc(OC)c1	c1c(OCC)c(cc(c1)OC)C#C	c1ccc2c(c1)C(=O)c1c2ccc(c1)C#C
c1cnc(c2c1nccn2)	c1c(C)cc(c(c1)C)C=CC#C	c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1
C(C(=O)1)C(=O)C1	c1c(C(F)(F)F)sc(C#N)c1	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)
c1cnc(cc1)C=CC#C	c(c(CC)1)ccc1c(c1)ccc1	c1sc(c2c1c(C(F)(F)F)ccc2OC)C#C
c1cc2cc(o)cc2cc1	c1c(OCC)cc(CC)c(OCC)c1	C1=[S]C(=O)C2=C1C(=O)[S]=C2C#C
Nc1sc(c2c1nccn2)	c1sc(c2c1sc(C#N)c2)C#C	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)
c(s1)c(O)c(NN)c1	C1=CC2=C(C1)N=C(C2)C#C	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)
C1=C(Cc2cscc2C1)	c1sc(c2c1[nH]c(C#N)c2)	c1c2c(sc(C(=O)OCC)c2F)c(s1)C#C
c1oc(cc1C(=O)OC)	c1c([CH2])c(c2c1ccsc2)	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)
c1sc(c(c1C=C)OC)	c1sc(c2c1cc(C)c(OC)c2)	c1sc(cc1)c1ccc(s1)c1sc(cc1)C#C
c1scc(C(=O)OC)c1	C1=CC2=C(C1)C=C(C2)C#C	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)
c1c(C)sc(C)c1C#C	c1c(OC)cc(c(c1)OCC)C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)
c1sc(c2c1OCCCC2)	C=C=C1CC2=C(C1)C(=CC2)	c1csc2c1c(CC)c(CC)c1c(csc21)C#C
c1sc(c(c1O)S)C#C	c1sc(c2c1oc(C#N)c2)C#C	c1sc(c2c1c(C(F)(F)F)ccc2C#N)C#C
c1sc(c(c1F)F)C#C	c1oc(cc1C(=O)C(F)(F)F)	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)
c1oc(c(c1OC)C#N)	c1c2c(ncs2)c(c2c1ncs2)	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)
C1=CC=C(C1=O)C#C	c1scc2c1C(=C[S@@]2[O])	c1scc2c1nc1c3cc(sc3c3sccc3c1n2)
N1C(=O)C=C(C1=O)	c(s1)c2C(NCC)OC(O)c2c1	c1n(CC)c(=O)c2c1c(=O)n(CC)c2C#C
C1=CC=C(C1=S)C#C	C1=c2ccccc2=C(C1=C)C#C	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)
c1oc(c(c1C)C)C#C	c1sc(c(c1C(=O)O)OC)C#C	c1cc2c(c3c1nsn3)cc(c1c2nsn1)C#C
c1c(C)cc(c(c1)C)	c1sc(c2c1c(S)ccc2O)C#C	C1C2CC(=C)O[C@H](C1C=CC2C)CCCC#C
c1sc(c2c1OCCCS2)	c1c(OCCC)c(cc(c1)OCCC)	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)C#C
c1c(CC)c(ccc1)C#C	c1cc(cc2c1ccc1c2cccc1)	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)C#C
c1c2ccccc2c(c1=O)	c1sc(c2c1c(C)ccc2C)C#C	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1
c(s1)c2NCOc2c1C#C	c1c(OCC)c(OCC)c(OCC)cc1	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3

Table B14: List of SMILES for the 908 monomer data set. (Part 5 of 8)

c(s1)c2ncsc2c1C#C	c1cc2c(s1)c1c(C2)cc(s1)	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)
c1occ(N(=O)=O)c1N	c1sc(c2c1cc(C(=O)C)cc2)	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)
c1sc(c2c1OCO2)C#C	c1sc(c2c1c(C#N)ccc2C#N)	c1sc2c(ccc3c2ccc2c3ccc3c2scc3)c1
c1sc(c(c1O)OC)C#C	c1sc(c2c1S(=O)(=O)CCC2)	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)C#C
c1c2C(=O)OCc2ccc1	c1sc(c2c1c(C#N)ccc2)C#C	c(c1)c2nccnc2c1c(c1)c2nccnc2c1C#C
c1scc2c1sc1c2csc1	c1sc2c(c3ccoc3cc2c1)C#C	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)C#C
c1sc(c2c1SCS2)C#C	c1sc(c2=CC=C(c12)[CH2])	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)C#C
c1sc(c(c1)C#N)C#C	c1sc(c2c1CCS(=O)(=O)C2)	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1C#C
c1sc(c2c1nco2)C#C	c1sc(c2c1cc(C=O)cc2)C#C	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)C#C
c1c2c(nccc2)c(s1)	c1sc(c2c1C(=O)CC2=O)C#C	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)C#C
c1oc2cc(oc2c1)C#C	c1sc(c2c1cc(C#N)cc2)C#C	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)C#C
c1scc(N(=O)=O)c1N	c1ccc2c(c1)oc1c2ccc(c1)	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)C#C
c1sc(C(F)(F)F)cc1	c1c(OCC)c(cc(c1)OCC)C#C	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)
c1c2c(cccc2)c(s1)	c1csc2c1cc(c1c2scc1)C#C	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)
c1sc(c2c1scn2)C#C	c1sc(c(c1SCC)SCC)C=CC#C	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)C#C
c1scc2c1C=C(C2=O)	N1c2cscc2N(CC(=O)C1)C#C	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)C#C
c1sc(c2c1occ2)C#C	c1c2nc(CC)c(CC)nc2c(s1)	c1sc(c2c1nc1c3cccnc3c3ncccc3c1n2)
c(c1)cc(cc)cc1C#C	C1=C(c2c3c(cccc13)ccc2)	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)C#C
c1oc(cc1C(F)(F)F)	c1sc(c2c1cc(C=C)c(C)c2)	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)C#C
N(CC)c1cscc1N(CC)	C1=C2OCCSC2=C([S@@]1NC)	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)C#C
c1oc(c(c1C#N)C#N)	c1sc(c2c1sc(=O)c(=O)s2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)
C1C([CH2])CCC1C#C	c1c(CC)c(CC)c(CC)cc1C#C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)
c1sc(c2c1scc2)C#C	c1sc(c2c1c(C=C)ccc2)C#C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)
c1sc(c(c1)C=O)C#C	c1sc(c2c1nc(CN)c(CN)n2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)
c1sc(C(=O)O)c(c1)	c1c(F)c(F)c(c(c1F)F)C#C	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)C#C
c1sc(c2c1[nH]cn2)	c1sc(c2c1oc(N(=O)=O)c2)	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)C#C
C1=CC=C(S1(=O)=O)	c1sc(c2c1SCC(=O)CO2)C#C	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)
N1c2cscc2N(C1)C#C	c1sc(c2c1C(=O)OC2=O)C#C	c1scc2c1nc1c3cc(sc3c3sccc3c1n2)C#C
c1sc(c2c1cc[nH]2)	c1sc(c2c1cc(C=C)cc2)C#C	c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)
Nc1sc(N(=O)=O)cc1	c1scc2c1nc(c(NC)n2)N(C)	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)
c1sc(c(c1C#N)C#N)	c1sc(c2c1sc(C(=O)CC)c2)	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3C#C
c1oc2cc(sc2c1)C#C	c1sc2c(c1CC)sc(c2CC)C#C	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)
c1cc(C)cc(c1C)C#C	c1sc(NC)c(c1C(=O)OC)C#C	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)C#C
c1sc(c2c1CCC2)C#C	c1sc(c2c1cc(OC)c(OC)c2)	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)C#C
c1sc2nc(sc2c1)C#C	c1c(OCC)cc(c(c1)OCC)C=C	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)
c1sc(c2c1ncs2)C#C	c(s1)c2cc(OC)c(OC)cc2c1	c(s1)cc(c(=C(CN)(CN))2)c1c(s3)c2cc3
C(=C)c1sc(C)c(c1)	c1csc2c1cc1c(ccs1)c2C#C	c1sc2c(c1)C([C](C#N)C#N)=c1c2sc(c1)
c1c(OC)c(ccc1)C#C	c1sc(c2c1oc(=O)c(=O)s2)	C1=CC(C#N)=C/C/1=C/1\C(=CC(=C1)C#N)
c1n(CC)nc(c1CC)CC	c1sc(c2c1C(=O)NC2=O)C#C	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1C#C
C=NNc1ccc(cc1)C#C	c1nc2c(c(=O)c3c2ncc3)c1	c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)

Table B15: List of SMILES for the 908 monomer data set. (Part 6 of 8)

C1=C([CH]S1)C(=C)	c1sc(c2c1SCC(=O)CS2)C#C	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2
c1sc(c2c1ocn2)C#C	c1sc(c2c1cc(C(=O)OC)s2)	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)C#C
c1scc2c1c(ccc2OC)	c1sc(c2c1oc(=O)c(=O)o2)	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)
c1sc(c2c1[nH]nc2)	c1sc(c2c1c(C(=O)C)ccc2)	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)
c1sc(c(c1)C(=O)C)	c1sc(c2c1C(=O)SC2=O)C#C	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)C#C
c1sc(c2c1cc(C)s2)	c1sc(c2c1c(C=O)ccc2)C#C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)C#C
c1cc(CN)ccc1CNC#C	c1sc(c(c1O)C(F)(F)F)C#C	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)C#C
c1c(C)c(c2c1ccco2)	C1S=C(/C(=N/N)/C1=S)C#C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)C#C
c1sc(c(c1O)C#N)C#C	c1sc(c2c1sc(N(=O)=O)c2)	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)C#C
C1Oc2cscc2OC(C1=O)	c(c(N)1)ccc1c(c(N)1)ccc1	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)C#C
c1sc(c2c1OSCS2)C#C	C1=[S]C(=C(O1)[N]C=O)C#C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)
c1c2nccnc2c(s1)C#C	c1scc2c1C=C(S2(=O)=O)C#C	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3
c1sc(c2c1OCCS2)C#C	c1sc(c2c1sc(C(=O)CC)c2F)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)C#C
c1oc(cc1C(=O)C)C#C	c1c(C(F)(F)F)sc(OC)c1C#C	c(s1)cc(c2)c1c(s3)c2c(c4)c3c(s5)c4cc5
C#Cc1sc(cc1)C#CC#C	c1ccc(OCC)c2c1c(OCC)ccc2	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)C#C
c1sc(c2c1SCCS2)C#C	c1sc(c2c1cc(C)c(C)c2)C#C	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)
c1sc(c(n1)C(=O)OC)	c1sc(c2c1cc(S)c(O)c2)C#C	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)C#C
c1sc(c2c1ncnc2)C#C	c1sc(c2c1C(=O)CCC2=O)C#C	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)
c1sc(c(c1CC)CC)C=C	c1sc(c2c1sc(C(=O)OCC)c2)	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)C#C
c1c(cc2c(c1)nccn2)	c1csc2c1c(CC)c1c(n2)scc1	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)C#C
c1c(oc2cscc2s1)C#C	c1sc(c2c1nc(N)c(N)n2)C#C	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2C#C
c1sc(c2c1c(C)ccc2)	c1c2c(ncc(CC)n2)c(s1)C#C	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)
c1c2nsnc2c(cc1)C#C	c1sc(c2c1cc(F)c(F)c2)C#C	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3C#C
c1c2cccc(c2cs1)C#C	c1sc(c2c1c(C=C)ccc2C)C#C	c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2
N1c2cscc2N(CC1)C#C	c1sc(c2c1[nH]c(=S)[nH]2)	c1c(C(=O)c2cscc2C(=O)c2ccccc2)ccc(c1)C#C
c1sc(c(c1O)C(=O)O)	c1sc(c2c1nc(OCC)c(CN)n2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)C#C
c1sc(c2c1sc(=S)o2)	c1scc2c1c1c(s2)cc(s1)C#C	C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1
c1sc(c(c1)N(=O)=O)	c1c2c(nccn2)c(c2c1nccn2)	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)C#C
c1sc(c2c1OCCO2)C#C	c1sc(c(c1OC)C(F)(F)F)C#C	c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)
c1sc(c2c1sc(=O)s2)	c1sc(c(c1CC)OCCS(O)(O)O)	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)
c1c2nonc2c(cc1)C#C	C1=C2C(C(=O)N1)=C(OC2=O)	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)C#C
c1sc(CC)c(c1CC)C#C	c1ccc(c2c1nc(CC)c(CC)n2)	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)
c1c(OC)c(cc(c1)OC)	c1sc(c(c1)C(=O)C(F)(F)F)	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)
c1sc(c2c1ncnc2C#N)	c1c2nc(C)c(C)nc2c(s1)C#C	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)
c1c2cscc2c(nn1)C#C	c1sc(c2c1c(C(=O)OC)ccc2)	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)
c(cc1)c2cs(c)cc2c1	c1sc(c2c1[nH]c(=O)[nH]2)	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)
c(oc1c(F)s2)csc1c2	c1scc2c1N(CCN2C(=O)O)C#C	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)
c1sc(c(c1C)S[CH2])	c1sc(c2c1c(N(=O)=O)ccc2)	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)
c1sc(c2c1cc(S)cc2)	c1sc(c(c1N(=O)=O)C#N)C#C	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)C#C
c1c2cscc2c(N)c(c1)	C1=C2C(C(=O)S1)=CN(C2=O)	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)C#C

Table B16: List of SMILES for the 908 monomer data set. (Part 7 of 8)

c1scc2c1c(c(cc2)O)	c1c2c(cccc2)cc2c1c(ccc2)	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)C#C
c1sc(c(c1CC)CC)C#C	c1sc2c(c1)sc1cc(sc21)C#C	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)C#C
c1sc(c2c1oc(=O)o2)	c1sc(c2c1nc(CCO)c(CN)n2)	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)C#C
c1sc(c2c1sc(=S)s2)	c1c(F)c(F)c(c2c1nsn2)C#C	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)C#C
c1sc(c2c1ccc(F)c2)	C1=C2C(C(=O)O1)=C(OC2=O)	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)C#C
c1cnc(c2c1nsn2)C#C	c1oc(c(c1C(F)(F)F)OC)C#C	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)
c1oc(c(c1)N(=O)=O)	c1sc(c2c1cc(N(=O)=O)cc2)	c(s1)cc(c(OC(CCC)CCC)2)c1c(OC(CCC)CCC)c(c3)c2sc3
c(s(c)1)c2ccccc2c1	c1sc2c(c1)c1c(scc1)c(c2)	c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)
c1sc(CC)c2c1nc(s2)	c1csc2c1c(C)c(C)c1c2scc1	C1C[C@H]2[C@H](CC1)N(C[CH2])[C@H]1[C@H](S2)CCCC1
c1sc(c(c1OC)OC)C#C	c1sc2c(c1)n(C)c1c2sc(c1)	c1ccc2c3cc4n(CC)c5cc6c(cc5c4cc3Cc2c1)Cc1c6ccc(c1)
c1c(C(=O)C)scc1C#C	c1csc2c1n(CC)c1c2scc1C#C	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)C#C
c1c(ncc2c1non2)C#C	c1oc2c(n1)cc1c(c2)oc(n1)	C1C[C@H]2[C@H](CC1)N(C[CH2])[C@H]1[C@H](S2)CCCC1C#C
n1c(C)cc(c1C=C)C#C	c1sc(CC)c2c1c(c[nH]2)C#C	c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)
c1c(OC)sc(OC)c1C#C	c1sc(c2c1c(C(=O)O)ccc2O)	c(s1)cc(c(OC(CCC)CCC)2)c1c(OC(CCC)CCC)c(c3)c2sc3C#C
c1sc(c2c1oc(=S)o2)	C1=Cc2cc3=CC(=Cc3cc2=C1)	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)
c1sc(c2c1cc(C)cc2)	C1=[S]C(=C(S1)C(=O)C)C#C	c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)C#C
c1sc(c2c1OCSS2)C#C	c1sc2c(c1)C(=O)C(=C2)C#C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)C#C
c1scc2c1c(c(cc2)N)	c1sc(c2c1cc(C#N)c(OC)c2)	c1cc2c(cc1)c1c(=C2C(CN)CN)cc2c(c1)=C(C(CN)CN)c1c2ccc(c1)
c1scc2c1C=C(C2)C#C	c1sc(c2c1cc(C(=O)OC)cc2)	c(s1)c(CCCCCC)cc1c(s1)c2c(=O)n(C)c(=O)c2c1c(s1)c(CCCCCC)cc1
C1=C(Oc2c(csc2)O1)	c1sc(c2c1SC(=O)CC(=O)O2)	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)
c1c(C)cc(CC)c(C)c1	n1c2cscc2n(c(=O)c1=O)C#C	c1c2n(CC)c3c(cc4c(c3)c3c(C4)cccc3)c2cc2c1c1c(C2)cc2c(c1)C(CC)(CC)c1c2cccc1
c1cc(c2c1[nH]cnc2)	c1sc2c(c1)oc1c2sc(c1)C#C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)C#C
c1sc(c2c1CCCC2)C#C	c1c([CH2])c(c2c1ccsc2)C#C	c1c2c(ncs2)c(c2c1ncs2)C#C

Table B17: List of SMILES for the 908 monomer data set. (Part 8 of 8)

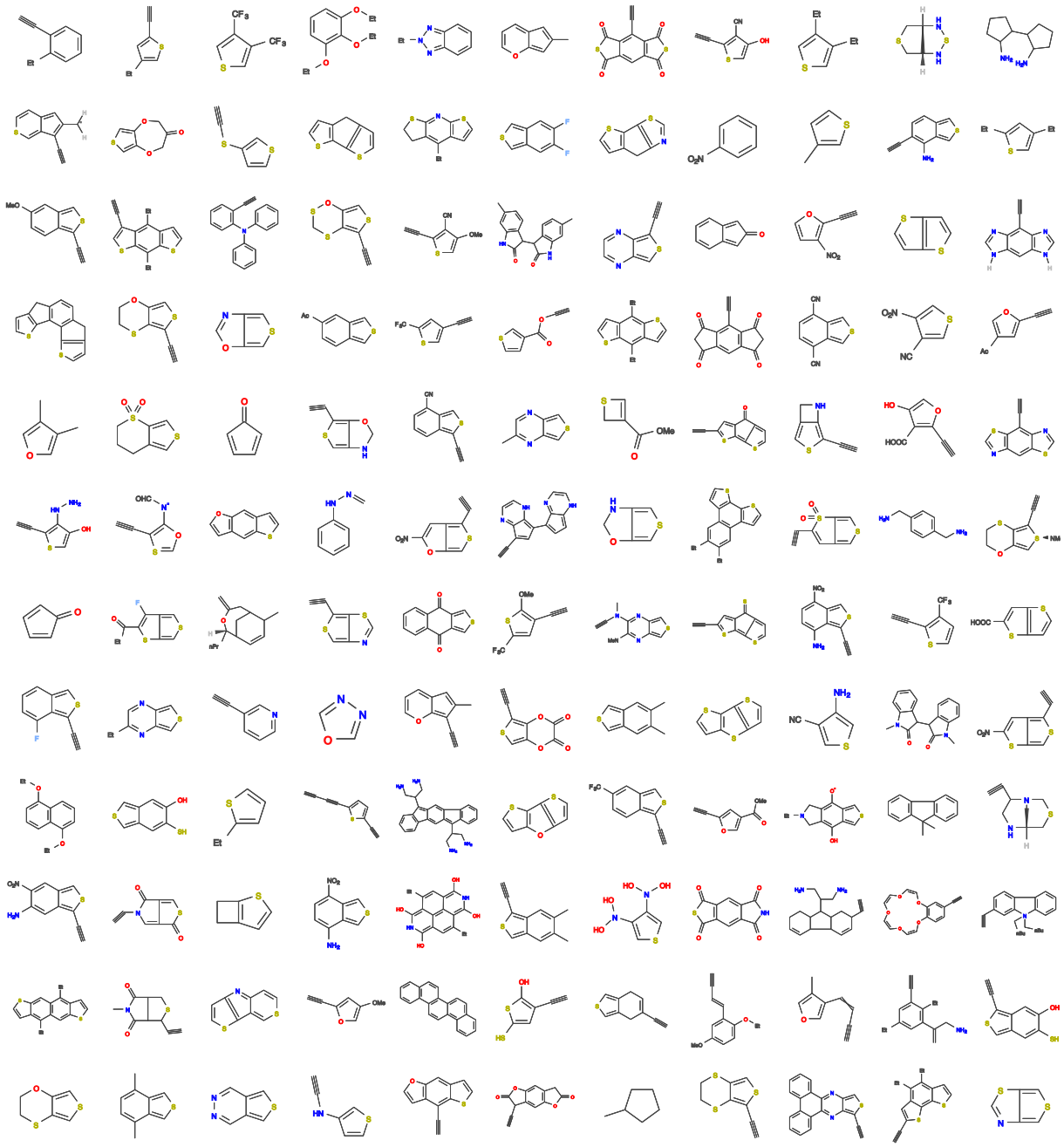


Figure B11: Molecules in the 908 monomer dataset (1 of 7).

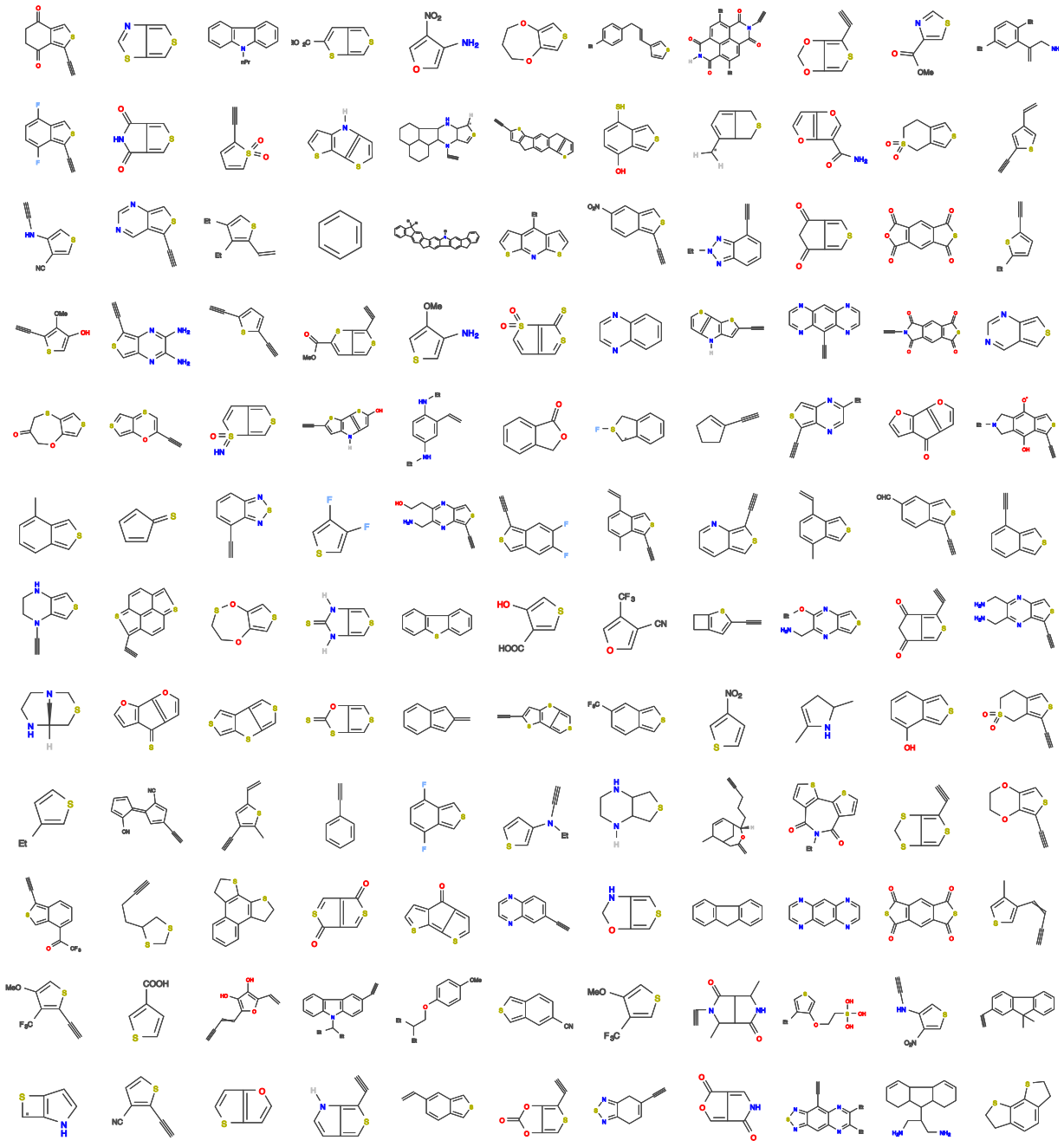


Figure B12: Molecules in the 908 monomer dataset (2 of 7).

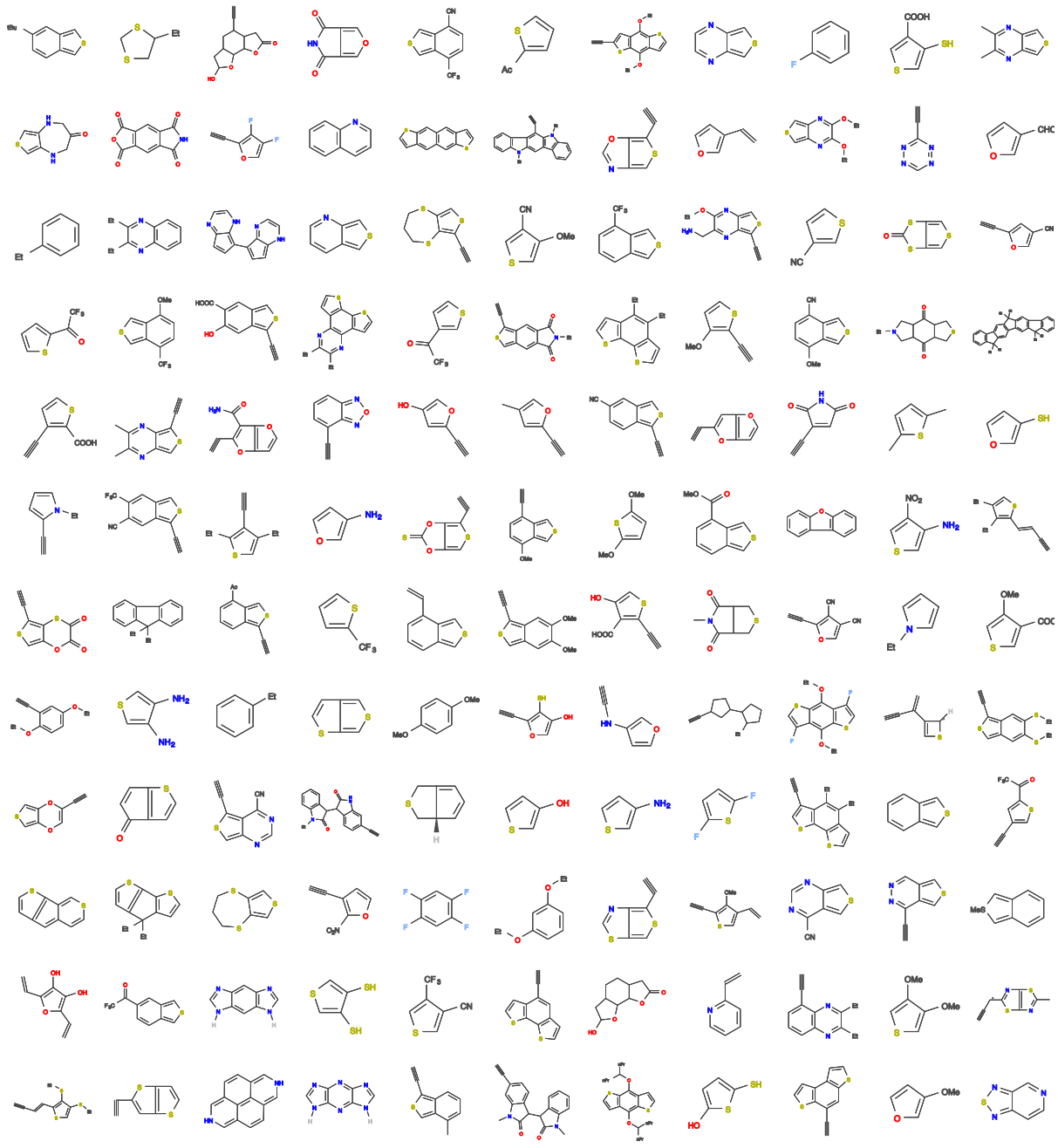


Figure B13: Molecules in the 908 monomer dataset (3 of 7).

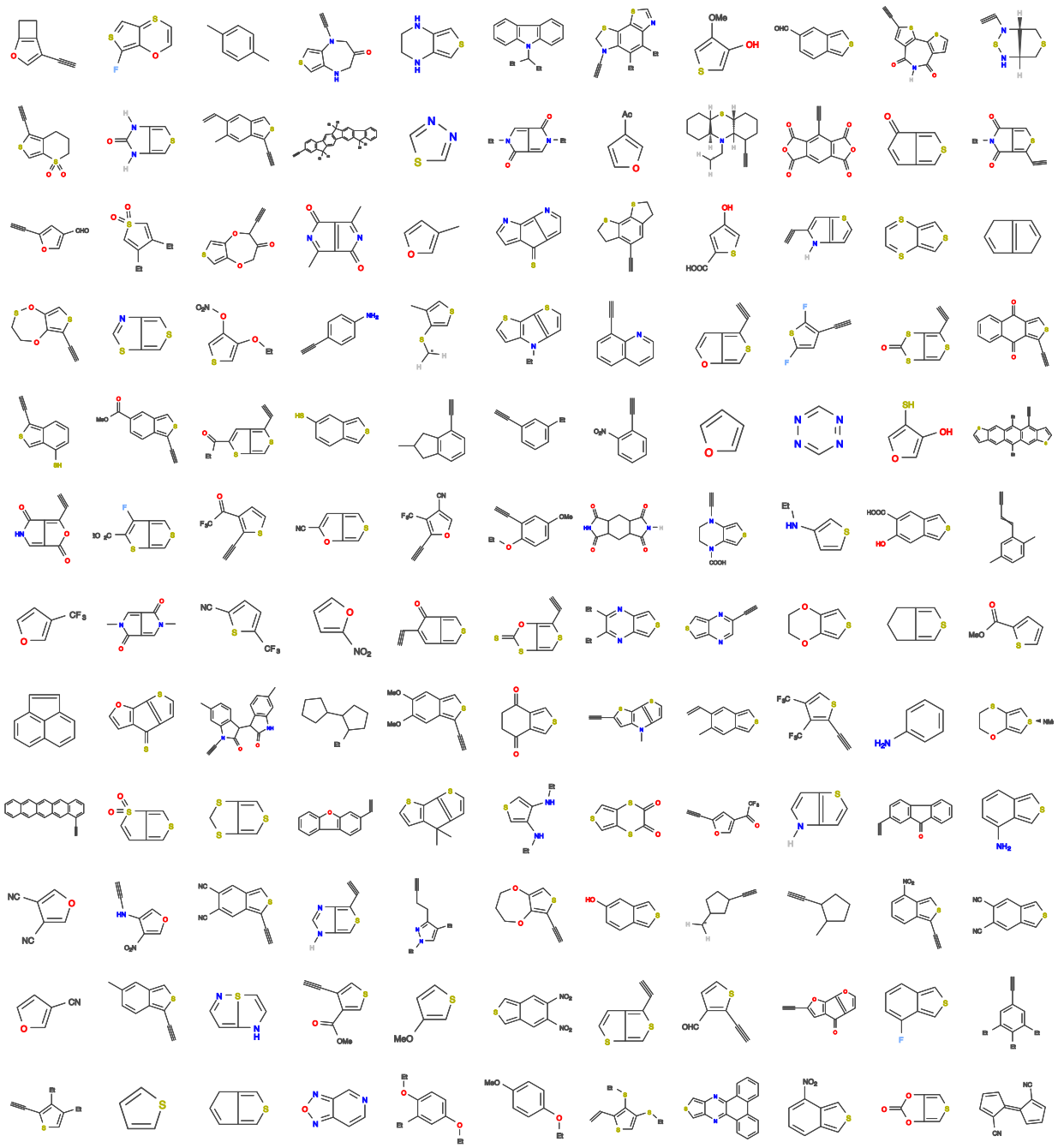


Figure B14: Molecules in the 908 monomer dataset (4 of 7).

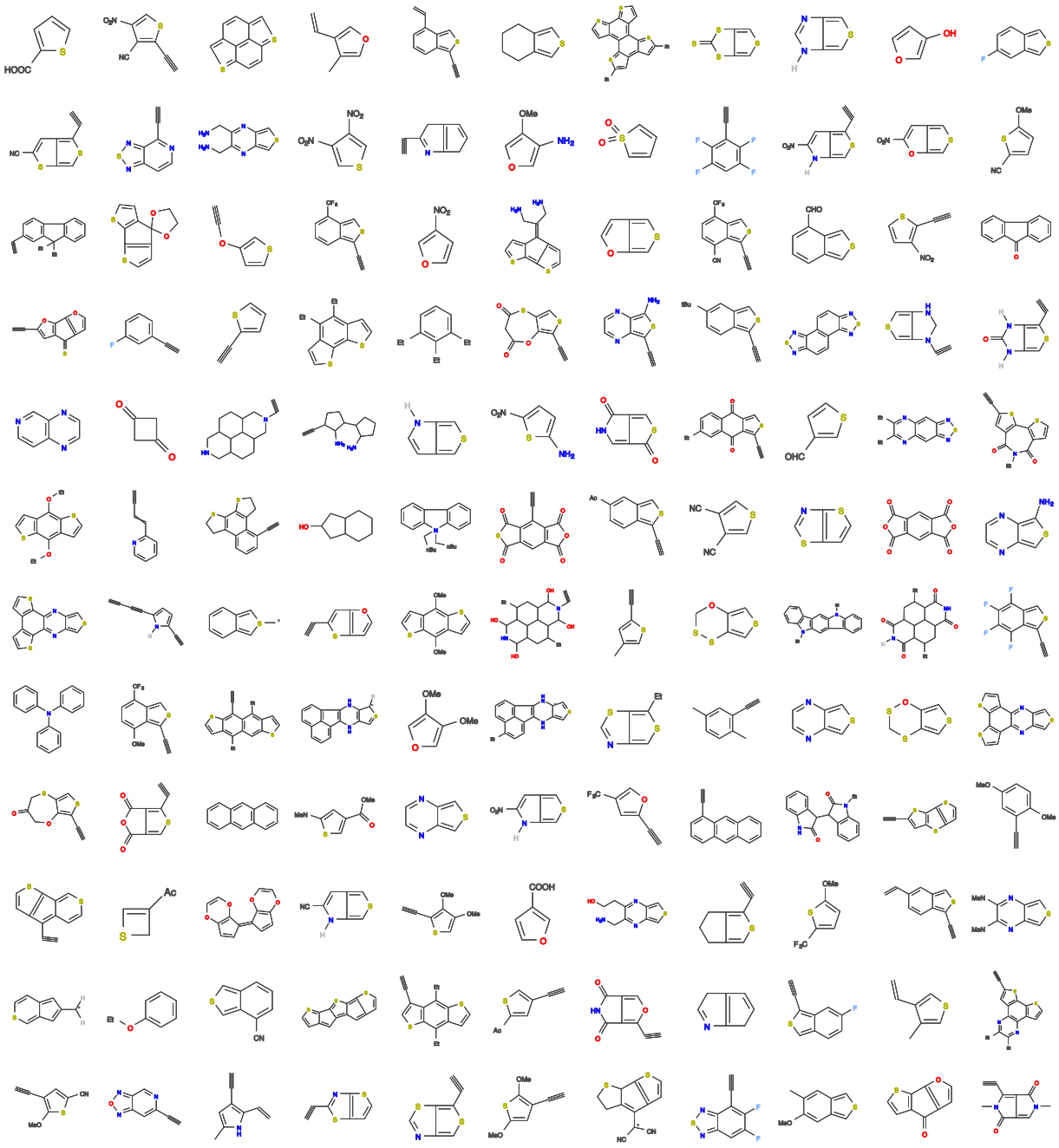


Figure B15: Molecules in the 908 monomer dataset (5 of 7).

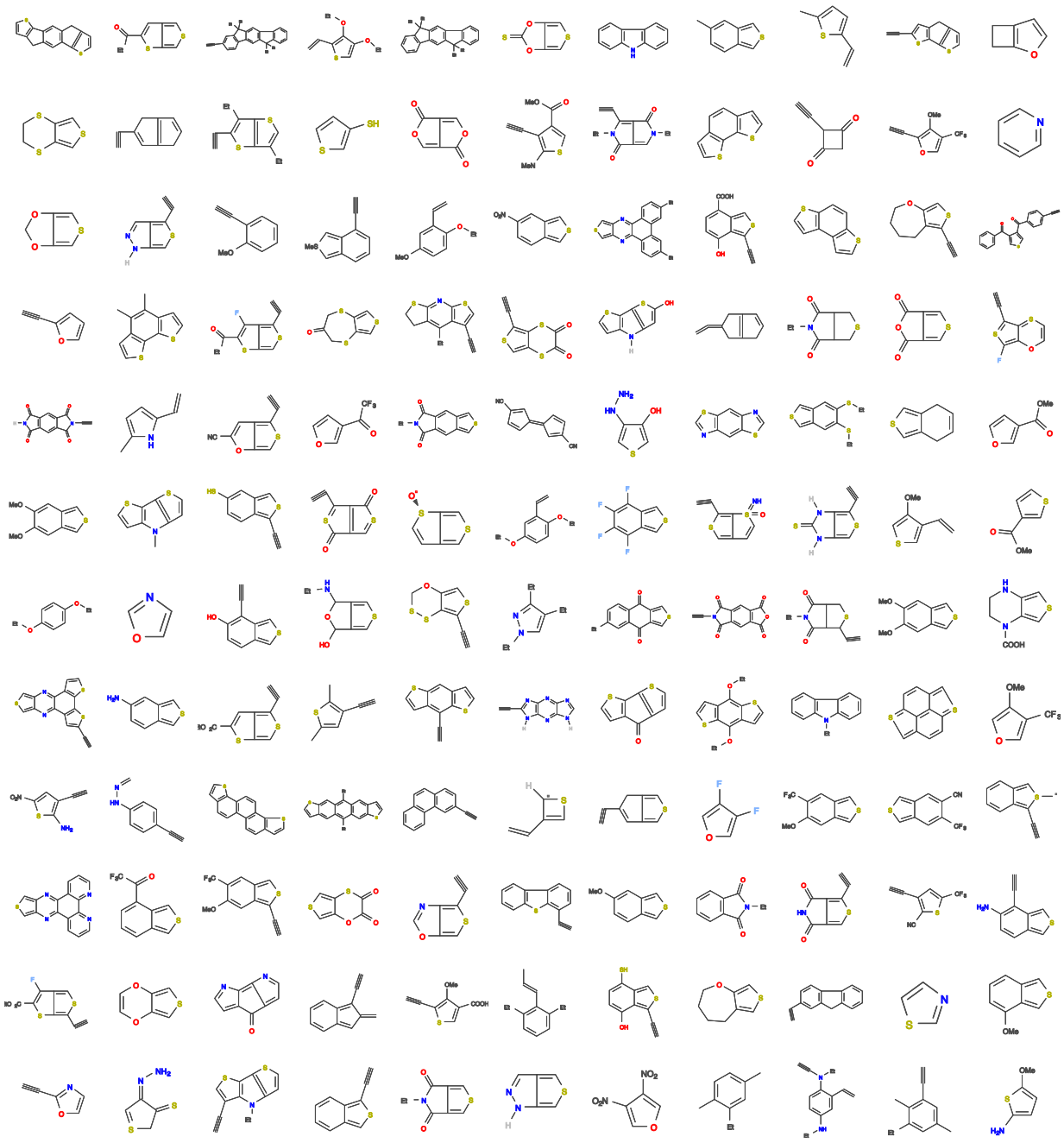


Figure B16: Molecules in the 908 monomer dataset (6 of 7).

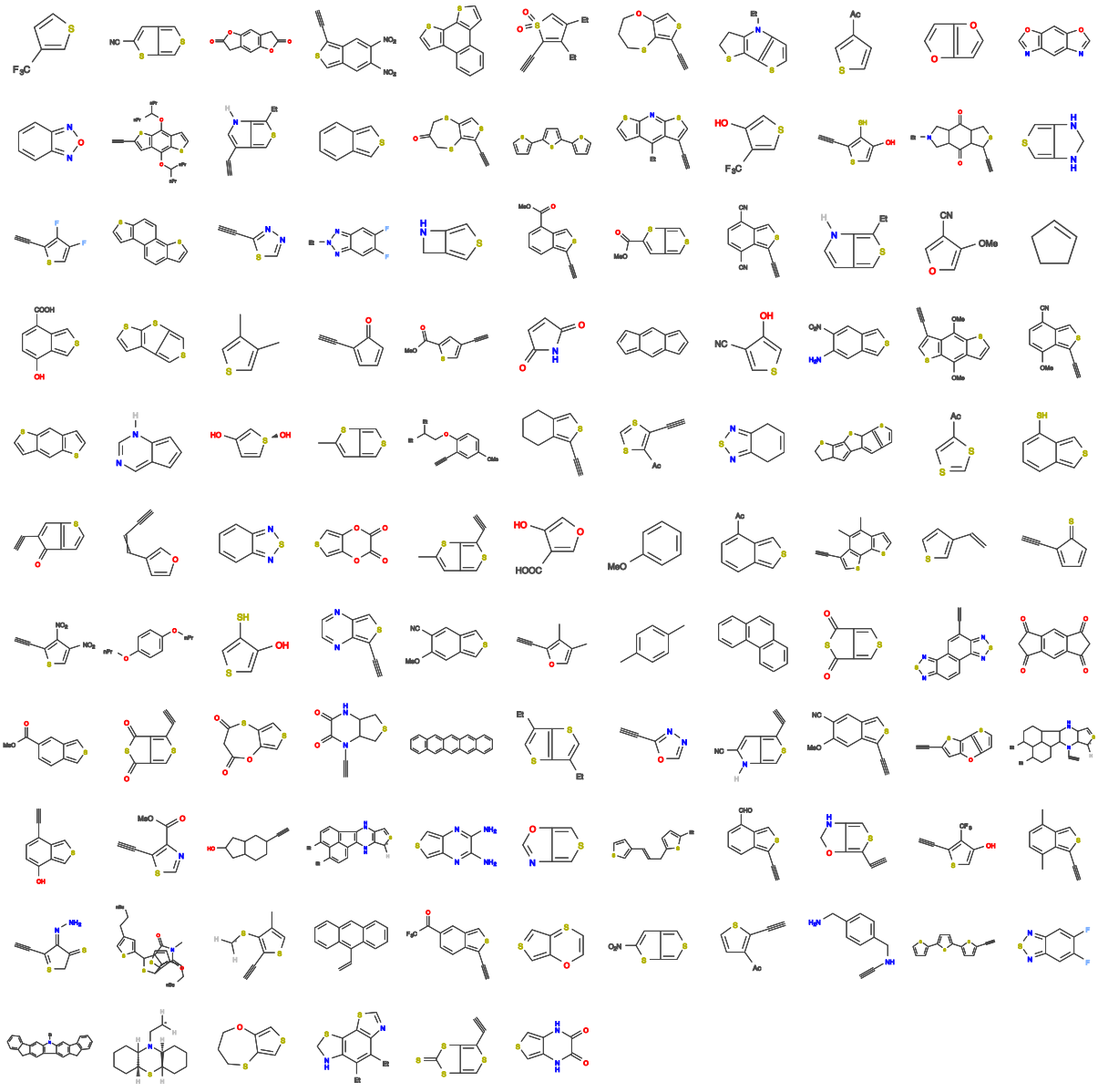


Figure B17: Molecules in the 908 monomer dataset (7 of 7).

C=C	C1=[S]c2cscc2[S]=C1	c1c(OCC)cc(c(c1)OCC)C=CC=C
C#CC#C	c1scc2c1c(ccc2O)C=C	c1sc(c2c1CC[C@@H](C)C2)C=C
C1OCC1	n1c2cscc2[nH]cc1C=C	c1sc(c2c1[nH]c(N(=O)=O)c2)
C=CC=C	C1=C(C(=C[S@@]1O)O)	c1sc(c2c1nc(CN)c(CN)n2)C=C
c1cCcc1	c1cc2cc(o)cc2cc1C=C	c1sc(c2c1sc(C(=O)CC)c2)C=C
C(=C)N=N	c1sc(c2c1sc(C#N)c2)	c1sc(c2c1CCC[C@H]2N(=O)=O)
c1occc1S	c1[nH]c(C=C)c(c1)C=C	n1c(C)c2c(c1=O)c(C)n(c2=O)
c1occc1N	C(=C)c1sc(C)c(c1)C=C	c(s1)c2c(=O)n(CC)c(=O)c2c1
c1sccc1O	c1c2C(=O)OCc2ccc1C=C	c1c2c3c(s1)ccc1c3c(cc2)sc1
c1sccc1S	c1sc(c2c1CCC[C@H]2O)	c1sc(c2c1cc(C(=O)O)c(O)c2)
c1SCSc1cc	c1sc2c(c3ccoc3cc2c1)	c1sc(c2c1cc(OC)c(OC)c2)C=C
Sc1cscc1S	c1n(CC)nc(c1CC)CCC=C	N1CN[C@@H]2[C@H]1N(CN2)C=C
C1OCC1C=C	c1sc2cc(C(=O)O)sc2c1	c1sc(c2c1c(C#N)ccc2C#N)C=C
c1ccc(s1)	N(CC)c1cscc1N(CC)C=C	c1sc(c2c1oc(=O)c(=O)o2)C=C
c1oc(nn1)	C1=CC=C(S1(=O)=O)C=C	c1sc(c(c1C(F)(F)F)C(F)(F)F)
c1oc(cc1)	c1[nH]c2cc([nH]c2c1)	c1sc(c2c1[C@H](C=C)CCC2)C=C
C=CC=CC=C	c1sc(c2c1C(=O)NC2=O)	C1=C2C(C(=O)S1)=CN(C2=O)C=C
c1sc(nn1)	c1sc(c2c1C(=O)CC2=O)	c1sc(c(c1)C(=O)C(F)(F)F)C=C
Nc1cscc1N	c1c(OCC)cc(OCC)c(c1)	c1sc(c2c1C(=O)c1ccccc1C2=O)
c1oc(nc1)	c1sc(c2c1SCC(=O)CO2)	c1sc2c(c1)n(C)c1c2sc(c1)C=C
c1scc(n1)	c1oc(c(c1C#N)C#N)C=C	C1=C2C(C(=O)O1)=C(OC2=O)C=C
C#CC#CC=C	c1occ(N(=O)=O)c1NC=C	c1c2c(nccn2)c(c2c1nccn2)C=C
c1[nH]cnc1	c1c2c(cccc2)c(s1)C=C	c(c(CC)c1)c(c(=C)cn)c(CC)c1
c1ccc(cc1)	c1sc(c2c1cc(C=C)cc2)	c1sc2c(c1)c1c(scc1)c(c2)C=C
c1cCcc1C=C	c1sc(c2c1c(C=C)ccc2)	C1=[S]C(=O)C2=C1C(=O)[S]=C2
c1occc1C=C	C1=C2CSC[C@H]2C(=C1)	c1sc2c(c1)C(C)(C)c1c2sc(c1)
c1cc(cnc1)	c1c2c(ncc(C)n2)c(s1)	C1=C2C(C(=O)N1)=C(OC2=O)C=C
c1oc(cc1C)	c1c(F)c(F)c(c(c1F)F)	c1sc(c2c1[C@H](C)CC[C@H]2C)
c1nnc(nn1)	c1sc(c(c1)C(=O)C)C=C	C1=Cc2cc3=CC(=Cc3cc2=C1)C=C
c1oc(cc1O)	c1sc(NC)c(c1C(=O)OC)	c1sc(c2c1cc(N(=O)=O)c(N)c2)
c1sc(cc1)S	c1scc2c1sc1c2csc1C=C	c1ccc(c2c1nc(CC)c(CC)n2)C=C
c1sc(cc1)N	C1=CC(=C2[C@H]1CCS2)	c1c(=O)oc2c1c(cc1cc(o)oc21)
c1ccc(=O)c1	c1sc(c2c1cc[nH]2)C=C	c1sc(c2c1cc(C#N)c(OC)c2)C=C
c1sccc1SC=C	c1[nH]c(c(c1)C(=O)C)	c1ccc2c(c1)C(=O)c1c2ccc(c1)
c1occc1SC=C	Nc1sc(N(=O)=O)cc1C=C	c1sc(c2c1[nH]c(=S)[nH]2)C=C
c1c2CCc2sc1	c1oc(cc1C(F)(F)F)C=C	c1c2c(sc(C(=O)OCC)c2F)c(s1)
c1c(F)cccc1	c1sc(c(c1SCC)SCC)C=C	c1sc(c2c1[C@H](C(=O)C)CCC2)
c1oc(cc1OC)	c1c2ccccc2c(c1=O)C=C	c1sc(c2c1C[C@H](C(=O)C)CC2)
c1occc1NC=C	c1sc(c2c1c(C#N)ccc2)	c1sc(c2c1[C@H](S)CC[C@H]2O)
C=C(C)CNC=C	c1c(CC)c(CC)c(CC)cc1	c1sc(c2c1nc(CCO)c(CN)n2)C=C
c1cc(SC)sc1	c1sc(c2c1cc(C=O)cc2)	c1sc(c2c1[C@H](F)CC[C@H]2F)
c1sccc1NC=C	c1scc(N(=O)=O)c1NC=C	c1sc(c2c1cc(N(=O)=O)cc2)C=C
c1sccc1OC=C	c1sc(c(c1OCC)OCC)C=C	c1sc(c2c1nc(OCC)c(CN)n2)C=C
c1oc2CCc2c1	N1c2cscc2N(CC(=O)C1)	c1sc(c2c1cc(C(=O)OC)cc2)C=C
C(=C)C#CC=C	c1csc2c1cc(c1c2scc1)	c1sc(c2c1SC(=O)CC(=O)O2)C=C

Table B18: List of SMILES for the 1235 monomer data set. (Part 1 of 10)

C(=C)N=NC=C	c1c2c(nccc2)c(s1)C=C	c(s1)cc(c(=S)2)c1c(s3)c2cc3
c1c(C)cc(s1)	c1sc(c2c1[nH]nc2)C=C	c1csc2c1c(C)c(C)c1c2scc1C=C
c(c(C)1)ccc1	c1sc(c2c1C(=O)OC2=O)	c1sc(c2c1c(C(=O)O)ccc2O)C=C
Sc1cscc1SC=C	c1sc(C(=O)O)c(c1)C=C	c1csc2c1c(CC)c1c(n2)scc1C=C
c1sc(cc1)C#C	c1scc2c1C=C(C2=O)C=C	c1oc2c(n1)cc1c(c2)oc(n1)C=C
c1oc(nc1)C=C	c1sc(c(c1C#N)C#N)C=C	c1sc(c2c1[nH]c(=O)[nH]2)C=C
c1scc(OC)c1N	c1c(OCC)c(cc(c1)OCC)	c1sc(c2c1CC[C@@H](OC)C2)C=C
c1sc(O)c(c1)	C1S=C(/C(=N/N)/C1=S)	c1sc(c2c1sc(C(=O)OCC)c2)C=C
c1oc(nn1)C=C	c1sc(c2c1cc(C#N)cc2)	c(c(N)1)ccc1c(c(N)1)ccc1C=C
Nc1cscc1NC=C	c1sc(c(c1O)C(F)(F)F)	c1sc(c2c1[C@H](N)CC[C@H]2N)
c1oc(cc1C=O)	c1scc2c1c(ccc2OC)C=C	c(s1)cc(c(=O)2)c1c(s3)c2cc3
c1ccc(s1)C=C	c1sc(c2c1[nH]cn2)C=C	c1sc(c2c1c(C(F)(F)F)ccc2OC)
c1sc(C)c(c1)	c1[nH]c2[nH]c(nc2n1)	c1sc(c2c1c(C(=O)OC)ccc2)C=C
c1sccc1N(CC)	c1sc(c2c1C(=O)SC2=O)	c1sc(cc1)c1ccc(s1)c1sc(cc1)
c1oc(cc1C#N)	c1sc2c(c1CC)sc(c2CC)	c1sc(c2c1sc(C(=O)CC)c2F)C=C
c1sc(nn1)C=C	c1sc(c2c1SCC(=O)CS2)	C1=C2C(C(=O)S1)=C(OC2=O)C=C
c1[nH]c(cc1)	c1sc(C(F)(F)F)cc1C=C	c1ccc(OCC)c2c1c(OCC)ccc2C=C
c1nc2CCNS2c1	c1sc(c2c1cc(C)s2)C=C	c1sc(c2c1c(N(=O)=O)ccc2)C=C
c1SCSc1ccC=C	c1csc2c1cc1c(ccs1)c2	c1sc(c(c1CC)OCCS(O)(O)O)C=C
c1oc(cc1)C=C	c1sc(c2c1c(C=O)ccc2)	c1sc(c(c1N(=O)=O)N(=O)=O)C=C
c1scc(n1)C=C	c1[nH]c(c2c1nccn2)C=C	c1c(F)c(F)c(c2c1nn(CC)n2)C=C
c1sc(cc1)C=C	c1sc(c2c1cc(F)c(F)c2)	c1cc2c(cc1)c1c(C2)cc(cc1)C=C
C=Cc1csc(c1)	C1=C(Oc2c(csc2)O1)C=C	N(CC)c1ccc(c(c1)C=C)N(CC)C=C
c1oc(c(c1F)F)	c1c2c(ncc(CC)n2)c(s1)	c1c(OC)cc(c(c1)OCC(CC)CC)C=C
C1=CC=C(C1=O)	c1sc2c(c1)sc1cc(sc21)	c1sc(c2c1cc(C#N)c(C#N)c2)C=C
c1scc(C#N)c1N	c1sc(c2c1cc(S)c(O)c2)	c1sc(c2c1c(C(F)(F)F)ccc2C#N)
c1c(CC)sc(c1)	C/C=C\1/OCCC(OCC1)C=C	c1csc2c1c(CC)c(CC)c1c(csc21)
c1ccc(cc1)C=C	c1sc(c2c1cc(S)cc2)C=C	c1oc2c(c1)C(=O)c1c2oc(c1)C=C
C1=CC=C(C1=S)	c1sc(c(c1CC)CC)C=CC=C	c1sc(c2c1OCC[C@@H](SC)O2)C=C
c1c(C)Nc(C)c1	c1sc(c2c1sc(=O)s2)C=C	c1sc(c2c1C[C@H](C(=O)OC)CC2)
c1c(CC)cc(s1)	c(cc1)c2cs(c)cc2c1C=C	c1cc2c(C(=O)N(C2=O)CC)cc1C=C
c1oc(c(c1C)C)	C1=C(Cc2c(C1)nsn2)C=C	c1cc2c(c3c1nsn3)cc(c1c2nsn1)
c1cc(cnc1)C=C	c1[nH]cc2c1[nH]cc2C=C	c1sc(c2c1cc(C(C)(C)C)cc2)C=C
c1c(C)sc(C)c1	c1c(cc2c(c1)nccn2)C=C	c1n(CC)c(=O)c2c1c(=O)n(CC)c2
c1n(CC)c(cc1)	c1sc(c2c1c(C=C)ccc2C)	c1sc2c(c1)C(=O)c1c2sc(c1)C=C
c1c(F)sc(F)c1	C1=C(CC2=C1C[CH]2)C=C	c1sc(c2c1[C@H](C(=O)OC)CCC2)
c1occc1C=CC=C	c1sc(c2c1cc(C)c(C)c2)	c1c(O)sc2c1[nH]c1c2sc(c1)C=C
c1cnc(cc1)C=C	c1c(C)c(c2c1ccco2)C=C	c1c2ccsc2cc2c1c(c1c(c2)scc1)
c1c(S)sc(O)c1	c1sc(c(c1OC)C(F)(F)F)	c(s1)c2c(=O)n(C)c(=O)c2c1C=C
c1oc(cc1C)C=C	C#Cc1[nH]c(cc1)C#CC=C	c1scc2c1C[C@@H](CC2)C(=C)C=C
c1oc(c(c1O)S)	c1scc2c1c(c(cc2)O)C=C	c1sc(c2c1cc(C(F)(F)F)cc2)C=C
c1c(N)ccc(c1)	c1c2nc(C)c(C)nc2c(s1)	c1c2c3c(s1)ccc1c3c(cc2)c(s1)
c1sc(c(c1)OC)	c1c(OC)c(cc(c1)OC)C=C	c1cc(c2c(c1)c1c(s2)cccc1)C=C
c1occ(C)c1C=C	c(oc1c(F)s2)csc1c2C=C	c1oc2c(c1)C(=S)c1c2oc(c1)C=C
c1sc(cc1)SC=C	c1oc(c(c1)N(=O)=O)C=C	c1sc2c(c1)C(=S)c1c2oc(c1)C=C

Table B19: List of SMILES for the 1235 monomer data set. (Part 2 of 10)

c1c(nc2cSc12)	c(s(c)1)c2ccccc2c1C=C	C1=C(CC)C(CC)=C(S1(=O)=O)C=C
c1oc(cc1O)C=C	c1sc2c(c1)C(=O)C(=C2)	c1c2n(CC)c3c(c2ccc1)cccc3C=C
c1nc2cscc2nc1	c1sc(c2c1ccc(F)c2)C=C	c1oc2c(c1)C(=O)c1c2sc(c1)C=C
C1SC=C1C(=O)C	c1cnc2c1c(=S)c1c2ncc1	c1sc(c2c1c(F)c(F)c(F)c2F)C=C
c1cscc1C(=O)O	c1sc(c(c1N(=O)=O)C#N)	c1sc(c2c1cc(SCC)c(SCC)c2)C=C
c1sc(c(c1O)S)	c1csc2c1n(CC)c1c2scc1	c1sc(c2c1CC[C@@H](C=O)C2)C=C
c1[nH]cnc1C=C	c1sc(c(c1O)C(=O)O)C=C	c1sc(c2c1CC[C@@H](C(=O)O)C2)
c1sc(c2c1CN2)	c1sccc2c1[C@@H](CCC2)N	c1oc(c(c1N(=O)=O)N(=O)=O)C=C
c1sc(c(c1C)C)	c1sc(c2c1C(=O)CCC2=O)	c1sc(c2c1c(C(F)(F)F)ccc2)C=C
c1nnc(nn1)C=C	c1scc2c1C=C(S2(=O)=O)	c1ccc(c2c1c1c(c3c2ccs3)scc1)
c1sc(cc1)NC=C	c1sc(c2c1sc(=S)o2)C=C	c1c2c(ccc3c2ccs3)c2c(ccs2)c1
c1c(F)cccc1C=C	c1c(F)c(F)c(c2c1nsn2)	C(=[N])/N=C(\C1=CSC=[S]1)C=C
c1c2CCc2sc1C=C	C1S[C@@H]2[C@@H]1NCC2	c1sc(c2c1nc(OCC)c(OCC)n2)C=C
C1SC=C1C(=O)OC	c1sc(CC)c2c1nc(s2)C=C	c1sc(c2c1c(N(=O)=O)ccc2N)C=C
C=NNc1ccc(cc1)	c1cc(c2c1[nH]cnc2)C=C	c1sc(c2c1[C@H](N)CC[C@H]2OC)
c(s1)c2NCOc2c1	c1sc2c(n1)Cc1c2sc(c1)	N1[C@@H]2CSC(=S)[C@H]2N(CC1)
c1cc(CN)ccc1CN	c1sc(c(c1)N(=O)=O)C=C	c1sc(c2c1[nH]c(N(=O)=O)c2)C=C
c(s1)c2scnc2c1	c1c(C(F)(F)F)sc(OC)c1	C1C2CC(=C)O[C@H](C1C=CC2C)CCC
c1ccc(=O)c1C=C	c1cccc2c1n(c1c2cccc1)	C1=C(C=C/C/1=C\1/C=CC(=C1)N)N
c1sc(c2c1ncs2)	c1sc(c2c1ncnc2C#N)C=C	C1=[S]C(=S)C2=C1C=C(S2(=O)=O)
c(s1)c2ncsc2c1	c1sc(c2c1oc(=S)o2)C=C	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)
c1sc(c2c1SCS2)	c1scc2c1c(c(cc2)N)C=C	c1csc2c1c(CC)c1c(c2CC)scc1C=C
c1sc(c2c1ocn2)	c1oc(c(c1O)C(=O)O)C=C	c1scc2c1C(=[S@@](OC)C(=C2)OC)
c1oc2cc(sc2c1)	c1sc(c(n1)C(=O)OC)C=C	c1sc(c2c1CC[C@@H](N(=O)=O)C2)
c1sc2nc(sc2c1)	c1scc2c1N(CCN2C(=O)O)	c1cc(CC)c(c(c1)CC)/C=C(/C)C=C
c1c(CC)c(ccc1)	c1oc(c(c1C(F)(F)F)OC)	c1c(c2c(s1)nc1c(c2CC)CCS1)C=C
c(c1)cc(cc)cc1	c1sc(c2c1cc(C)cc2)C=C	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)
c1sc(c2c1scc2)	c1c2cscc2c(N)c(c1)C=C	c(s1)c2c(=O)N(CC)c(=O)c2c1C=C
c1cc(SC)sc1C=C	c1sc(c2c1c(C)ccc2)C=C	c1c2[nH]cnc2c(c2c1[nH]cn2)C=C
C1=C(c2cscc12)	c1csc2c1C(S)=c1c2scc1	c1sc(c2c1[C@H](C(F)(F)F)CCC2)
c1oc(cc1OC)C=C	C1SC[C@](C1)(C(=O)N)N	c1n(C)c(=O)c2c1c(=O)n(C)c2C=C
c1sc(c2c1CCC2)	c1sc(c2c1c(S)ccc2)C=C	c1[nH]cc2c1S(=O)(=O)C(=C2)C=C
c1sc(c(c1O)OC)	c1sc(c2c1sc(=S)s2)C=C	c1cc2c(s1)c1c(C32OCCO3)cc(s1)
c1sc(C=O)c(c1)	C1=Cc2[nH]c(cc2C1)C=C	c1sccc1/C=C(/Cc1sc(CC)cc1)C=C
c1oc2CCc2c1C=C	c1sc(CC)c2c1c(c[nH]2)	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)
c1sc(c2c1scn2)	c1scc2c1c1c(s2)sc(c1)	c1[nH]c(=O)c2c1c(=O)[nH]c2C=C
C1C([CH2])CCC1	C1=CC2=CC(=NC2=C1)C=C	c1sc(c2c1cc(C(=O)O)c(O)c2)C=C
c1sc(c2c1OCO2)	c1c(C)cc(CC)c(C)c1C=C	c1sc(c2c1[C@H](C)CC[C@H]2C=C)
c1cc(C)cc(c1C)	C1=[S]C(=C(S1)C(=O)C)	c1cc2c(s1)c1c([nH]2)cc(s1)C=C
c1sc(c2c1occ2)	c1sc(c2c1oc(=O)o2)C=C	n1c(C)c2c(c1=O)c(C)n(c2=O)C=C
c1sc(c(c1)C=O)	c1scc2c1c1c(s2)cc(s1)	c1[nH]c2nc3[nH]c(nc3nc2n1)C=C
N1c2cscc2N(C1)	n1c2cscc2n(c(=O)c1=O)	c1sc(c2c1C[C@H](C(F)(F)F)CC2)
c1oc2cc(oc2c1)	N1[C@H]2CSC[C@H]2N(S1)	c1c2n(CCC)c3c(c2ccc1)cccc3C=C
c1c(OC)c(ccc1)	c1sc(c2c1C=[S]C=C2)C=C	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)
c1sc(c(c1)C#N)	c1sc(OC)c(C(=O)C)c1C=C	c1c2c3CCSc3c3SCCc3c2c(cc1)C=C

Table B20: List of SMILES for the 1235 monomer data set. (Part 3 of 10)

c1sc(c2c1nco2)	c1c(C(F)(F)F)sc(C#N)c1	c1sc(c2c1[C@H](OC)CC[C@H]2OC)
c1c(sc(OC)c1)N	c1c(C)cc(c(c1)C)C=CC=C	c(s1)c2c(=O)n(CC)c(=O)c2c1C=C
c1sc(c(c1CC)CC)	c1c2c(SCC2)c2c(CCS2)c1	c1c2c3c(s1)ccc1c3c(cc2)sc1C=C
c1sc(CC)c(c1CC)	c1sc(c2c1C[C@H](S)CC2)	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)
c1sc(c2c1OCCO2)	c1oc2c(C(=O)N)c(oc2c1)	c1sc(c2c1C[C@H](F)[C@@H](F)C2)
c1c2cscc2c(nn1)	c(s1)c2ccs(=N)(=O)c2c1	c(c(CC)c1)c(c(=C)cn)c(CC)c1C=C
C=Cc1csc(c1)C=C	c1sc(c2c1[C@H](C)CCC2)	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1
c1[nH]c(cc1)C#C	c1oc(c(c1C#N)C(F)(F)F)	c1sc(c2c1[C@H](C)CC[C@H]2C)C=C
c1scc(OC)c1NC=C	c1sc(c2c1c(F)ccc2F)C=C	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)
c1oc(cc1C#N)C=C	C1=[S]c2cscc2[S]=C1C=C	c1c2c(sc(C(=O)OCC)c2F)c(s1)C=C
c1c(OC)sc(OC)c1	C(=S)[CH]C1=[S]C=C(O1)	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)
c1sc(O)c(c1)C=C	c(o1)c2c(=O)Nc(=O)c2c1	c1sc(c2c1[C@H](S)CC[C@H]2O)C=C
C#Cc1sc(cc1)C#C	c1cc(cc2c1ccc1c2cccc1)	c(c1)c2nccnc2c1c(c1)c2nccnc2c1
c1sc(c2c1ncnc2)	c1c(C(=O)C(F)(F)F)scc1	c1sc(c2c1c(C(F)(F)F)ccc2OC)C=C
c1sc(C)c(c1)C=C	c1sc(c2c1cc(OC)cc2)C=C	c1sc(c2c1[C@H](C(=O)C)CCC2)C=C
c1oc(cc1C=O)C=C	c1c2c(ncs2)c(c2c1ncs2)	c1sc(c2c1C[C@H](C(=O)C)CC2)C=C
c1c2nsnc2c(cc1)	c1cc2c(cs1)c1c(ccs1)c2	c1c(=O)oc2c1c(cc1cc(o)oc21)C=C
c1c(C(=O)C)scc1	c1sc(c(c1C#N)C(F)(F)F)	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)
c1sccc1N(CC)C=C	c1scc2c1C(=C[S@@]2[O])	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)
c1occ(OC)c1NC=C	c1sc(c(c1C(=O)O)OC)C=C	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)
c1c(OCC)c(ccc1)	c1sc(c(c1OCC)ON(=O)=O)	c1sc(c(c1C(F)(F)F)C(F)(F)F)C=C
c1sc(c(c1OC)OC)	c1sc(c2c1c(C)ccc2C)C=C	c1sc(c2c1[C@H](F)CC[C@H]2F)C=C
N1c2cscc2N(CC1)	C1=CC2=C(C1)C=C(C2)C=C	c1sc(c2c1[C@H](N)CC[C@H]2N)C=C
c(c(C)1)ccc1C=C	c1c(OCC)cc(CC)c(OCC)c1	c1c2c(ccs2)c(c2c1cc1c(c2)scc1)
c1oc(cc1C(=O)C)	c1sc(c2c1c(S)ccc2O)C=C	c1sccc1/C=C(/Cc1ccc(CC)cc1)C=C
c1sc(c2c1OCCS2)	C=Cc1oc(c(c1O)O)C=CC=C	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)
c1c(C)cc(s1)C=C	C1=C(C(=C[S@@]1O)O)C=C	c1sc(c2c1C[C@H](S)[C@@H](O)C2)
c1[nH]c(cc1)C=C	C1=CC2=C(C1)N=C(C2)C=C	c1sc(c2c1C[C@H](C)[C@@H](C)C2)
c1scc2c1C=C(C2)	c1sc2C3=C([CH]S3)Cc2c1	c1sc(cc1)c1ccc(s1)c1sc(cc1)C=C
c1c(ncc2c1non2)	c1c2c(ccs1)nc1c2sc(c1)	c(s1)cc(c(=S)2)c1c(s3)c2cc3C=C
c1cnc(c2c1nsn2)	c1sc(c2c1cc(C)c(OC)c2)	c1scc2c1C[C@@H]([C@@H](OC)C2)N
c1c2nonc2c(cc1)	c1ccc(c2c1nn(CC)n2)C=C	c(s1)cc(c(=O)2)c1c(s3)c2cc3C=C
c1oc(cc1C(=O)O)	c1scc(N(O)O)c1N(O)OC=C	c1ccc2c(c1)C(=O)c1c2ccc(c1)C=C
c1sc(c2c1OCSS2)	c(c(CC)1)ccc1c(c1)ccc1	C1=[S]C(=O)C2=C1C(=O)[S]=C2C=C
c1oc(c(c1OC)OC)	c1c(OC)cc(c(c1)OCC)C=C	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)
c1sc(c2c1OSCS2)	c1sc(c2c1[nH]c(C#N)c2)	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)
c1sc(c2c1nccn2)	c1oc(cc1C(=O)C(F)(F)F)	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)
c1sc(c(c1O)C#N)	c1c(OCC)c(cc(c1)OC)C=C	c1sc(c2c1C(=O)c1ccccc1C2=O)C=C
c1c2nccnc2c(s1)	c1sc(c2c1CCC[C@H]2C=O)	c1sc2c(c1)C(C)(C)c1c2sc(c1)C=C
c1sc(c2c1SCCS2)	c(s1)c2C(NCC)OC(O)c2c1	c1sc(c2c1cc(N(=O)=O)c(N)c2)C=C
n1c(C)cc(c1C=C)	c1sc(c2c1sc(C#N)c2)C=C	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)
c1sc(cc1)C#CC=C	c1sc(c2c1[C@H](S)CCC2)	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)
c1c2cccc(c2cs1)	c1sc2c(c1)[nH]c(c2)C=C	c1sc(c2c1[C@H](C(=O)OC)CCC2)C=C
c1sc(c2c1CCCC2)	c(s1)c(CCN)c(CCN)c1C=C	c1cc2c(c3c1nsn3)cc(c1c2nsn1)C=C
c1c(oc2cscc2s1)	c1cc(N(=O)=O)c(cc1)C=C	c1csc2c1c(CC)c(CC)c1c(csc21)C=C

Table B21: List of SMILES for the 1235 monomer data set. (Part 4 of 10)

c1sc(cc1)C=CC=C	c1c(OCCC)c(cc(c1)OCCC)	c1c2c(ccc3c2ccs3)c2c(ccs2)c1C=C
N1C(=S)C=C(C1=S)	c1sc(c2c1c(C#N)ccc2OC)	c1sc(c2c1CC[C@@H](C(=O)O)C2)C=C
c1sc(c2c1SCCCS2)	C=C=C1CC2=C(C1)C(=CC2)	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)
c1cc2cc(o)cc2cc1	NCc1cc(OCC)c(cc1OCC)CN	c1ccc(c2c1c1c(c3c2ccs3)scc1)C=C
c1sc(c2c1OCCCO2)	c1c(OCC)c(OCC)c(OCC)cc1	C1=C(F)C=C/C/1=C\1/C(=CC(=C1)F)
c1cnc(c2c1nccn2)	c1cc2c(s1)c1c(C2)cc(s1)	c1sc(c2c1[C@H](C#N)CC[C@H]2C#N)
c1sc(c(c1C)C)C=C	c1sc(c2c1cc(C(=O)C)cc2)	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)
c1c(C#N)sc(OC)c1	c1sc(c2c1c(C#N)ccc2C#N)	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)
c1c2cS(F)cc2ccc1	c1sc(c2c1SCC(=O)CS2)C=C	c1scc2c1nc1c3cc(sc3c3sccc3c1n2)
c1n(CC)c(cc1)C=C	c1sc(c2c1SCC(=O)CO2)C=C	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)
C1SCN2[C@H]1NCC2	c1sc(c2c1c(C#N)ccc2)C=C	c1sc(c2c1C[C@H](C(=O)OC)CC2)C=C
c1scc2c1c(ccc2O)	c1sc(c2=CC=C(c12)[CH2])	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)
C1=CC=C(C1=S)C=C	c1sc(c2c1CCC[C@H]2O)C=C	c1n(CC)c(=O)c2c1c(=O)n(CC)c2C=C
n1c2cscc2[nH]cc1	c1[nH]c2[nH]c(nc2n1)C=C	c1c2c3c(s1)ccc1c3c(cc2)c(s1)C=C
c1c2c(OCN2)c(s1)	c1sc(c2c1cc(C=C)cc2)C=C	N1[C@@H]2CSC(=S)[C@H]2N(CC1)C=C
C1SC=C1C(=O)CC=C	c1sc(c(c1SCC)SCC)C=CC=C	c1sc(c2c1C[C@H](C(F)(F)F)CC2)C=C
c1scc(C(=O)O)c1S	c1sc(c2c1C(=O)NC2=O)C=C	c1scc2c1C(=[S@@](OC)C(=C2)OC)C=C
c1scc(C)c1C=CC=C	c1sc(c2c1S(=O)(=O)CCC2)	c1sc(c2c1C[C@H](OC)[C@@H](OC)C2)
c1ccc(c2c1cccn2)	c1sc(c2c1c(C=C)ccc2)C=C	c1cc2c(s1)c1c(C2(CC)CC)cc(s1)C=C
c1sc(c(c1OC)C#N)	c1c(CC)c(CC)c(CC)cc1C=C	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1
c1nc2cscc2nc1C=C	c1sc(c2c1CC[C@@H](N)C2)	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)C=C
c(s1)c(O)c(NN)c1	c1c2c(ncc(C)n2)c(s1)C=C	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)C=C
c1[nH]cc(c1C)C=C	c1sc(NC)c(c1C(=O)OC)C=C	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)C=C
c1c(CC)cc(s1)C=C	c1c(OCC)cc(OCC)c(c1)C=C	c1sc(c2c1[C@H](OC)CC[C@H]2OC)C=C
c1sc(C(=O)O)cc1O	c1[nH]c2cc([nH]c2c1)C=C	c1sc(c2c1[C@H](C(F)(F)F)CCC2)C=C
c1c(N)ccc(c1)C=C	c1[nH]c(c(c1)C(=O)C)C=C	c1cc2c(s1)c1c(C32OCCO3)cc(s1)C=C
c1c(C)cc(c(c1)C)	c1sc(c2c1C(=O)SC2=O)C=C	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)C=C
c1cnc(cc1)C=CC=C	c1sc(c2c1CCS(=O)(=O)C2)	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3
C1=CC=C(C1=O)C=C	c1ccc2c(c1)oc1c2ccc(c1)	c1sc(c2c1CC[C@@H](N(=O)=O)C2)C=C
c1coc(N(=O)=O)c1	c1sc(c2c1CC[C@@H](O)C2)	c1sc2c(ccc3c2ccc2c3ccc3c2scc3)c1
c1c(nc2cSc12)C=C	c1csc2c1cc(c1c2scc1)C=C	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)C=C
c1sc(c2c1cccc2F)	c1c(F)c(F)c(c(c1F)F)C=C	c1sc(c2c1[C@H](C)CC[C@H]2C=C)C=C
c1scc(C(=O)OC)c1	c1sc(c2c1CC(=O)C(=O)C2)	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)
c1sc(c(c1O)S)C=C	c1sc(c2c1oc(=O)c(=O)s2)	C1=C(C=C/C/1=C\1/C=CC(=C1)N)NC=C
c1c(S)sc(O)c1C=C	N1c2cscc2N(CC(=O)C1)C=C	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)
c1c(CC)sc(c1)C=C	c1csc2c1cc1c(ccs1)c2C=C	C1C2CC(=C)O[C@H](C1C=CC2C)CCCC=C
C(C(=O)1)C(=O)C1	c1sc(c2c1cc(C=C)c(C)c2)	c1cc2c(s1)c1c(cc2)c2c(cc1)cc(s2)
c1c(C)sc(C)c1C=C	C1=C2OCCSC2=C([S@@]1NC)	C1=[S]C(=S)C2=C1C=C(S2(=O)=O)C=C
c1c(C)Nc(C)c1C=C	c1sc(c2c1sc(=O)c(=O)s2)	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1C=C
Nc1sc(c2c1nccn2)	c1sc(c2c1C(=O)CC2=O)C=C	c1sc(c2c1nc1c3cccnc3c3ncccc3c1n2)
c1occ(C)c1C=CC=C	c1sc(c2c1cc(C=O)cc2)C=C	c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1C=C
c1oc(c(c1C)C)C=C	c1scc2c1CCC[C@H]2C(=O)O	c1c2c(ccs2)c(c2c1cc1c(c2)scc1)C=C
c1sc(N(=O)=O)cc1	c1sc(c2c1nc(CN)c(CN)n2)	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)C=C
c1oc(c(c1O)S)C=C	c1sc(c2c1oc(N(=O)=O)c2)	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)C=C
c1sc(C(=O)OC)cc1	c1sc(c2c1C(=O)OC2=O)C=C	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)C=C

Table B22: List of SMILES for the 1235 monomer data set. (Part 5 of 10)

c1oc(c(c1F)F)C=C	c1c(OCC)c(cc(c1)OCC)C=C	c1sc(c2c1C[C@H](F)[C@@H](F)C2)C=C
C1=C(Cc2cscc2C1)	C1=CC(=C2[C@H]1CCS2)C=C	C1=CC(OC)=C(/C/1=C\1/C=C(OC)C=C1)
c1oc(cc1C(=O)OC)	c1c2nc(CC)c(CC)nc2c(s1)	c1scc2c1C[C@@H]([C@@H](OC)C2)NC=C
c1sc(c(c1C=C)OC)	C1S=C(/C(=N/N)/C1=S)C=C	c(c1)c2nccnc2c1c(c1)c2nccnc2c1C=C
c1sc(c2c1OCCCC2)	c1scc2c1nc(c(NC)n2)N(C)	c1sc(c2c1C[C@H](C)[C@@H](C)C2)C=C
c1cscc1C(=O)OC=C	c1sc2c(c1CC)sc(c2CC)C=C	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)C=C
c1c(F)sc(F)c1C=C	N1CN[C@@H]2[C@H]1N(CN2)	c1sc(c2c1C[C@H](C#N)[C@@H](OC)C2)
c1sc(c2c1CN2)C=C	c1sc(c2c1sc(C(=O)CC)c2)	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)C=C
c1oc(c(c1OC)C#N)	c1sc(c2c1cc(C#N)cc2)C=C	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)
N1C(=O)C=C(C1=O)	c1sc(c2c1CC[C@@H](C)C2)	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)
c1sc(c(c1F)F)C=C	c1sc(c(c1OCC)OCC)C=CC=C	C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)
c1scc(C#N)c1NC=C	c1c(OCC)cc(c(c1)OCC)C=C	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)C=C
c1sc(c2c1OCCCS2)	c1[nH]c(cc1)C(C(=O)N)=C	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)C=C
c1c2ccccc2c(c1=O)	C1=C(c2c3c(cccc13)ccc2)	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)C=C
c1sc(c(c1)C#N)C=C	c1sc(c2c1cc(OC)c(OC)c2)	c1sc(c2c1C[C@H](S)[C@@H](O)C2)C=C
c1sc(c2c1nco2)C=C	c1sc2c(c3ccoc3cc2c1)C=C	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)C=C
c1occ(N(=O)=O)c1N	c1nc2c(c(=O)c3c2ncc3)c1	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)C=C
c1cc(C)cc(c1C)C=C	c1sc(c2c1[C@H](OC)CCC2)	c1scc2c1C[C@H](C)[C@H](C2)C(=C)C=C
c(c1)cc(cc)cc1C=C	c(s1)c2cc(OC)c(OC)cc2c1	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)C=C
c1sc(c(c1)C=O)C=C	c1sc(c2c1cc(C(=O)OC)s2)	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)C=C
c1sc(c2c1ocn2)C=C	c1sc(c2c1sc(N(=O)=O)c2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)
c1c(CC)c(ccc1)C=C	c1sc(c2c1oc(=O)c(=O)o2)	c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)
c1c2C(=O)OCc2ccc1	c1sc(c2c1c(C(=O)C)ccc2)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)
c1sc(c2c1SCS2)C=C	C1=C2CSC[C@H]2C(=C1)C=C	C1=C(F)C=C/C/1=C\1/C(=CC(=C1)F)C=C
c1oc2cc(sc2c1)C=C	c(c(N)1)ccc1c(c(N)1)ccc1	c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)
C=NNc1ccc(cc1)C=C	c1sc(c(c1N(=O)=O)C#N)C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)
c1[nH]c(C=C)c(c1)	C1=C2C(C(=O)O1)=C(OC2=O)	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)C=C
c1c2c(nccc2)c(s1)	c1sc(c2c1sc(C(=O)CC)c2F)	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)C=C
c1scc2c1sc1c2csc1	c1scc2c1N(CCN2C(=O)O)C=C	c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)
c1scc(N(=O)=O)c1N	c1scc2c1C=C(S2(=O)=O)C=C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)
c1sc(C(F)(F)F)cc1	c1sc(c2c1sc(C(=O)OCC)c2)	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)C=C
c1c2c(cccc2)c(s1)	c1sc(c2c1c(C(=O)OC)ccc2)	c1sc(c2c1[C@H](C#N)CC[C@H]2C#N)C=C
c(s1)c2scnc2c1C=C	c1csc2c1c(CC)c1c(n2)scc1	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)
N1c2cscc2N(C1)C=C	C1=C2C(C(=O)S1)=C(OC2=O)	c1scc2c1nc1c3cc(sc3c3sccc3c1n2)C=C
c1oc(cc1C(F)(F)F)	c1sc(c2c1[C@H](C=C)CCC2)	c1sc(c2c1C[C@H](C(=O)C(F)(F)F)CC2)
N(CC)c1cscc1N(CC)	c1sc(c2c1[nH]c(=S)[nH]2)	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)
c1oc(c(c1C#N)C#N)	c1sc(c2c1nc(OCC)c(CN)n2)	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)
c1sc(c(c1OCC)OCC)	c1sc2c(c1)C(=O)C(=C2)C=C	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1C=C
C1=CC=C(S1(=O)=O)	n1c2cscc2n(c(=O)c1=O)C=C	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)
c1sc(c2c1[nH]cn2)	c1csc2c1C(S)=c1c2scc1C=C	c1sc(c2c1C[C@H](OC)[C@@H](OC)C2)C=C
c1scc2c1C=C(C2=O)	C1S[C@@H]2[C@@H]1NCC2C=C	c1sc(c2c1[C@H](C(=O)O)CC[C@H]2O)C=C
c1sc(c2c1ncs2)C=C	C1=C2C(C(=O)N1)=C(OC2=O)	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)C=C
c1sc(c(c1C#N)C#N)	c1c(F)c(F)c(c2c1nsn2)C=C	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)
c(s1)c2ncsc2c1C=C	c1sc2c(c1)oc1c2sc(c1)C=C	C1=CC(C#N)=C/C/1=C/1\C(=CC(=C1)C#N)
c1c(sc(OC)c1)NC=C	c1cnc2c1c(=S)c1c2ncc1C=C	c1sc2c(ccc3c2ccc2c3ccc3c2scc3)c1C=C

Table B23: List of SMILES for the 1235 monomer data set. (Part 6 of 10)

c(nc1c2)cs1nc2C=C	c1sc(c(c1OC)C(F)(F)F)C=C	c1cc2c(s1)c1c(cc2)c2c(cc1)cc(s2)C=C
c1sc(c(c1O)OC)C=C	c1sc(c(c1)C(=O)C(F)(F)F)	c(s1)cc(c(=C(CN)(CN))2)c1c(s3)c2cc3
c1sc(c2c1scc2)C=C	C1=[S]C(=C(S1)C(=O)C)C=C	c1scc2c1[C@@H](CC[C@H]2N(=O)=O)NC=C
C(=C)c1sc(C)c(c1)	c1c(C(F)(F)F)sc(OC)c1C=C	c1sc2c(c1)C([C](C#N)C#N)=c1c2sc(c1)
c1sc(C(=O)O)c(c1)	c1sc(c2c1nc(CCO)c(CN)n2)	C1=CC(=[S]C1)c1ccc(s1)C1=[S]CC(=C1)
c(s1)c2NCOc2c1C=C	c1sc(c2c1[nH]c(=O)[nH]2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2OC)
c1oc2cc(oc2c1)C=C	c1sc(c2c1cc(S)c(O)c2)C=C	c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)
c1sc(c2c1CCC2)C=C	c1c2c(ncc(CC)n2)c(s1)C=C	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)C=C
c1sc(c2c1occ2)C=C	c1c2c(nccn2)c(c2c1nccn2)	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3C=C
c1sc(c2c1cc[nH]2)	c1sc(c2c1c(C=C)ccc2C)C=C	c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)
Nc1sc(N(=O)=O)cc1	c1ccc(OCC)c2c1c(OCC)ccc2	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2
c1n(CC)nc(c1CC)CC	c1sc2c(c1)sc1cc(sc21)C=C	c1sc(c2c1nc1c3cccnc3c3ncccc3c1n2)C=C
C1=C([CH]S1)C(=C)	c1sc(c2c1nc(N)c(N)n2)C=C	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)
c1sc(c2c1scn2)C=C	c1csc2c1n(CC)c1c2scc1C=C	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)C=C
c1c(OC)c(ccc1)C=C	c1sc2c(c1)c1c(scc1)c(c2)	C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)C=C
c1sc(c2c1OCO2)C=C	c1sc(c2c1c(N(=O)=O)ccc2)	c1sc(c2c1C[C@H](N)[C@@H](N(=O)=O)C2)
c1sc(C=O)c(c1)C=C	C1=C2C(C(=O)S1)=CN(C2=O)	c1sc(c2c1C[C@H](C#N)[C@@H](OC)C2)C=C
c1scc2c1c(ccc2OC)	c1c2nc(C)c(C)nc2c(s1)C=C	C1=CC(OC)=C(/C/1=C\1/C=C(OC)C=C1)C=C
c1sc(c2c1[nH]nc2)	c1scc2c1c1c(s2)sc(c1)C=C	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)C=C
C1=C(c2cscc12)C=C	c1c2c(cccc2)cc2c1c(ccc2)	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)
c1sc(c(c1)C(=O)C)	c1sc(c2c1cc(C#N)c(OC)c2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C#N)
c1cc(CN)ccc1CNC=C	C1SC[C@](C1)(C(=O)N)NC=C	c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)
c1sc(c2c1cc(C)s2)	c1sc(c2c1cc(N(=O)=O)cc2)	c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)
C1SC=C1C(=O)OCC=C	c1scc2c1[C@@H](CCC2)NC=C	c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)C=C
c1sc2nc(sc2c1)C=C	c1csc2c1c(C)c(C)c1c2scc1	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)C=C
c1oc(cc1C(=O)C)C=C	C1=[S]C(=C(O1)[N]C=O)C=C	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)C=C
c1c(C)c(c2c1ccco2)	c1sc2c(c1)n(C)c1c2sc(c1)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)
C1Oc2cscc2OC(C1=O)	c1sc(c2c1cc(F)c(F)c2)C=C	c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)C=C
c1c2nonc2c(cc1)C=C	c1cccc2c1n(c1c2cccc1)C=C	c1sc(c2c1C[C@H](C(=O)C(F)(F)F)CC2)C=C
c1c(C(=O)C)scc1C=C	c1scc2c1c1c(s2)cc(s1)C=C	c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)C=C
c1c(OC)sc(OC)c1C=C	c1sc(c2c1CC[C@@H](OC)C2)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)C=C
c1sc(c2c1c(C)ccc2)	c1sc(c2c1cc(C)c(C)c2)C=C	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)C=C
c1sc(c2c1OCCO2)C=C	c1ccc(c2c1nc(CC)c(CC)n2)	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3
c1sc(c2c1OCSS2)C=C	c1oc2c(n1)cc1c(c2)oc(n1)	c1nc2c(s1)[C@H]1[C@H](C2(CC)CC)N(CS1)
c1sc(CC)c(c1CC)C=C	c1sc(c2c1c(C(=O)O)ccc2O)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)C=C
c1c2nsnc2c(cc1)C=C	c1oc(c(c1C(F)(F)F)OC)C=C	c1c(C(=O)c2cscc2C(=O)c2ccccc2)ccc(c1)
c1c(cc2c(c1)nccn2)	c1sc2c(n1)Cc1c2sc(c1)C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)C=C
c1sc(c(c1O)C(=O)O)	C1=Cc2cc3=CC(=Cc3cc2=C1)	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)
c1sc(c(c1O)C#N)C=C	c1sc(c2c1cc(C(=O)OC)cc2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2OC)C=C
c1sc(c2c1sc(=S)o2)	c1sc(c2c1SC(=O)CC(=O)O2)	C1=CC(C#N)=C/C/1=C/1\C(=CC(=C1)C#N)C=C
c1sc(c2c1cc(S)cc2)	c1sc(c2c1C(=O)CCC2=O)C=C	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)
C1=CC2=CC(=NC2=C1)	c1cc2c(cs1)c1c(ccs1)c2C=C	c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)C=C
c1c(OCC)c(ccc1)C=C	c1c(O)sc2c1[nH]c1c2sc(c1)	C1=CC(=[S]C1)c1ccc(s1)C1=[S]CC(=C1)C=C
c1c(OC)c(cc(c1)OC)	c1c(C(F)(F)F)sc(C#N)c1C=C	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)C=C
c(cc1)c2cC(C)cc2c1	c1oc2c(C(=O)N)c(oc2c1)C=C	c(s1)cc(c(=C(CN)(CN))2)c1c(s3)c2cc3C=C

Table B24: List of SMILES for the 1235 monomer data set. (Part 7 of 10)

c1sc(c2c1sc(=O)s2)	c1sc(c2c1c(N(=O)=O)ccc2N)	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)C=C
c1sc(c(c1)N(=O)=O)	c1sc(c2c1[C@H](C)CCC2)C=C	N1CN[C@@H]2N[C@H]3[C@@H](N[C@H]12)NCN3
C1=Cc2[nH]c(cc2C1)	c1c(C(=O)C(F)(F)F)scc1C=C	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)
c1c2cccc(c2cs1)C=C	c1oc(cc1C(=O)C(F)(F)F)C=C	C1=CC(N(=O)=O)C(C2C=CC=C2N(=O)=O)=C1C=C
c1sc(c2c1ncnc2C#N)	c1sc(c2c1c(F)c(F)c(F)c2F)	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)C=C
c(cc1)c2cs(c)cc2c1	N(CC)c1ccc(c(c1)C=C)N(CC)	c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)
C1=C(CC2=C1C[CH]2)	c1oc2c(c1)C(=O)c1c2oc(c1)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C#N)C=C
c(oc1c(F)s2)csc1c2	c1cc(c2c(c1)c1c(s2)cccc1)	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2C=C
c1sc(c2c1CCCC2)C=C	c1oc2c(c1)C(=S)c1c2oc(c1)	c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)C=C
C/C=C\1/OCCC(OCC1)	c1sc(c2c1c(C#N)ccc2OC)C=C	c1cc2c(s1)c1c(N2)c(cc2c1c1c(N2)ccs1)C=C
c1oc(c(c1OC)OC)C=C	c1sc(c2c1cc(C(F)(F)F)cc2)	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C#N)C2)
c1[nH]c(c2c1nccn2)	c1c(OCCC)c(cc(c1)OCCC)C=C	c1sc(c2c1C[C@H](N)[C@@H](N(=O)=O)C2)C=C
c1c2cscc2c(N)c(c1)	c1cc2c(cc1)c1c(C2)cc(cc1)	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)C=C
C#Cc1sc(cc1)C#CC=C	c1c(OC)cc(c(c1)OCC(CC)CC)	c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)C=C
c1scc2c1c(c(cc2)O)	c1sc(c2c1cc(C(C)(C)C)cc2)	c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2
c1sc(c2c1ncnc2)C=C	c1sc(c2c1nc(OCC)c(OCC)n2)	N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)
c1c2nccnc2c(s1)C=C	c1sc(c2c1c(C(F)(F)F)ccc2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)C=C
c1sc(c2c1sc(=S)s2)	c(c(CC)1)ccc1c(c1)ccc1C=C	c1c(C(=O)c2cscc2C(=O)c2ccccc2)ccc(c1)C=C
c1sc(c2c1ccc(F)c2)	c1c(OCC)cc(CC)c(OCC)c1C=C	c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)C=C
c1oc(c(c1)N(=O)=O)	c1sc(c2c1C[C@H](S)CC2)C=C	N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)
c1c(oc2cscc2s1)C=C	c1sc(c2c1[C@H](S)CCC2)C=C	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)C=C
c1c(C)cc(CC)c(C)c1	c(s1)c2ccs(=N)(=O)c2c1C=C	C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1
c(s(c)1)c2ccccc2c1	c1sc(c(c1OCC)ON(=O)=O)C=C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](OC)C2)C=C
c1sc(c2c1c(S)ccc2)	C(=S)[CH]C1=[S]C=C(O1)C=C	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C(F)(F)F)
c1c(ncc2c1non2)C=C	c1sc(c(c1N(=O)=O)N(=O)=O)	N1CN[C@@H]2N[C@H]3[C@@H](N[C@H]12)NCN3C=C
c1sc(c2c1oc(=O)o2)	c(o1)c2c(=O)Nc(=O)c2c1C=C	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)C=C
N1c2cscc2N(CC1)C=C	c1sc(c(c1C#N)C(F)(F)F)C=C	C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1
c1sc(c2c1OSCS2)C=C	c1sc(c2c1CCC[C@H]2C=O)C=C	c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)
c1c2cscc2c(nn1)C=C	C1=C(CC)C(CC)=C(S1(=O)=O)	N1[C@H]2[C@H](N[C@@H]3[C@@H](CSC3)N2)N(S1)
c1sc(c(c1OC)OC)C=C	c(s1)c2C(NCC)OC(O)c2c1C=C	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)
c1scc2c1C=C(C2)C=C	c1sc(c2c1cc(C)c(OC)c2)C=C	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)
c1sc(c2c1oc(=S)o2)	c1sc2c(c1)C(=S)c1c2oc(c1)	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)C=C
c1sc(c2c1cc(C)cc2)	c1c([CH2])c(c2c1ccsc2)C=C	c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)
c1sc(CC)c2c1nc(s2)	c1c2c(SCC2)c2c(CCS2)c1C=C	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)
c1sc(c2c1nccn2)C=C	c1sc(c2c1cc(C#N)c(C#N)c2)	c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)
n1c(C)cc(c1C=C)C=C	c1sc2C3=C([CH]S3)Cc2c1C=C	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)
c1oc(cc1C(=O)O)C=C	NCc1cc(OCC)c(cc1OCC)CNC=C	c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)C=C
c1sc(c2c1OCCS2)C=C	c1cc(cc2c1ccc1c2cccc1)C=C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C#N)C2)C=C
c1scc2c1c(c(cc2)N)	N1[C@H]2CSC[C@H]2N(S1)C=C	N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)C=C
C1=C(Oc2c(csc2)O1)	c1sc(c2c1OCC[C@@H](SC)O2)	N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)C=C
c1[nH]cc2c1[nH]cc2	c1c2n(CC)c3c(c2ccc1)cccc3	c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2C=C
c1[nH]c(cc1)C#CC=C	c1scc2c1C[C@@H](CC2)C(=C)	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)
c1cc(c2c1[nH]cnc2)	c1oc2c(c1)C(=O)c1c2sc(c1)	C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1C=C
C1=C(Cc2c(C1)nsn2)	C(=[N])/N=C(\C1=CSC=[S]1)	C1=C[C@@H]2[C@H](C1)[C@@H]1[C@@H](C=CC1)C2=C
c1oc(c(c1O)C(=O)O)	c(s1)c2c(=O)n(C)c(=O)c2c1	C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1C=C

Table B25: List of SMILES for the 1235 monomer data set. (Part 8 of 10)

c1cnc(c2c1nsn2)C=C	c1c2c(ccs1)nc1c2sc(c1)C=C	c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)C=C
c1sc(c(n1)C(=O)OC)	C=C=C1CC2=C(C1)C(=CC2)C=C	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)
c1sc(c(c1CC)CC)C=C	c1sc(c2c1CC[C@@H](C=O)C2)	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C(F)(F)F)C2)
c1sc(c2c1SCCS2)C=C	c1oc(c(c1C#N)C(F)(F)F)C=C	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C(F)(F)F)C=C
c1ccc(c2c1nn(CC)n2)	c1sc2c(c1)C(=O)c1c2sc(c1)	C1C(=O)O[C@@H]2[C@H]1CC[C@@H]1[C@H]2OC(=O)C1
c1sc(C(=O)OC)cc1C=C	c1c(OC)cc(c(c1)OCC)C=CC=C	C1C(=O)O[C@H]2[C@@H]1CC[C@H]1[C@@H]2OC(=O)C1
c1cc(N(=O)=O)c(cc1)	c1cc2c(C(=O)N(C2=O)CC)cc1	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)
c1ccc(c2c1cccn2)C=C	c1oc(c(c1N(=O)=O)N(=O)=O)	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)C=C
c1oc(c(c1OC)C#N)C=C	c1c(c2c(s1)c1c(n2CC)CCS1)	N1[C@H]2[C@H](N[C@@H]3[C@@H](CSC3)N2)N(S1)C=C
C1=c2ccccc2=C(C1=C)	c1c2c(ncs2)c(c2c1ncs2)C=C	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)C=C
N1C(=O)C=C(C1=O)C=C	c1sc(c2c1[nH]c(C#N)c2)C=C	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)C=C
c1cnc(c2c1nccn2)C=C	c1scc2c1C(=C[S@@]2[O])C=C	c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)C=C
c1scc(N(O)O)c1N(O)O	c1sc(c2c1cc(SCC)c(SCC)c2)	c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)C=C
c1c(C#N)sc(OC)c1C=C	c1c(F)c(F)c(c2c1nn(CC)n2)	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)C=C
c1cc2c(s1)cc(s2)C=C	c1sc(c2c1cc(C=C)c(C)c2)C=C	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)C=C
N1C(=S)C=C(C1=S)C=C	c1sc(c2c1CC(=O)C(=O)C2)C=C	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)C=C
c1sc(c2c1c(S)ccc2O)	c1c(c2c(s1)nc1c(c2CC)CCS1)	C1C(=O)O[C@@H]2[C@H]1CC[C@@H]1[C@H]2OC(=O)C1C=C
c1sc(N(=O)=O)cc1C=C	c1cc2occoccoccoccoc2cc1C=C	C1C(=O)O[C@H]2[C@@H]1CC[C@H]1[C@@H]2OC(=O)C1C=C
c1sc(c2c1OCCCC2)C=C	c1c(OCC)c(OCC)c(OCC)cc1C=C	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)
c1sc(c2c1OCCCO2)C=C	c(s1)c2cc(OC)c(OC)cc2c1C=C	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)C=C
c1c2cS(F)cc2ccc1C=C	c1sc(c2c1sc(=O)c(=O)s2)C=C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C(F)(F)F)C2)C=C
c1sc(c2c1c(F)ccc2F)	c1[nH]c(=O)c2c1c(=O)[nH]c2	C1=C[C@@H]2[C@H](C1)[C@@H]1[C@@H](C=CC1)C2=CC=C
c1sc(c2c1C=[S]C=C2)	C1=C2OCCSC2=C([S@@]1NC)C=C	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)C=C
c1sc(C(=O)O)cc1OC=C	c1sc(c2=CC=C(c12)[CH2])C=C	C1C[C@@H]2[C@@H](CC1)N(CC)[C@@H]1[C@@H](S2)CCCC1
c1[nH]cc(c1C)C=CC=C	c1nc2c(c(=O)c3c2ncc3)c1C=C	c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)
c1sc(c(c1)C(F)(F)F)	C1=C(c2c3c(cccc13)ccc2)C=C	c(s1)cc(c(OC(CCC)CCC)2)c1c(OC(CCC)CCC)c(c3)c2sc3
c1sc(c2c1OSCCO2)C=C	c1c2n(CCC)c3c(c2ccc1)cccc3	C1C[C@H]2[C@H](CC1)N(C[CH2])[C@H]1[C@H](S2)CCCC1
c1sc(c(c1C(=O)O)OC)	c1cc2c(s1)c1c([nH]2)cc(s1)	c1ccc2c3cc4n(CC)c5cc6c(cc5c4cc3Cc2c1)Cc1c6ccc(c1)
c1scc(C(=O)OC)c1C=C	c1cc(CC)c(c(c1)CC)/C=C(/C)	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)C=C
C=Cc1oc(c(c1O)O)C=C	c1sc(c2c1c(C(=O)C)ccc2)C=C	C1C[C@@H]2[C@@H](CC1)N(CC)[C@@H]1[C@@H](S2)CCCC1C=C
c1sc(c2c1OCCCS2)C=C	c1sc(c2c1cc(C(=O)OC)s2)C=C	c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)C=C
C1SCN2[C@H]1NCC2C=C	c1c2c3CCSc3c3SCCc3c2c(cc1)	c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)
c1coc(N(=O)=O)c1C=C	c1sc(c2c1[C@H](OC)CCC2)C=C	c1ccc2c3cc4n(CC)c5cc6c(cc5c4cc3Cc2c1)Cc1c6ccc(c1)C=C
c1c(C)cc(c(c1)C)C=C	c1cc2c(s1)c1c(C2)cc(s1)C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)
Nc1sc(c2c1nccn2)C=C	c1sc(c2c1sc(N(=O)=O)c2)C=C	N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1
C1=CC2=C(C1)C=C(C2)	c1scc2c1nc(c(NC)n2)N(C)C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)C=C
c1sc(c(c1C=C)OC)C=C	c1sc(c2c1oc(=O)c(=O)s2)C=C	c1cc2c(cc1)c1c(=C2C(CN)CN)cc2c(c1)=C(C(CN)CN)c1c2ccc(c1)
c1sc(c2c1oc(C#N)c2)	c1c2[nH]cnc2c(c2c1[nH]cn2)	N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1C=C
c1oc(cc1C(=O)OC)C=C	c1sc(c2c1S(=O)(=O)CCC2)C=C	c1cc2c(cc1)c1c(=C2C(CN)CN)cc2c(c1)=C(C(CN)CN)c1c2ccc(c1)C=C
c1sc(c2c1cccc2F)C=C	c1[nH]c2nc3[nH]c(nc3nc2n1)	c(s1)c(CCCCCC)cc1c(s1)c2c(=O)n(C)c(=O)c2c1c(s1)c(CCCCCC)cc1
C1=C(Cc2cscc2C1)C=C	c1sc(c2c1oc(N(=O)=O)c2)C=C	c(s1)c(CCCCCC)cc1c(s1)c2c(=O)n(C)c(=O)c2c1c(s1)c(CCCCCC)cc1C=C
c1scc(C(=O)O)c1SC=C	c1sc(c2c1CC[C@@H](O)C2)C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)
c1sc2c(c1)[nH]c(c2)	c(s1)c2c(=O)N(CC)c(=O)c2c1	c1c2n(CC)c3c(cc4c(c3)c3c(C4)cccc3)c2cc2c1c1c(C2)cc2c(c1)C(CC)(CC)c1c2cccc1
c(s1)c(O)c(NN)c1C=C	c1sccc1/C=C(/Cc1sc(CC)cc1)	c1c(SC)c(SC)c(c2c1nc1c(n2)c2c(nc3cc(SC)c(SC)cc3n2)c2c1nc1cc(SC)c(SC)cc1n2)
c(s1)c(CCN)c(CCN)c1	c1n(C)c(=O)c2c1c(=O)n(C)c2	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)C=C

Table B26: List of SMILES for the 1235 monomer data set. (Part 9 of 10)

C1=CC2=C(C1)N=C(C2)	c1scc2c1CCC[C@H]2C(=O)OC=C	c1c(SC)c(SC)c(c2c1nc1c(n2)c2c(nc3cc(SC)c(SC)cc3n2)c2c1nc1cc(SC)c(SC)cc1n2)C=C
c1sc(c2c1cc(OC)cc2)	c1c2nc(CC)c(CC)nc2c(s1)C=C	c1c2n(CC)c3c(cc4c(c3)c3c(C4)cccc3)c2cc2c1c1c(C2)cc2c(c1)C(CC)(CC)c1c2cccc1C=C
c1c2c(OCN2)c(s1)C=C	c1ccc2c(c1)oc1c2ccc(c1)C=C	c1cc2c(=O)n(C)c(=O)c3c2c2c1C1=C4[C@@H](c2cc3)C=CC2=C4[C@@H](C(=O)N(C2=O)C)C(=C1)
c1sc(c(c1OC)C#N)C=C	c1sc(c2c1CCS(=O)(=O)C2)C=C	c1cc2c(=O)n(C)c(=O)c3c2c2c1C1=C4[C@@H](c2cc3)C=CC2=C4[C@@H](C(=O)N(C2=O)C)C(=C1)C=C
c1sc(c2c1SCCCS2)C=C	c1csc2c1c(CC)c1c(c2CC)scc1	C1=C(C)C(C)=C(C2=N[C@H]3N4[C@@H](N[C@@H]5C[C@H](C)[C@H](C[C@@H]35)C)[C@H]3[C@H](N[C@@H]4[C@H]12)C)[C@H](C3)C)C
c1sc(c2c1c(C)ccc2C)	c1sc(c2c1CC[C@@H](N)C2)C=C	C1=C(C)C(C)=C(C2=N[C@H]3N4[C@@H](N[C@@H]5C[C@H](C)[C@H](C[C@@H]35)C)[C@H]3[C@H](N[C@@H]4[C@H]12)C)[C@H](C3)C)C
c1sc(OC)c(C(=O)C)c1	c1[nH]cc2c1S(=O)(=O)C(=C2)	c1c(OCC)cc(c(c1)OCC)C=CC=C

Table B27: List of SMILES for the 1235 monomer data set. (Part 10 of 10)

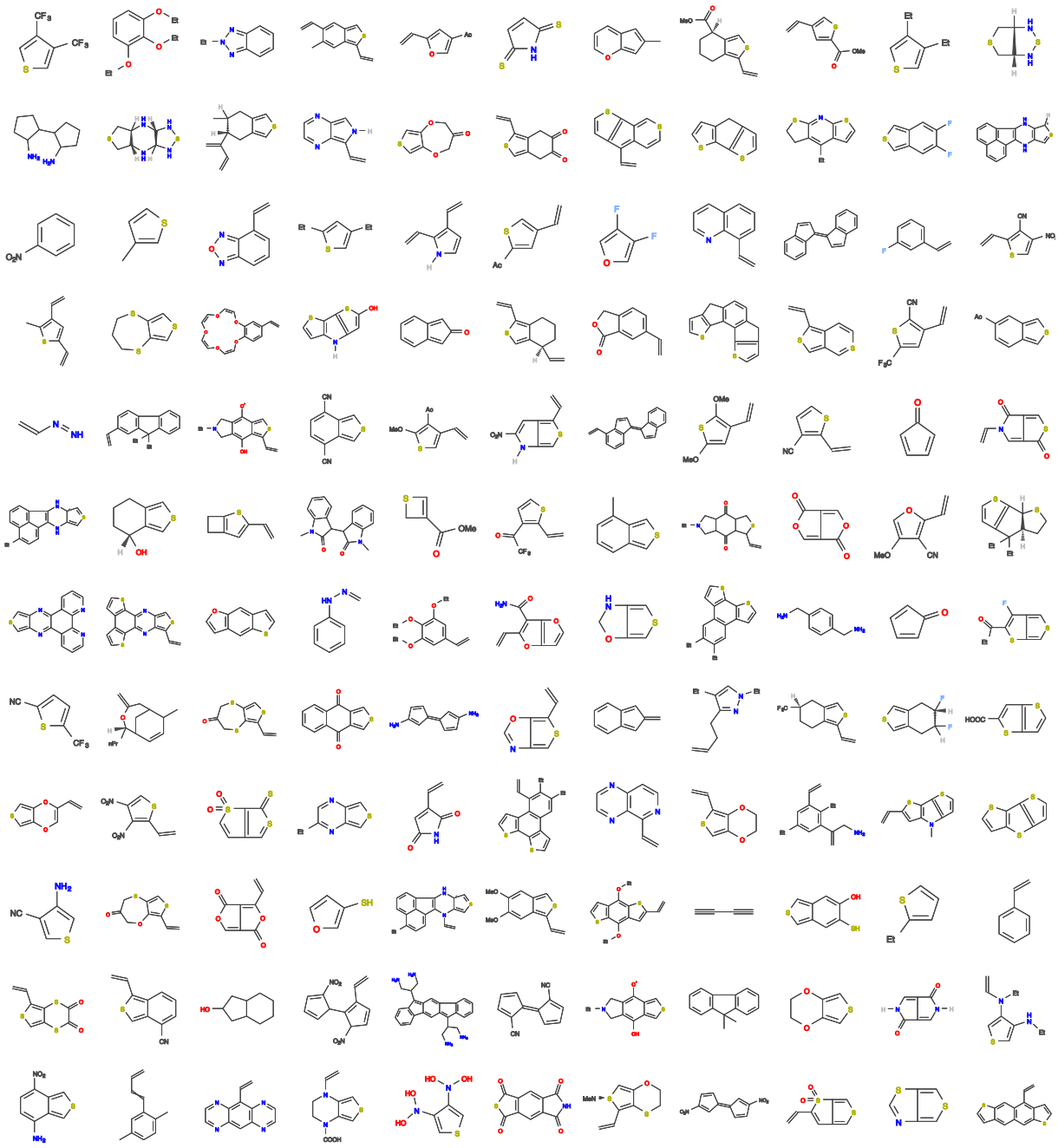


Figure B18: Molecules in the 1235 monomer dataset (1 of 10).

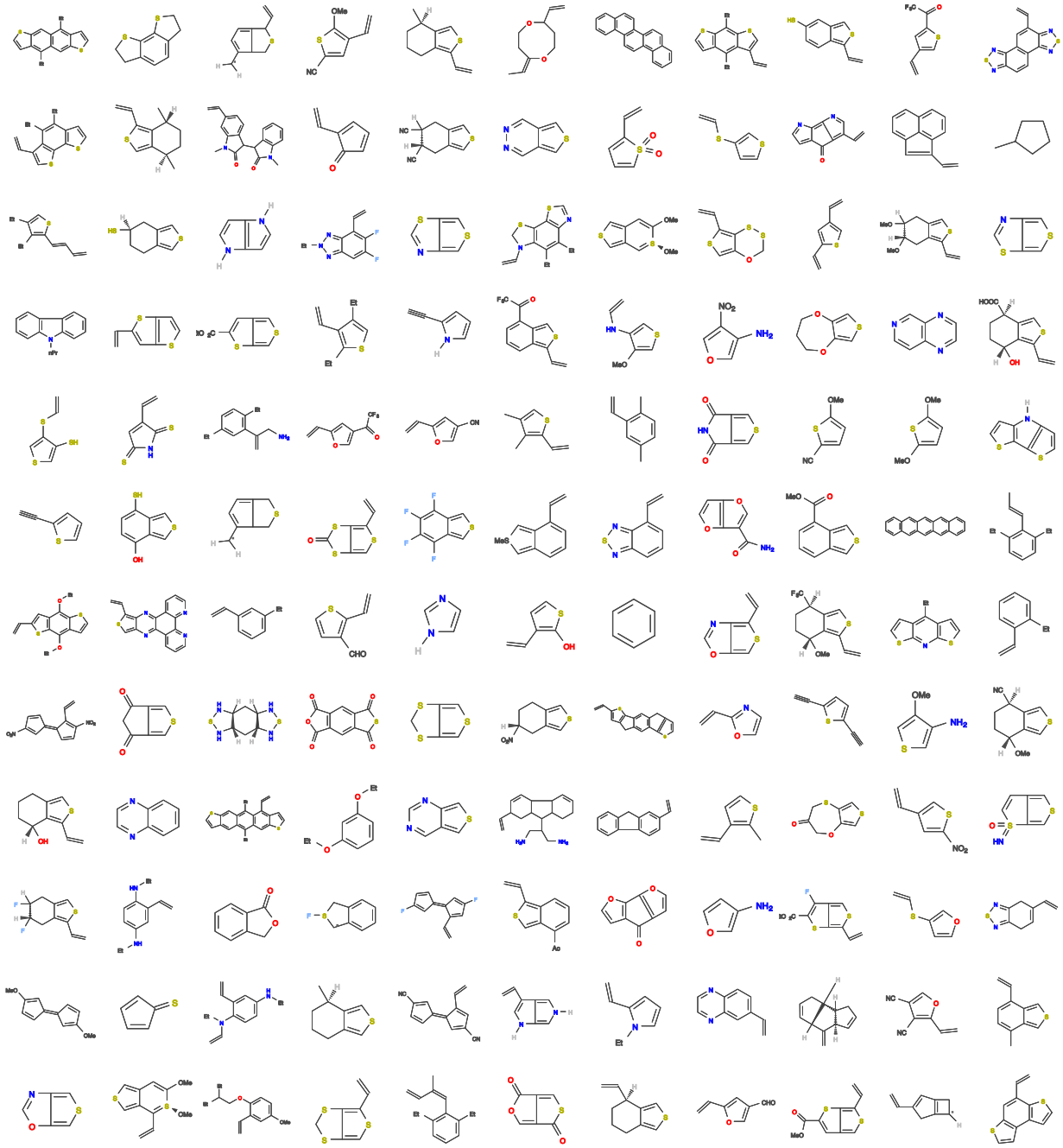


Figure B19: Molecules in the 1235 monomer dataset (2 of 10).

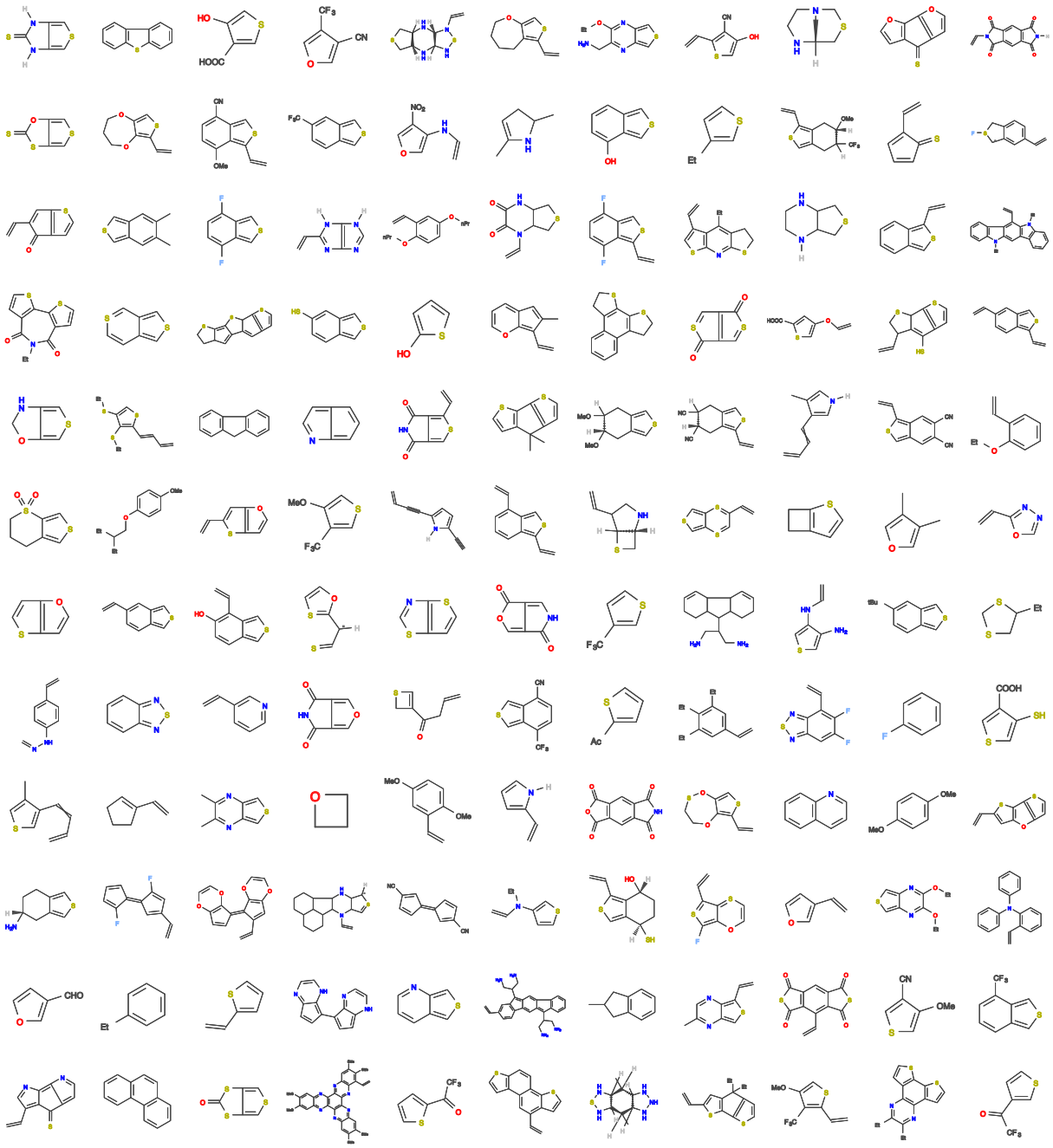


Figure B20: Molecules in the 1235 monomer dataset (3 of 10).

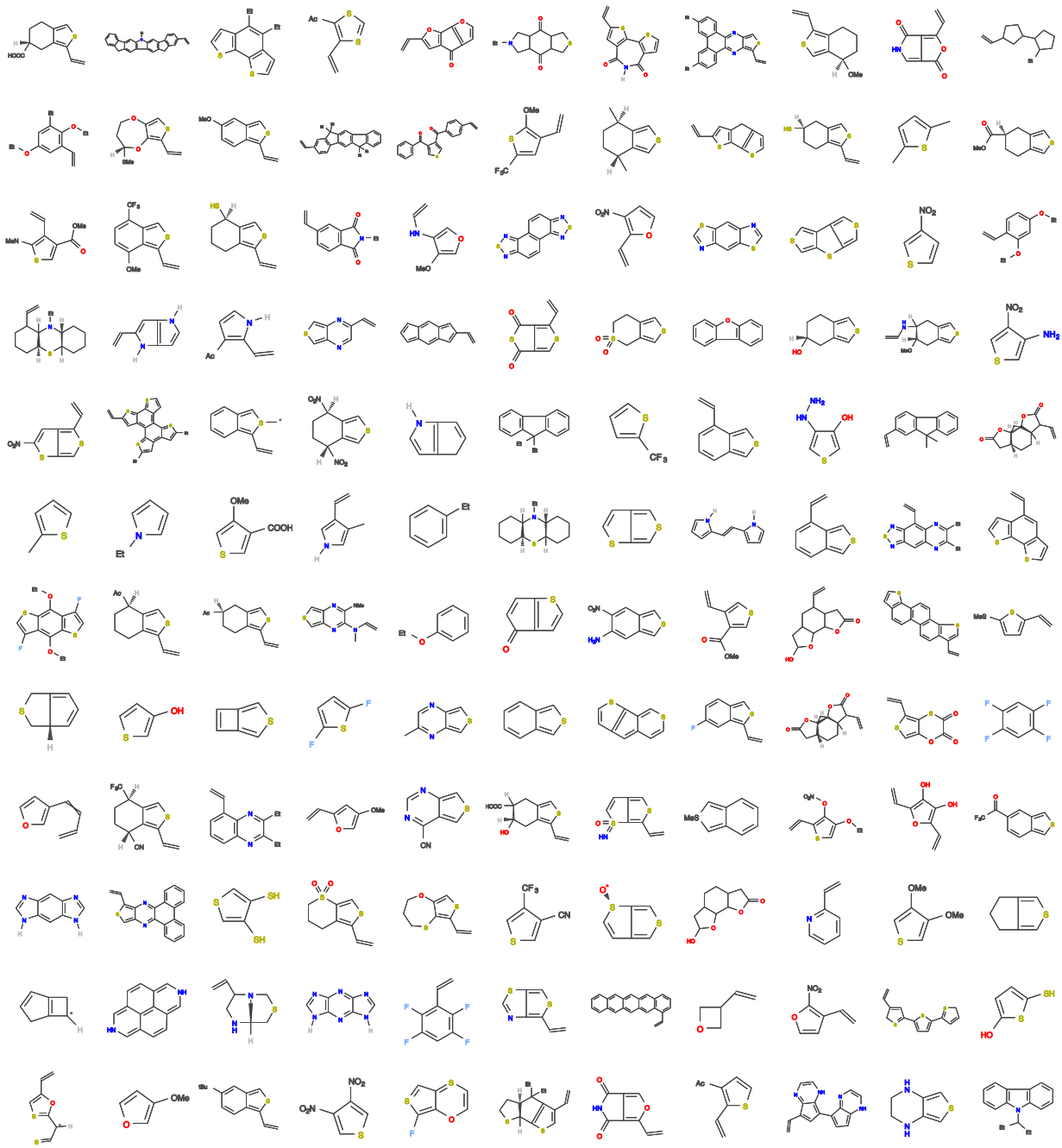


Figure B21: Molecules in the 1235 monomer dataset (4 of 10).

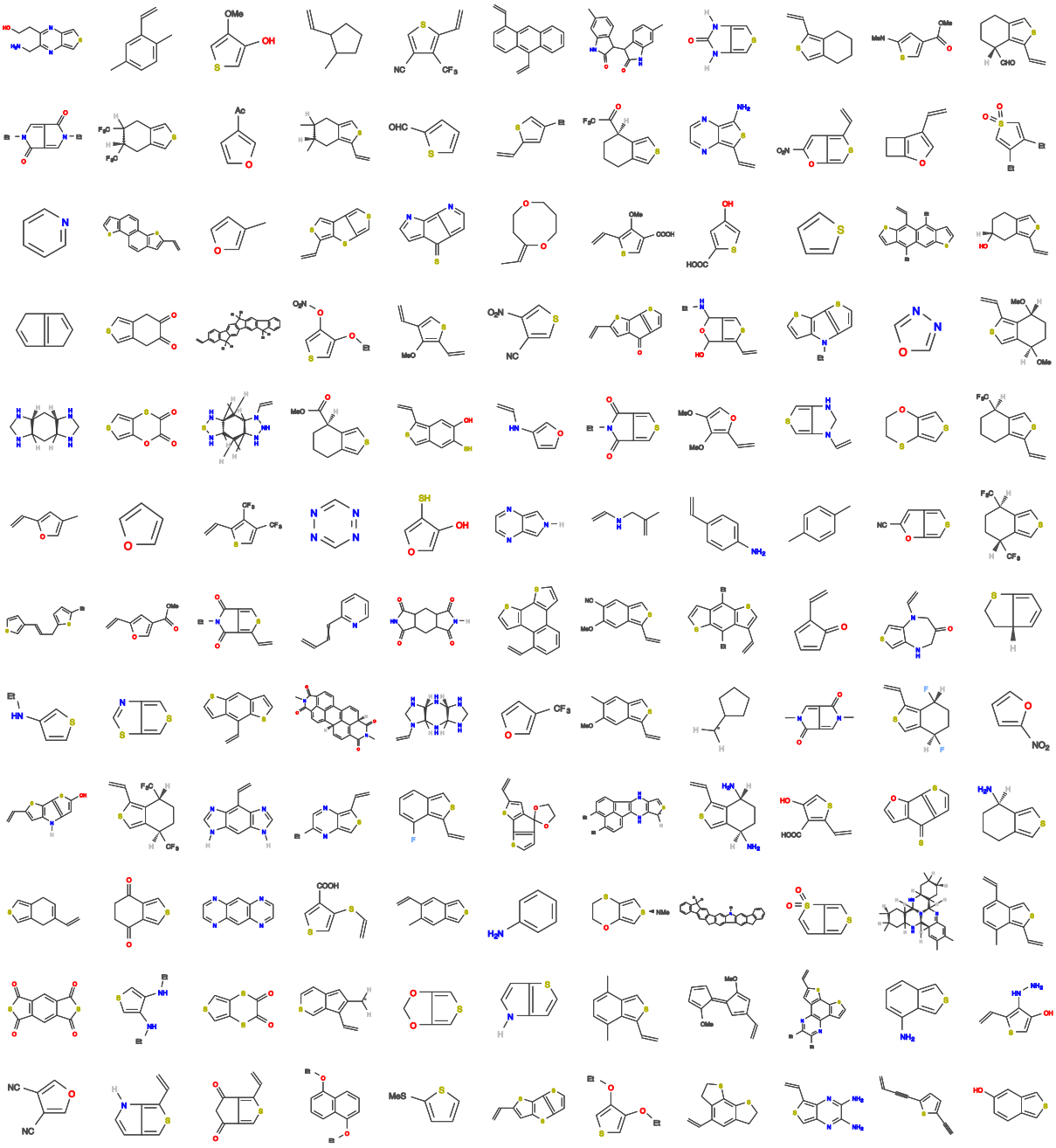


Figure B22: Molecules in the 1235 monomer dataset (5 of 10).

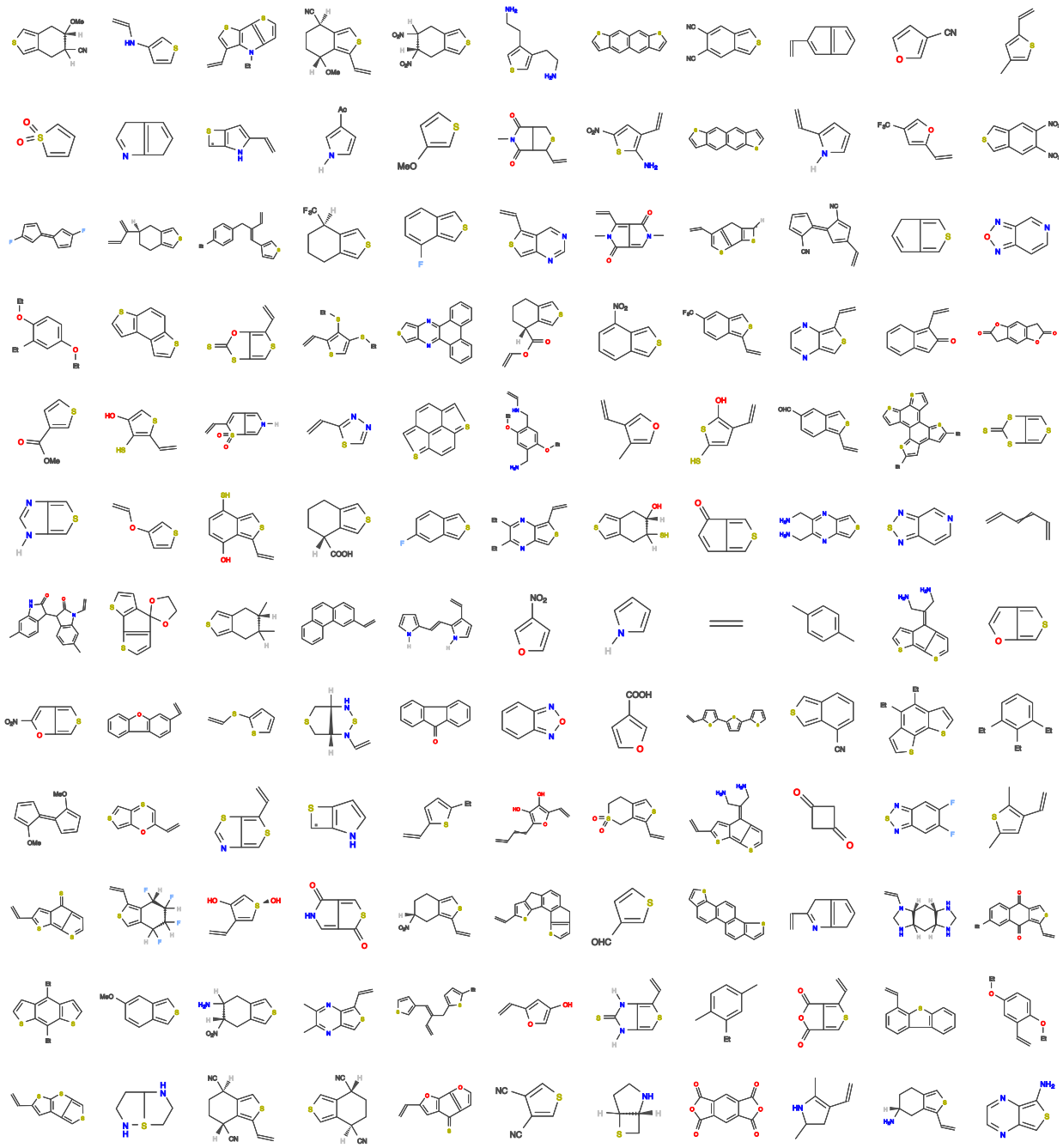


Figure B23: Molecules in the 1235 monomer dataset (6 of 10).

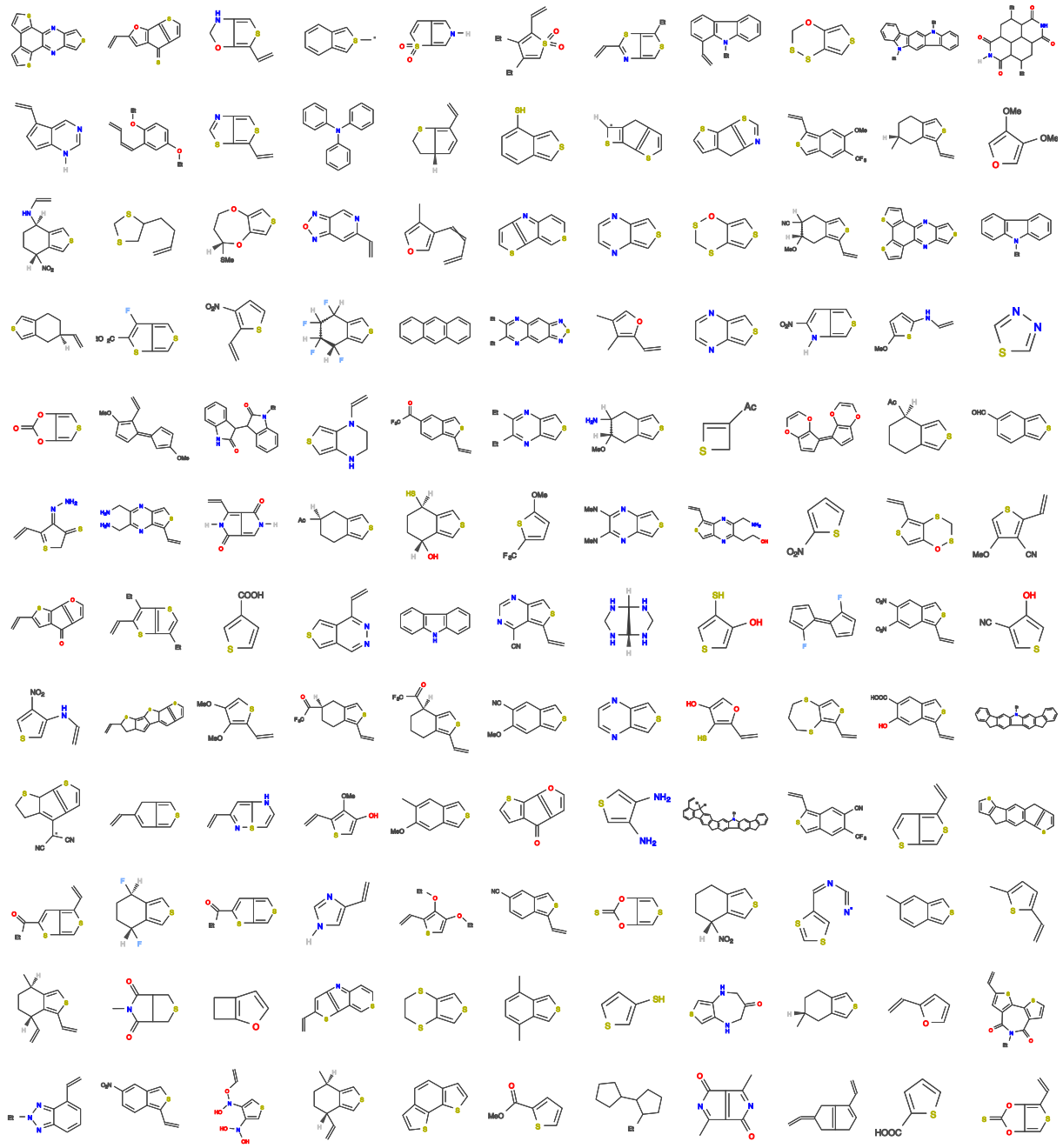


Figure B24: Molecules in the 1235 monomer dataset (7 of 10).

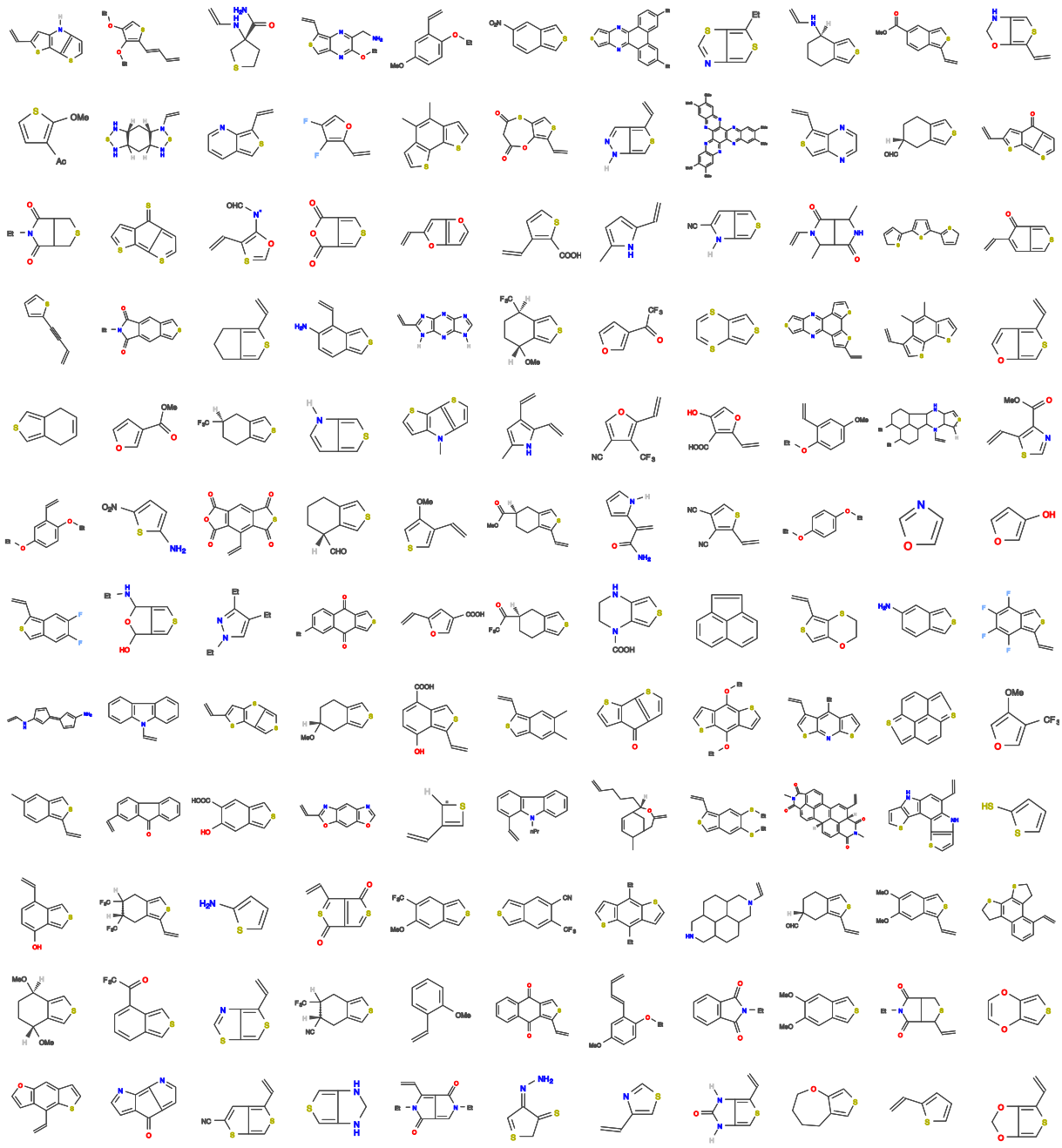


Figure B25: Molecules in the 1235 monomer dataset (8 of 10).

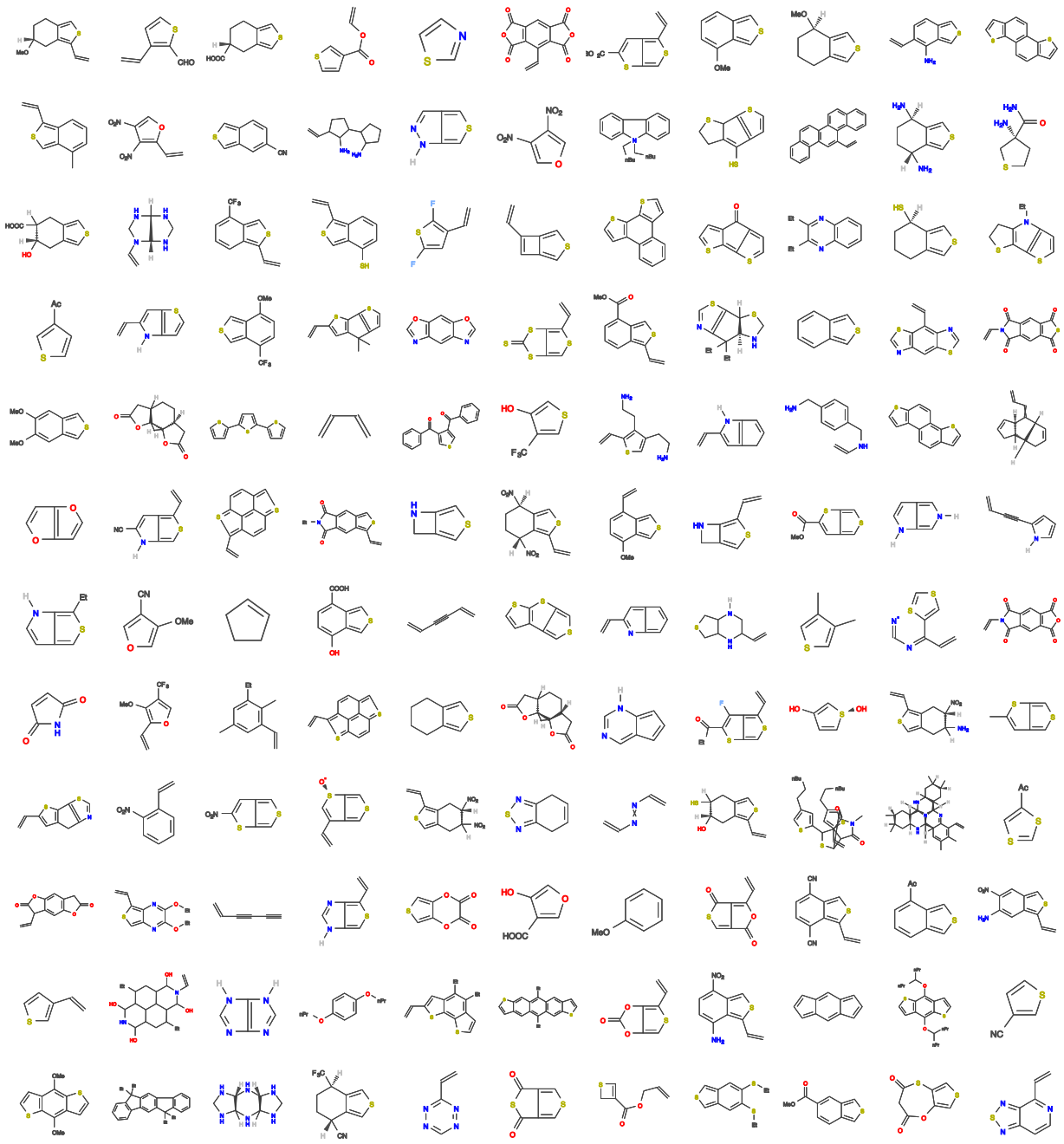


Figure B26: Molecules in the 1235 monomer dataset (9 of 10).

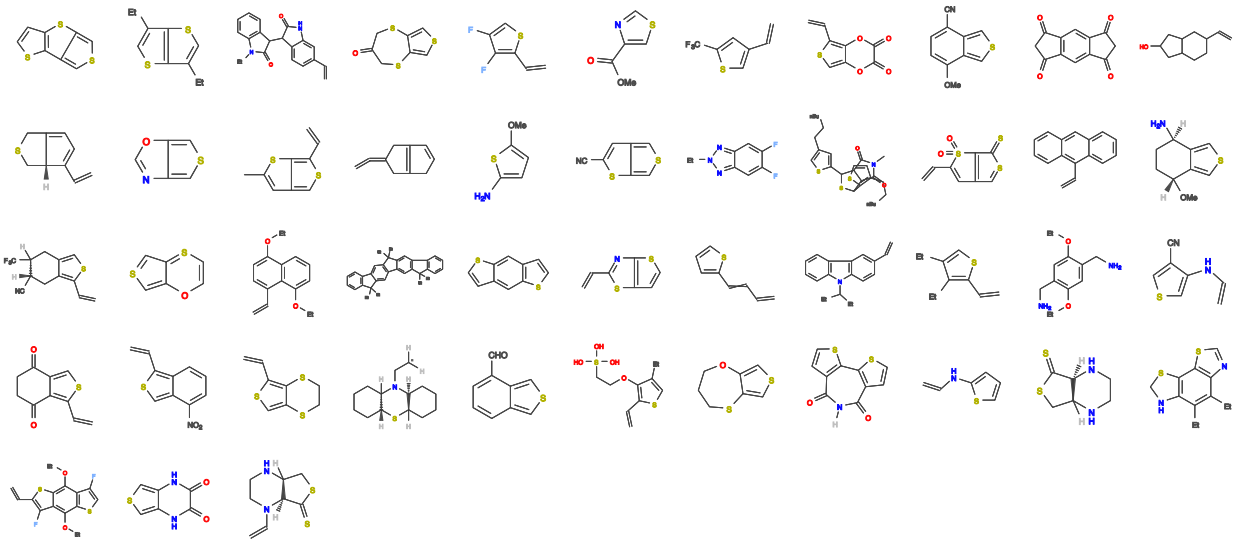


Figure B27: Molecules in the 1235 monomer dataset (10 of 10).

C=C	c1occ(N(=O)=O)c1NC#C	c1sc(c2c1cc(N(=O)=O)cc2)C#C
C=CC#C	c1sc(c2c1cc(C=C)cc2)	c1sc(c2c1c(C(=O)O)ccc2O)C=C
C1OCC1	c1c(F)c(F)c(c(c1F)F)	c1oc2c(n1)cc1c(c2)oc(n1)C=C
C#CC#C	c1scc2c1sc1c2csc1C=C	c1sc(c2c1nc(OCC)c(CN)n2)C#C
C=CC=C	c1c2c(nccc2)c(s1)C#C	C1=C2C(C(=O)N1)=C(OC2=O)C=C
c1cCcc1	Nc1sc(N(=O)=O)cc1C=C	c1sc(c(c1)C(=O)C(F)(F)F)C#C
C(=C)N=N	c1c(OCC)cc(OCC)c(c1)	c1ccc(c2c1nc(CC)c(CC)n2)C=C
C(=C)C#C	c1c2ccccc2c(c1=O)C=C	c1sc(c2c1C[C@H](C(=O)C)CC2)
c1occc1S	c1sc(c2c1c(C#N)ccc2)	c1sc(c2c1[C@H](S)CC[C@H]2O)
c1occc1N	c1c(CC)c(CC)c(CC)cc1	c1sc2c(c1)c1c(scc1)c(c2)C=C
c1sccc1O	c1[nH]c(C=C)c(c1)C#C	c1sc(c2c1c(N(=O)=O)ccc2)C=C
c1sccc1N	c1sc2c(c1CC)sc(c2CC)	c1sc(c2c1[nH]c(=S)[nH]2)C#C
c1sccc1S	c1csc2c1cc1c(ccs1)c2	c1sccc1/C=C(/Cc1ccc(CC)cc1)
c1oc(nn1)	c1[nH]c(C=C)c(c1)C=C	c1sc(cc1)c1ccc(s1)c1sc(cc1)
Sc1cscc1S	c1c2C(=O)OCc2ccc1C=C	C1=Cc2cc3=CC(=Cc3cc2=C1)C#C
c1oc(nc1)	c1sc(c2c1SCC(=O)CO2)	c1sc(c(c1)C(=O)C(F)(F)F)C=C
C#CC#CC=C	c1sc(c2c1cc[nH]2)C#C	c1sc(c2c1C(=O)c1ccccc1C2=O)
C#CC#CC#C	c1sc(c2c1cc[nH]2)C=C	c1sc2c(c1)n(C)c1c2sc(c1)C=C
c1oc(cc1)	c1scc2c1C=C(C2=O)C=C	C1=C2C(C(=O)S1)=CN(C2=O)C#C
c1sc(nn1)	c1sc(c2c1[nH]nc2)C=C	c1c2c(nccn2)c(c2c1nccn2)C#C
C1OCC1C#C	c1[nH]c(c(c1)C(=O)C)	c1sc(c2c1nc(CCO)c(CN)n2)C#C
Nc1cscc1N	c1sc(c(c1SCC)SCC)C=C	c1sc(c2c1nc(CCO)c(CN)n2)C=C
c1scc(n1)	c1scc(N(=O)=O)c1NC=C	C1=[S]C(=O)C2=C1C(=O)[S]=C2
C=CC=CC#C	c1csc2c1cc(c1c2scc1)	c1csc2c1c(CC)c1c(n2)scc1C=C
c1SCSc1cc	c1sc(C(=O)O)c(c1)C=C	c1c2c(nccn2)c(c2c1nccn2)C=C
C1OCC1C=C	c1c(OCC)c(cc(c1)OCC)	c1sc2c(c1)c1c(scc1)c(c2)C#C
C=CC=CC=C	C1=C([CH]S1)C(=C)C#C	C1=C2C(C(=O)N1)=C(OC2=O)C#C
c1ccc(s1)	c1sc(c(c1O)C(F)(F)F)	c1sc(c2c1SC(=O)CC(=O)O2)C#C
c1cCcc1C#C	C1=C([CH]S1)C(=C)C=C	c1sc(c2c1sc(C(=O)CC)c2F)C#C
c1cCcc1C=C	c1sc(C(F)(F)F)cc1C=C	c(c(N)1)ccc1c(c(N)1)ccc1C=C
c1oc(cc1O)	c1[nH]c2[nH]c(nc2n1)	C1=Cc2cc3=CC(=Cc3cc2=C1)C=C
c1ccc(cc1)	c1oc(cc1C(F)(F)F)C=C	c1csc2c1c(CC)c1c(n2)scc1C#C
c1sc(cc1)S	c1sc(c2c1C(=O)SC2=O)	c1csc2c1c(C)c(C)c1c2scc1C#C
c1occc1C=C	C(=C)c1sc(C)c(c1)C=C	c1sc(c2c1cc(C#N)c(OC)c2)C#C
c1sc(cc1)N	c1sc(C(F)(F)F)cc1C#C	c1sc(c(c1C(F)(F)F)C(F)(F)F)
c1[nH]cnc1	N(CC)c1cscc1N(CC)C=C	C1=C2C(C(=O)O1)=C(OC2=O)C=C
c1cc(cnc1)	C1=CC=C(S1(=O)=O)C=C	c1sc2c(c1)C(C)(C)c1c2sc(c1)
c1nnc(nn1)	c1sc(NC)c(c1C(=O)OC)	c1sc(c2c1[C@H](C)CC[C@H]2C)
c1occc1SC=C	c1oc(c(c1C#N)C#N)C#C	c1ccc(c2c1nc(CC)c(CC)n2)C#C
c1occc1NC#C	c1scc(N(=O)=O)c1NC#C	c1sc(c2c1[C@H](N)CC[C@H]2N)
c1sccc1OC#C	N1c2cscc2N(CC(=O)C1)	c1sc(c2c1[C@H](C=C)CCC2)C#C
c1oc2CCc2c1	c1sc(C(=O)O)c(c1)C#C	c1sc(c2c1c(C(=O)O)ccc2O)C#C
c1sccc1NC#C	c1scc2c1c(ccc2OC)C#C	C1=C2C(C(=O)S1)=C(OC2=O)C=C
c1occc1SC#C	c1sc(c2c1c(C=C)ccc2)	c1ccc2c(c1)C(=O)c1c2ccc(c1)
c1c2CCc2sc1	C1=C2CSC[C@H]2C(=C1)	c1sc(c2c1[nH]c(=O)[nH]2)C#C

Table B28: List of SMILES for the 1759 monomer data set. (Part 1 of 14)

c1sccc1SC#C	c1sc(c2c1[nH]cn2)C=C	c1c2c(cccc2)cc2c1c(ccc2)C#C
C(=C)C#CC#C	c1sc(c(c1C#N)C#N)C=C	c1sc(c2c1[C@H](C(=O)C)CCC2)
c1oc(cc1OC)	C1=CC(=C2[C@H]1CCS2)	c(s1)cc(c(=S)2)c1c(s3)c2cc3
c1cc(SC)sc1	c1sc(c2c1[nH]cn2)C#C	c1sc(c2c1sc(C(=O)OCC)c2)C#C
C=C(C)CNC#C	c1sc(c2c1c(C=O)ccc2)	c1sc(c2c1c(C(=O)OC)ccc2)C=C
C(=C)C#CC=C	c1oc(cc1C(F)(F)F)C#C	c1sc(c2c1sc(C(=O)CC)c2F)C=C
C(=C)N=NC=C	c1sc(c2c1C(=O)OC2=O)	c1ccc(OCC)c2c1c(OCC)ccc2C=C
c1ccc(=O)c1	C1=CC=C(S1(=O)=O)C#C	c1sc(c2c1[C@H](C=C)CCC2)C=C
c1sccc1SC=C	C1S=C(/C(=N/N)/C1=S)	c1sc2c(c1)n(C)c1c2sc(c1)C#C
c1sccc1NC=C	c1sc(c2c1cc(C)s2)C#C	c(c(CC)c1)c(c(=C)cn)c(CC)c1
c1c(F)cccc1	c1sc(c2c1SCC(=O)CS2)	c(s1)cc(c(=O)2)c1c(s3)c2cc3
c1sccc1OC=C	c1sc(c2c1cc(C)s2)C=C	c1sc(c2c1CC[C@@H](OC)C2)C#C
C=C(C)CNC=C	c1sc(c(c1)C(=O)C)C#C	c1c(=O)oc2c1c(cc1cc(o)oc21)
C(=C)N=NC#C	c1[nH]c(c2c1nccn2)C=C	c1sc(c2c1c(C(=O)OC)ccc2)C#C
c1occc1NC=C	c1sc(c(c1N(=O)=O)C#N)	c1sc(c2c1cc(C(=O)OC)cc2)C#C
Nc1cscc1NC=C	c1sc(c(c1CC)CC)C=CC=C	c1c2c(sc(C(=O)OCC)c2F)c(s1)
c1occ(OC)c1N	c1sc(c2c1c(C=C)ccc2C)	c1sc(c2c1c(N(=O)=O)ccc2)C#C
c1[nH]c(cc1)	C(=N)/N=C(\C1=CS=CS1)	c1sc(c2c1[nH]c(=S)[nH]2)C=C
c1sc(nn1)C#C	C#Cc1[nH]c(cc1)C#CC=C	c1sc(c2c1[C@H](F)CC[C@H]2F)
c1scc(n1)C=C	c(oc1c(F)s2)csc1c2C=C	c1sc(c2c1nc(OCC)c(CN)n2)C=C
c1sc(cc1)C#C	c1sc(c2c1oc(=S)o2)C=C	c1sc(c2c1SC(=O)CC(=O)O2)C=C
c1oc(nc1)C=C	c1oc(c(c1)N(=O)=O)C=C	c1csc2c1c(C)c(C)c1c2scc1C=C
c1oc(cc1C=O)	C1=C(Oc2c(csc2)O1)C#C	c1sc(c2c1[nH]c(=O)[nH]2)C=C
Nc1cscc1NC#C	C1=Cc2[nH]c(cc2C1)C#C	c1sc(c2c1CC[C@@H](OC)C2)C=C
c1SCSc1ccC=C	c1sc(c2c1oc(=O)o2)C=C	c1sc(c2c1c(C(F)(F)F)ccc2OC)
c1oc(cc1)C=C	c1c(cc2c(c1)nccn2)C#C	c1sc(c2c1cc(N(=O)=O)c(N)c2)
c1sc(O)c(c1)	c1sc2c(n1)Cc1c2sc(c1)	C1=C2C(C(=O)S1)=CN(C2=O)C=C
c1sc(cc1)C=C	c1sc(c2c1ccc(F)c2)C#C	c1sc(c2c1sc(C(=O)OCC)c2)C=C
c1nc2CCNS2c1	c1sc(c2c1ncnc2C#N)C=C	c1sc(c(c1CC)OCCS(O)(O)O)C=C
c1sc(nn1)C=C	c1sc(c2c1cc(S)cc2)C#C	c1sc(c(c1N(=O)=O)N(=O)=O)C=C
c1c(C)cc(s1)	c1oc(c(c1C(F)(F)F)OC)	c1scc2c1C[C@@H](CC2)C(=C)C=C
c(c(C)1)ccc1	c1sc(c2c1cc(C)cc2)C=C	c1sc(c2c1cc(C(F)(F)F)cc2)C#C
c1scc(OC)c1N	c1oc(c(c1O)C(=O)O)C#C	c1sc(c2c1OCC[C@@H](SC)O2)C#C
c1oc(nn1)C=C	c1c2cscc2c(N)c(c1)C=C	c1c(O)sc2c1[nH]c1c2sc(c1)C#C
c1ccc(s1)C=C	C1=CC2=CC(=NC2=C1)C=C	c1sc(c2c1c(C(F)(F)F)ccc2C#N)
c1sc(C)c(c1)	c(cc1)c2cs(c)cc2c1C=C	c1sc(c2c1C[C@H](C(=O)OC)CC2)
Sc1cscc1SC#C	c1sc(c2c1cc(F)c(F)c2)	c1cc2c(C(=O)N(C2=O)CC)cc1C=C
c1oc(cc1)C#C	c1c(C)c(c2c1ccco2)C#C	c1sc2c(c1)C(=O)c1c2sc(c1)C=C
c1oc(nc1)C#C	c1sc(c2c1cc(C)c(C)c2)	c1sc(c2c1nc(OCC)c(OCC)n2)C#C
C=Cc1csc(c1)	c1sc(c2c1oc(=O)o2)C#C	c1sc(c2c1cc(C(C)(C)C)cc2)C=C
Sc1cscc1SC=C	c1sc(c(c1CC)CC)C=CC#C	c(s1)c2c(=O)n(C)c(=O)c2c1C=C
c1oc(cc1C#N)	c1sc(c(c1O)C(=O)O)C#C	c1cc2c(c3c1nsn3)cc(c1c2nsn1)
c1sccc1N(CC)	c1sc(c2c1sc(=S)o2)C#C	c1cc(c2c(c1)c1c(s2)cccc1)C=C
c1ccc(s1)C#C	c1sc(c2c1c(C)ccc2)C#C	c1oc2c(c1)C(=O)c1c2oc(c1)C=C
c1oc(nn1)C#C	c1sc(c2c1sc(=O)s2)C#C	c1sc(c2c1nc(OCC)c(OCC)n2)C=C

Table B29: List of SMILES for the 1759 monomer data set. (Part 2 of 14)

c1c(C)sc(C)c1	c1scc2c1C=C(S2(=O)=O)	c1sc(c(c1N(=O)=O)N(=O)=O)C#C
c1sc(c(c1F)F)	c1sc(c2c1cc(C)cc2)C#C	c1sc(c2c1c(N(=O)=O)ccc2N)C=C
C1=CC=C(C1=O)	c1sc(c(c1)N(=O)=O)C#C	N1[C@@H]2CSC(=S)[C@H]2N(CC1)
c1c(F)sc(F)c1	c(oc1c(F)s2)csc1c2C#C	c(s1)c2c(=O)n(C)c(=O)c2c1C#C
c1c(S)sc(O)c1	c(cc1)c2cC(C)cc2c1C=C	N(CC)c1ccc(c(c1)C=C)N(CC)C=C
c1sc(c(c1)OC)	c1c(C)cc(CC)c(C)c1C#C	c1c(OC)cc(c(c1)OCC(CC)CC)C=C
c1occ(C)c1C=C	C1SC[C@](C1)(C(=O)N)N	c1sc(c2c1cc(C#N)c(C#N)c2)C#C
c1ccc(cc1)C=C	c1scc2c1c1c(s2)sc(c1)	c1c2n(CC)c3c(c2ccc1)cccc3C=C
c1scc(C)c1C=C	c1c(F)c(F)c(c2c1nsn2)	c1sc(c2c1c(F)c(F)c(F)c2F)C#C
c1c(C)Nc(C)c1	c1oc(c(c1)N(=O)=O)C#C	c1oc2c(c1)C(=O)c1c2sc(c1)C=C
c1c(CC)cc(s1)	C1Oc2cscc2OC(C1=O)C=C	c1cc(c2c(c1)c1c(s2)cccc1)C#C
c1oc(cc1O)C#C	c1c2c(ncc(CC)n2)c(s1)	c1sc(c2c1CC[C@@H](C=O)C2)C=C
c1cnc(cc1)C=C	c1sc(c2c1cc(S)c(O)c2)	C1=C(CC)C(CC)=C(S1(=O)=O)C#C
c1oc(c(c1F)F)	C/C=C\1/OCCC(OCC1)C=C	c1c2c(ccc3c2ccs3)c2c(ccs2)c1
c1[nH]cnc1C#C	c1c(cc2c(c1)nccn2)C=C	c1sc2c(c1)C(=S)c1c2oc(c1)C=C
c1sc(c2c1CN2)	c1c(C)c(c2c1ccco2)C=C	c1c(F)c(F)c(c2c1nn(CC)n2)C=C
c1nnc(nn1)C#C	c1scc2c1c(c(cc2)O)C=C	c1cc2c(cc1)c1c(C2)cc(cc1)C#C
c1sc(cc1)NC=C	c1[nH]c(c2c1nccn2)C#C	c1cc2c(cc1)c1c(C2)cc(cc1)C=C
c1oc(c(c1C)C)	c1c2nc(C)c(C)nc2c(s1)	c1sc(c2c1cc(C(C)(C)C)cc2)C#C
c1cc(cnc1)C#C	c1c(OC)c(cc(c1)OC)C=C	c1csc2c1c(CC)c(CC)c1c(csc21)
c1occc1C=CC#C	c1sc(c2c1ncnc2C#N)C#C	c1sc(c2c1OCC[C@@H](SC)O2)C=C
C1SC=C1C(=O)C	c1sc(c2c1ccc(F)c2)C=C	c1sc(c2c1cc(SCC)c(SCC)c2)C#C
c1ccc(cc1)C#C	c1scc2c1c(c(cc2)N)C=C	c1sc(c2c1cc(SCC)c(SCC)c2)C=C
c1oc(cc1C)C#C	C1Oc2cscc2OC(C1=O)C#C	c1c2ccsc2cc2c1c(c1c(c2)scc1)
c1sc(cc1)SC#C	c1cc(c2c1[nH]cnc2)C=C	c1oc2c(c1)C(=O)c1c2oc(c1)C#C
c1oc(c(c1O)S)	c1sc(c2c1c(S)ccc2)C#C	c1oc2c(c1)C(=S)c1c2oc(c1)C#C
c1scc(C#N)c1N	c1sc(c2c1C(=O)CCC2=O)	C1=C(CC)C(CC)=C(S1(=O)=O)C=C
c1c(N)ccc(c1)	C1S[C@@H]2[C@@H]1NCC2	c1oc(c(c1N(=O)=O)N(=O)=O)C#C
c1sc(cc1)NC#C	c1sc(c(c1)N(=O)=O)C=C	c1c(c2c(s1)c1c(n2CC)CCS1)C=C
c1cscc1C(=O)O	c1sc(CC)c2c1nc(s2)C=C	N(CC)c1ccc(c(c1)C=C)N(CC)C#C
c1occc1C=CC=C	c1c(OC)c(cc(c1)OC)C#C	c1sc(c2c1c(C(F)(F)F)ccc2)C=C
C1=CC=C(C1=S)	c(cc1)c2cs(c)cc2c1C#C	c1scc2c1C[C@@H](CC2)C(=C)C#C
c1c(nc2cSc12)	c1sc(c(n1)C(=O)OC)C=C	c1c(O)sc2c1[nH]c1c2sc(c1)C=C
c1n(CC)c(cc1)	c1scc2c1N(CCN2C(=O)O)	c1sc(c2c1cc(C#N)c(C#N)c2)C=C
c1oc(cc1O)C=C	c(s(c)1)c2ccccc2c1C#C	c1sc(c2c1CC[C@@H](C=O)C2)C#C
c1nc2cscc2nc1	c1sc(c2c1c(S)ccc2)C=C	c1sc(c2c1cc(C(F)(F)F)cc2)C=C
c1[nH]cnc1C=C	c1sc(c2c1sc(=S)s2)C=C	c1c2c3c(s1)ccc1c3c(cc2)c(s1)
c1sc(c(c1C)C)	c1scc2c1c(c(cc2)N)C#C	C(=[N])/N=C(\C1=CSC=[S]1)C#C
c1sc(c(c1O)S)	c1sc(c2c1nc(N)c(N)n2)	c1sc(c2c1c(C(F)(F)F)ccc2)C#C
c1nnc(nn1)C=C	n1c2cscc2n(c(=O)c1=O)	c1oc2c(c1)C(=S)c1c2oc(c1)C=C
c1sc(c2c1ncs2)	c1c2cscc2c(N)c(c1)C#C	c1sc(c2c1c(F)c(F)c(F)c2F)C=C
C1=C(c2cscc12)	c1sc2c(c1)sc1cc(sc21)	c1sc(c2c1c(N(=O)=O)ccc2N)C#C
c1oc(cc1OC)C=C	c1sc(c2c1cc(S)cc2)C=C	c1sc(c2c1CC[C@@H](C(=O)O)C2)
c1cc(C)cc(c1C)	c1sc(c2c1sc(=O)s2)C=C	c1c(F)c(F)c(c2c1nn(CC)n2)C#C
c1oc2CCc2c1C=C	C1=C(CC2=C1C[CH]2)C=C	c1oc(c(c1N(=O)=O)N(=O)=O)C=C

Table B30: List of SMILES for the 1759 monomer data set. (Part 3 of 14)

c1sc(c2c1OCO2)	c1sc(c(c1OC)C(F)(F)F)	c1ccc(c2c1c1c(c3c2ccs3)scc1)
c(nc1c2)cs1nc2	C1=C(Cc2c(C1)nsn2)C#C	c1sc(c2c1[C@H](C(=O)OC)CCC2)
C1SC=C1C(=O)OC	C/C=C\1/OCCC(OCC1)C#C	c1sc(c2c1[C@H](N)CC[C@H]2OC)
c1oc(cc1OC)C#C	C1=C(Oc2c(csc2)O1)C=C	c1csc2c1c(CC)c1c(c2CC)scc1C=C
c1ccc(=O)c1C=C	c1sc(c2c1c(C)ccc2)C=C	c1cc2c(s1)c1c([nH]2)cc(s1)C#C
c(s1)c2ncsc2c1	c1sc(c2c1oc(=S)o2)C#C	n1c(C)c2c(c1=O)c(C)n(c2=O)C#C
c1c2CCc2sc1C#C	c(s(c)1)c2ccccc2c1C=C	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)
c1sc(c(c1)C#N)	C1=C(Cc2c(C1)nsn2)C=C	c(s1)c2c(=O)N(CC)c(=O)c2c1C#C
c(c1)cc(cc)cc1	c1sc(c2c1sc(=S)s2)C#C	c(s1)c2c(=O)N(CC)c(=O)c2c1C=C
c1sc(c2c1scn2)	c1cnc2c1c(=S)c1c2ncc1	c1c2[nH]cnc2c(c2c1[nH]cn2)C=C
c1sc(c2c1SCS2)	c1csc2c1n(CC)c1c2scc1	c1cc2c(s1)c1c([nH]2)cc(s1)C=C
c1oc2cc(oc2c1)	c1sc(CC)c2c1c(c[nH]2)	c1n(C)c(=O)c2c1c(=O)n(C)c2C=C
N1c2cscc2N(C1)	c(cc1)c2cC(C)cc2c1C#C	c1[nH]cc2c1S(=O)(=O)C(=C2)C=C
c1sc(c2c1nco2)	c1sc(c(c1O)C(=O)O)C=C	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)
c1sc(c2c1ocn2)	c1sc(c(c1C)S[CH2])C=C	n1c(C)c2c(c1=O)c(C)n(c2=O)C=C
c1sc(C=O)c(c1)	c1sc(c2c1sc(=S)o2)C=C	c1sc(c2c1C[C@H](C(F)(F)F)CC2)
c1sc(c2c1scc2)	c1csc2c1C(S)=c1c2scc1	c1[nH]c2nc3[nH]c(nc3nc2n1)C#C
c1c(CC)c(ccc1)	c1c(C)cc(CC)c(C)c1C=C	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)
c1sc(c(c1O)OC)	c1cccc2c1n(c1c2cccc1)	c1[nH]c(=O)c2c1c(=O)[nH]c2C#C
c1c(F)cccc1C#C	c1oc(c(c1O)C(=O)O)C=C	C1C2CC(=C)O[C@H](C1C=CC2C)CCC
c1sc(c(c1)C=O)	c1[nH]cc2c1[nH]cc2C=C	c1c2[nH]cnc2c(c2c1[nH]cn2)C#C
c1cc(SC)sc1C=C	C1=[S]C(=C(O1)[N]C=O)	c1sc(c2c1[nH]c(N(=O)=O)c2)C=C
c1c(OC)c(ccc1)	C1=Cc2[nH]c(cc2C1)C=C	c1scc2c1C(=[S@@](OC)C(=C2)OC)
c1c(F)cccc1C=C	c1scc2c1c(c(cc2)O)C#C	c1c(c2c(s1)nc1c(c2CC)CCS1)C=C
c1c2CCc2sc1C=C	c1scc2c1c1c(s2)cc(s1)	c1sc(c2c1[nH]c(N(=O)=O)c2)C#C
C=NNc1ccc(cc1)	c1sc2c(c1)C(=O)C(=C2)	c1c2c3CCSc3c3SCCc3c2c(cc1)C#C
c(s1)c2NCOc2c1	c1sc(c(n1)C(=O)OC)C#C	c1c2c3CCSc3c3SCCc3c2c(cc1)C=C
c1cc(CN)ccc1CN	c1c2c(ccs1)nc1c2sc(c1)	c1sc(c2c1cc(C(=O)O)c(O)c2)C=C
c(s1)c2scnc2c1	c1oc2c(C(=O)N)c(oc2c1)	c1n(C)c(=O)c2c1c(=O)n(C)c2C#C
c1oc2cc(sc2c1)	C(=S)[CH]C1=[S]C=C(O1)	c1[nH]cc2c1S(=O)(=O)C(=C2)C#C
c1sc(c2c1CCC2)	c1sc(OC)c(C(=O)C)c1C#C	c1cc2c(s1)c1c(C2(CC)CC)cc(s1)
c1oc2CCc2c1C#C	c1sc(c2c1cc(OC)cc2)C=C	c1c2c3c(s1)ccc1c3c(cc2)sc1C=C
c1cc(SC)sc1C#C	C1=c2ccccc2=C(C1=C)C=C	c1sc(c2c1CCC[C@H]2N(=O)=O)C=C
c1sc(c2c1occ2)	c1scc2c1C(=C[S@@]2[O])	c1sc(c2c1CCC[C@H]2N(=O)=O)C#C
c1sc2nc(sc2c1)	c1c(OCC)cc(CC)c(OCC)c1	c1sc(c2c1CC[C@@H](N(=O)=O)C2)
c1c(sc(OC)c1)N	c1c([CH2])c(c2c1ccsc2)	C1=[S]C(=S)C2=C1C=C(S2(=O)=O)
c1c2nsnc2c(cc1)	c(c(CC)1)ccc1c(c1)ccc1	c1cc(CC)c(c(c1)CC)/C=C(/C)C=C
c1sc(cc1)C#CC=C	c1c(OC)cc(c(c1)OCC)C=C	c1sc(c2c1cc(C(=O)O)c(O)c2)C#C
c1sc(c(c1OC)OC)	c1sc(c2c1[nH]c(C#N)c2)	c1sc(c2c1[C@H](OC)CC[C@H]2OC)
c1c(ncc2c1non2)	c1sc(c(c1OCC)ON(=O)=O)	c1[nH]c(=O)c2c1c(=O)[nH]c2C=C
c1c2nonc2c(cc1)	c1sc(c2c1c(S)ccc2O)C#C	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)
c1oc(cc1C(=O)O)	c1c2c(ncs2)c(c2c1ncs2)	c1sc(c2c1[C@H](C)CC[C@H]2C=C)
c1sc(c2c1OCSS2)	c1sc(c2c1c(C)ccc2C)C#C	c(s1)c2c(=O)n(CC)c(=O)c2c1C#C
c1sc(c2c1SCCS2)	NCc1cc(OCC)c(cc1OCC)CN	c1c2c3c(s1)ccc1c3c(cc2)sc1C#C
c1sc(c2c1nccn2)	N1[C@H]2CSC[C@H]2N(S1)	c1csc2c1c(CC)c1c(c2CC)scc1C#C

Table B31: List of SMILES for the 1759 monomer data set. (Part 4 of 14)

c1sc(cc1)C#CC#C	c1sc(OC)c(C(=O)C)c1C=C	C1=C(C=C/C/1=C\1/C=CC(=C1)N)N
c1scc(OC)c1NC=C	c1sc(c2c1C[C@H](S)CC2)	c1sc(c2c1[C@H](C(F)(F)F)CCC2)
c1sc(O)c(c1)C#C	c(s1)c2ccs(=N)(=O)c2c1	c1sccc1/C=C(/Cc1sc(CC)cc1)C=C
c1oc(cc1C#N)C=C	c(o1)c2c(=O)Nc(=O)c2c1	c1cc2c(s1)c1c(C32OCCO3)cc(s1)
c1sc(O)c(c1)C=C	c(s1)c(CCN)c(CCN)c1C#C	c1[nH]c2nc3[nH]c(nc3nc2n1)C=C
C#Cc1sc(cc1)C#C	c1sc(c(c1C#N)C(F)(F)F)	c1c2n(CCC)c3c(c2ccc1)cccc3C=C
c1oc(cc1C=O)C=C	c1sc(c2c1oc(C#N)c2)C=C	c(s1)c2c(=O)n(CC)c(=O)c2c1C=C
c1sc(C)c(c1)C#C	c1c(C(F)(F)F)sc(C#N)c1	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)
c1c(oc2cscc2s1)	c1sc(c2c1c(C)ccc2C)C=C	c(c(CC)c1)c(c(=C)cn)c(CC)c1C=C
c1scc2c1C=C(C2)	C1=CC2=C(C1)C=C(C2)C=C	c1sc(c2c1cc(N(=O)=O)c(N)c2)C#C
c1oc(c(c1OC)OC)	c1sc(c2c1sc(C#N)c2)C#C	c(c(CC)c1)c(c(=C)cn)c(CC)c1C#C
c1sc(c2c1OSCS2)	C1=[S]c2cscc2[S]=C1C=C	c1sc(c2c1[C@H](C(=O)C)CCC2)C=C
c1oc(cc1C=O)C#C	c(s1)c(CCN)c(CCN)c1C=C	c1sc(c2c1[C@H](F)CC[C@H]2F)C=C
c1sc(cc1)C=CC#C	c1sc(c2c1cc(C)c(OC)c2)	c1sc(c2c1C[C@H](C(=O)C)CC2)C=C
c1nc2CCNS2c1C=C	c1sc(c(c1)C(F)(F)F)C=C	c1sc(c(c1C(F)(F)F)C(F)(F)F)C#C
c1sc(c2c1OCCO2)	c1sc(c2c1CCC[C@H]2C=O)	c1sc(c2c1c(C(F)(F)F)ccc2OC)C#C
c1sc(c(c1CC)CC)	c(s1)c2C(NCC)OC(O)c2c1	c(s1)cc(c(=O)2)c1c(s3)c2cc3C=C
c1c2nccnc2c(s1)	C1=CC2=C(C1)N=C(C2)C=C	c1sc(c2c1C[C@H](F)[C@@H](F)C2)
c1sccc1N(CC)C=C	c1cc(cc2c1ccc1c2cccc1)	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)
c1oc(cc1C#N)C#C	c1sc(c2c1cc(OC)cc2)C#C	c1sc(c2c1[C@H](C)CC[C@H]2C)C#C
c1c(OCC)c(ccc1)	c1sc(c2c1[C@H](C)CCC2)	C1=[S]C(=O)C2=C1C(=O)[S]=C2C=C
c1c(OC)sc(OC)c1	c1sc(c2c1sc(C#N)c2)C=C	c(s1)cc(c(=S)2)c1c(s3)c2cc3C#C
c(c(C)1)ccc1C#C	c1sc(c2c1c(F)ccc2F)C#C	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1
N1c2cscc2N(CC1)	c1sc(c2c1c(F)ccc2F)C=C	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)
c1occ(OC)c1NC#C	c1c(C(=O)C(F)(F)F)scc1	c1sc(c2c1[C@H](N)CC[C@H]2N)C=C
c1[nH]c(cc1)C=C	c1sc(c2c1C=[S]C=C2)C#C	c1sc(c2c1[C@H](N)CC[C@H]2N)C#C
c1c(C)cc(s1)C#C	c1cc2c(cs1)c1c(ccs1)c2	c1sccc1/C=C(/Cc1ccc(CC)cc1)C=C
c1sc(c2c1CCCC2)	c1sc(c2c1c(S)ccc2O)C=C	c1sc(c2c1C[C@H](C)[C@@H](C)C2)
c1sc(c2c1OCCS2)	C1=CC2=C(C1)N=C(C2)C#C	c(s1)cc(c(=S)2)c1c(s3)c2cc3C=C
C=Cc1csc(c1)C=C	c1sc(c2c1oc(C#N)c2)C#C	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)
c1sc(cc1)C=CC=C	c1sc2c(c1)[nH]c(c2)C#C	c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1
c1c2cccc(c2cs1)	C1=CC2=C(C1)C=C(C2)C#C	c1sc(c2c1C[C@H](C(=O)C)CC2)C#C
c(c(C)1)ccc1C=C	c1oc(cc1C(=O)C(F)(F)F)	c(c1)c2nccnc2c1c(c1)c2nccnc2c1
c1sc(CC)c(c1CC)	c1c(OCC)c(cc(c1)OC)C=C	c1sc(c2c1cc(N(=O)=O)c(N)c2)C=C
c1occ(OC)c1NC=C	c1sc(c(c1C(=O)O)OC)C#C	c1sc(c2c1C(=O)c1ccccc1C2=O)C=C
c1c2cscc2c(nn1)	c1sc(c2c1[C@H](S)CCC2)	c1c2c(sc(C(=O)OCC)c2F)c(s1)C#C
c1sc(C)c(c1)C=C	c1cc(N(=O)=O)c(cc1)C=C	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)
c1[nH]c(cc1)C#C	c1scc(N(O)O)c1N(O)OC=C	c1sc(c2c1[C@H](C(=O)C)CCC2)C#C
C=Cc1csc(c1)C#C	c1c2c(SCC2)c2c(CCS2)c1	c1sc(c2c1[C@H](C)CC[C@H]2C)C=C
c1sc(c2c1ncnc2)	c1sc(c2c1C=[S]C=C2)C=C	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)
c1cnc(c2c1nsn2)	c1c(C)cc(c(c1)C)C=CC=C	c1c(=O)oc2c1c(cc1cc(o)oc21)C#C
c1oc(cc1C(=O)C)	c1oc(c(c1C#N)C(F)(F)F)	c1sc(c2c1[C@H](S)CC[C@H]2O)C=C
c1c(C(=O)C)scc1	C=Cc1oc(c(c1O)O)C=CC#C	c1sc(c2c1c(C(F)(F)F)ccc2OC)C=C
n1c(C)cc(c1C=C)	c1ccc(c2c1nn(CC)n2)C=C	c1sc(c2c1C(=O)c1ccccc1C2=O)C#C
c1sc(c(c1O)C#N)	c1sc(c2c1c(C#N)ccc2OC)	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)

Table B32: List of SMILES for the 1759 monomer data set. (Part 5 of 14)

c1sc(c(c1C)C)C=C	C=Cc1oc(c(c1O)O)C=CC=C	c1ccc2c(c1)C(=O)c1c2ccc(c1)C#C
c1c(CC)sc(c1)C#C	c1sc(c(c1C(=O)O)OC)C=C	c1c2c(ccs2)c(c2c1cc1c(c2)scc1)
C1=CC=C(C1=S)C=C	c1cc(N(=O)=O)c(cc1)C#C	c1scc2c1C[C@@H]([C@@H](OC)C2)N
c1oc(c(c1F)F)C#C	C1=C(C(=C[S@@]1O)O)C=C	c1sc(c2c1[C@H](S)CC[C@H]2O)C#C
c1sc(c(c1OC)C#N)	c1c(C)cc(c(c1)C)C=CC#C	c1c2c(sc(C(=O)OCC)c2F)c(s1)C=C
c1[nH]cc(c1C)C=C	c1sc(c(c1)C(F)(F)F)C#C	c(s1)cc(c(=O)2)c1c(s3)c2cc3C#C
c1c(C)cc(c(c1)C)	c1ccc(c2c1nn(CC)n2)C#C	c1ccc2c(c1)C(=O)c1c2ccc(c1)C=C
c1cnc(cc1)C=CC=C	C=C=C1CC2=C(C1)C(=CC2)	c1sc2c(c1)C(C)(C)c1c2sc(c1)C=C
c1cc2cc(o)cc2cc1	C1=c2ccccc2=C(C1=C)C#C	c1c(=O)oc2c1c(cc1cc(o)oc21)C=C
c1sc(c(c1O)S)C=C	c1sc2c(c1)[nH]c(c2)C=C	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)
c1c(C#N)sc(OC)c1	c1c(OCCC)c(cc(c1)OCCC)	c1sc(cc1)c1ccc(s1)c1sc(cc1)C=C
c1c(C)sc(C)c1C=C	c1cc2c(s1)c1c(C2)cc(s1)	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)
c1oc(cc1C(=O)OC)	c1sc(c2c1c(C#N)ccc2C#N)	c1sc(c2c1[C@H](F)CC[C@H]2F)C#C
c1cscc1C(=O)OC=C	c1csc2c1cc(c1c2scc1)C#C	c1sc(c(c1C(F)(F)F)C(F)(F)F)C=C
C1=CC=C(C1=O)C#C	c1sc(c2c1c(C#N)ccc2)C=C	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)
c1scc(C#N)c1NC=C	c1c(OCC)c(cc(c1)OCC)C=C	C1=[S]C(=O)C2=C1C(=O)[S]=C2C#C
c1sc(c2c1SCCCS2)	c1sc2c(c3ccoc3cc2c1)C#C	c1sc(cc1)c1ccc(s1)c1sc(cc1)C#C
c1cnc(c2c1nccn2)	c1sc(c2c1cc(C=C)cc2)C=C	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)
c1cc2c(s1)cc(s2)	c1sc(c2c1C(=O)NC2=O)C=C	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)
c1n(CC)c(cc1)C=C	c1sc(c2c1CC[C@@H](N)C2)	c1sc(c2c1[C@H](C(=O)OC)CCC2)C=C
c1scc2c1c(ccc2O)	c1sc(NC)c(c1C(=O)OC)C=C	c1csc2c1c(CC)c(CC)c1c(csc21)C=C
n1c2cscc2[nH]cc1	c1[nH]c2cc([nH]c2c1)C=C	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)
c1scc(C)c1C=CC#C	c1sc(c2c1c(C=O)ccc2)C=C	c1sc(c2c1c(C(F)(F)F)ccc2C#N)C=C
c1scc(C)c1C=CC=C	c1cc2occoccoccoccoc2cc1	c1sc(c2c1CC[C@@H](C(=O)O)C2)C#C
c(s1)c(O)c(NN)c1	c1sc(c2c1oc(=O)c(=O)s2)	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)
c1oc(c(c1O)S)C#C	C1=C(c2c3c(cccc13)ccc2)	C1=C(F)C=C/C/1=C\1/C(=CC(=C1)F)
c1sc(c2c1cccc2F)	C1=C2OCCSC2=C([S@@]1NC)	c1scc2c1nc1c3cc(sc3c3sccc3c1n2)
c1scc(C(=O)OC)c1	c1sc(c2c1sc(=O)c(=O)s2)	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)
c1c(N)ccc(c1)C#C	c1sc(c2c1C(=O)CC2=O)C=C	c1sc(c2c1CC[C@@H](C(=O)O)C2)C=C
c1nc2cscc2nc1C=C	c1sc(c2c1oc(N(=O)=O)c2)	c1n(CC)c(=O)c2c1c(=O)n(CC)c2C#C
c1coc(N(=O)=O)c1	c1sc(c2c1C(=O)OC2=O)C=C	c1sc(c2c1[C@H](C#N)CC[C@H]2C#N)
c1sc(C(=O)O)cc1O	c1sc(c2c1C(=O)OC2=O)C#C	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)
c1c(S)sc(O)c1C=C	c1[nH]c2[nH]c(nc2n1)C#C	c1sc(c2c1[C@H](N)CC[C@H]2OC)C=C
c1sc(c(c1F)F)C#C	c1sc(c2c1sc(C(=O)CC)c2)	c1c2c3c(s1)ccc1c3c(cc2)c(s1)C=C
c1c(C)Nc(C)c1C=C	c1sc2c(c3ccoc3cc2c1)C=C	c1sc(c2c1C[C@H](C(=O)OC)CC2)C#C
Nc1sc(c2c1nccn2)	C1S=C(/C(=N/N)/C1=S)C#C	c1sc(c2c1c(C(F)(F)F)ccc2C#N)C#C
C1=C(Cc2cscc2C1)	c1csc2c1cc1c(ccs1)c2C=C	c1sc(c2c1[C@H](N)CC[C@H]2OC)C#C
c1sc(c2c1CN2)C=C	c1[nH]c2[nH]c(nc2n1)C=C	N1[C@@H]2CSC(=S)[C@H]2N(CC1)C#C
c1oc(c(c1O)S)C=C	c1sc(c(c1SCC)SCC)C=CC=C	c1ccc(c2c1c1c(c3c2ccs3)scc1)C=C
C1=CC=C(C1=S)C#C	c1c2c(ncc(C)n2)c(s1)C=C	c1n(CC)c(=O)c2c1c(=O)n(CC)c2C=C
N1C(=S)C=C(C1=S)	c1c(OCC)cc(OCC)c(c1)C=C	c1cc2c(c3c1nsn3)cc(c1c2nsn1)C=C
c1cscc1C(=O)OC#C	c1[nH]c(c(c1)C(=O)C)C=C	N1[C@@H]2CSC(=S)[C@H]2N(CC1)C=C
c1c(S)sc(O)c1C#C	c1sc(c2c1C(=O)SC2=O)C=C	c1c2c(ccc3c2ccs3)c2c(ccs2)c1C=C
c1scc(C#N)c1NC#C	c1sc(c2c1CC[C@@H](O)C2)	c1csc2c1c(CC)c(CC)c1c(csc21)C#C
c1c2cS(F)cc2ccc1	c1sc(c2c1sc(N(=O)=O)c2)	c1sc(c2c1[C@H](C(=O)OC)CCC2)C#C

Table B33: List of SMILES for the 1759 monomer data set. (Part 6 of 14)

c1nc2cscc2nc1C#C	c1c(F)c(F)c(c(c1F)F)C=C	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)
c1sc(c2c1OSCCO2)	c1sc(NC)c(c1C(=O)OC)C#C	c1scc2c1C[C@H](C)[C@H](C2)C(=C)
c1oc(c(c1OC)C#N)	N1c2cscc2N(CC(=O)C1)C=C	c1cc2c(c3c1nsn3)cc(c1c2nsn1)C#C
c1sc(c(c1O)S)C#C	c1sc(c2c1c(C=C)ccc2)C#C	c1c2ccsc2cc2c1c(c1c(c2)scc1)C=C
C(C(=O)1)C(=O)C1	c1sc(c2c1cc(C=C)c(C)c2)	c1sc(c2c1C[C@H](C(=O)OC)CC2)C=C
c1cnc(cc1)C=CC#C	c1sc(c2c1CC(=O)C(=O)C2)	c1sc(c2c1C[C@H](C(F)(F)F)CC2)C=C
c1sc(c(c1)OC)C=C	C1=CC(=C2[C@H]1CCS2)C=C	c1c2nsnc2c(c2c1nc(CC)c(CC)n2)C#C
c1sc(c(c1C=C)OC)	c1c(CC)c(CC)c(CC)cc1C#C	c1sc(c2c1[C@H](OC)CC[C@H]2OC)C#C
c1sc(c2c1OCCCC2)	c1csc2c1cc1c(ccs1)c2C#C	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)C#C
N1C(=O)C=C(C1=O)	c1sc(c2c1C(=O)SC2=O)C#C	c1sc2c(ccc3c2ccc2c3ccc3c2scc3)c1
c1c(C)sc(C)c1C#C	c1sc(c2c1cc(C=O)cc2)C=C	C1C2CC(=C)O[C@H](C1C=CC2C)CCCC=C
c1oc(c(c1C)C)C#C	c1sc(c2c1CCS(=O)(=O)C2)	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)C#C
c1sc(c2c1CN2)C#C	c1sc(c2c1C(=O)CC2=O)C#C	c1scc2c1[C@@H](CC[C@H]2N(=O)=O)N
c1oc(c(c1F)F)C=C	c1c(CC)c(CC)c(CC)cc1C=C	C1=C(C=C/C/1=C\1/C=CC(=C1)N)NC#C
C1SCN2[C@H]1NCC2	c1sc(c2c1oc(=O)c(=O)o2)	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)C#C
c1c2c(OCN2)c(s1)	c1csc2c1cc(c1c2scc1)C=C	c1cc2c(cc1)c1c(C2(C)C)cc(cc1)C=C
C1SC=C1C(=O)CC=C	c1sc(c(c1SCC)SCC)C=CC#C	c1sc(c2c1[C@H](OC)CC[C@H]2OC)C=C
c1scc(C(=O)O)c1S	c1[nH]c2cc([nH]c2c1)C#C	c1sc(c2c1[C@H](C)CC[C@H]2C=C)C#C
c1ccc(c2c1cccn2)	c1c2nc(CC)c(CC)nc2c(s1)	c1sc(c2c1CC[C@@H](N(=O)=O)C2)C=C
c1sc(c(c1)OC)C#C	c1c(F)c(F)c(c(c1F)F)C#C	c1cc2c(s1)c1c(C2(CC)CC)cc(s1)C=C
c1sc(c(c1F)F)C=C	c1sc(c2c1SCC(=O)CO2)C#C	c1sc(c2c1C[C@H](C(F)(F)F)CC2)C#C
c1c(CC)cc(s1)C=C	c1sc(c2c1cc(OC)c(OC)c2)	c1scc2c1C(=[S@@](OC)C(=C2)OC)C=C
c1sc(N(=O)=O)cc1	C1S=C(/C(=N/N)/C1=S)C=C	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)C=C
c1c(F)sc(F)c1C#C	c1sc(c2c1cc(C#N)cc2)C=C	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1
c1n(CC)c(cc1)C#C	N1c2cscc2N(CC(=O)C1)C#C	c1sc(c2c1CC[C@@H](N(=O)=O)C2)C#C
c1c(CC)sc(c1)C=C	c1sc(c(c1OCC)OCC)C=CC=C	c1c2c(C=C)c3c(cc2ccc1)c(ccc3)C=C
c1oc(c(c1C)C)C=C	c(s1)c2cc(OC)c(OC)cc2c1	c1sc(c2c1[C@H](C(F)(F)F)CCC2)C=C
c1c(N)ccc(c1)C=C	c1nc2c(c(=O)c3c2ncc3)c1	c1sc(c2c1[C@H](C(=O)O)CC[C@H]2O)
C1=CC=C(C1=O)C=C	c1sc(c2c1[C@H](OC)CCC2)	C1=C(C=C/C/1=C\1/C=CC(=C1)N)NC=C
c1sc(C(=O)OC)cc1	c1sc(c2c1c(C(=O)C)ccc2)	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)
c1c(F)sc(F)c1C=C	C1=C2CSC[C@H]2C(=C1)C=C	c1sc(c2c1[C@H](C(F)(F)F)CCC2)C#C
c1occ(C)c1C=CC#C	c1sc(c(c1O)C(F)(F)F)C#C	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3
c1scc2c1sc1c2csc1	c1c(OCC)c(OCC)c(OCC)cc1	c1sc(c2c1[C@H](C)CC[C@H]2C=C)C=C
c1c2ccccc2c(c1=O)	c1sc(c2c1cc(C(=O)C)cc2)	C1=[S]C(=S)C2=C1C=C(S2(=O)=O)C=C
c1cc(C)cc(c1C)C=C	c1sc(c2c1S(=O)(=O)CCC2)	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)
C1SC=C1C(=O)OCC#C	c1sc(c2c1c(C#N)ccc2)C#C	c1cc2c(s1)c1c(C32OCCO3)cc(s1)C=C
c1sc(c2c1ocn2)C=C	c1sc(c2c1SCC(=O)CO2)C=C	c1sc(c2c1cc(C(F)(F)F)c(OC)c2)C=C
c1sc(c2c1SCS2)C=C	c1sc(c2c1CCC[C@H]2O)C#C	c1[nH]c(/C=C/c2[nH]ccc2)c(c1)C#C
c1[nH]c(C=C)c(c1)	c1sc(c2c1CCC[C@H]2O)C=C	c(s1)c2c(o)c3cN(CC)cc3c(o)c2c1C=C
c(s1)c2scnc2c1C#C	c1sc(c2c1cc(C=O)cc2)C#C	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)C#C
c1oc2cc(oc2c1)C#C	c1scc2c1CCC[C@H]2C(=O)O	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)
c1sc(c2c1scn2)C#C	c1sc(c2c1c(C=C)ccc2)C=C	C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)
c(s1)c2scnc2c1C=C	c1sc(c2c1cc(C#N)cc2)C#C	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)C=C
c1sc(c2c1ncs2)C#C	c1ccc2c(c1)oc1c2ccc(c1)	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)C#C
N1c2cscc2N(C1)C=C	c1c(OCC)c(cc(c1)OCC)C#C	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)C=C

Table B34: List of SMILES for the 1759 monomer data set. (Part 7 of 14)

N(CC)c1cscc1N(CC)	c1sc(c(c1O)C(F)(F)F)C=C	c1c(c2c(s1)c(CC)c1c(c2CC)scc1)C#C
c(s1)c2ncsc2c1C=C	N1CN[C@@H]2[C@H]1N(CN2)	C1=CC(OC)=C(/C/1=C\1/C=C(OC)C=C1)
c1cc(C)cc(c1C)C#C	c1sc2cc(C(=O)O)sc2c1C=C	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)C=C
c1sc(c2c1OCO2)C#C	c1sc(c2c1nc(CN)c(CN)n2)	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)C#C
c1sc(c2c1CCC2)C#C	c1sc2cc(C(=O)O)sc2c1C#C	c1sc(c2c1C[C@H](C)[C@@H](C)C2)C#C
c(s1)c2NCOc2c1C=C	c1sc(c2c1cc(C=C)cc2)C#C	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)C#C
c1sc(c2c1CCC2)C=C	c1scc2c1nc(c(NC)n2)N(C)	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)C=C
c1n(CC)nc(c1CC)CC	c1sc2c(c1CC)sc(c2CC)C=C	c(c1)c2nccnc2c1c(c1)c2nccnc2c1C=C
c1scc2c1c(ccc2OC)	c1sc2c(c1CC)sc(c2CC)C#C	c1c(c2c(s1)c(OC)c1c(c2OC)scc1)C#C
c1sc(c(c1)C(=O)C)	c1[nH]c(c(c1)C(=O)C)C#C	c1sc(c2c1C[C@H](S)[C@@H](O)C2)C=C
c(c1)cc(cc)cc1C=C	c1c(OCC)cc(c(c1)OCC)C=C	C1C(=O)Oc2c1cc1c(c2)C(C(=O)O1)C=C
c1sc(c2c1scn2)C=C	c1sc(c2c1CC[C@@H](C)C2)	c1sc(c2c1C[C@H](S)[C@@H](O)C2)C#C
C1=C([CH]S1)C(=C)	c1[nH]c(cc1)C(C(=O)N)=C	c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1C=C
c(s1)c2ncsc2c1C#C	c1sc(c2c1C(=O)NC2=O)C#C	c1nc2c(s1)c1c(c(c2CC)CC)N(CS1)C=C
c1sc(c2c1cc(C)s2)	c1sc(c2c1SCC(=O)CS2)C#C	c1sc(c2c1c(C(=O)C(F)(F)F)ccc2)C=C
c1sc(c(c1)C=O)C=C	c1sc(c2c1cc(C(=O)OC)s2)	c1sc(c2c1C[C@H](F)[C@@H](F)C2)C=C
c1sc(c2c1SCS2)C#C	c1sc(c2c1c(C=O)ccc2)C#C	c1sc(c2c1C[C@H](C)[C@@H](C)C2)C=C
c(c1)cc(cc)cc1C#C	c(c(N)1)ccc1c(c(N)1)ccc1	c1cc2c(C(CN)CN)c3cC(ccc3c2cc1)C=C
c1oc(c(c1C#N)C#N)	c1sc(CC)c2c1c(c[nH]2)C=C	c1c2c(ccs2)c(c2c1cc1c(c2)scc1)C=C
C1C([CH2])CCC1C#C	c1sc2c(c1)C(=O)C(=C2)C=C	c1sc(c2c1C[C@H](C#N)[C@@H](OC)C2)
c1c2c(nccc2)c(s1)	c1sc(c2c1[C@H](C=C)CCC2)	c1scc2c1C[C@@H]([C@@H](OC)C2)NC=C
c1sc(c2c1[nH]cn2)	C1=C2C(C(=O)N1)=C(OC2=O)	c1sc(c2c1C[C@H](F)[C@@H](F)C2)C#C
c1sc(c(c1O)OC)C=C	c1sc(c2c1c(C(=O)OC)ccc2)	c1cc(cc2c1n(c1c2cccc1)C(CC)CC)C=C
c1sc(c2c1ncs2)C=C	c1c(C(F)(F)F)sc(OC)c1C=C	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)C=C
Nc1sc(N(=O)=O)cc1	C(=N)/N=C(\C1=CS=CS1)C#C	c(c1)c2nccnc2c1c(c1)c2nccnc2c1C#C
c1c(sc(OC)c1)NC=C	c1sc2c(c1)sc1cc(sc21)C=C	c1c2c(OCC)c3c(ccs3)c(OCC)c2sc1C#C
c1sc(c2c1scc2)C=C	c1sc(c2c1c(N(=O)=O)ccc2)	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)
c1sc(c(c1OCC)OCC)	c1sc(c2c1nc(CCO)c(CN)n2)	n1cc2c3c(c1)ccc1c3c(cc2)cn(c1)C#C
c1sc(c2c1[nH]nc2)	c1scc2c1[C@@H](CCC2)NC=C	c1sc(c2c1cc(C(=O)C(F)(F)F)cc2)C#C
c1sc(c2c1ocn2)C#C	C1=[S]C(=C(O1)[N]C=O)C=C	c1sc(c2c1cc(C(F)(F)F)c(C#N)c2)C=C
c1sc(c2c1OCO2)C=C	c1scc2c1[C@@H](CCC2)NC#C	c1sc(c2c1nc1c3cccnc3c3ncccc3c1n2)
c1sc2nc(sc2c1)C#C	c1sc(c2c1CC[C@@H](OC)C2)	c1sc(c2c1[C@H](C#N)CC[C@H]2OC)C#C
c1sc(C=O)c(c1)C=C	c1sc(c2c1sc(C(=O)CC)c2F)	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)C=C
c1c(CC)c(ccc1)C#C	C(=N)/N=C(\C1=CS=CS1)C=C	C1=C(F)C=C/C/1=C\1/C(=CC(=C1)F)C=C
c1sc(C(F)(F)F)cc1	c1sc(c2c1sc(C(=O)OCC)c2)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)
c1sc(c2c1nco2)C=C	c1csc2c1c(CC)c1c(n2)scc1	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)C=C
C1C([CH2])CCC1C=C	c1scc2c1c1c(s2)cc(s1)C#C	c1sc(c2c1C[C@H](C(=O)C(F)(F)F)CC2)
c1occ(N(=O)=O)c1N	c1sc(c2c1c(C=C)ccc2C)C#C	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)
c1sc(c(c1O)OC)C#C	c1csc2c1n(CC)c1c2scc1C#C	c1cc2c(cc1)c1c(C2(CC)CC)cc(cc1)C#C
C=NNc1ccc(cc1)C=C	c1sc(c(c1OC)C(F)(F)F)C#C	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)
c1sc(C(=O)O)c(c1)	c1sc2c(c1)oc1c2sc(c1)C=C	c1scc2c1nc1c3cc(sc3c3sccc3c1n2)C=C
c1sc(C=O)c(c1)C#C	c1sc(c(c1)C(=O)C(F)(F)F)	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)
c1c2c(cccc2)c(s1)	c1sc(c(c1OC)C(F)(F)F)C=C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)
c1sc(c(c1)C=O)C#C	c1c2c(nccn2)c(c2c1nccn2)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)
c1oc(cc1C(F)(F)F)	C1=C2C(C(=O)S1)=CN(C2=O)	c1cc(c(cc1)N(c1ccccc1)c1ccccc1)C=C

Table B35: List of SMILES for the 1759 monomer data set. (Part 8 of 14)

C1=CC=C(S1(=O)=O)	c1csc2c1C(S)=c1c2scc1C=C	c1scc2c1C[C@H](C)[C@H](C2)C(=C)C#C
N1c2cscc2N(C1)C#C	c1sc(c2c1cc(C#N)c(OC)c2)	C1=C/C(/C(=C1)F)=C\1/C=C(C=C1F)C#C
c1oc2cc(sc2c1)C#C	c1c(F)c(F)c(c2c1nsn2)C=C	c1sc(c2c1[C@H](C#N)CC[C@H]2C#N)C=C
C(=C)c1sc(C)c(c1)	c1c(F)c(F)c(c2c1nsn2)C#C	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)C#C
C=NNc1ccc(cc1)C#C	C1SC[C@](C1)(C(=O)N)NC=C	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)C=C
c1c(OC)c(ccc1)C=C	c1ccc(c2c1nc(CC)c(CC)n2)	c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)
C1SC=C1C(=O)OCC=C	c1oc2c(n1)cc1c(c2)oc(n1)	c1sc(c2c1[C@H](C#N)CC[C@H]2C#N)C#C
c1cc(CN)ccc1CNC#C	c1sc(c2c1c(C(=O)O)ccc2O)	c1scc2c1C[C@H](C)[C@H](C2)C(=C)C=C
c1sc(c2c1occ2)C#C	C1=Cc2cc3=CC(=Cc3cc2=C1)	c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)
c(s1)c2NCOc2c1C#C	c1sc(c2c1SC(=O)CC(=O)O2)	c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)
c1sc(c2c1scc2)C#C	c1sc(c2c1C(=O)CCC2=O)C#C	c1sc(c2c1nc1c3sccc3c3ccsc3c1n2)C#C
c1c(CC)c(ccc1)C=C	c1sc(c(c1N(=O)=O)C#N)C=C	c1sc(c2c1C(=O)c1ccc(CC)cc1C2=O)C=C
c1c2C(=O)OCc2ccc1	c1scc2c1C=C(S2(=O)=O)C#C	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)
c1oc2cc(sc2c1)C=C	c1c(C(F)(F)F)sc(OC)c1C#C	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1C=C
c1sc(c2c1nco2)C#C	c1sc(c2c1cc(C)c(C)c2)C#C	c1c2c(OCC)c3c(c(c2sc1)OCC)cc(s3)C=C
c1scc(N(=O)=O)c1N	c1scc2c1C=C(S2(=O)=O)C=C	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)
c1scc2c1C=C(C2=O)	c1sc(c2c1nc(OCC)c(CN)n2)	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3C=C
c1sc(c2c1cc[nH]2)	n1c2cscc2n(c(=O)c1=O)C=C	c(s1)cc(c(=C(CN)(CN))2)c1c(s3)c2cc3
c1sc(c(c1C#N)C#N)	C1S[C@@H]2[C@@H]1NCC2C=C	c1scc2c1[C@@H](CC[C@H]2N(=O)=O)NC=C
c1sc(c2c1occ2)C=C	c1sc(c2c1[nH]c(=O)[nH]2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2OC)
c1oc2cc(oc2c1)C=C	c1sc(c2c1c(C=C)ccc2C)C=C	c1sc(c2c1[C@H](C(=O)O)CC[C@H]2O)C=C
c1c(sc(OC)c1)NC#C	c1csc2c1n(CC)c1c2scc1C=C	C1=CC(=[S]C1)c1ccc(s1)C1=[S]CC(=C1)
c1sc(c(c1)C#N)C#C	c1sc(c(c1N(=O)=O)C#N)C#C	c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)
C1=C(c2cscc12)C=C	c1cccc2c1n(c1c2cccc1)C=C	c1sc2c(ccc3c2ccc2c3ccc3c2scc3)c1C=C
c1c(OC)c(ccc1)C#C	c1sc2c(c1)sc1cc(sc21)C#C	c(s1)cc(c(CC)c(CC)2)c1c(s3)c2cc3C#C
c1sc2nc(sc2c1)C=C	c1scc2c1c1c(s2)cc(s1)C=C	c1sc(c2c1[C@H](C(=O)O)CC[C@H]2O)C#C
c1c(C)c(c2c1ccco2)	c1oc(c(c1C(F)(F)F)OC)C#C	c1sc2c(c1)C([C](C#N)C#N)=c1c2sc(c1)
c1sc(c2c1OSCS2)C#C	c1sc(c2c1cc(N(=O)=O)cc2)	C1=CC(C#N)=C/C/1=C/1\C(=CC(=C1)C#N)
c1c2nccnc2c(s1)C#C	c1csc2c1c(C)c(C)c1c2scc1	c(s1)c2c(=O)c3cn(CC)cc3c(=O)c2c1C#C
c1sc(c2c1OCCO2)C=C	c1c2nc(C)c(C)nc2c(s1)C=C	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)
c1sc(c2c1ncnc2)C#C	c1sc(c2c1cc(F)c(F)c2)C=C	c1sc(c2c1C[C@H](OC)[C@@H](OC)C2)C=C
C1=CC2=CC(=NC2=C1)	c1sc(CC)c2c1c(c[nH]2)C#C	c1cc2c(s1)c1c(cc2)c2c(cc1)cc(s2)C=C
C1=Cc2[nH]c(cc2C1)	c1sc2c(c1)C(=O)C(=C2)C#C	c1sc(c2c1cc1C(=O)N(CC)C(=O)c1c2)C#C
c1c2cccc(c2cs1)C=C	c1sc2c(c1)oc1c2sc(c1)C#C	C1=CC(N(=O)=O)C(C2C=CC=C2N(=O)=O)=C1
c1c2cscc2c(nn1)C#C	c1sc(c2c1cc(S)c(O)c2)C=C	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)
c1sc(c2c1cc(C)cc2)	c1ccc(OCC)c2c1c(OCC)ccc2	c1sc(c2c1nc1c3ccccc3c3ccccc3c1n2)C=C
c1c2cscc2c(N)c(c1)	c1sc(c2c1nc(N)c(N)n2)C=C	c1sc(c2c1C[C@H](C#N)[C@@H](OC)C2)C#C
c1scc2c1c(c(cc2)O)	c1scc2c1N(CCN2C(=O)O)C=C	C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)C#C
c1sc(c2c1ncnc2)C=C	c1sc(c2c1cc(S)c(O)c2)C#C	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)C#C
c1c2nccnc2c(s1)C=C	c1c2c(ncc(CC)n2)c(s1)C#C	c1sc(c2c1nc1c3cccnc3c3ncccc3c1n2)C=C
c1c(ncc2c1non2)C=C	c1sc(c2c1nc(N)c(N)n2)C#C	c1sc(c2c1C[C@H](C#N)[C@@H](OC)C2)C=C
c1sc(c2c1OCSS2)C=C	c1sc(c2c1C(=O)CCC2=O)C=C	c1sc(c2c1cc(N(=O)=O)c(N(=O)=O)c2)C=C
C1Oc2cscc2OC(C1=O)	c1sc(c2c1cc(F)c(F)c2)C#C	C1=C/C(/C(=C1)OC)=C\1/C=C(C=C1OC)C=C
c1c(oc2cscc2s1)C=C	C1=C2C(C(=O)S1)=C(OC2=O)	C1=CC(OC)=C(/C/1=C\1/C=C(OC)C=C1)C#C
c1c2nonc2c(cc1)C=C	c1sc(c2c1[nH]c(=S)[nH]2)	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2

Table B36: List of SMILES for the 1759 monomer data set. (Part 9 of 14)

c1c2nsnc2c(cc1)C=C	c1sc2c(n1)Cc1c2sc(c1)C=C	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C#N)
c1sc(c2c1c(C)ccc2)	c1cnc2c1c(=S)c1c2ncc1C=C	c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)
c1sc(c2c1CCCC2)C#C	C1=C2C(C(=O)O1)=C(OC2=O)	c1sc(c2c1C[C@H](N)[C@@H](N(=O)=O)C2)
c1sc(c(c1O)C(=O)O)	c1c2nc(C)c(C)nc2c(s1)C#C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)C#C
c1sc(c2c1OCCO2)C#C	c1scc2c1N(CCN2C(=O)O)C#C	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)C#C
c(cc1)c2cC(C)cc2c1	c1c2c(ncc(CC)n2)c(s1)C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)C#C
c1c(oc2cscc2s1)C#C	c1scc2c1c1c(s2)sc(c1)C=C	c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)C=C
c1sc(c2c1oc(=O)o2)	c1c2c(cccc2)cc2c1c(ccc2)	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)C=C
c(oc1c(F)s2)csc1c2	c1sc2c(c1)c1c(scc1)c(c2)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)C=C
c1[nH]c(c2c1nccn2)	c1sc2c(c1)n(C)c1c2sc(c1)	c1c(C(=O)c2cscc2C(=O)c2ccccc2)ccc(c1)
c(cc1)c2cs(c)cc2c1	c1sc(c2c1cc(C)c(C)c2)C=C	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)C=C
c1sc(c(c1CC)CC)C#C	c1sc(c2c1cc(C(=O)OC)cc2)	c1cc2c(s1)c1c(c(=O)[nH]c2=O)cc(s1)C#C
c1sc(c2c1sc(=S)s2)	n1c2cscc2n(c(=O)c1=O)C#C	c1c2c(=O)oc(=O)c2c(c2c1c(=O)oc2=O)C#C
c1oc(c(c1)N(=O)=O)	c1oc(c(c1C(F)(F)F)OC)C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)
c(s(c)1)c2ccccc2c1	c1c([CH2])c(c2c1ccsc2)C#C	c1sc(c2c1C[C@H](C(=O)C(F)(F)F)CC2)C#C
c1sc(c2c1OSCS2)C=C	c1c(C(F)(F)F)sc(C#N)c1C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)C#C
c1c(ncc2c1non2)C#C	c1c2c(ncs2)c(c2c1ncs2)C#C	c1nc2c(s1)[C@H]1[C@H](C2(CC)CC)N(CS1)
c1scc2c1C=C(C2)C=C	c1oc2c(C(=O)N)c(oc2c1)C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)sc2=O)C=C
c1sc(c(c1C)S[CH2])	c1c(C(=O)C(F)(F)F)scc1C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)oc2=O)C=C
c1sc(c2c1nccn2)C=C	c1c2c(SCC2)c2c(CCS2)c1C#C	c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)C#C
C1=C(Oc2c(csc2)O1)	N(CC)c1ccc(c(c1)C=C)N(CC)	c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)
c1cnc(c2c1nsn2)C=C	c1oc2c(c1)C(=O)c1c2oc(c1)	c1c2c(=O)sc(=O)c2c(c2c1c(=O)sc2=O)C=C
c1sc(c(c1O)C#N)C#C	c1c(OCCC)c(cc(c1)OCCC)C=C	c1c2C(=O)CC(=O)c2c(c2c1C(=O)CC2=O)C=C
c1sc(c2c1OCCS2)C#C	c(s1)c2C(NCC)OC(O)c2c1C#C	c1sc(c2c1[C@H](C(=O)C(F)(F)F)CCC2)C=C
c1c(OC)sc(OC)c1C=C	c(c(CC)1)ccc1c(c1)ccc1C=C	c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)
c1sc(CC)c(c1CC)C#C	c1cc(cc2c1ccc1c2cccc1)C=C	c1sc(c2c1C[C@H](C#N)[C@@H](C#N)C2)C#C
c1[nH]cc2c1[nH]cc2	c1sc(c(c1C#N)C(F)(F)F)C=C	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3
c1c2nsnc2c(cc1)C#C	c1sc(c2c1CCC[C@H]2C=O)C=C	c1c2c(=O)sc(=O)c2c(c2c1c(=O)oc2=O)C#C
c1sc(c2c1cc(S)cc2)	c1sc(c(c1N(=O)=O)N(=O)=O)	c1sc(c2c1C[C@H](C(=O)C(F)(F)F)CC2)C=C
c1c(OCC)c(ccc1)C=C	c(s1)c2c(=O)n(C)c(=O)c2c1	c1sc2c(c1)C=c1c2sc2=c3sc(cc3C=c12)C=C
c1sc(c2c1ccc(F)c2)	C1=C(CC)C(CC)=C(S1(=O)=O)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2OC)C=C
c1sc(c2c1ncnc2C#N)	c1cc(cc2c1ccc1c2cccc1)C#C	c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)C#C
c1c2cscc2c(nn1)C=C	c1c2n(CC)c3c(c2ccc1)cccc3	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)C=C
C/C=C\1/OCCC(OCC1)	c1cc2c(cs1)c1c(ccs1)c2C=C	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2OC)C#C
c1sc(c(c1O)C#N)C=C	c1sc(c2c1c(N(=O)=O)ccc2N)	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)
C#Cc1sc(cc1)C#CC=C	c1c(O)sc2c1[nH]c1c2sc(c1)	c1sc(c2c1C[C@H](C(=O)O)[C@@H](O)C2)C=C
c1cnc(c2c1nsn2)C#C	c1cc2c(cc1)c1c(C2)cc(cc1)	c1sc2c(c1)C([C](C#N)C#N)=c1c2sc(c1)C=C
c1c(C(=O)C)scc1C#C	c1c(OC)cc(c(c1)OCC(CC)CC)	c1sc2c(c1)C([C](C#N)C#N)=c1c2sc(c1)C#C
c1sc(c2c1OCCS2)C=C	C(=S)[CH]C1=[S]C=C(O1)C#C	N1CN[C@@H]2N[C@H]3[C@@H](N[C@H]12)NCN3
c1sc(CC)c(c1CC)C=C	c1sc(c(c1OCC)ON(=O)=O)C=C	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)
n1c(C)cc(c1C=C)C=C	c1oc(c(c1C#N)C(F)(F)F)C=C	C1=CC(C#N)=C/C/1=C/1\C(=CC(=C1)C#N)C=C
c1[nH]c(cc1)C#CC#C	c(s1)c2C(NCC)OC(O)c2c1C=C	c(s1)cc(c(=C(CN)(CN))2)c1c(s3)c2cc3C=C
c1oc(cc1C(=O)O)C=C	c1oc(c(c1C#N)C(F)(F)F)C#C	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)C#C
c1scc2c1C=C(C2)C#C	c1sc2c(c1)C(=S)c1c2oc(c1)	c1cc2c(s1)c1c(c(=O)n(c2=O)CC)cc(s1)C=C
C1=C(Cc2c(C1)nsn2)	NCc1cc(OCC)c(cc1OCC)CNC=C	C1=CC(=[S]C1)c1ccc(s1)C1=[S]CC(=C1)C=C

Table B37: List of SMILES for the 1759 monomer data set. (Part 10 of 14)

c1sc(c(c1CC)CC)C=C	c1sc(c2c1OCC[C@@H](SC)O2)	C1=C/C(/C(=C1)C#N)=C\1/C=C(C=C1C#N)C#C
c1oc(cc1C(=O)C)C=C	c(o1)c2c(=O)Nc(=O)c2c1C#C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](OC)C2)
c1c(C(=O)C)scc1C=C	c1oc2c(c1)C(=O)c1c2sc(c1)	C1=CC(=[S]C1)c1ccc(s1)C1=[S]CC(=C1)C#C
c1oc(cc1C(=O)C)C#C	c(s1)c2ccs(=N)(=O)c2c1C#C	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)C#C
c1c(cc2c(c1)nccn2)	c1sc2c(c1)C(=O)c1c2sc(c1)	c1cc2c(s1)c1c(C2)cc2c(c1)Cc1c2sc(c1)C=C
C#Cc1sc(cc1)C#CC#C	c1sc(c2c1[nH]c(C#N)c2)C#C	c1c2c(ccs2)c(CC)c2c1c(CC)c1c(scc1)c2C=C
c1sc(c2c1c(S)ccc2)	c1c(OC)cc(c(c1)OCC)C=CC=C	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C#N)C#C
c1sc(c2c1SCCS2)C#C	c1oc(cc1C(=O)C(F)(F)F)C=C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C#N)C2)
c1sc(c2c1OCSS2)C#C	c1c(c2c(s1)c1c(n2CC)CCS1)	c1sc(c2c1C[C@H](N)[C@@H](N(=O)=O)C2)C=C
c1sc(c2c1SCCS2)C=C	c1sc(c2c1[nH]c(C#N)c2)C=C	c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)
c1sc(c2c1sc(=S)o2)	c1sc(c2c1[C@H](C)CCC2)C#C	c1cc2c(s1)c1c(C2)ccc2c1c1c(C2)cc(s1)C=C
c1c2cccc(c2cs1)C#C	c1sc(c2c1[C@H](S)CCC2)C#C	c1cc2c(s1)c1c(N2)c(cc2c1c1c(N2)ccs1)C=C
c1sc(c2c1sc(=O)s2)	c(c(CC)1)ccc1c(c1)ccc1C#C	c1sc2c(c1F)c(OCC)c1c(c2OCC)c(F)c(s1)C=C
c1c2nonc2c(cc1)C#C	c1cc(c2c(c1)c1c(s2)cccc1)	C1=CC(N(=O)=O)C(C2C=CC=C2N(=O)=O)=C1C=C
c1c(OC)c(cc(c1)OC)	c1sc(c2c1cc(C(F)(F)F)cc2)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C#N)C=C
C1=C(CC2=C1C[CH]2)	c1sc(c2c1nc(OCC)c(OCC)n2)	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)
c1sc(c2c1CCCC2)C=C	c1sc(c2c1C[C@H](S)CC2)C=C	c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)C=C
c1oc(c(c1OC)OC)C=C	c(s1)c2ccs(=N)(=O)c2c1C=C	c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)C#C
c1oc(cc1C(=O)O)C#C	c(o1)c2c(=O)Nc(=O)c2c1C=C	c1nc2c(s1)[C@H]1[C@H](C2(CC)CC)N(CS1)C=C
c1oc(c(c1O)C(=O)O)	c1sc(c(c1OCC)ON(=O)=O)C#C	N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)
N1c2cscc2N(CC1)C=C	c1sc(c2c1cc(C)c(OC)c2)C=C	c1c(C(=O)c2cscc2C(=O)c2ccccc2)ccc(c1)C=C
c1sc(c(c1OC)OC)C#C	c1c2c(SCC2)c2c(CCS2)c1C=C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)C#C
c1c(OC)sc(OC)c1C#C	c1sc(c2c1cc(C#N)c(C#N)c2)	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3C=C
N1c2cscc2N(CC1)C#C	c1c2c(ccs1)nc1c2sc(c1)C=C	c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2
c1sc(c2c1oc(=S)o2)	c1sc(c2c1[C@H](C)CCC2)C=C	c1c(c2c(s1)[C@H]1[C@H](C2(CC)CC)CCS1)C=C
c1scc2c1c(c(cc2)N)	c1cc2c(cs1)c1c(ccs1)c2C#C	c1cc2c(C=C/C/2=C\2/C=Cc3c2cccc3)c(c1)C#C
c1sc(c(c1OC)OC)C=C	c1sc(c2c1CCC[C@H]2C=O)C#C	N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)
c1sc(CC)c2c1nc(s2)	c1sc(c2c1cc(SCC)c(SCC)c2)	c(c1)ccc2c1N(CCCCC)(CCCCC)c(c3)c2ccc3C#C
c1cc(c2c1[nH]cnc2)	c1c(C(F)(F)F)sc(C#N)c1C#C	c1c2c(=O)n(c(=O)c2cc2c1c(=O)[nH]c2=O)C=C
c1c(OCC)c(ccc1)C#C	c1c2c(ncs2)c(c2c1ncs2)C=C	N1CN[C@@H]2N[C@H]3[C@@H](N[C@H]12)NCN3C=C
c1sc(c2c1nccn2)C#C	c1sc(c2c1C[C@H](S)CC2)C#C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](OC)C2)C#C
c1sc(c(n1)C(=O)OC)	c1sc(c(c1C#N)C(F)(F)F)C#C	c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)
c1ccc(c2c1cccn2)C=C	c1oc2c(c1)C(=S)c1c2oc(c1)	C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1
C(C(=O)1)C(=O)C1C=C	c1sc(c2c1cc(C(C)(C)C)cc2)	C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1
c1oc(c(c1OC)C#N)C=C	C(=[N])/N=C(\C1=CSC=[S]1)	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)C=C
C=Cc1oc(c(c1O)O)C=C	c1oc2c(C(=O)N)c(oc2c1)C#C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](OC)C2)C=C
c1sc(N(=O)=O)cc1C=C	c1sc(c2c1c(C#N)ccc2OC)C=C	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C(F)(F)F)
c1sc(c2c1C=[S]C=C2)	c1c(C(=O)C(F)(F)F)scc1C#C	c1cc2c(s1)c1c(ccs1)c1c2c(c(CC)c(CC)c1)C#C
c1sc(c2c1OSCCO2)C=C	N1[C@H]2CSC[C@H]2N(S1)C#C	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)C=C
c1sc(c2c1SCCCS2)C#C	c1oc(cc1C(=O)C(F)(F)F)C#C	c1cc2c(s1)c1c(c3c2nc(CC)c(CC)n3)cc(s1)C#C
c1sc(c2c1cccc2F)C=C	N1[C@H]2CSC[C@H]2N(S1)C=C	N1CN[C@@H]2N[C@H]3[C@@H](N[C@H]12)NCN3C#C
c1scc(C(=O)OC)c1C=C	c1sc(c2c1c(F)c(F)c(F)c2F)	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)C=C
c1cc2c(s1)cc(s2)C#C	c1scc2c1C[C@@H](CC2)C(=C)	c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)
c1coc(N(=O)=O)c1C=C	c1c(OCCC)c(cc(c1)OCCC)C#C	N1[C@H]2[C@H](N[C@@H]3[C@@H](CSC3)N2)N(S1)
C1=CC2=C(C1)C=C(C2)	c1sc(c2c1c(C(F)(F)F)ccc2)	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)

Table B38: List of SMILES for the 1759 monomer data set. (Part 11 of 14)

c1sc(c2c1OCCCO2)C#C	C=C=C1CC2=C(C1)C(=CC2)C=C	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)
c(s1)c(CCN)c(CCN)c1	c1sc(c2c1CC[C@@H](C=O)C2)	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)
c1oc(cc1C(=O)OC)C#C	c1cc2c(C(=O)N(C2=O)CC)cc1	c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)C=C
C(C(=O)1)C(=O)C1C#C	c1oc(c(c1N(=O)=O)N(=O)=O)	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C#N)C2)C#C
Nc1sc(c2c1nccn2)C=C	c1c(F)c(F)c(c2c1nn(CC)n2)	n1c2[CH][S]=Cc2n(c2c1c1c3c(ccc1)cccc23)C#C
N1C(=O)C=C(C1=O)C=C	c1sc(c2c1c(C#N)ccc2OC)C#C	c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)
c1sc(C(=O)OC)cc1C#C	c1cc2occcccccccoc2cc1C=C	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)
c1sc(c(c1)C(F)(F)F)	c(s1)c2cc(OC)c(OC)cc2c1C=C	c1sc(c2c1[C@H](N(=O)=O)CC[C@H]2N(=O)=O)C#C
C1=C(C(=C[S@@]1O)O)	c1sc(c2c1sc(N(=O)=O)c2)C#C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C#N)C2)C=C
c1cc2cc(o)cc2cc1C#C	c1cc2occcccccccoc2cc1C#C	c1c2c(cccc2)c2c(c1)c1c(c3c(cc1)cccc3)cc2C=C
c1scc2c1c(ccc2O)C=C	c1sc(c2c1nc(CN)c(CN)n2)C#C	N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)C#C
c1sc(c(c1OC)C#N)C#C	c1sc(c2c1CCS(=O)(=O)C2)C#C	N1[C@@H]2C[C@@H]3NSN[C@@H]3C[C@@H]2N(S1)C=C
c1[nH]cc(c1C)C=CC#C	c1sc(c2c1oc(=O)c(=O)s2)C#C	N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)C#C
c1sc(c(c1C=C)OC)C=C	c1sc(c2c1c(C(=O)C)ccc2)C#C	N1CN[C@@H]2C[C@H]3[C@@H](C[C@H]12)N(CN3)C=C
c1sc(c2c1OCCCC2)C=C	c1[nH]c2nc3[nH]c(nc3nc2n1)	c1sc(c2c1nc1c3ccc(CC)cc3c3cc(CC)ccc3c1n2)C=C
c1c2cS(F)cc2ccc1C=C	c1sc(c2c1CCC[C@H]2N(=O)=O)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C(F)(F)F)C=C
c1sc(C(=O)OC)cc1C=C	c1sc(c2c1S(=O)(=O)CCC2)C#C	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)
c1scc(C(=O)OC)c1C#C	c1ccc2c(c1)oc1c2ccc(c1)C#C	C1C(=O)O[C@@H]2[C@H]1CC[C@@H]1[C@H]2OC(=O)C1
c1sc(c(c1C(=O)O)OC)	c1ccc2c(c1)oc1c2ccc(c1)C=C	C1C(=O)O[C@H]2[C@@H]1CC[C@H]1[C@@H]2OC(=O)C1
C1=[S]c2cscc2[S]=C1	c1sc(c2c1CCS(=O)(=O)C2)C=C	C1=C[C@@H]2[C@H](C1)[C@@H]1[C@@H](C=CC1)C2=C
c1sc2c(c1)[nH]c(c2)	c1[nH]cc2c1S(=O)(=O)C(=C2)	C1=CC2=C(OC=CO2)/C/1=C/1\C2=C(OC=CO2)C=C1C=C
c1sc(c2c1c(S)ccc2O)	c1sc(c2c1[nH]c(N(=O)=O)c2)	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C(F)(F)F)C2)
c1sc(OC)c(C(=O)C)c1	c1sc(c2c1sc(C(=O)CC)c2)C=C	C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1C=C
c1sc(c2c1OCCCC2)C#C	c1cc(CC)c(c(c1)CC)/C=C(/C)	c1sc(c2c1[C@H](C(F)(F)F)CC[C@H]2C(F)(F)F)C#C
c1c2c(OCN2)c(s1)C=C	c1sc(c2c1[C@H](OC)CCC2)C=C	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)
c1c2c(OCN2)c(s1)C#C	c1sc(c2c1c(C#N)ccc2C#N)C=C	C1=C(N(=O)=O)C=C/C/1=C\1/C=CC(N(=O)=O)=C1C#C
cc(sc1n2)nc1sc2cC#C	c1sc(c2c1[C@H](OC)CCC2)C#C	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)
c1sc(c2c1cccc2F)C#C	c1sc(c2c1cc(C(=O)OC)s2)C#C	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)
c1c(C#N)sc(OC)c1C=C	c1sc(c2c1cc(C(=O)O)c(O)c2)	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)C=C
C1=C(Cc2cscc2C1)C#C	c1scc2c1CCC[C@H]2C(=O)OC#C	c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)C=C
c1cc2c(s1)cc(s2)C=C	N1CN[C@@H]2[C@H]1N(CN2)C#C	N1[C@H]2[C@H](N[C@@H]3[C@@H](CSC3)N2)N(S1)C=C
c1cnc(c2c1nccn2)C=C	c1sc(c2c1oc(=O)c(=O)s2)C=C	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)C#C
c1sc(c2c1OCCCO2)C=C	n1c(C)c2c(c1=O)c(C)n(c2=O)	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)C#C
C1=c2ccccc2=C(C1=C)	c1sc(c2c1CC[C@@H](O)C2)C=C	c1c2c(c3c(cccc3)n2CC)c(c2c1c1c(n2CC)cccc1)C=C
c1sc(c2c1c(F)ccc2F)	c1n(C)c(=O)c2c1c(=O)n(C)c2	c1sc(c2c1C[C@H](N(=O)=O)[C@@H](N(=O)=O)C2)C#C
c1sc(C(=O)O)cc1OC=C	c1sc(c2c1CC[C@@H](C)C2)C=C	c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)C=C
c1[nH]cc(c1C)C=CC=C	c1sc(c2c1cc(OC)c(OC)c2)C#C	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)C#C
c1oc(cc1C(=O)OC)C=C	c1sc(c2c1c(C(=O)C)ccc2)C=C	c1sc(c2c1[C@H](F)[C@H](F)[C@@H](F)[C@H]2F)C#C
c1sc(c2c1SCCCS2)C=C	c1sc(c2c1cc(C(=O)C)cc2)C#C	c1cc2c(cc1)C(C(=O)N2CC)C1C(=O)Nc2cc(ccc12)C=C
c1sc(c2c1oc(C#N)c2)	c1cc2c(s1)c1c(C2)cc(s1)C#C	c1cccc2c1N(C)C(=O)C2C1C(=O)N(C)c2cc(ccc12)C=C
C1=C(Cc2cscc2C1)C=C	c1csc2c1c(CC)c1c(c2CC)scc1	c1c(c2c(cc1)cc1c(cc3c(c1)cc1c(c3)cccc1)c2)C#C
C1SCN2[C@H]1NCC2C=C	N1CN[C@@H]2[C@H]1N(CN2)C=C	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)C=C
c1c(OCC)c(cc(c1)OC)	c1[nH]c(=O)c2c1c(=O)[nH]c2	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)
Nc1sc(c2c1nccn2)C#C	c1sc(c2c1CC[C@@H](O)C2)C#C	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)C#C
c1sc(c(c1OC)C#N)C=C	c1sc(c2c1CC(=O)C(=O)C2)C=C	c1sc(c2c1C[C@H](C(F)(F)F)[C@@H](C(F)(F)F)C2)C#C

Table B39: List of SMILES for the 1759 monomer data set. (Part 12 of 14)

c1c(C#N)sc(OC)c1C#C	c1c2c3CCSc3c3SCCc3c2c(cc1)	N1[C]2C=[S]C=C2N(C2=C1c1c3c(ccc1)c(CC)ccc23)C#C
N1C(=S)C=C(C1=S)C#C	c1sc(c2c1oc(N(=O)=O)c2)C#C	C1C(=O)O[C@@H]2[C@H]1CC[C@@H]1[C@H]2OC(=O)C1C=C
N1C(=S)C=C(C1=S)C=C	c1sc(c2c1sc(=O)c(=O)s2)C=C	C1C(=O)O[C@H]2[C@@H]1CC[C@H]1[C@@H]2OC(=O)C1C=C
c1sc(c2c1OCCCS2)C#C	c1cc2c(s1)c1c([nH]2)cc(s1)	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)C=C
c1ccc(c2c1cccn2)C#C	c1sc(c2c1cc(C(=O)OC)s2)C=C	C1=C[C@@H]2[C@H](C1)[C@@H]1[C@@H](C=CC1)C2=CC=C
c1ccc(c2c1nn(CC)n2)	c1sc(c2c1CC[C@@H](C)C2)C#C	c1c2c(ccs2)c(c2c1c(c1c(c2CC)cc2c(c1)ccs2)CC)C=C
c(s1)c(O)c(NN)c1C#C	c1cc2c(s1)c1c(C2)cc(s1)C=C	c1sc(c2c1C[C@H](C(F)(F)F)F)[C@@H](C(F)(F)F)C2)C=C
c1cc(N(=O)=O)c(cc1)	c1sc(c2c1sc(N(=O)=O)c2)C=C	C1C(=O)O[C@H]2[C@@H]1CC[C@H]1[C@@H]2OC(=O)C1C#C
C1SCN2[C@H]1NCC2C#C	c(s1)c2cc(OC)c(OC)cc2c1C#C	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)C=C
c1scc(N(O)O)c1N(O)O	c1[nH]c(cc1)C(C(=O)N)=CC#C	N1C(=O)C(c2c1cc(cc2)C)C1C(=O)N(c2c1ccc(C)c2)C#C
c1sc(c2c1c(C)ccc2C)	c1sc(c2c1sc(C(=O)CC)c2)C#C	n1c(O)c2c3c(c1O)cc(CC)c1c3c(cc2CC)c(O)n(c1O)C#C
cc(sc1n2)nc1sc2cC=C	c1sc(c2c1oc(N(=O)=O)c2)C=C	C1C(=O)O[C@@H]2[C@H]1CC[C@@H]1[C@H]2OC(=O)C1C#C
c1cc2cc(o)cc2cc1C=C	c(s1)c2c(=O)N(CC)c(=O)c2c1	C1C[C@H]2[C@H](CC1)N(C[CH2])[C@H]1[C@H](S2)CCCC1
N1C(=O)C=C(C1=O)C#C	c1sccc1/C=C(/Cc1sc(CC)cc1)	c(s1)cc(c(OC(CCC)CCC)2)c1c(OC(CCC)CCC)c(c3)c2sc3
c1sc(c(c1C=C)OC)C#C	c1sc(c2c1cc(C=C)c(C)c2)C#C	C1C[C@@H]2[C@@H](CC1)N(CC)[C@@H]1[C@@H](S2)CCCC1
c1sc(c2c1OCCCS2)C=C	c1nc2c(c(=O)c3c2ncc3)c1C=C	c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)
c1c(C)cc(c(c1)C)C=C	c(s1)c2c(=O)n(CC)c(=O)c2c1	c1ccc2c3cc4n(CC)c5cc6c(cc5c4cc3Cc2c1)Cc1c6ccc(c1)
c1scc(C(=O)O)c1SC=C	c1c2c3c(s1)ccc1c3c(cc2)sc1	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)C=C
c1oc(c(c1OC)C#N)C#C	c1c2n(CCC)c3c(c2ccc1)cccc3	n1c2C=[S][CH]c2n(c2c1c1c3c(c(cc1)CC)c(CC)ccc23)C#C
c(s1)c(O)c(NN)c1C=C	C1=C(c2c3c(cccc13)ccc2)C=C	C1C[C@H]2[C@H](CC1)N(C[CH2])[C@H]1[C@H](S2)CCCC1C#C
c1sc(c2c1OSCCO2)C#C	c1sc(c2c1cc(C=C)c(C)c2)C=C	c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)
c1sc(N(=O)=O)cc1C#C	c1sc(c2c1cc(OC)c(OC)c2)C=C	c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)C#C
C1=CC2=C(C1)N=C(C2)	C1=C2OCCSC2=C([S@@]1NC)C#C	C1C[C@@H]2[C@@H](CC1)N(CC)[C@@H]1[C@@H](S2)CCCC1C=C
c1sc(c2c1cc(OC)cc2)	c1sc(c2c1oc(=O)c(=O)o2)C#C	c1cc2c(s1)c1c(c3c2c2sc(CC)cc2c2cc(CC)sc32)cc(s1)C=C
c1coc(N(=O)=O)c1C#C	c1[nH]c(cc1)C(C(=O)N)=CC=C	c(s1)cc(c(OC(CCC)CCC)2)c1c(OC(CCC)CCC)c(c3)c2sc3C=C
c1scc2c1c(ccc2O)C#C	C1=C2OCCSC2=C([S@@]1NC)C=C	c1ccc2c3cc4n(CC)c5cc6c(cc5c4cc3Cc2c1)Cc1c6ccc(c1)C=C
c1sc(c2c1sc(C#N)c2)	c1sc(c2=CC=C(c12)[CH2])C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)
c1sc(c2c1CCC[C@H]2O)	c1c(c2c(s1)nc1c(c2CC)CCS1)	N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1
c1sc2cc(C(=O)O)sc2c1	c1c(OCC)c(OCC)c(OCC)cc1C=C	c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)C#C
c1oc(c(c1C#N)C#N)C=C	c1sc(c2c1CC[C@@H](N)C2)C#C	c1c2c3c4c(c1CC)c(=O)[nH]c(=O)c4cc(CC)c3c(=O)n(c2=O)C=C
c1c2c(cccc2)c(s1)C=C	c1scc2c1nc(c(NC)n2)N(C)C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)C#C
c1c2c(ncc(C)n2)c(s1)	c1c2[nH]cnc2c(c2c1[nH]cn2)	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2ccc(c1)C=C
c1sc(c2c1cc(C#N)cc2)	c1sc(c2c1S(=O)(=O)CCC2)C=C	c1cc2c(cc1)c1c(=C2C(CN)CN)cc2c(c1)=C(C(CN)CN)c1c2ccc(c1)
c1scc2c1C=C(C2=O)C#C	c1sc(c2c1cc(C(=O)C)cc2)C=C	N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1C#C
c1sc(c2c1[nH]nc2)C#C	c1c(OCC)cc(c(c1)OCC)C=CC=C	N1N[C@H]2[C@@H](C)[C@@H]3[C@@H](NSN3)[C@H](C)[C@H]2N1C=C
c1sc(c(c1)C(=O)C)C=C	c1scc2c1CCC[C@H]2C(=O)OC=C	c(s1)c(CCCCCC)cc1c(s1)c2c(=O)n(C)c(=O)c2c1c(s1)c(CCCCCC)cc1
c1c2c(cccc2)c(s1)C#C	c1c2nc(CC)c(CC)nc2c(s1)C=C	c1cc2c(cc1)c1c(=C2C(CN)CN)cc2c(c1)=C(C(CN)CN)c1c2ccc(c1)C=C
c1sc(c2c1cc(C=O)cc2)	c1sc(c2c1CC[C@@H](N)C2)C=C	c(s1)c(CCCCCC)cc1c(s1)c2c(=O)n(C)c(=O)c2c1c(s1)c(CCCCCC)cc1C#C
c1sc(c(c1OCC)OCC)C=C	c1sc(c2c1nc(CN)c(CN)n2)C=C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)
c1c2c(nccc2)c(s1)C=C	c1sc(c2c1CC(=O)C(=O)C2)C#C	c1c2n(CC)c3c(cc4c(c3)c3c(C4)cccc3)c2cc2c1c1c(C2)cc2c(c1)C(CC)(CC)c1c2cccc1
Nc1sc(N(=O)=O)cc1C#C	c1scc2c1nc(c(NC)n2)N(C)C#C	c1cc2c(cc1)c1c(C2(CC)CC)cc2c(c1)C(CC)(CC)c1c2cc2c(c1)c1c(C2(CC)CC)cc(cc1)C#C
C(=C)c1sc(C)c(c1)C#C	c1sc(c2c1sc(=O)c(=O)s2)C#C	c1c(SC)c(SC)c(c2c1nc1c(n2)c2c(nc3cc(SC)c(SC)cc3n2)c2c1nc1cc(SC)c(SC)cc1n2)C=C
c1sc2c(c3ccoc3cc2c1)	c1sc(c2c1oc(=O)c(=O)o2)C=C	c1c2n(CC)c3c(cc4c(c3)c3c(C4)cccc3)c2cc2c1c1c(C2)cc2c(c1)C(CC)(CC)c1c2cccc1C#C
c1sc(c2c1C(=O)NC2=O)	c1sc(c2c1c(C#N)ccc2C#N)C#C	c1occ(N(=O)=O)c1NC=C
c1sc(c2c1C(=O)CC2=O)	c1sc(c2c1cc(C(=O)OC)cc2)C=C	c1sc(c2c1cc(N(=O)=O)cc2)C=C

Table B40: List of SMILES for the 1759 monomer data set. (Part 13 of 14)

C1=C(C)C(C)=C(C2=N[C@H]3N4[C@@H](N[C@@H]5C[C@H](C)[C@H](C[C@@H]35)C)[C@H]3[C@H](N[C@@H]4[C@H]12)C[C@H]([C@H](C3)C)C)
C1=C(C)C(C)=C(C2=N[C@H]3N4[C@@H](N[C@@H]5C[C@H](C)[C@H](C[C@@H]35)C)[C@H]3[C@H](N[C@@H]4[C@H]12)C[C@H]([C@H](C3)C)C)C=C

c1cc2c(=O)n(C)c(=O)c3c2c2c1C1=C4[C@@H](c2cc3)C=CC2=C4[C@@H](C(=O)N(C2=O)C)C(=C1)C=C

c1cc2c(=O)n(C)c(=O)c3c2c2c1C1=C4[C@@H](c2cc3)C=CC2=C4[C@@H](C(=O)N(C2=O)C)C(=C1)

Table B41: List of SMILES for the 1759 monomer data set. (Part 14 of 14)

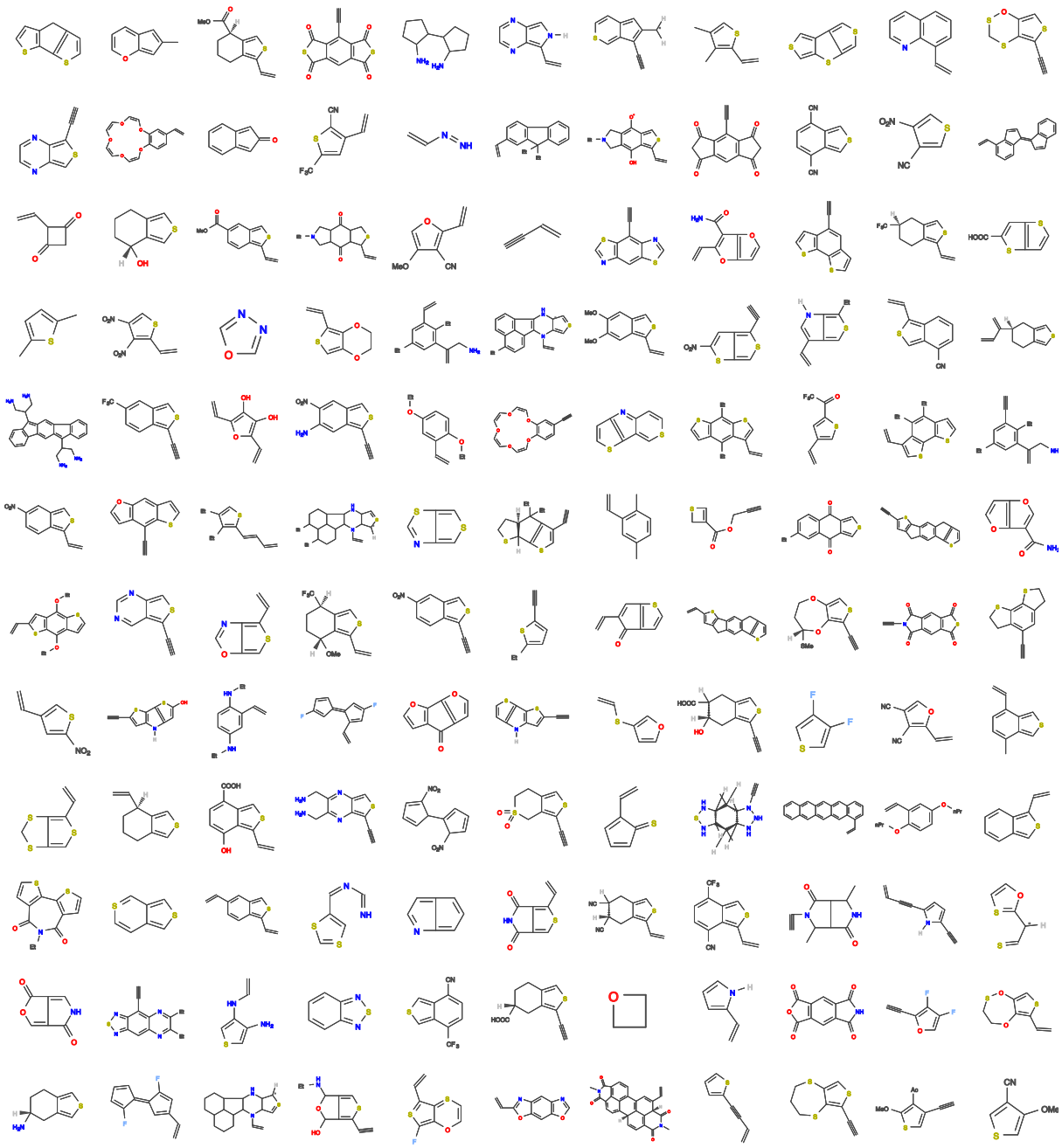


Figure B28: Molecules in the 1759 monomer dataset (1 of 14).

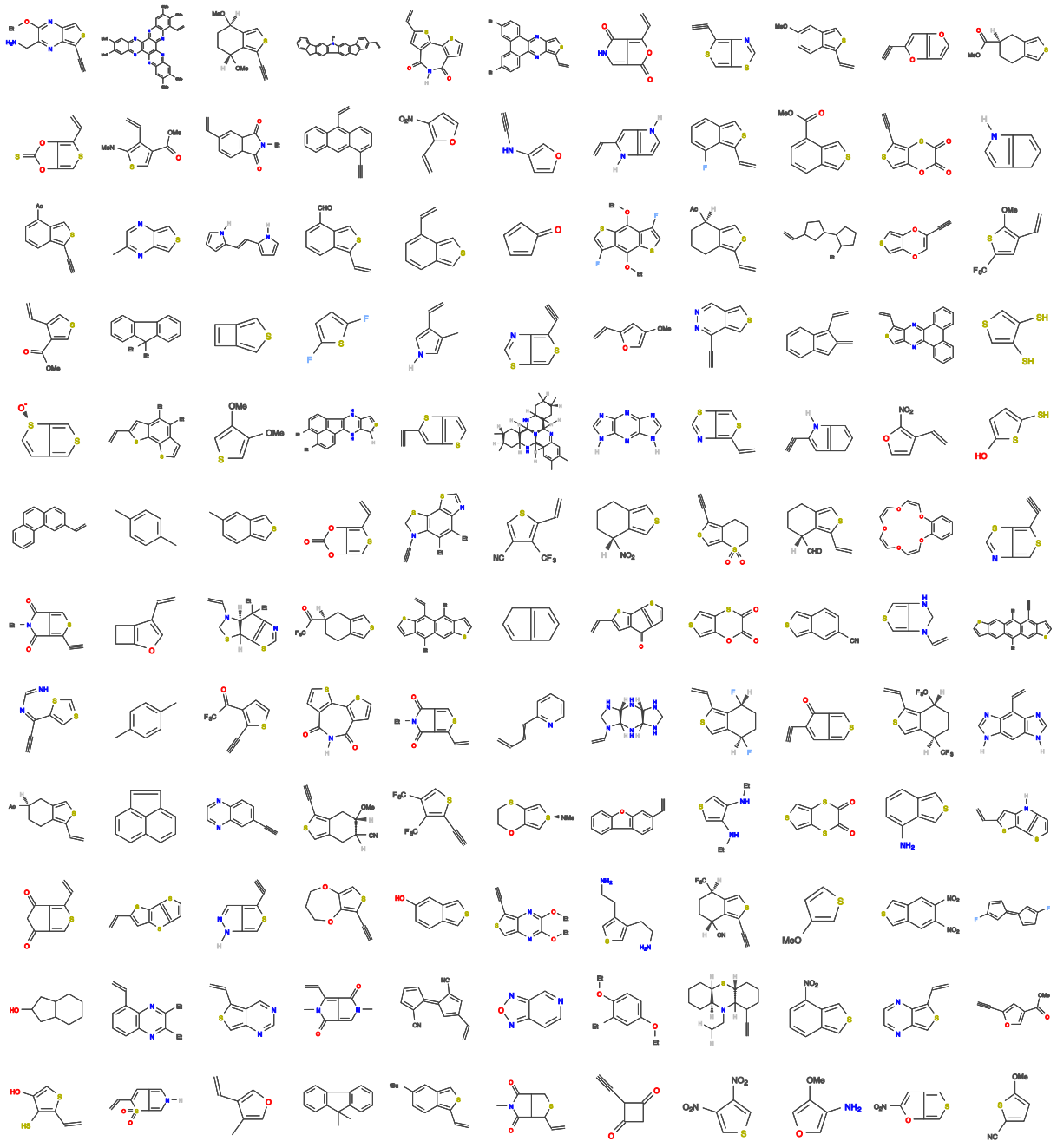


Figure B29: Molecules in the 1759 monomer dataset (2 of 14).

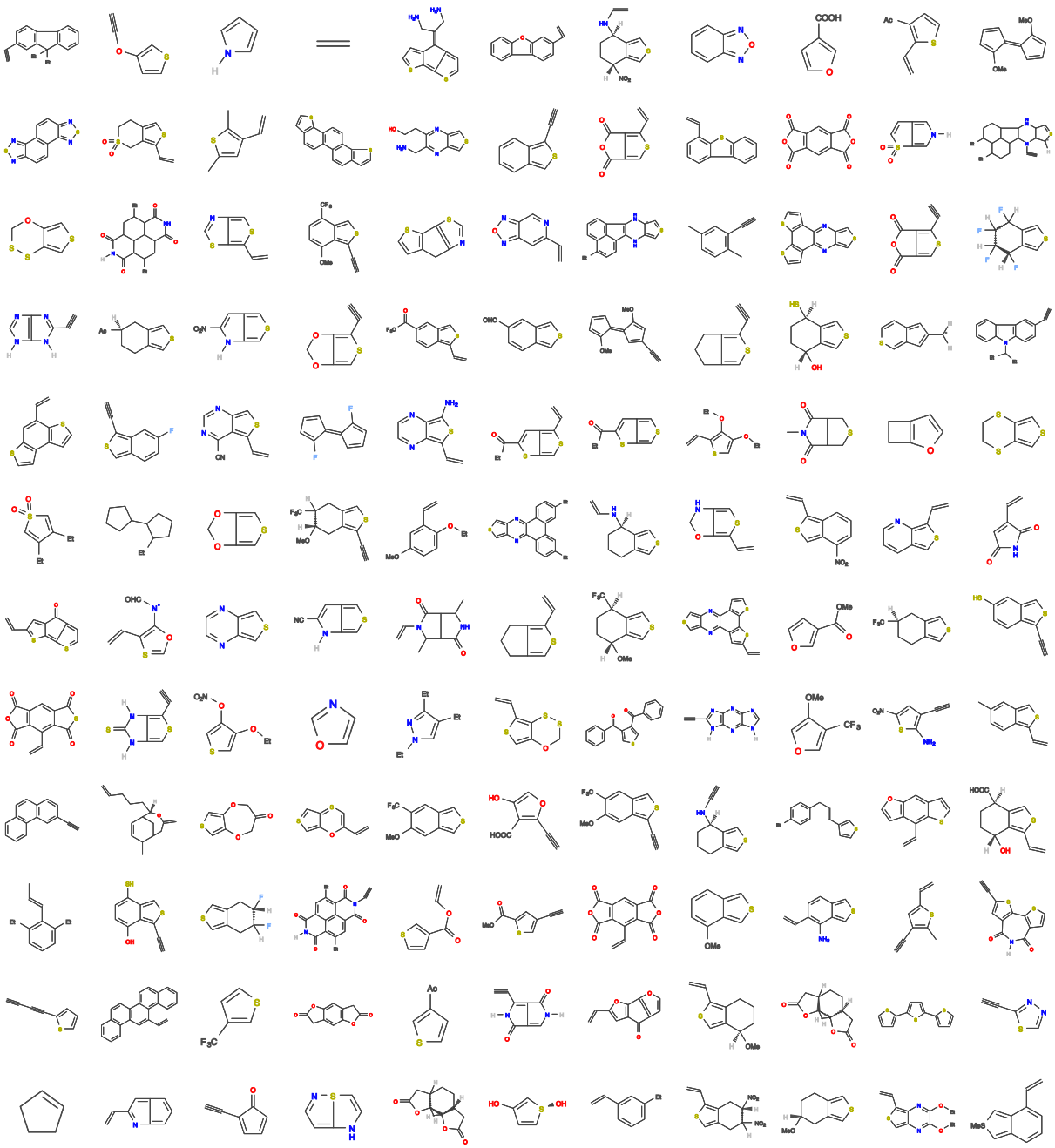


Figure B30: Molecules in the 1759 monomer dataset (3 of 14).

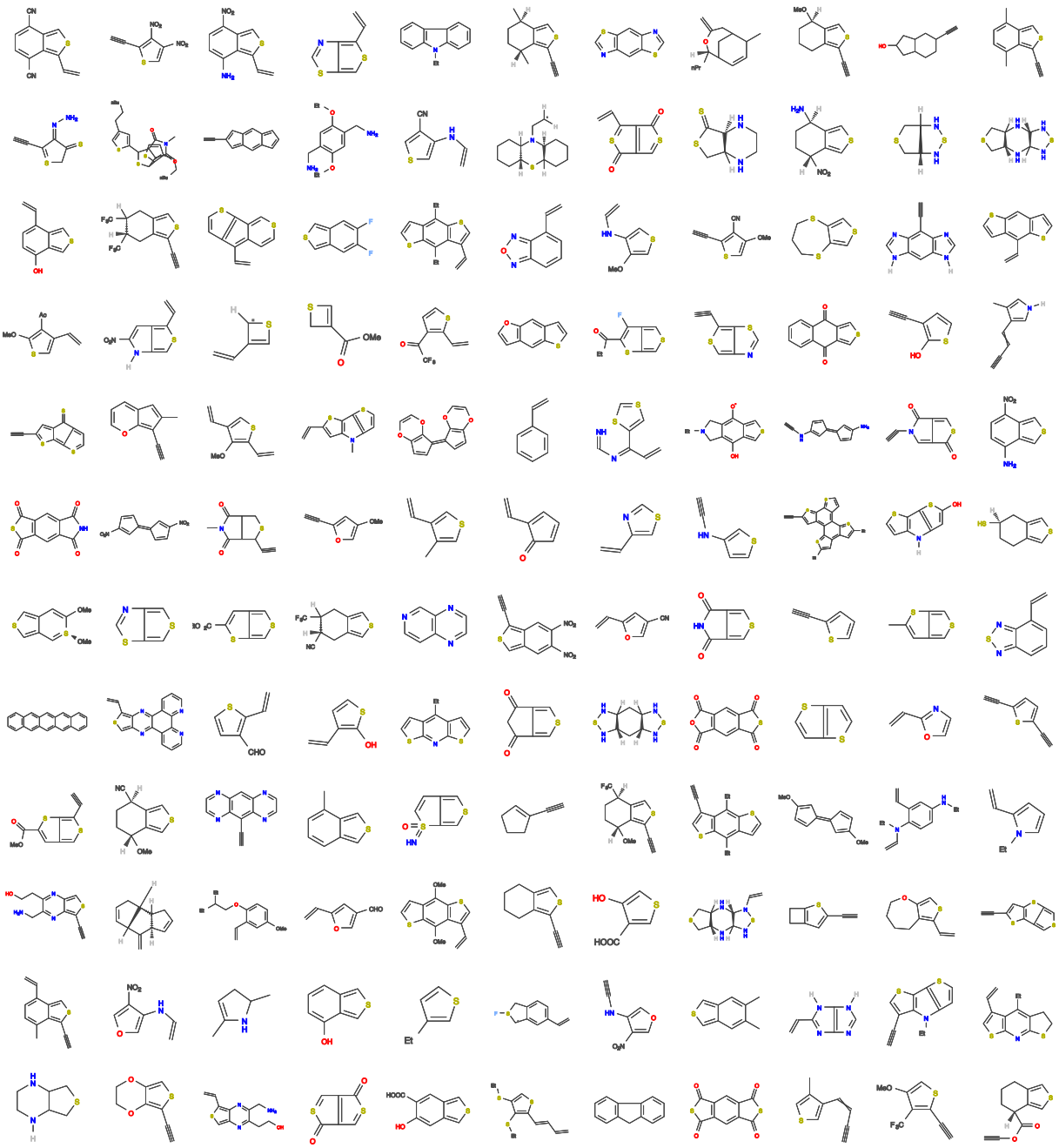


Figure B31: Molecules in the 1759 monomer dataset (4 of 14).

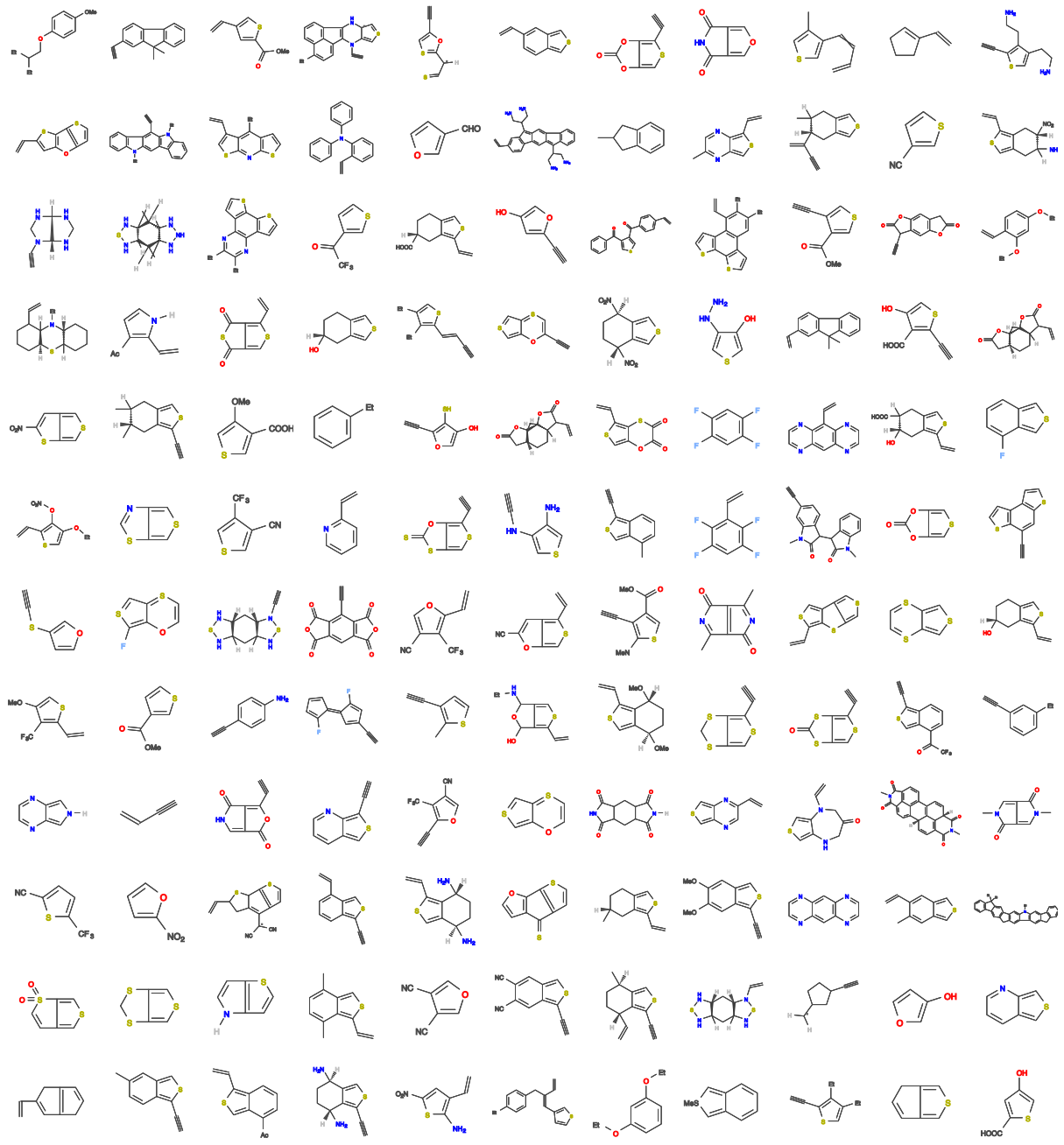


Figure B32: Molecules in the 1759 monomer dataset (5 of 14).

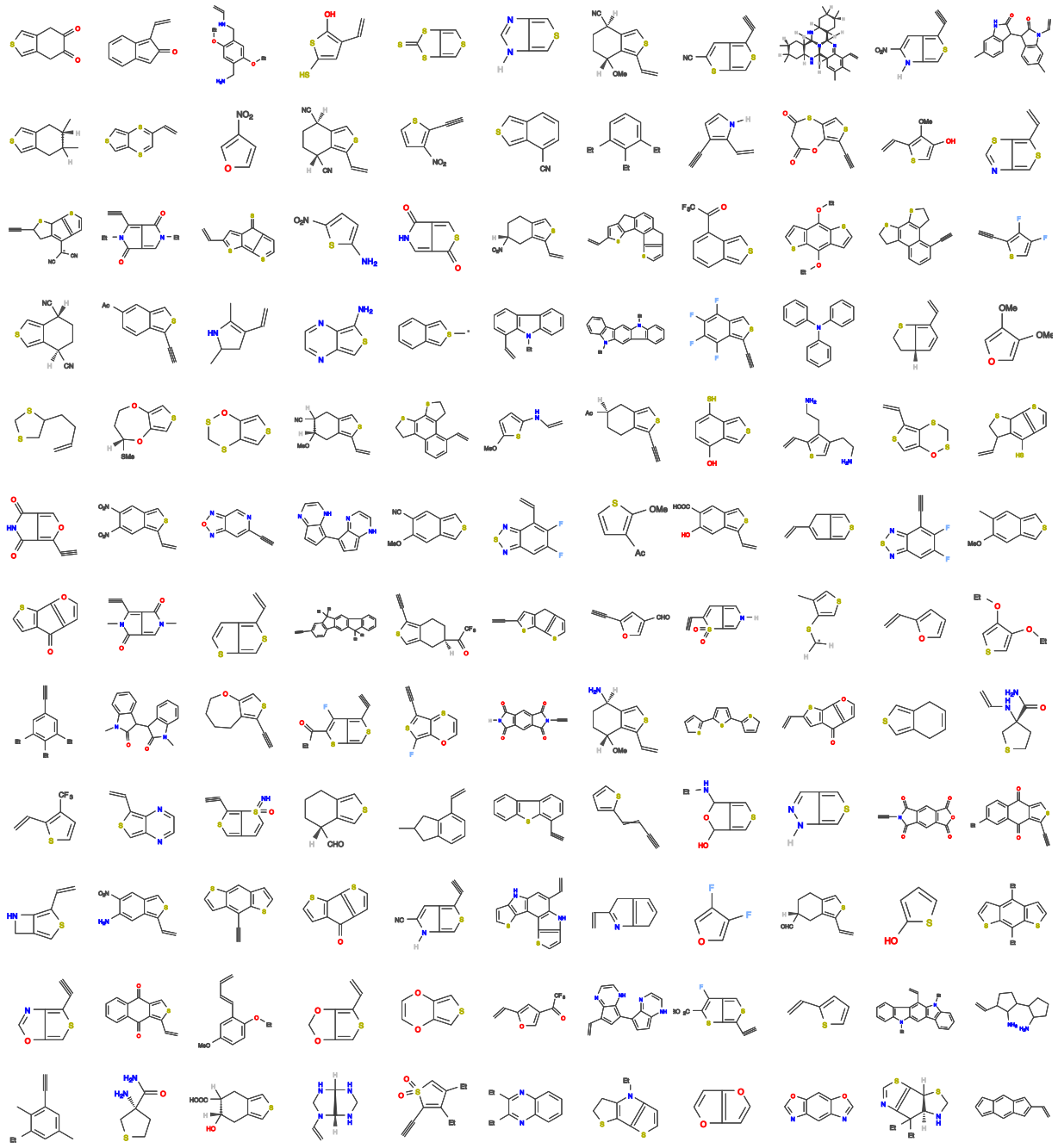


Figure B33: Molecules in the 1759 monomer dataset (6 of 14).

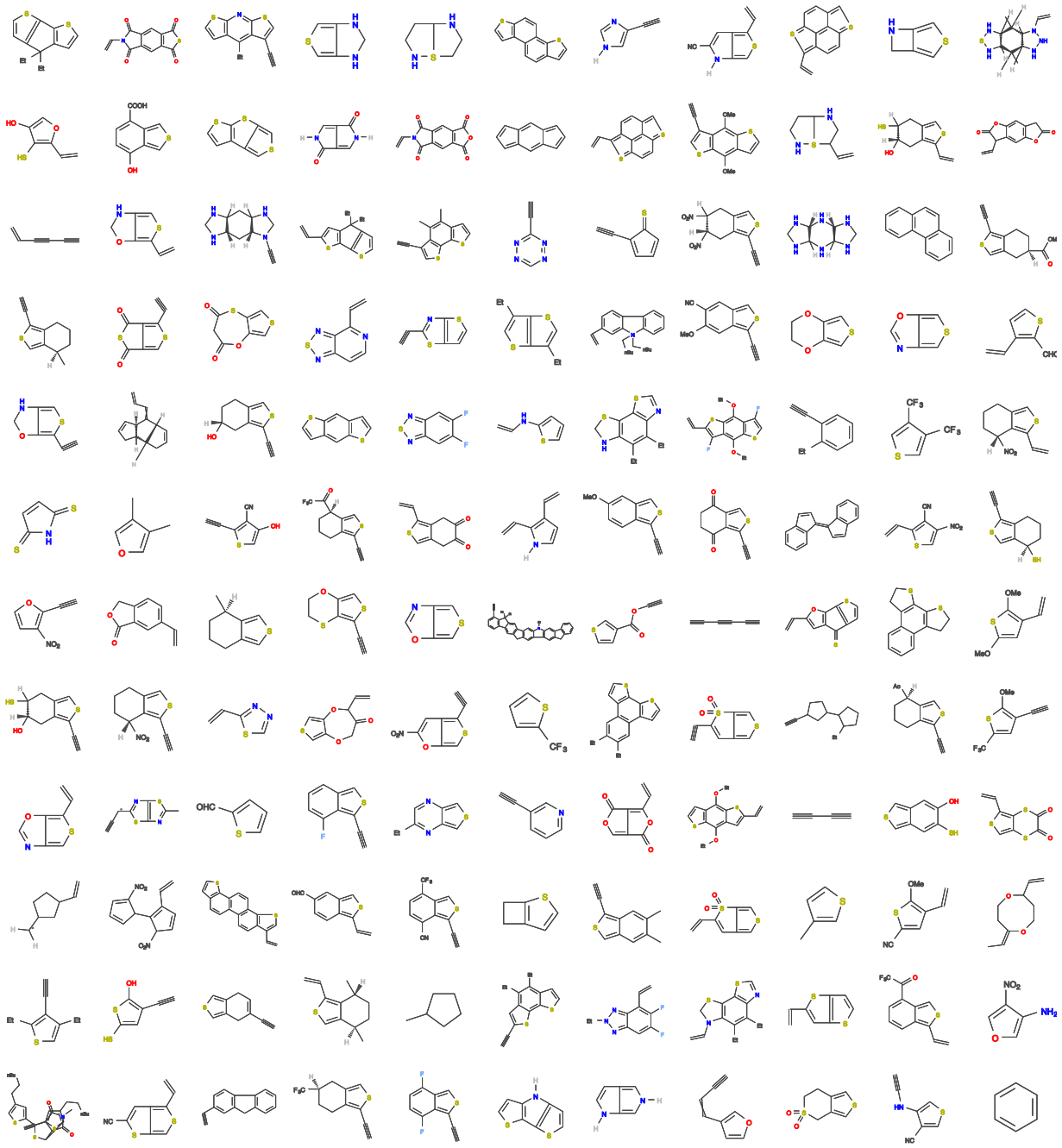


Figure B34: Molecules in the 1759 monomer dataset (7 of 14).

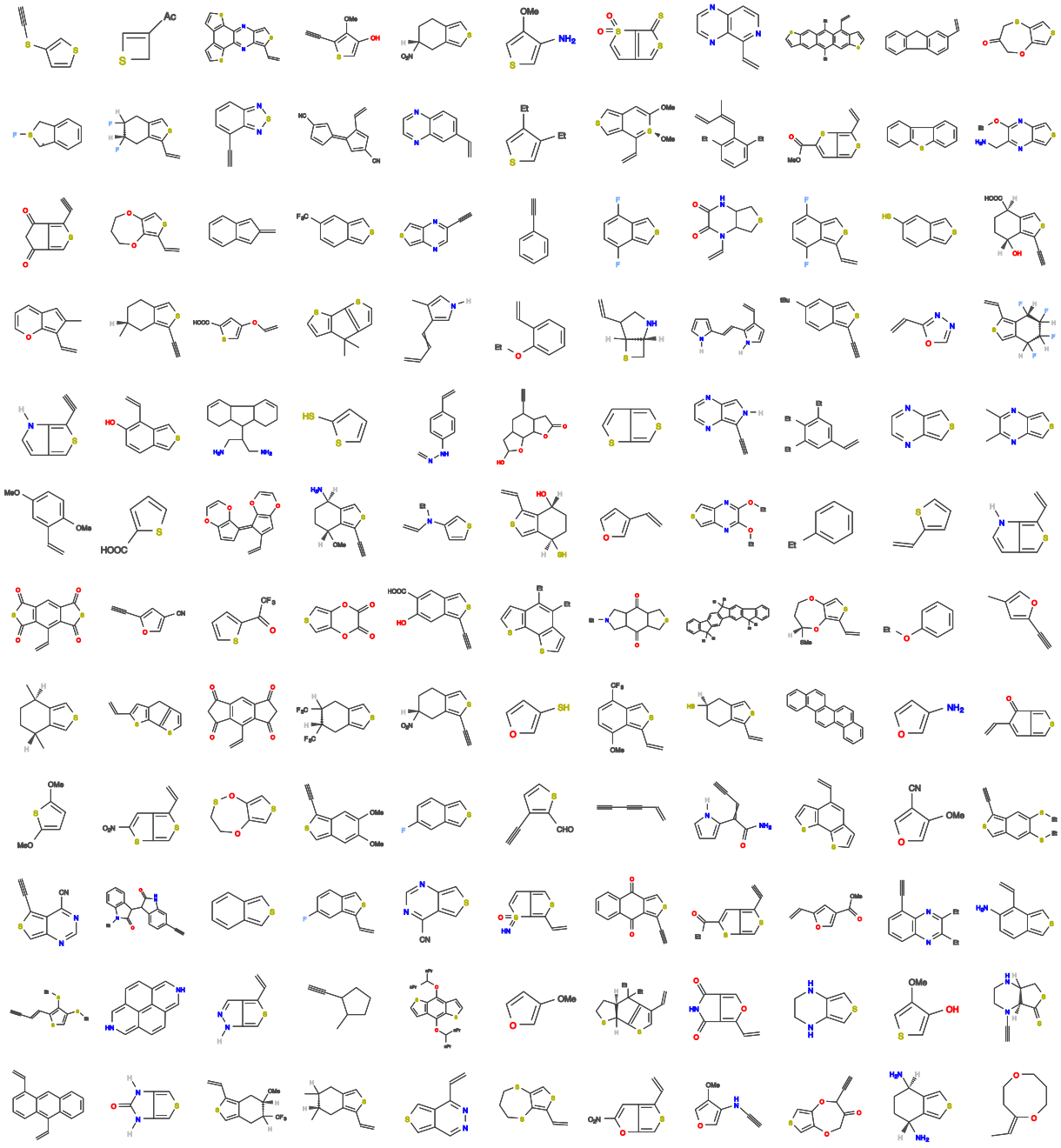


Figure B35: Molecules in the 1759 monomer dataset (8 of 14).

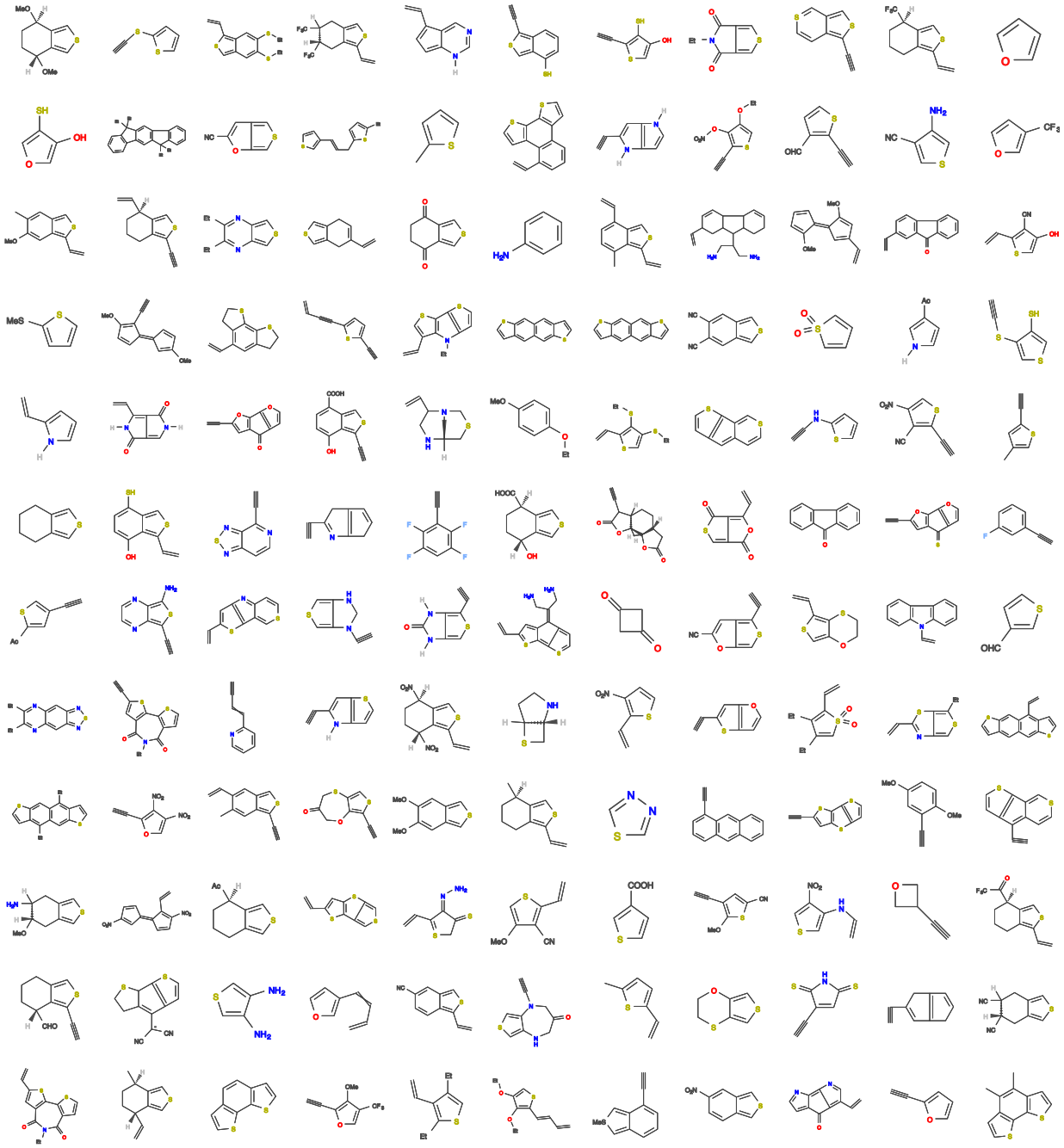


Figure B36: Molecules in the 1759 monomer dataset (9 of 14).

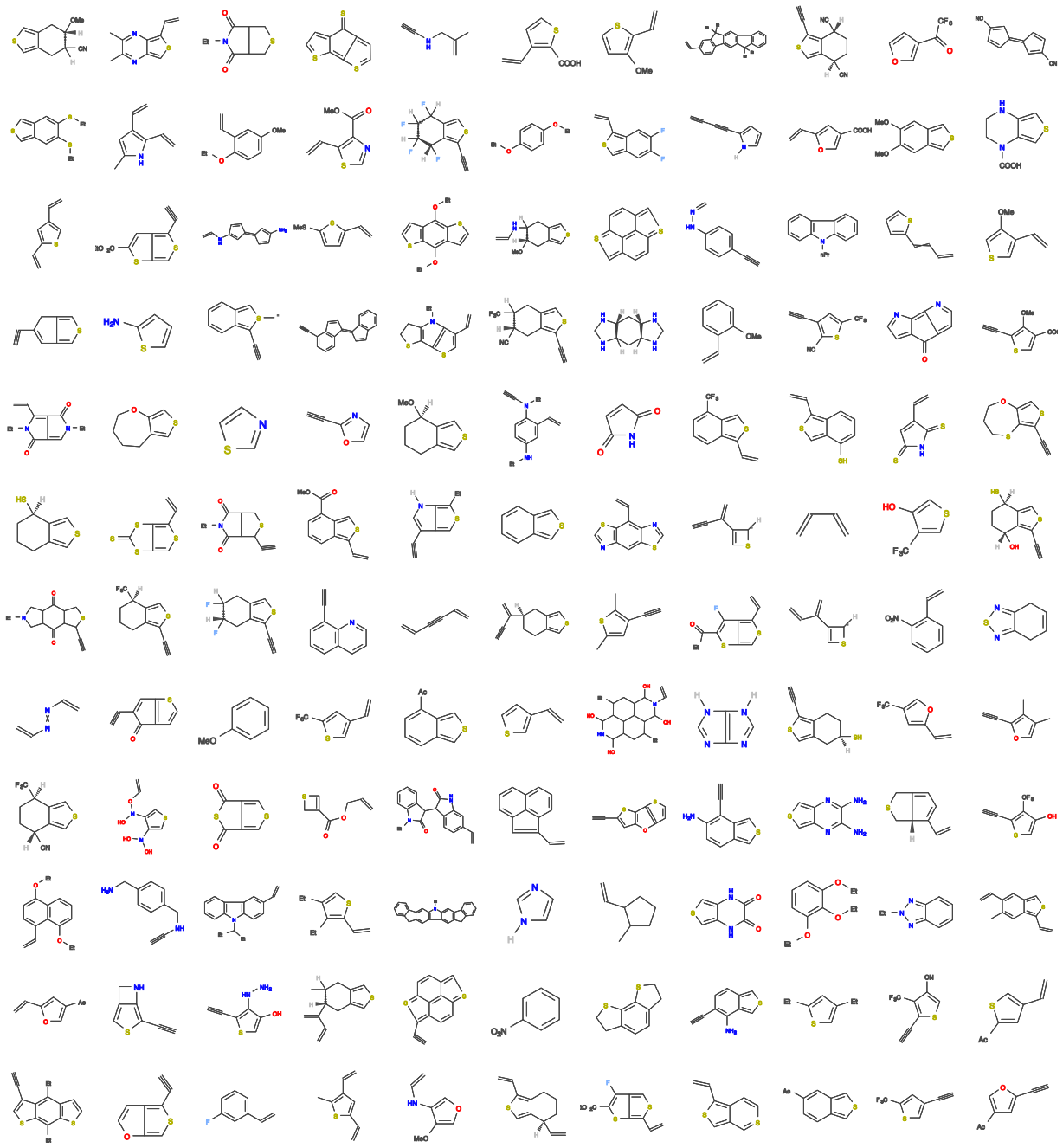


Figure B37: Molecules in the 1759 monomer dataset (10 of 14).

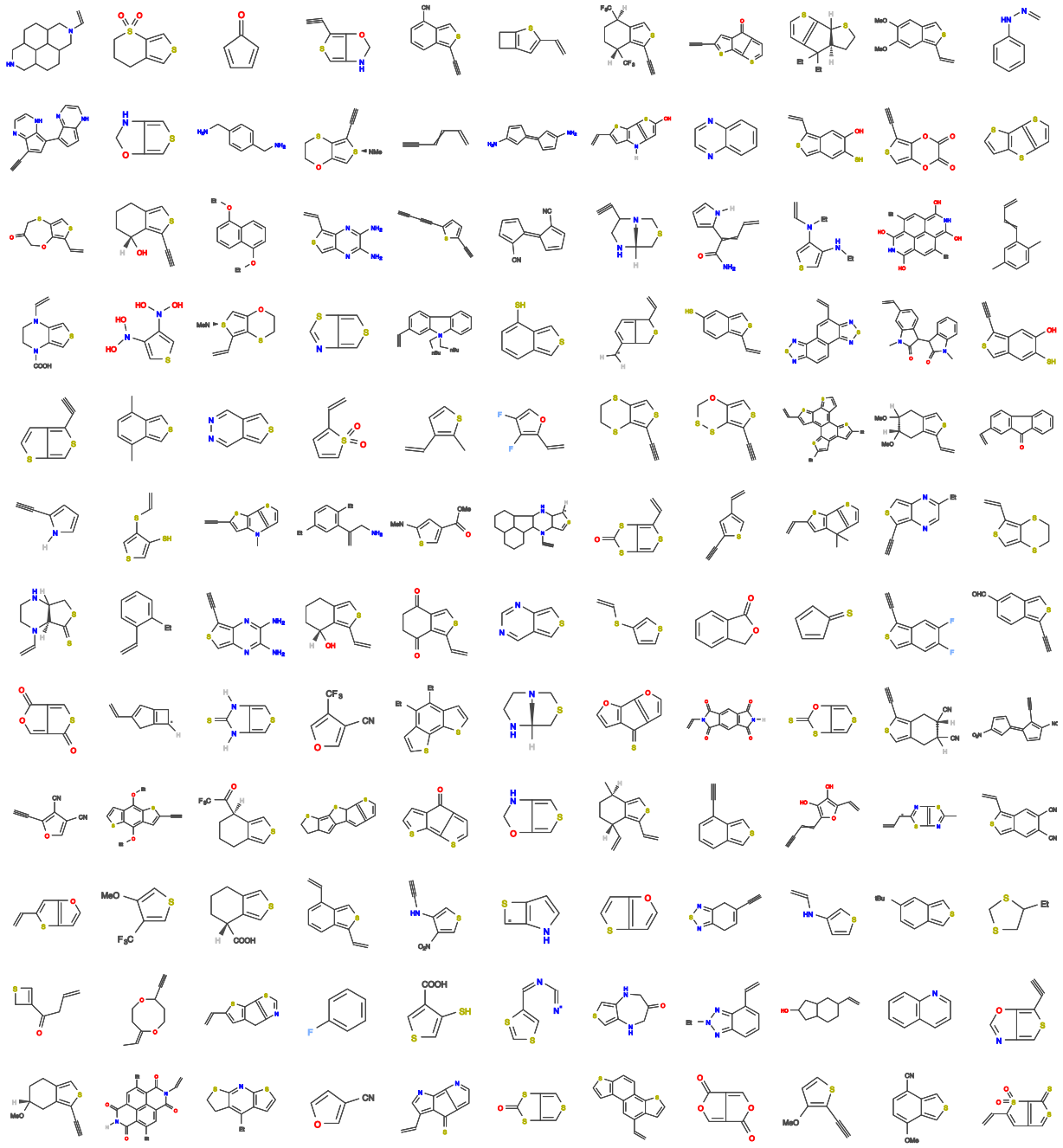


Figure B38: Molecules in the 1759 monomer dataset (11 of 14).

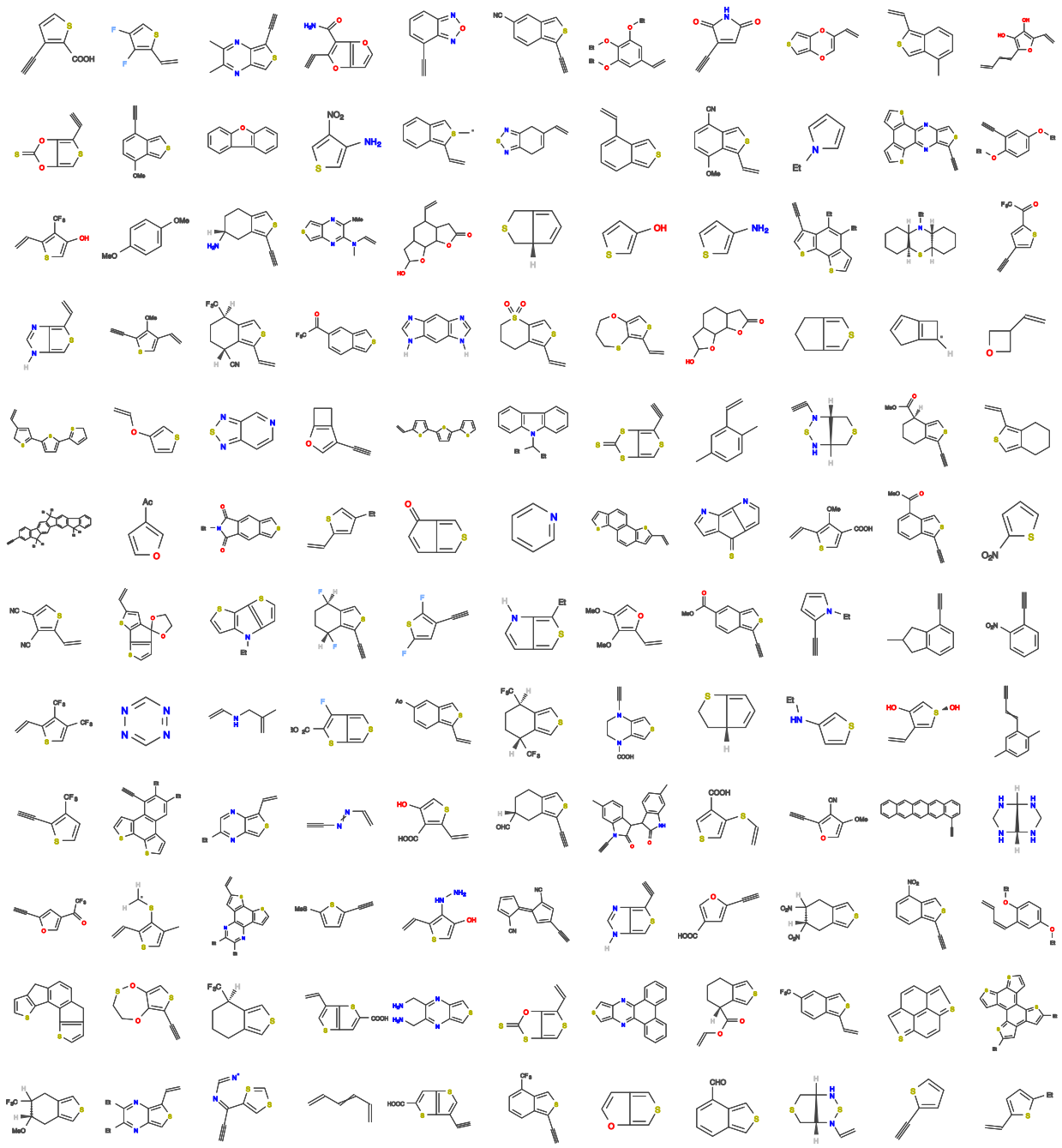


Figure B39: Molecules in the 1759 monomer dataset (12 of 14).

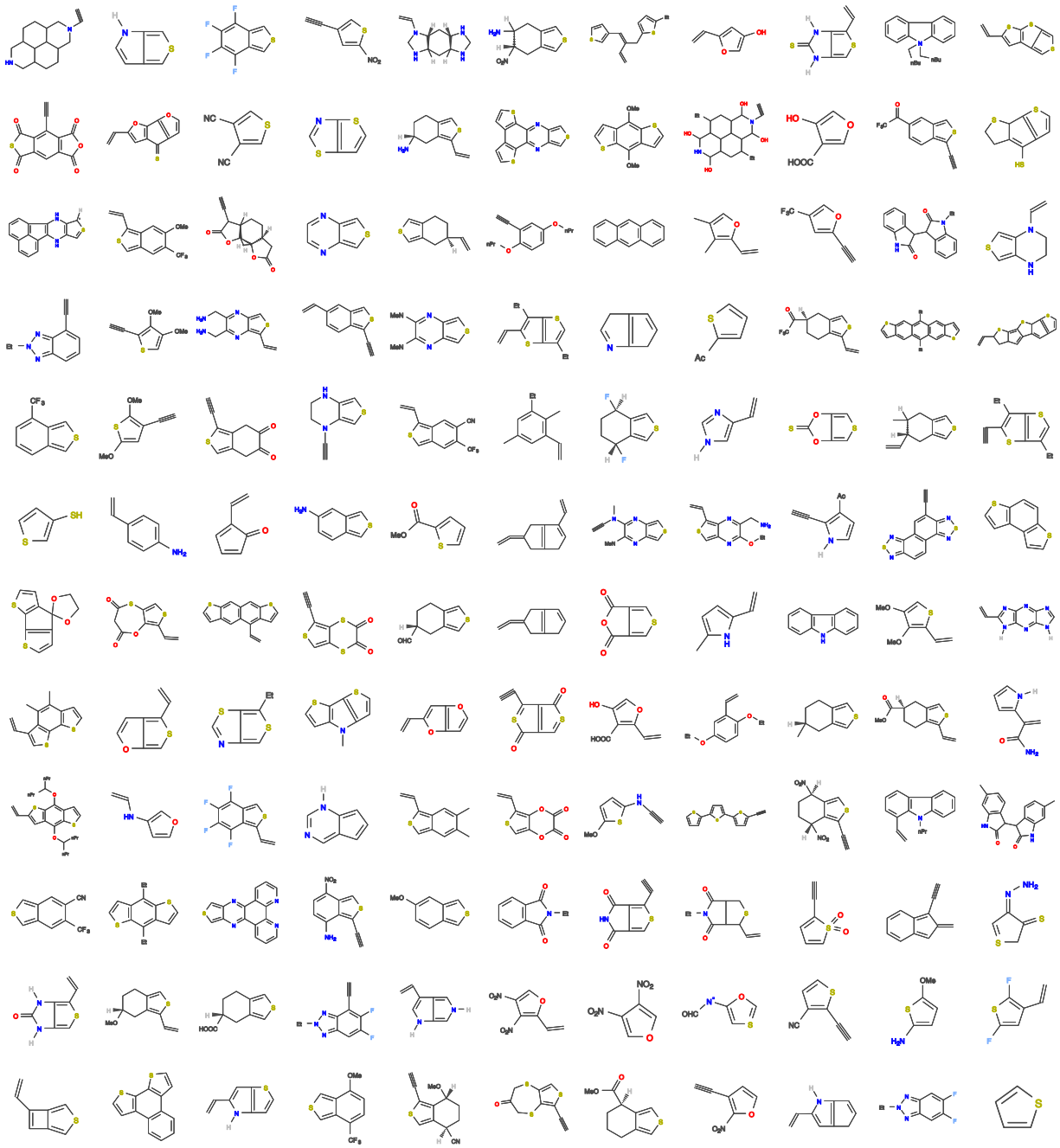


Figure B40: Molecules in the 1759 monomer dataset (13 of 14).

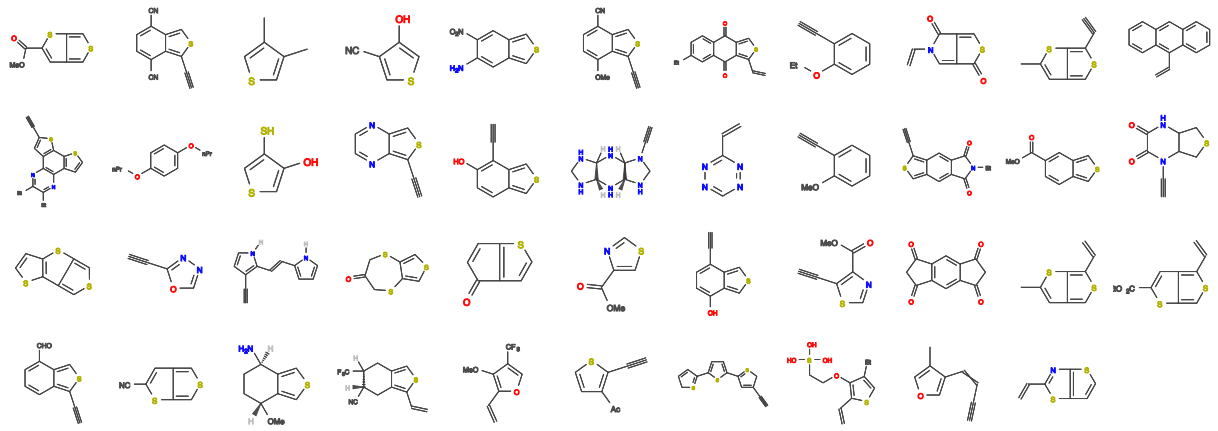


Figure B41: Molecules in the 1759 monomer dataset (14 of 14).

APPENDIX C

ZINDO SELENIUM PARAMETERIZATION

C.1 INTRODUCTION

As stated previously, conjugated thiophenes are promising materials for OPV materials. Periodic table trends indicate that elements down a period can be expected to have similar properties. Therefore, the element under sulfur, which is in thiophene materials, is of interest for OPV materials. Selenium therefore is of interest to improve on the OPV materials that have already been developed. Although selenium is a promising element to include in the production of OPV materials, it is difficult to include in computational screening studies. It is a challenging problem to obtain appropriate parameters for selenium. Although not the most accurate computational method, ZINDO can be useful as an initial screening method when beginning to run the Genetic Algorithm (GA) due to its ability to quickly give an estimate for the excited state energies in pi-conjugated molecules. The problem with using ZINDO in our experiment is that it has not been parameterized for selenium which is present in many of the monomers in our initial group. If ZINDO were to be used without the proper parameters, then the estimates for monomers containing selenium would be very inaccurate and many potentially good materials could be eliminated too early in the process.

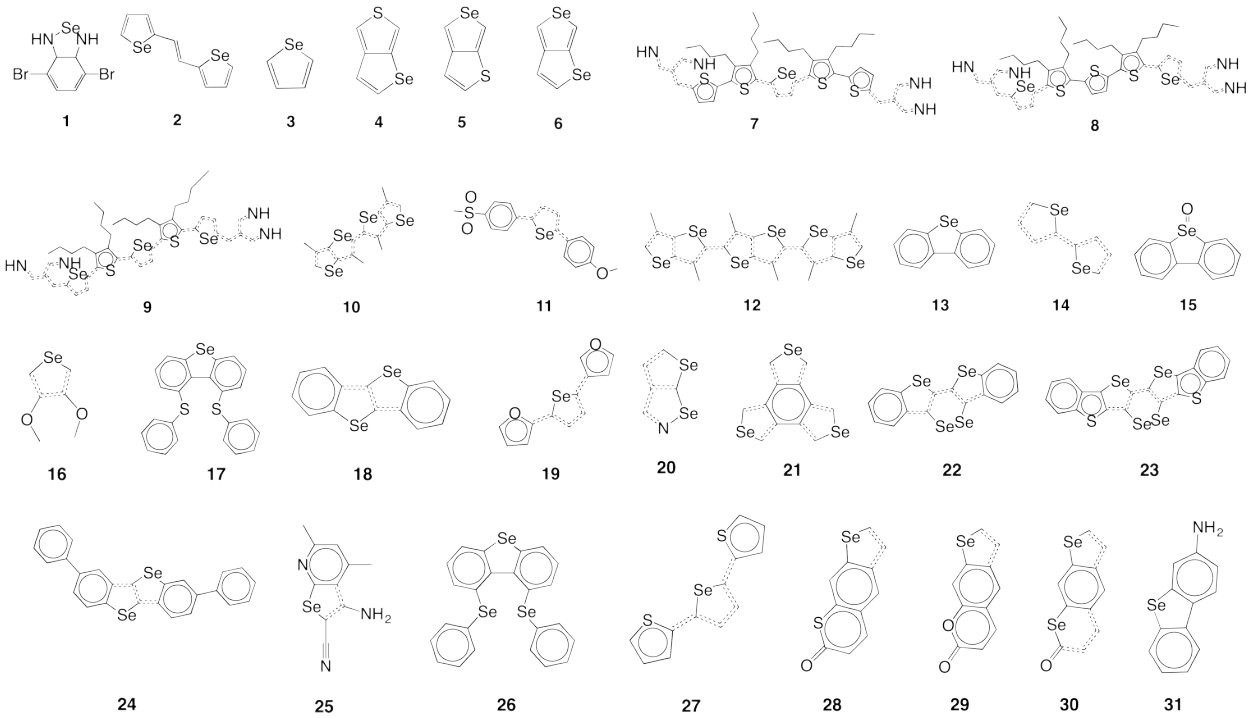


Figure C1: Structure of the molecules used in the selenium paramaterization

C.2 EXPERIMENT

Determination of appropriate selenium parameters involved several steps including performing a literature search of known selenium compounds, perform calculations on these compound and then compare the calculations with the known values.

A literature and patent search was performed to identify a diverse selection of selenium compounds which have been synthesized. Experimental wavelength (lambda) values of the each selenium compound were compiled. This literature search produced 31 selenium species with experimental lambda values. The structures of the compounds are shown in Figure C1 and are demonstrated to be a diverse set of structures. A diverse set of molecules ensures that the calculated parameters will work for a molecularly diverse set of selenium molecules when applied to new molecules which are not in the test set. For each of the 31 selenium containing species, SMILES (Table C1) were used to generate input files to calculate energies with PM6, B3LYP/6-31G* and TD-DFT methods. The energy values from each calculation method were compared with the experimentally known lambda values. The B3LYP/6-31G* and PM6 calculations do not show a strong correlation with the expected values (Figure C2(left)), but the TD-DFT calculations showed a high correlation with the experimental values (Figure C2(right)). Since the TD-DFT data seemed to match the experimental values most closely, for the set of selenium molecules in the data set, these values were then used to optimize the gamma and exponent parameters to find their minimum values.

Since the gamma and exponent parameters included with the ZINDO package are known to be inaccurate for selenium, these parameters were varied to determine new values for these parameters. The exponent value was varied from 2.420 to 2.460. Average error in the calculated values was compared (C3 (left)) and an exponent value of 2.439 was determined to produce the smallest average error. Optimization of the gamma parameter was performed by varying the parameter from 5.55 to 7.78 at the fixed exponent value of 2.439, the optimal calculated value of the exponent. The optimal gamma value was determined by the value which produced the smallest average error and corresponds with a gamma value of 6.425 (C3 (right)).

1	Brc(cc1)c2n[Se]nc2c1Br
2	c([Se]1)ccc1ccc(cc2)[Se]c2
3	c([Se]1)ccc1
4	c(s1)c2[Se]ccc2c1
5	c([Se]1)c2sccc2c1
6	c([Se]1)c2[Se]ccc2c1
7	ncc(cn)cc(s1)ccc1c(s1)c(CCCC)c(CCCC)c1c([Se]1)ccc1c(s1)c(CCCC)c(CCCC)c1c(s1)ccc1cc(cn)cn
8	ncc(cn)cc([Se]1)ccc1c(s1)c(CCCC)c(CCCC)c1c(s1)ccc1c(s1)c(CCCC)c(CCCC)c1c([Se]1)ccc1cc(cn)cn
9	ncc(cn)cc([Se]1)ccc1c(s1)c(CCCC)c(CCCC)c1c([Se]1)ccc1c(s1)c(CCCC)c(CCCC)c1c([Se]1)ccc1cc(cn)cn
10	c([Se]1)c(C)c([Se]2)c1c(C)c2c([Se]1)c(C)c([Se]2)c1c(C)c2
11	CS(=O)(=O)c(cc1)ccc1c([Se]1)ccc1c(cc1)ccc1OC
12	c([Se]1)c(C)c([Se]2)c1c(C)c2c([Se]1)c(C)c([Se]2)c1c(C)c2c([Se]1)c(C)c([Se]2)c1c(C)c2
13	c(c1)ccc([Se]2)c1c(cc3)c2cc3
14	c([Se]1)ccc1c(cc1)[Se]c1
15	c(c1)ccc([Se](=O)2)c1c(cc3)c2cc3
16	c([Se]1)c(OC)c(OC)c1
17	c(c(Sc1ccccc1)1)ccc([Se]2)c1c(c(Sc1ccccc1)c3)c2cc3
18	c(c1)ccc2c1[Se]c3c2[Se]c(cc4)c3cc4
19	c(co1)cc1c([Se]1)ccc1c(co1)cc1
20	c([Se]1)cc(c2)c1[Se]n2
21	c([Se]1)c2c3c[Se]cc3c4c[Se]cc4c2c1
22	c(c1)ccc2c1[Se]c3c2[Se][Se]c4c3[Se]c(c5)c4ccc5
23	c(c1)ccc(s2)c1c([Se]3)c2c([Se][Se]4)c3c([Se]5)c4c(s6)c5c(cc7)c6cc7
24	c(cc1)ccc1c(c1)ccc2c1[Se]c3c2[Se]c(cc4)c3cc4c(cc1)ccc1
25	c(c(C)1)c(C)nc([Se]2)c1c(N)c2C#N
26	c(c1)cc([Se]c4ccccc4)c2c1[Se]c(cc3)c2c([Se]c4ccccc4)c3
27	c(s1)ccc1c([Se]1)ccc1c(s1)ccc1
28	c(c1)c(=O)sc(c2)c1cc(c3)c2[Se]c3
29	c(c1)c(=O)oc(c2)c1cc(c3)c2[Se]c3
30	c(c1)c(=O)[Se]c(c2)c1cc(c3)c2[Se]c3
31	c(c1)c(N)cc([Se]2)c1c(cc3)c2cc3

Table C1: SMILES of the molecules used in the selenium paramaterization

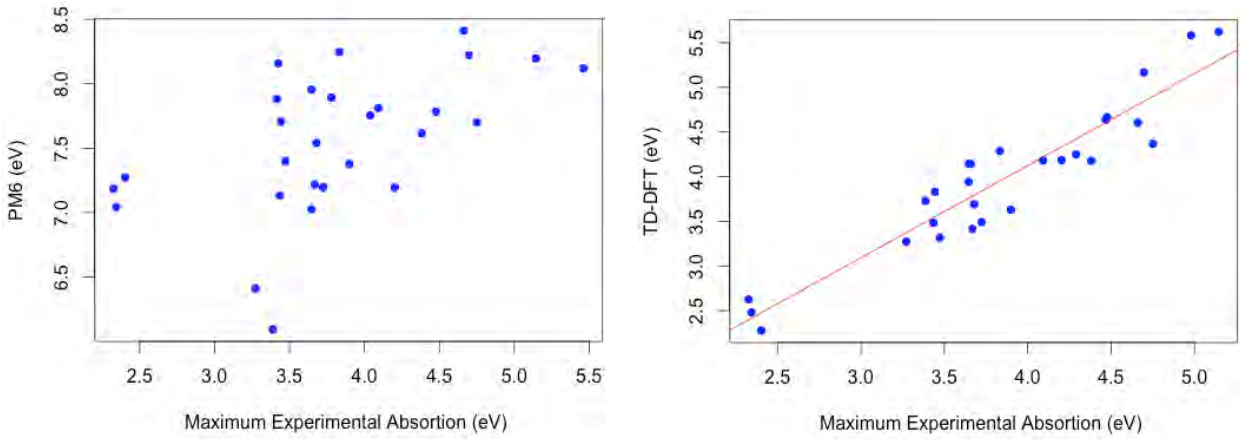


Figure C2: Selenium molecule calculations using PM6 and TD-DFT compared to the experimental values for the sample set of selenium compounds.

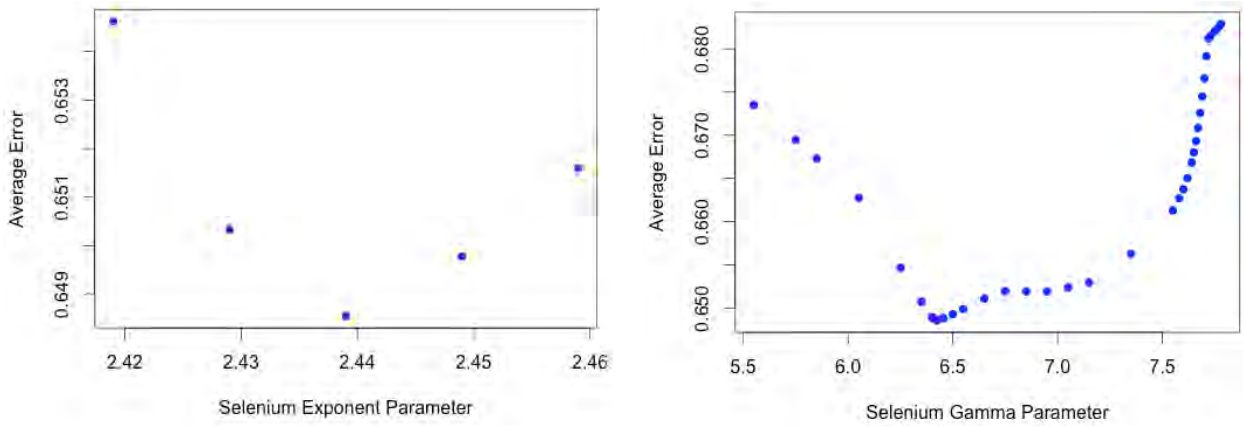


Figure C3: Selenium molecule calculations using PM6 and TD-DFT compared to the experimental values for the sample set of selenium compounds.

C.3 CONCLUSION

When this parameterization is compared with other parameters determined for Se monomers, the optimization results in different sets of parameters. If one set of parameters are not available for a method then the method is by nature inaccurate. We therefore conclude that ZINDO is not an appropriate method to use when incorporating d orbital electrons, as are present in selenium. During GA screening and other experiments, we will therefore omit compounds containing selenium since they will not be calculated appropriately. In addition, experimentalists prefer not to work on selenium containing molecules due to the toxicity. Since it is our goal to identify different types of compounds as good OPV material candidates, based on this parameterization, we have not included selenium containing molecules.

BIBLIOGRAPHY

[1] Kanal, I. Y.; Owens, S. G.; Bechtel, J. S.; Hutchison, G. R. *The Journal of Physical Chemistry Letters* **2013**, 1613–1623.

[2] NREL, S. R. **2004**, 1–2.

[3] Energy, E. I. A. U. S. D. O. **2004**, 1–390.

[4] Li, G.; Zhu, R.; Yang, Y. **2012**, *6*, 153–161.

[5] Duan, C.; Huang, F.; Cao, Y. *Journal of Materials Chemistry* **2012**, *22*, 10416–10434.

[6] Tang, C. W. *Applied Physics Letters* **1986**, *48*, 183.

[7] Yu, G.; Gao, J.; Hummelen, J. C.; Wudl, F.; Heeger, A. J. *Science* **1995**, *270*, 1789–1790.

[8] Liang, Y.; Yu, L. *Accounts of Chemical Research* **2010**, *43*, 1227–1236.

[9] He, Z.; Zhong, C.; Su, S.; Xu, M.; Wu, H.; Cao, Y. *Nature Photonics* **2012**, *6*, 593–597.

[10] Janssen, R. A. J.; Nelson, J. *Advanced Materials* **2012**, *25*, 1847–1858.

[11] Zhou, H.; Yang, L.; You, W. *Macromolecules* **2012**, *45*, 607–632.

[12] Facchetti, A. *Chemistry of Materials* **2011**, *23*, 733–758.

[13] Wang, Y.; Wei, W.; Liu, X.; Gu, Y. *Solar Energy Materials and Solar Cells* **2012**, *98*, 129–145.

[14] Sommer, M.; Huettner, S.; Thelakkat, M. *Journal of Materials Chemistry* **2010**, *20*, 10788–10797.

[15] Baranovskii, S. D.; Wiemer, M.; Nenashev, A. V.; Jansson, F.; Gebhard, F. *The Journal of Physical Chemistry Letters* **2012**, *3*, 1214–1221.

[16] Bakulin, A. A.; Rao, A.; Pavelyev, V. G.; van Loosdrecht, P. H. M.; Pshenichnikov, M. S.; Niedzialek, D.; Cornil, J.; Beljonne, D.; Friend, R. H. *Science* **2012**, *335*, 1340–1344.

[17] Beljonne, D.; Cornil, J.; Muccioli, L.; Zannoni, C.; Brédas, J.-L.; Castet, F. *Chemistry of Materials* **2011**, *23*, 591–609.

[18] Troisi, A. *Organic Electronics* **2011**, *12*, 1988–1991.

[19] Li, Y. *Accounts of Chemical Research* **2012**, *45*, 723–733.

[20] Lyons, B. P.; Clarke, N.; Groves, C. *Energy & Environmental Science* **2012**, *5*, 7657.

[21] Vehoff, T.; Baumeier, B.; Troisi, A.; Andrienko, D. *Journal of the American Chemical Society* **2010**, *132*, 11702–11708.

[22] McMahon, D. P.; Troisi, A. *ChemPhysChem* **2010**, *11*, 2067–2074.

[23] Jailaubekov, A. E.; Willard, A. P.; Tritsch, J. R.; Chan, W.-L.; Sai, N.; Gearba, R.; Kaake, L. G.; Williams, K. J.; Leung, K.; Rossky, P. J.; Zhu, X.-Y. *Nature Materials* **2012**, *12*, 66–73.

[24] Kaake, L.; Dang, X.-D.; Leong, W. L.; Zhang, Y.; Heeger, A.; Nguyen, T.-Q. *Advanced Materials* **2012**, *25*, 1706–1712.

[25] Maurano, A.; Hamilton, R.; Shuttle, C. G.; Ballantyne, A. M.; Nelson, J.; O'Regan, B.; Zhang, W.; McCulloch, I.; Azimi, H.; Morana, M.; Brabec, C. J.; Durrant, J. R. *Advanced Materials* **2010**, *22*, 4987–4992.

[26] Fischer, S. A.; Isborn, C. M.; Prezhdo, O. V. *Chemical Science* **2011**, *2*, 400–406.

[27] Massip, S.; Oberhumer, P. M.; Tu, G.; Albert-Seifried, S.; Huck, W. T. S.; Friend, R. H.; Greenham, N. C. *The Journal of Physical Chemistry C* **2011**, *115*, 25046–25055.

[28] *Materials Genome Initiative for Global Competitiveness*; 2012.

[29] Rajan, K. *Annual Review of Materials Research* **2008**, *38*, 299–322.

[30] Sokolov, A. N.; Atahan-Evrenk, S.; Mondal, R.; Akkerman, H. B.; nchez Carrera, R. S. S. a.; Granados-Focil, S.; Schrier, J.; Mannsfeld, S. C. B.; Zoombelt, A. P.; Bao, Z.; aacute n Aspuru-Guzik, A. *Nature Communications* **2011**, *2*, 437–438.

[31] Giri, G.; Verploegen, E.; Mannsfeld, S. C. B.; Atahan-Evrenk, S.; Kim, D. H.; Lee, S. Y.; Becerril, H. A.; Aspuru-Guzik, A.; Toney, M. F.; Bao, Z. *Nature* **2011**, *480*, 504–508.

[32] Sánchez-Carrera, R. S.; Atahan, S.; Schrier, J.; Aspuru-Guzik, A. *The Journal of Physical Chemistry C* **2010**, *114*, 2334–2340.

[33] Castelli, I. E.; Olsen, T.; Datta, S.; Landis, D. D.; Dahl, S.; Thygesen, K. S.; Jacobsen, K. W. *Energy & Environmental Science* **2012**, *5*, 5814.

[34] Castelli, I. E.; Landis, D. D.; Thygesen, K. S.; Dahl, S.; Chorkendorff, I.; Jaramillo, T. F.; Jacobsen, K. W. *Energy & Environmental Science* **2012**, *5*, 9034.

[35] Wu, Y.; Lazic, P.; Hautier, G.; Persson, K.; Ceder, G. *Energy & Environmental Science* **2012**, *6*, 157.

[36] Burkhardt, S. E.; Lowe, M. A.; Conte, S.; Zhou, W.; Qian, H.; Rodríguez-Calero, G. G.; Gao, J.; Hennig, R. G.; Abruña, H. D. *Energy & Environmental Science* **2012**, *5*, 7176.

[37] Kazakov, A.; McLinden, M. O.; Frenkel, M. *Industrial & Engineering Chemistry Research* **2012**, 120917100332001.

[38] Martsinovich, N.; Troisi, A. *The Journal of Physical Chemistry C* **2011**, *115*, 11781–11792.

[39] Kim, J.; Lin, L.-C.; Martin, R. L.; Swisher, J. A.; Haranczyk, M.; Smit, B. *Langmuir* **2012**, *28*, 11914–11919.

[40] Kim, J.; Lin, L.-C.; Swisher, J. A.; Haranczyk, M.; Smit, B. *Journal of the American Chemical Society* **2012**, *134*, 18940–18943.

[41] Lin, L.-C. *Nature Materials* **2012**, *11*, 633–641.

[42] Martin, R. L.; Willems, T. F.; Lin, L.-C.; Kim, J.; Swisher, J. A.; Smit, B.; Haranczyk, M. *ChemPhysChem* **2012**, *13*, 3595–3597.

[43] Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. **2012**, 1–7.

[44] De Vleeschouwer, F.; Yang, W.; Beratan, D. N.; Geerlings, P.; De Proft, F. *Physical Chemistry Chemical Physics* **2012**, *14*, 16002.

[45] Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. *Physical Review Letters* **2012**, *108*, 058301.

[46] von Lilienfeld, O.; Lins, R.; Rothlisberger, U. *Physical Review Letters* **2005**, *95*, 153002.

[47] von Lilienfeld, O. A.; Tuckerman, M. E. *The Journal of Chemical Physics* **2006**, *125*, 154104.

[48] Keinan, S.; Hu, X.; Beratan, D. N.; Yang, W. *The Journal of Physical Chemistry A* **2007**, *111*, 176–181.

[49] Hu, X.; Beratan, D. N.; Yang, W. *The Journal of Chemical Physics* **2008**, *129*, 064102.

[50] Xiao, D.; Yang, W.; Beratan, D. N. *The Journal of Chemical Physics* **2008**, *129*, 044106.

[51] Wang, M.; Hu, X.; Beratan, D. N.; Yang, W. *Journal of the American Chemical Society* **2006**, *128*, 3228–3232.

[52] von Lilienfeld, O. A.; Tuckerman, M. E. *Journal of Chemical Theory and Computation* **2007**, *3*, 1083–1090.

[53] Olivares-Amaya, R.; Amador-Bedolla, C.; Hachmann, J.; Atahan-Evrenk, S.; Sánchez-Carrera, R. S.; Vogt, L.; Aspuru-Guzik, A. *Energy & Environmental Science* **2011**, *4*, 4849.

[54] Hachmann, J.; Olivares-Amaya, R.; Atahan-Evrenk, S.; Amador-Bedolla, C.; Sánchez-Carrera, R. S.; Gold-Parker, A.; Vogt, L.; Brockway, A. M.; Aspuru-Guzik, A. *The Journal of Physical Chemistry Letters* **2011**, *2*, 2241–2251.

[55] Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. *COMPUTATIONAL MATERIALS SCIENCE* **2013**, *68*, 314–319.

[56] Fink, T.; Bruggesser, H.; Reymond, J.-L. *Angewandte Chemie International Edition* **2005**, *44*, 1504–1508.

[57] Reymond, J.-L.; van Deursen, R.; Blum, L. C.; Ruddigkeit, L. *MedChemComm* **2010**, *1*, 30.

[58] Fink, T.; Reymond, J.-L. *Journal of Chemical Information and Modeling* **2007**, *47*, 342–353.

[59] Pillai, A. D.; Rani, S.; Rathod, P. D.; Xavier, F. P.; Vasu, K. K.; Padh, H.; Sudarsanam, V. *Bioorganic & Medicinal Chemistry* **2005**, *13*, 1275–1283.

[60] Liao, S. Y.; Chen, T. J.; Miao, T. F.; Qian, L.; Zheng, K. C. *Chemical Biology & Drug Design* **2009**, *74*, 289–296.

[61] De Vleeschouwer, F.; Chankisjijev, A.; Yang, W.; Geerlings, P.; De Proft, F. *The Journal of Organic Chemistry* **2013**, *78*, 3151–3158.

[62] Suh, C.; Rajan, K. *Materials Science and Technology* **2009**, *25*, 466–471.

[63] O'Boyle, N. M.; Campbell, C. M.; Hutchison, G. R. *The Journal of Physical Chemistry C* **2011**, *115*, 16200–16210.

[64] HANSCH, C.; MALONEY, P. P.; FUJITA, T.; MUIR, R. M. *Nature* **1962**, *194*, 178–180.

[65] Cherkasov, A. et al. *Journal of Medicinal Chemistry* **2014**, *57*, 4977–5010.

[66] Sumpter, B. G.; Meunier, V. *Journal of Polymer Science Part B: Polymer Physics* **2012**, *50*, 1071–1089.

[67] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *Journal of the American Chemical Society* **1985**, *107*, 3902–3909.

[68] Stewart, J. J. P. *Journal of Molecular Modeling* **2007**, *13*, 1173–1213.

[69] Ridley, J.; Zerner, M. *Theoretica Chimica Acta* **1973**, *32*, 111–134.

[70] Hutchison, G. R.; Ratner, M. A.; Marks, T. J. *The Journal of Physical Chemistry A* **2002**, *106*, 10596–10605.

[71] Scharber, M. C.; Mühlbacher, D.; Koppe, M.; Denk, P.; Waldauf, C.; Heeger, A. J.; Brabec, C. J. *Advanced Materials* **2006**, *18*, 789–794.

[72] Weininger, D. *Journal of chemical information and computer sciences* **1988**, *28*, 31–36.

[73] Riniker, S.; Landrum, G. A. *Journal of Chemical Information and Modeling* **2015**, *55*, 2562–2574.

[74] O'Boyle, N. M.; Vandermeersch, T.; Flynn, C. J.; Maguire, A. R.; Hutchison, G. R. *Journal of Cheminformatics* **2011**, *3*, 1–9.

[75] Leite, T. B.; Gomes, D.; Miteva, M. A.; Chomilier, J.; Villoutreix, B. O.; Tuffery, P. *Nucleic Acids Research* **2007**, *35*, W568–W572.

[76] Landrum, G.

[77] Inc, C. C. G.

[78] Ebejer, J.-P.; Morris, G. M.; Deane, C. M. *Journal of Chemical Information and Modeling* **2012**, *52*, 1146–1158.

[79] Schwab, C. H. *Drug Discovery Today: Technologies* **2010**, *7*, e245–e253.

[80] Colman, R. F. *Advances in enzyme regulation* **2001**, *13*, 413–433.

[81] Perdew, J. P.; Burke, K.; Ernzerhof, M. *Physical Review Letters* **1996**,

[82] Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 490–519.

[83] Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 520–552.

[84] Halgren, T. A. *Journal of Computational Chemistry* **1996**, *17*, 553–586.

[85] Halgren, T. A.; Nachbar, R. B. *Journal of Computational Chemistry* **1996**, *17*, 587–615.

[86] Halgren, T. A. *Journal of Computational Chemistry* **2004**, *17*, 616–641.

[87] Cramer, C. J. *Essentials of Computational Chemistry*, 2nd ed.; John Wiley & Sons: England, 2004.

[88] Burke, K.; friends, **2007**, 1–104.

[89] Bredow, T.; Jug, K. *Theoretical Chemistry Accounts* **2004**, *113*, 1–14.

[90] Jensen, F. *Introduction to Computational Chemistry*, 2nd ed.; John Wiley & Sons: West Sussex, England, 2007.

[91] Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. *Journal of Cheminformatics* **2012**, *4*, 17.

[92] Kanal, I. Y.; Bechtel, J. S.; Hutchison, G. R. *Sequence Matters: Determining the Sequence Effect of Electronic Structure Properties in π-Conjugated Polymers*; American Chemical Society: Washington, DC, 2014; pp 379–393.

[93] Wang, C.; Dong, H.; Hu, W.; Liu, Y.; Zhu, D. *Chemical Reviews* **2012**, *112*, 2208–2267.

[94] Skabara, P. J. *Chemical Communications* **2013**, *49*, 9242–9244.

[95] Szarko, J. M.; Guo, J.; Rolczynski, B. S.; Chen, L. X. *Journal of Materials Chemistry* **2011**, *21*, 7849.

[96] Zhao, X.; Zhan, X. *Chemical Society Reviews* **2011**, *40*, 3728.

[97] Risko, C.; McGehee, M. D.; Brédas, J.-L. *Chemical Science* **2011**, *2*, 1200–1218.

[98] Gong, X.; Tong, M.; Brunetti, F. G.; Seo, J.; Sun, Y.; Moses, D.; Wudl, F.; Heeger, A. J. *Advanced Materials* **2011**, *23*, 2272–2277.

[99] Mei, J.; Bao, Z. *Chemistry of Materials* **2014**, *26*, 604–615.

[100] Lutz, J. F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. *Science* **2013**, *341*, 1238149–1238149.

[101] Blankenship, R. E. et al. *Science* **2011**, *332*, 805–809.

[102] Norris, B. N.; Zhang, S.; Campbell, C. M.; Auletta, J. T.; Calvo-Marzal, P.; Hutchison, G. R.; Meyer, T. Y. *Macromolecules* **2013**, *46*, 1384–1392.

[103] Lutz, J.-F. *Polymer Chemistry* **2010**, *1*, 55–62.

[104] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *Journal of Cheminformatics* **2011**, *3*, 33.

[105] O'Boyle, N. M.; Morley, C.; Hutchison, G. R. *Chemistry Central Journal* **2008**, *2*, 5.

[106] O'Boyle, N. M. et al. *Journal of Cheminformatics* **2011**, *3*, 37.

[107] Team, R. C.

[108] RStudio, **2012**,

[109] Savoie, B. M.; Jackson, N. E.; Marks, T. J.; Ratner, M. A. *Physical Chemistry Chemical Physics* **2013**, *15*, 4538.

[110] Becke, A. D. *The Journal of Chemical Physics* **1993**, *98*, 5648.

[111] Lee, C.; Yang, W.; Parr, R. G. *Physical review. B, Condensed matter* **1988**, *37*, 785–789.

[112] Nelson, J. *Materials Today* **2011**, *14*, 462–470.

[113] Kuhn, H. *The Journal of Chemical Physics* **1949**, *17*, 1198.

[114] Torras, J.; Casanovas, J.; Alemán, C. *The Journal of Physical Chemistry A* **2012**, *116*, 7571–7583.

[115] Zhang, S.; Bauer, N. E.; Kanal, I. Y.; You, W.; Hutchison, G. R.; Meyer, T. Y. *Macromolecules* **2017**, *50*, 151–161.

[116] Grimsdale, A. C.; Chan, K. L.; Martin, R. E.; Jokisz, P. G.; Holmes, A. B. *Chemical Reviews* **2009**, *109*, 897–1091.

[117] Facchetti, A. *Chemistry of Materials* **2010**, *23*, 733–758.

[118] Henson, Z. B.; Müllen, K.; Bazan, G. C. *Nature chemistry* **2012**, *4*, 699–704.

[119] Yassar, A.; Miozzo, L.; Gironda, R.; Horowitz, G. *Progress in Polymer Science* **2013**,

[120] Duan, L.; Hou, L.; Lee, T. W.; Qiao, J.; Zhang, D. *Journal of Materials ...* **2010**,

[121] Ellinger, S.; Graham, K. R.; Shi, P.; Farley, R. T. *Chemistry of ...* **2011**,

[122] Wang, C.; Dong, H.; Hu, W.; Liu, Y.; Zhu, D. *Chemical Reviews* **2011**,

[123] Mishra, A.; Bäuerle, P. *Angewandte Chemie International Edition* **2012**, *51*, 2020–2067.

[124] Chen, Y.; Wan, X.; Long, G. *Accounts of Chemical Research* **2013**, *46*, 2645–2655.

[125] Rochat, S.; Swager, T. M. *ACS Applied Materials & Interfaces* **2013**, *5*, 4488–4502.

[126] Zhang, L.; Colella, N. S.; Cherniawski, B. P.; Mannsfeld, S. C. B.; Briseño, A. L. *ACS Applied Materials & Interfaces* **2014**, *6*, 5327–5343.

[127] Lu, L.; Zheng, T.; Wu, Q.; Schneider, A. M.; Zhao, D.; Yu, L. *Chemical Reviews* **2015**, *115*, 12666–12731.

[128] Beaujuge, P. M.; Amb, C. M. *Accounts of chemical ...* **2010**,

[129] Park, Y. S.; Kale, T. S.; Nam, C.-Y.; Choi, D.; Grubbs, R. B. *Chemical communications (Cambridge, England)* **2014**, *50*, 7964–7967.

[130] Lin, L.-Y.; Chen, Y.-H.; Huang, Z.-Y.; Lin, H.-W.; Chou, S.-H.; Lin, F.; Chen, C.-W.; Liu, Y.-H.; Wong, K.-T. *Journal of the American Chemical Society* **2011**, *133*, 15822–15825.

[131] Ward, R. E.; Meyer, T. Y. *Macromolecules* **2003**,

[132] Copenhafer, J. E.; Walters, R. W.; Meyer, T. Y. *Macromolecules* **2008**,

[133] Stayshich, R. M.; Meyer, T. Y. *Journal of the American Chemical Society* **2010**, *132*, 10920–10934.

[134] Li, J.; Stayshich, R. M.; Meyer, T. Y. *Journal of the American Chemical Society* **2011**, *133*, 6910–6913.

[135] Li, J.; Rothstein, S. N.; Little, S. R.; Edenborn, H. M.; Meyer, T. Y. *Journal of the American Chemical Society* **2012**, *134*, 16352–16359.

[136] Rosales, A. M.; Segalman, R. A.; Zuckermann, R. N. *Soft Matter* **2013**,

[137] De Bo, G.; Kuschel, S.; Leigh, D. A.; Lewandowski, B.; Papmeyer, M.; Ward, J. W. *Journal of the American Chemical Society* **2014**, *136*, 5811–5814.

[138] Edwardson, T. G. W.; Carneiro, K. M. M.; Serpell, C. J.; Sleiman, H. F. *Angewandte Chemie International Edition* **2014**, *53*, 4567–4571.

[139] Schulz, M. D.; Wagener, K. B. *Macromolecular Chemistry and ...* **2014**,

[140] Rowan, S. J.; Barner-Kowollik, C.; Klumperman, B. *ACS Macro ...* **2015**,

[141] Fitzner, R.; Mena-Osteritz, E.; Mishra, A.; Schulz, G.; Reinold, E.; Weil, M.; Körner, C.; Ziehlke, H.; Elschner, C.; Leo, K.; Riede, M.; Pfeiffer, M.; Uhrich, C.; Bäuerle, P. *Journal of the American Chemical Society* **2012**, *134*, 11064–11067.

[142] Palermo, E. F.; McNeil, A. J. *Macromolecules* **2012**,

[143] Doval, D. A.; Molin, M. D.; Ward, S.; Fin, A.; Sakai, N. *Chemical ...* **2014**,

[144] Liang, L.; Wang, J. T.; Xiang, X.; Ling, J.; Zhao, F. G. *Journal of Materials ...* **2014**,

[145] Tsai, C.-H.; Fortney, A.; Qiu, Y.; Gil, R. R.; Yaron, D.; Kowalewski, T.; Noonan, K. J. T. *Journal of the American Chemical Society* **2016**, *138*, 6798–6804.

[146] Yamamoto, T.; Fang, Q.; Morikita, T. *Macromolecules* **2003**,

[147] Lu, S.; Yang, M.; Luo, J.; Cao, Y. *Synthetic Metals* **2004**,

[148] Li, X.; Zhang, Y.; Yang, R.; Huang, J. *Journal of Polymer ...* **2005**,

[149] Liu, L.; Pei, J.; Wen, S.; Li, J.; Yao, B. ... *Chemistry and Physics* **2013**,

[150] Kim, J.; Han, A. R.; Hong, J.; Kim, G.; Lee, J.; Shin, T. J. *Chemistry of ...* **2014**,

[151] Nguyen, T. L.; Choi, H.; Ko, S. J.; Uddin, M. A. *Energy & ...* **2014**,

[152] Zhang, S.; Hutchison, G. R. *Macromolecular rapid ...* **2016**,

[153] Jørgensen, M.; Krebs, F. C. *The Journal of Organic Chemistry* **2005**, *70*, 6004–6017.

[154] Frisch, M. J. et al.

[155] Tomasi, J.; Mennucci, B.; Cammi, R. *Chemical Reviews* **2005**, *105*, 2999–3094.

[156] Cossi, M.; Barone, V. *The Journal of Chemical Physics* **2001**, *115*, 4708.

[157] Zhou, C.; Liang, Y.; Liu, F.; Sun, C.; Huang, X. *Advanced Functional ...* **2014**,

[158] Li, W.; Albrecht, S.; Yang, L.; Roland, S.; Tumbleston, J. R.; McAfee, T.; Yan, L.; Kelly, M. A.; Ade, H.; Neher, D.; You, W. *Journal of the American Chemical Society* **2014**, *136*, 15566–15576.

[159] Bartelt, J. A.; Lam, D.; Burke, T. M. *Advanced Energy ...* **2015**,

[160] Proctor, C. M.; Love, J. A.; Nguyen, T. Q. *Advanced Materials* **2014**,

[161] Bartesaghi, D.; Pérez, I. D. C.; Kniepert, J.; Roland, S.; Turbiez, M.; Neher, D.; Koster, L. J. A. *Nature Communications* **2015**, *6*, 7083.

[162] Zhan, C.-G.; Nichols, J. A.; Dixon, D. A. *The Journal of Physical Chemistry A* **2003**, *107*, 4184–4195.

[163] Rienstra-Kiracofe, J. C.; Tschumper, G. S.; Schaefer, H. F.; Nandi, S.; Ellison, G. B. *Chemical Reviews* **2002**, *102*, 231–282.

[164] Perdew, J. P.; Levy, M. *Physical Review B* **1997**, *56*, 16021.

[165] Levy, M. *Physical Review A* **1995**, *52*, R4313.

[166] Winget, P.; Weber, E. J.; Cramer, C. J.; Truhlar, D. G. *Physical Chemistry Chemical Physics* **2000**, *2*, 1231–1239.

[167] Jaque, P.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *The Journal of Physical Chemistry C* **2007**, *111*, 5783–5799.

[168] Chu, P.-H.; Zhang, L.; Colella, N. S.; Fu, B.; Park, J. O.; Srinivasarao, M.; Briseño, A. L.; Reichmanis, E. *ACS Applied Materials & Interfaces* **2015**, *7*, 6652–6660.

[169] Reineke, S.; Lindner, F.; Schwartz, G.; Seidler, N.; Walzer, K.; Lüssem, B.; Leo, K. *Nature* **2009**, *459*, 234–238.

[170] Gwon, H.; Hong, J.; Kim, H.; Seo, D.-H.; Jeon, S.; Kang, K. *Energy & Environmental Science* **2014**, *7*, 538.

[171] Kuribara, K. et al. *Nature Communications* **2012**, *3*, 723–7.

[172] Leonat, L.; White, M. S.; Głowacki, E. D.; Scharber, M. C.; Zillger, T.; Rühling, J.; Hübler, A.; Sariciftci, N. S. *The Journal of Physical Chemistry C* **2014**, *118*, 16813–16817.

[173] Hutchison, G.; Zhao, Y.-J.; Delley, B.; Freeman, A.; Ratner, M.; Marks, T. *Physical Review B* **2003**, *68*, 035204.

[174] Hutchison, G. R.; Ratner, M. A.; Marks, T. J. *The Journal of Physical Chemistry B* **2005**, *109*, 3126–3138.

[175] Zamoshchik, N.; Sheynin, Y.; Bendikov, M. *Israel Journal of Chemistry* **2014**, *54*, 723–735.

[176] Rienstra-Kiracofe, J. C.; Barden, C. J.; Brown, S. T.; Schaefer, H. F. *The Journal of Physical Chemistry A* **2001**, *105*, 524–528.

[177] de Oliveira, G.; Martin, J.; De Proft, F.; Geerlings, P. *Physical Review A* **1999**, *60*, 1034–1045.

[178] De Proft, F.; Geerlings, P. *The Journal of Chemical Physics* **1997**, *106*, 3270.

[179] Muscat, J.; Wander, A.; Harrison, N. M. *Chemical physics letters* **2001**, *342*, 397–401.

[180] Montavon, G.; Rupp, M.; Gobre, V. *New Journal of ...* **2013**,

[181] Efron, B.; Tibshirani, R. J. An introduction to the bootstrap. 1994.

[182] Geisser, S. Predictive inference. 1993.

[183] Gagorik, A. G.; Mohin, J. W.; Kowalewski, T.; Hutchison, G. R. *Advanced Functional Materials* **2015**, *25*, 1996–2003.

[184] Kettle, J.; Waters, H.; Horie, M.; Chang, S.-W. *Journal of Physics D: Applied Physics* **2012**, *45*, 125102.

[185] Rupakheti, C.; Al-Saadon, R.; Zhang, Y.; Virshup, A. M.; Zhang, P.; Yang, W.; Beratan, D. N. *Journal of Chemical Theory and Computation* **2016**, *12*, 1942–1952.

[186] Virshup, A. M.; Contreras-García, J.; Wipf, P.; Yang, W.; Beratan, D. N. *Journal of the American Chemical Society* **2013**, 130502155645006.

[187] Rupakheti, C.; Virshup, A.; Yang, W.; Beratan, D. N. *Journal of Chemical Information and Modeling* **2015**, *55*, 529–537.

[188] Gómez-Bombarelli, R.; Duvenaud, D. *arXiv.org* **2016**,

[189] Rinderspacher, B. C.; Andzelm, J.; Rawlett, A.; Dougherty, J.; Beratan, D. N.; Yang, W. *Journal of Chemical Theory and Computation* **2009**, *5*, 3321–3329.

[190] O'Boyle, N. M.; Tenderholt, A. L.; Langner, K. M. *Journal of Computational Chemistry* **2008**, *29*, 839–845.

[191] Van Der Walt, S.; Colbert, S. C. *Computing in Science & ...* **2011**,

[192] McKinney, W. *Python for High Performance and Scientific Computing* **2011**,

[193] Rappé, A. K.; Casewit, C. J. *Molecular Mechanics Across Chemistry*; University Science Books, 1997.

[194] Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. *Pharmacological Reviews* **2013**, *66*, 334–395.

[195] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nature Reviews Drug Discovery* **2004**, *3*, 935–949.

[196] Zimmer, M. *Chemical Reviews* **1995**,

[197] Mohamadi, F.; Richards, N.; Guida, W. C. *Journal of ...* **1990**,

[198] Kaminský, J.; Jensen, F. *Journal of Chemical Theory and Computation* **2016**, *12*, 694–705.

[199] Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chemical Reviews* **2012**, *112*, 1839–1862.

[200] Kaminský, J.; Jensen, F. *Journal of Chemical Theory and Computation* **2007**, *3*, 1774–1788.

[201] Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. *Journal of Chemical Information and Modeling* **2010**, *50*, 572–584.

[202] Hartshorn, M. J.; Verdonk, M. L.; Chessari, G.; Brewerton, S. C.; Mooij, W. T. M.; Mortenson, P. N.; Murray, C. W. *Journal of Medicinal Chemistry* **2007**, *50*, 726–741.

[203] Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

[204] Stewart, J. J. P. *Journal of Molecular Modeling* **2012**, *19*, 1–32.

[205] Stewart, J. J. P.

[206] Neese, F. *Wiley Interdisciplinary Reviews: Computational Molecular Science 2*, 73–78.

[207] Becke, A. *Physical review. A, General physics* **1988**, *38*, 3098–3100.

[208] Schäfer, A.; Horn, H.; Ahlrichs, R. *The Journal of Chemical Physics* **1992**,

[209] Schäfer, A.; Huber, C.; Ahlrichs, R. *The Journal of Chemical Physics* **1994**, *100*, 5829.

[210] Kossmann, S.; Neese, F. *Journal of Chemical Theory and Computation* **2010**, *6*, 2325–2338.

[211] Grimme, S.; Ehrlich, S.; Goerigk, L. *Journal of Computational Chemistry* **2011**, *32*, 1456–1465.

[212] Forti, F.; Cavasotto, C. N.; Orozco, M.; Barril, X.; Luque, F. J. *Journal of Chemical Theory and Computation* **2012**, *8*, 1808–1819.

[213] Kristam, R.; Gillet, V. J.; Lewis, R. A.; Thorner, D. *Journal of Chemical Information and Modeling* **2005**, *45*, 461–476.

[214] Miteva, M. A.; Guyon, F.; Tuffery, P. *Nucleic Acids Research* **2010**, *38*, W622–W627.

[215] Vainio, M. J.; Johnson, M. S. *Journal of Chemical Information and Modeling* **2007**, *47*, 2462–2474.

[216] Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. *Journal of Chemical Theory and Computation* **2013**, *9*, 4046–4063.

[217] McDaniel, J. G.; Schmidt, J. R. *Annual Review of Physical Chemistry* **2016**, *67*, 467–488.