Accelerated Discovery of High-Refractive-Index Polyimides via First-Principles Molecular Modeling, Virtual High-Throughput Screening, and Data Mining

Mohammad Atif Faiz Afzal, Mojtaba Haghighatlari, Sai Prasad Ganesh, Chong Cheng, Johannes Hachmann

Submitted date: 05/02/2019 • Posted date: 06/02/2019
License: CC BY-NC-ND 4.0

Citation information: Afzal, Mohammad Atif Faiz; Haghighatlari, Mojtaba; Ganesh, Sai Prasad; Cheng, Chong; Hachmann, Johannes (2019): Accelerated Discovery of High-Refractive-Index Polyimides via First-Principles Molecular Modeling, Virtual High-Throughput Screening, and Data Mining. ChemRxiv. Preprint.

We present a high-throughput computational study to identify novel polyimides (PIs) with exceptional refractive index (RI) values for use as optic or optoelectronic materials. Our study utilizes an RI prediction protocol based on a combination of first-principles and data modeling developed in previous work, which we employ on a large-scale PI candidate library generated with the ChemLG code. We deploy the virtual screening software ChemHTPS to automate the assessment of this extensive pool of PI structures in order to determine the performance potential of each candidate. This rapid and efficient approach yields a number of highly promising leads compounds. Using the data mining and machine learning program package ChemML, we analyze the top candidates with respect to prevalent structural features and feature combinations that distinguish them from less promising ones. In particular, we explore the utility of various strategies that introduce highly polarizable moieties into the PI backbone to increase its RI yield. The derived insights provide a foundation for rational and targeted design that goes beyond traditional trial-and-error searches.

File list (2)

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accelerated Discovery of High-Refractive-Index Pol...</td>
<td>1.02 MiB</td>
</tr>
<tr>
<td>R1_R2_PI_screening_results.xlsx</td>
<td>10.96 MiB</td>
</tr>
</tbody>
</table>
Accelerated Discovery of High-Refractive-Index Polyimides via First-Principles Molecular Modeling, Virtual High-Throughput Screening, and Data Mining

Mohammad Atif Faiz Afzal,∗ Mojtaba Haghighatlari, Sai Prasad Ganesh, Chong Cheng, and Johannes Hachmann

1 Department of Chemical and Biological Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14260, United States
2 Computational and Data-Enabled Science and Engineering Graduate Program, University at Buffalo, The State University of New York, Buffalo, NY 14260, United States
3 New York State Center of Excellence in Materials Informatics, Buffalo, NY 14203, United States

ABSTRACT

We present a high-throughput computational study to identify novel polyimides (PIs) with exceptional refractive index (RI) values for use as optic or optoelectronic materials. Our study utilizes an RI prediction protocol based on a combination of first-principles and data modeling developed in previous work, which we employ on a large-scale PI candidate library generated with the ChemLG code. We deploy the virtual screening software ChemHTPS to automate the assessment of this extensive pool of PI structures in order to determine the performance potential of each candidate. This rapid and efficient approach yields a number of highly promising leads compounds. Using the data mining and machine learning program package ChemML, we analyze the top candidates with respect to prevalent structural features and feature combinations that distinguish them from less promising ones. In particular, we explore the utility of various strategies that introduce highly polarizable moieties into the PI backbone to increase its RI yield. The derived insights provide a foundation for rational and targeted design that goes beyond traditional trial-and-error searches.

I. INTRODUCTION

Polyimides (PIs), shown in Fig. 1, are synthesized by polycondensation of $R_1$-containing dianhydride with $R_2$-based diamine or diisocyanate [1]. They are an appealing class of organic materials due to their exceptional thermal stability and easy processability [2–4]. These properties are complemented by mechanical stability, flexibility, light weight, low cost, as well as flame and radiation resistance, and they thus hold much promise for a range of applications [5, 6].

![Polyimide (PI) core structure with residues $R_1$ and $R_2$.](image)

FIG. 1. Polyimide (PI) core structure with residues $R_1$ and $R_2$.

However, their generally low index of refraction (RI) undermines their utility for use in many optic and optoelectronic devices [7–11], such as (image) sensors [12, 13], displays [14], and light sources (including organic light-emitting diodes) [15], in which organic materials can be deployed in situ as micro lenses [16], waveguides [17], microresonators [18], interferometers [19], anti-reflective coatings [20], optical adhesives [21], and substrates [22]. Most of these applications demand large RI values, often upwards of 1.7 or 1.8. Typical carbon-based polymers – PIs included – only exhibit values in the range of 1.3–1.5 [11]. This situation provides a strong incentive to pursue novel high-RI PIs that are suitable for the aforementioned applications.

As the properties of organic materials can be tailored and tuned by controlling their molecular structure, they are a prime target for combinatorial search or rational design attempts [23–25]. The addition of highly polarizable moieties that do not have extensive $\pi$-electron conjugation can increase the RI of polymers without negatively affecting other optical properties. (Significant $\pi$-conjugation can lead to large optical dispersion and birefringence [26], as well as poor transparency and coloration.) For example, the incorporation of small aromatic rings, halogens, metals, and in particular sulfur have shown promise for this purpose [3, 27–31]. In 2007, Ueda et al. developed PI films with relatively high RI values of up to 1.75, but they, unfortunately, exhibited large birefringence [32]. In recent years, these findings have been improved upon by increasing the sulfur content of the PIs, yielding RI values of up to 1.76 and smaller birefringence [33].

In this paper, we present a computational and data-driven approach to study the RI values of PIs and rapidly identify promising lead compounds. We investigate different ways that introduce highly polarizable moieties into the PI framework in order to overcome the technical limits of existing compounds. Given the encouraging results from the earlier work discussed above, we put a focus on incorporating sulfur into the PIs, and in doing so,
create a new class of high-RI polymers. In Sec. II, we detail the methods employed in this work, i.e., we provide background on our data-driven in silico approach (Sec. II A), introduce our RI modeling protocol (Sec. II B), discuss the molecular design space we consider (Sec. II C), and outline our data mining and analysis techniques (Sec. II D). Sec. III presents and discusses the results of our study, and our findings are summarized in Sec. IV.

II. BACKGROUND, METHODS, AND COMPUTATIONAL DETAILS

A. Data-Driven In Silico Approach

The development of new materials such as PIs has traditionally been an experimentally-driven, trial-and-error process, guided by experience, intuition, and conceptual insights. This approach is, however, often costly, slow, biased towards certain domains of chemical space, and limited to relatively small-scale studies, which may easily miss promising leads (both on individual compounds as well as compound classes).

The study at hand instead embraces a data-driven in silico research paradigm that has gained considerable interest in the past few years [34, 35] for its promise to address the inherent complexities of structure-property relationships and the vastness of chemical space more efficiently (see, e.g., Refs. [36–43]). Our work brings together molecular modeling, high-throughput computational screening, and machine learning as well as a corresponding software ecosystem to support data-driven discovery and rational design [44, 45]. It has its greatest utility as part of integrated research pipelines with experimentalist partners, where it provides guidance for experimental efforts and mitigates some of their aforementioned shortcomings.

Our research approach and its rationale can be summarized as following: We first establish a computational modeling protocol that can make sufficiently accurate and fast predictions for the target property in the compound class(es) of interest – in the study at hand for the RI values of PIs (cf. Sec. II B). We then create a large-scale virtual library of candidates within that compound class (cf. Sec. II C), on which we apply this modeling protocol to evaluate the performance potential of each candidate compound. In addition to obtaining information about each individual compound, we can mine the screening results in their entirety to reveal underlying structure property-relationships (cf. Sec. II D). Our approach can thus identify both specific lead compounds as well as high-value molecular patterns and features for the de novo design of tailored PIs or the creation of additional screening libraries.

B. Refractive Index Modeling Protocol

According to the Lorentz-Lorenz equation, the RI \( n_r \) is a function of the polarizability \( \alpha \) and the number density \( N \), i.e.,

\[
n_r = \sqrt{\frac{1 + 2\alpha N/3\epsilon_0}{1 - \alpha N/3\epsilon_0}}.
\]

In previous work [46, 47], we developed a modeling protocol that allows us to accurately and efficiently predict the RI values of polymers within the Lorentz-Lorenz equation. This protocol is based on a combination of first-principles quantum chemistry calculations and data modeling. We compute (static) polarizability values \( \alpha \) using the coupled-perturbed self-consistent field equations within the Kohn-Sham density functional theory (DFT) framework with the PBE0 hybrid functional [48] and the double-\( \zeta \)-quality def2-SVP basis set by the Karlsruhe group [49]. We employ an all-electron, closed-shell approach and include Grimme’s D3 correction [50] to account for dispersion interactions. The polarizability calculations are performed on geometries optimized using the universal force field (UFF) [51] following a 3D conformational screening as implemented in the OpenBabel software [52]. All DFT calculations are carried out using the ORCA 3.0.2 quantum chemistry package [53]. We compute the number density \( N \) via the van der Waals volume \( V_{vdW} \) and packing factor \( K_p \) of the amorphous bulk polymer as

\[
N = \frac{K_p}{V_{vdW}}.
\]

We obtain \( V_{vdW} \) using Slonimskii’s method detailed in Ref. [54] and \( K_p \) using the support vector regression (SVR) machine learning model introduced in Ref. [46] (except during the prescreening of individual residues \( R_1 \) and \( R_2 \) discussed in the following section, for which we simply employ a constant \( K_p \) of 0.75 that is typically found in PIs [55]). We obtain the asymptotic limit for the polymers via a linear extrapolation scheme as outlined in Ref. [46]. (In exploratory calculations, we already observe near-perfect extensivity based on the monomer units, which we explain with the large size of the PI monomers (~150 atoms) and the finite correlation length exhibited in them. There is thus no need to perform expensive oligomer-sequence calculations.) We previously demonstrated on a set of 112 non-conjugated polymers with experimentally known RI values that this protocol can make rapid and accurate predictions [46, 47], and thus enable the high-throughput screening study at hand. We execute the latter using our automated virtual screening code ChemHTPS [56], which creates inputs, executes and monitors the calculations, parses and assesses the results, and which extracts and post-processes the information of interest. Where applicable, we report the mean absolute error (MAE) and the root mean squared error (RMSE) as well as their percentage errors (MAPE and...
RMSPE, respectively) compared to available experimental data.

C. Molecular Design Space

In Fig. 2, we show the RI heat map in the α/N parameter space and mark the positions of the 112 polymers we previously studied [46, 47]. The figure emphasizes the relative importance of the counteracting parameters and their feasibility in real-world compounds. In order to design high-RI systems, we in principle have to pursue compounds that simultaneously feature both large polarizability and number density values. The positioning of the known compounds with high-RI leans more towards high polarizability and relatively small number density rather than vice versa. One path to narrowing down compound space is to maximize the polarizability for systems from the same compound family, for which the number density is similar. Given that this is known to be the case for PIs, we pursue this path and focus our search on highly polarizable PI candidates [55].

![RI heat map](image)

FIG. 2. Refractive Index (RI) heat map showing the dependence on polarizability and number density expressed in the Lorentz-Lorenz equation. The dots mark 112 experimentally known polymers that were used to benchmark the employed RI model.

For this, we create a PI library based on 29 building blocks and bonding rules shown in Fig. 3, which we select following empirical expertise with regard to promise and synthetic feasibility. The building blocks can be divided into linkers (B1-B6) and (hetero-)aromatic moieties (B7-B29). Using our ChemLG molecular library generator code [57] with a systematic combinatorial linking scheme, we initially generate 38,619 $R_1$ structures and 171,172 for $R_2$. Combining all $R_1$ and $R_2$ to form PIs would lead to a total of 6.6 billion compounds. To restrict the search space to a more manageable number of candidates, we select only the most promising $R_1$ and $R_2$ groups (which ultimately distinguish the PI candidates) based on the computed RI values of the individual $R_1$ and $R_2$. After prescreening the residues, we develop PI candidates by introducing the top 100 $R_1$ and 100 $R_2$ structures into the PI core motif as shown in Fig. 1. We subsequently screen the 10,000 resulting PI candidates. ChemLG keeps a record of the list of building blocks and connections that are used to create a particular compound, and we use this information for the analysis of structure-property relationships.

![Building blocks](image)

FIG. 3. Molecular building blocks used to create the residues $R_1$ and $R_2$ of the PI screening library studied in this work. The “R” on the building blocks denote allowed sites for linking. The building blocks marked in green (B1-B6) are linkers and the blue ones (B7-B29) are (hetero-)aromatic moieties.

D. Data Mining and Pattern Analysis

In addition to identifying the best candidates from our high-throughput screening, i.e., those with the highest predicted RI values, we further analyze the collected data to develop a better understanding of the structural patterns that lead to high-RI PIs. For instance, we conduct a hypergeometric distribution analysis, in which we compute the $Z$-scores ($Z_i$) of each building block $i$ used in the creation of the screening library as

$$Z_i = \frac{k_i - mK_i}{\sigma_i},$$

with

$$\sigma_i = \left[ \frac{mK_i}{M} \times \left( \frac{M - K_i}{M} \right) \times \left( \frac{M - m}{M - 1} \right) \right]^{\frac{1}{2}},$$

where $M$ is the total number of compounds in the entire library, $m$ is the subset of compounds that is considered (e.g., the top compounds), $K_i$ is the number of
TABLE I. Comparison of RI results from the employed prediction model with the experimental values of ten known high-RI PIs. We report errors and percentage errors in each case. For the entire set, we obtain a mean absolute error (MAE) and root mean squared error (RMSE) of 0.021 (1.2%) and 0.025 (1.4%), respectively. (The values in brackets are the corresponding percentage errors, i.e., MAPE and RMSPE.)

<table>
<thead>
<tr>
<th>Ar&lt;sub&gt;a−j&lt;/sub&gt; Structure</th>
<th>Experiment value [11]</th>
<th>Calculated value</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>1.746</td>
<td>1.738</td>
<td>-0.008 (-0.5%)</td>
</tr>
<tr>
<td>b</td>
<td>1.753</td>
<td>1.739</td>
<td>-0.013 (-0.7%)</td>
</tr>
<tr>
<td>c</td>
<td>1.749</td>
<td>1.707</td>
<td>-0.042 (-2.4%)</td>
</tr>
<tr>
<td>d</td>
<td>1.748</td>
<td>1.751</td>
<td>0.003 (0.2%)</td>
</tr>
<tr>
<td>e</td>
<td>1.733</td>
<td>1.760</td>
<td>0.027 (1.6%)</td>
</tr>
<tr>
<td>f</td>
<td>1.758</td>
<td>1.779</td>
<td>0.021 (1.2%)</td>
</tr>
<tr>
<td>g</td>
<td>1.760</td>
<td>1.735</td>
<td>-0.025 (-1.4%)</td>
</tr>
<tr>
<td>h</td>
<td>1.726</td>
<td>1.741</td>
<td>0.015 (0.9%)</td>
</tr>
<tr>
<td>i</td>
<td>1.737</td>
<td>1.743</td>
<td>0.005 (0.3%)</td>
</tr>
<tr>
<td>j</td>
<td>1.769</td>
<td>1.724</td>
<td>-0.045 (-2.5%)</td>
</tr>
</tbody>
</table>

occurrences of building block i in M molecules and k, its occurrences in the subset of m molecules. A large Z-score thus indicates a statistical overexpression of the associated building block in the high-RI candidates relative to the overall screening library. By applying the Z-score analysis, we can thus identify the most important building blocks and the degree to which they correlate with large RI values. We perform a similar analysis of building block combinations to reveal synergistic effects similar to those known from push-pull or donor-acceptor copolymers. In addition, we present an analysis of Z-score trends for each building block in ranked candidate subsets as well as of the average RI values of the candidates derived from each building block. All data mining work was conducted using our ChemML code [58].

III. RESULTS AND DISCUSSION

We previously tested the RI modeling protocol described in Sec. II B for the before-mentioned set of 112 non-conjugated polymers and established its predictive performance. However, this benchmark set did not contain any PIs. To prove our protocol’s validity for this particular compound class, we perform additional calculations on ten PIs with experimentally known RI values. We compare the experimental and computational results in Tab. I. The MAE (MAPE) and RMSE (RMSPE) values for this comparison are 0.021 (1.2%) and 0.025 (1.4%), respectively, which suggests that our protocol can accurately predict the RI values of PIs as well.

The results of the $R_1$ and $R_2$ residue prescreening are summarized in Fig. 4 (a) and (b), respectively, and the results of the subsequent PI screening are given in Fig. 4 (c). The plots show histograms of the computed RI results. Most of the candidates for $R_1$ and $R_2$ have RI values between 1.5 and 1.7 with an average of (a) 1.600, and (b) 1.627, respectively. However, the PI candidates derived from the top $R_1$ and $R_2$ residues have significantly larger RI values with an average of 1.843. We note that either of these averages is well above the range of typical organic polymers (i.e., 1.3–1.5), which indicates
a good choice of building blocks. As per the objectives of this work, we are able to identify a sizable number of compounds with RI values greater than 1.7, i.e., (a) 2851 (7.4%), (b) 22388 (13.1%), and (c) 9985 (99.8%). At or above the critical threshold of 1.8, we still find (a) 131 (0.3%), (b) 1252 (0.7%), and (c) 6961 (69.6%) compounds. The shift to considerably larger RI values also demonstrates the success of our strategy to prescreen the $R_1$ and $R_2$ individually and build PI candidates based on the top residues. (Details of the prescreening and screening results are provided in the Supplementary Material.)

Fig. 5 shows the RI distribution of (a) $R_1$, (b) $R_2$, and (c) PI candidates containing each constituent building block (cf. the PI candidate library construction illustrated in Fig. 3). We find that $R_1$ and $R_2$ candidates containing building blocks $B_{28}$ (anthracene), $B_{25}$ (dibenzothiophene), and $B_{24}$ (thianthrene) have the highest RI values, while those containing building blocks $B_{17}$ (cyclopentadiene) and $B_{16}$ (1,3,5-triazinan) show the lowest RI. The ranking of the building blocks is very similar for both $R_1$ and $R_2$ structures, which is unsurprising considering their general similarity. The PI candidates do not contain building blocks $B_{10}$ (1,4-dithiane) and $B_{14}$ (toluene with linking in 2,4 position), as these were missing in the top 100 structures of $R_1$ and $R_2$ that were used in their construction. The average RI values for all the other building blocks – with the exception of $B_{26}$ (tetracyclohexane) – are greater than 1.8. However, it should be noted that the number of PI candidates containing a specific building block is very variable. The red plot in Fig. 5 (c) shows the count of each building block in the 10,000 PI structures. The most common building blocks in the PI library are $B_1$ (CH$_2$-linker), $B_{25}$, $B_2$ (S-linker), $B_{28}$, and $B_3$ (O-linker), with $B_{28}$ and $B_2$ occurring in almost all PI candidates. Building blocks $B_1$, $B_2$, and $B_3$ do not exhibit particularly large average RI values, however, given the construction template for the residues introduced in Fig. 3, they are statistically more likely to occur. The average RI value alone is thus not a sufficient metric to gauge the potential impact of each building block on the performance of PI candidates.

In the following, we analyze the contribution of the building blocks in the high-RI candidates of $R_1$ and $R_2$, using the hypergeometric distribution analysis detailed in Sec. II.D. We focus on the top 10% of the $R_1$ and $R_2$ candidates (with RI values greater than 1.687 and 1.711, respectively). We do not include PI candidates in this analysis as the selective construction scheme with its biased building block counts makes it less meaningful. Fig. 6 shows the resulting Z-scores, which identify the over- or underrepresentation of each building block in the high-RI $R_1$ and $R_2$ candidates compared to a random sample. The results point to $B_{28}$ and $B_{25}$ as the most prevalent building blocks in the top residues, and thus the most promising ones to consider for the design of high-RI polymers. This finding is in good agreement with the prior analysis of the RI averages and distribution. The ranking of the building blocks in both $R_1$ and $R_2$ are again largely the same, suggesting that the effect of the building blocks is similar for the somewhat different sequence in $R_1$ and $R_2$.

The above Z-score analysis only yields insights into the pervasiveness of building blocks in the top 10% candidates. In order to gain a more comprehensive picture, we now evaluate the Z-score of each building block in each 10% segment of the $R_1$ and $R_2$ libraries. For this, we sort the $R_1$ and $R_2$ libraries by increasing RI value, divide them into 10 subsets, and perform a hypergeometric distribution analysis for each building block in each subset. The results are shown in Fig. 7. We observe certain general trends for the building blocks making up $R_1$ and/or $R_2$:

1. The Z-score of some building blocks increases with increasing RI values, indicating a direct correlation (e.g., for $B_{28}$, $B_{25}$, $B_{24}$, and $B_{23}$).

2. The Z-score of some building blocks decreases with increasing RI values, indicating a negative correlation (e.g., for $B_{17}$, $B_{22}$, $B_{10}$, and $B_{16}$).

3. The Z-score of some building blocks goes through a maximum for intermediate RI values, indicating a corresponding correlation with average candidates (e.g., for $B_{11}$, $B_{26}$, $B_{19}$, $B_{18}$, $B_{20}$, and $B_{7}$).

4. The Z-score of some building blocks does not show a clear trend, indicating a lesser impact on and correlation with the RI values (e.g., for $B_1$ and $B_6$).

In the bottom right corner of Fig. 7, we plot the average RI values of the $R_1$ and $R_2$ candidates in each of the 10 subsets. We observe that the $R_2$ structures have somewhat higher RI values in comparison to $R_1$ structures. The principal difference in $R_1$ and $R_2$ candidates is the number of aromatic building blocks, i.e., two vs three, and the higher content of these moieties correlates with higher RI values.

In addition to analysing the influence of individual building blocks on the RI values, we also study the potential impact of building block pairs. For this we calculate the joint Z-scores of all possible building block combinations in the top 10% candidates. The results for $R_1$ and $R_2$ are shown in Fig. 8. This analysis reveals some dependence on particular building block combinations. For instance, $B_{23}$, $B_{24}$, and $B_{25}$ perform significantly worse when paired with $B_4$ and $B_5$, but exhibit large Z-scores in combination with $B_2$ and $B_3$. However, overall we find the impact of individual building blocks to be the dominant factor. For instance, $B_{28}$ has the largest positive Z-score, regardless of its counterpart.

Overall, we find anthracene ($B_{28}$) to be the most promising moiety, a surprising finding and somewhat contradictory to our initial hypothesis considering that it does not contain sulfur. However, it is in line with (and independently confirms) other efforts in the community that aim to integrate anthracene into high-RI optical polymers (see e.g., Ref. [59]). The other outstanding
FIG. 4. RI distribution histograms of (a) the 38,619 individual $R_1$ residues; (b) the 171,172 individual $R_2$ residues; and (c) the 10,000 PI structures resulting from the top $R_1$ and $R_2$ residues. We observe a distinct shift towards higher RI values in (c).

FIG. 5. RI value distribution around the respective average RI values (blue points) of the (a) $R_1$ residues; (b) $R_2$ residues; and (c) PI candidates containing each building block. The blue bands refer to one standard deviation. The red points in (c) show the building block counts.

FIG. 6. Z-score of each building block in the top 10% $R_1$ and $R_2$ candidates compared to the entire $R_1$ and $R_2$ candidate pool, respectively. Green color represents a positive Z-score, and negative ones are shown in red.

IV. CONCLUSIONS

In this study, we demonstrated that the data-driven in silico approach that is being advanced by us and others can rapidly and efficiently assess the properties and performance potential of high-RI candidate compounds, identify numerous leads for next-generation PIs, and elucidate structure-property relationships that form the foundation for rational design rules. By combining our RI prediction model with virtual high-throughput screening techniques, we characterized candidates on a large scale at a fraction of the time and cost of traditional studies. We identified high-value building blocks (e.g., anthracene, dibenzothiophene, thianthrene) and structural patterns (e.g., S- and O-linkers) that correlate with large RI values. Correspondingly, we identified regions in chemical space in which we can hope to maximize the RI values of PIs. These guidelines allow us to target specific molecular motifs and create polymers with exceptional optical properties. In future experimen-
tal work, we will utilize these guidelines and pursue the promising candidates that have emerged.

SUPPLEMENTARY MATERIAL

Electronic supplementary material accompanies this paper and is available through the journal website. It provides details of the computational data displayed in the figures throughout this paper or that were used in the statistical analysis. We also give detailed definitions of all statistical metrics used in this work.

COMPETING FINANCIAL INTERESTS

The authors declare to have no competing financial interests.

ACKNOWLEDGMENTS

This work was supported by start-up funds provided through the University at Buffalo (UB), the National Science Foundation (NSF) CAREER program (grant No. OAC-1751161), and the New York State Center of Excellence in Materials Informatics (grant No. CMI-1148092-8-75163). Computing time on the high-performance computing clusters 'Rush', 'Alpha', 'Beta', and 'Gamma' was provided by the UB Center for Computational Research (CCR). The work presented in this paper is part of MAFA’s PhD thesis [60]. MH gratefully acknowledges support by Phase-I and Phase-II Software Fellowships (grant No. ACI-1547580-479590) of the NSF Molecular Sciences Software Institute (grant No. ACI-1547580) at Virginia Tech [61, 62].
FIG. 8. Z-scores of building block combinations in the top 10% candidates of $R_1$ and $R_2$. The size of the circles correspond to the magnitude of the Z-score value. Green circles indicate positive Z-scores and red circles negative values.


[40] Carlos Amador-Bedolla, Roberto Olivares-Amaya, Johannes Hachmann, and Alán Aspuru-Guzik, “Organic Photovoltaics,” in Informatics for Materials Science and
Other files

R1_R2_PI_screening_results.xlsx (10.96 MiB)