Designing Molecules with Optimal Properties Using the Linear Combination of Atomic Potentials Approach in an AM1 Semiempirical Framework

Shahar Keinan, Xiangqian Hu, David N. Beratan,* and Weitao Yang*

Department of Chemistry, Duke University, Durham, North Carolina 27708

Received: July 20, 2006; In Final Form: November 2, 2006

The linear combination of atomic potentials (LCAP) approach is implemented in the AM1 semiempirical framework and is used to design molecular structures with optimized properties. The optimization procedure uses property derivative information to search molecular space and thus avoid direct enumeration and evaluation of each molecule in a library. Two tests are described: the optimization of first hyperpolarizabilities of substituted aromatics and the optimization of a figure of merit for n-type organic semiconductors.

I. Introduction

Chemical space is vast, with an estimated 10^{65} stable molecules accessible with a molecular weight below 850.[1] For example, there are more then 10^{29} possible derivatives of *n*-hexane.[2] Designing optimal new molecular materials with specified properties requires scanning chemical space. Exploration by direct enumeration and evaluation is prohibitively costly. Traditional methods to design new structures are often based on structure−activity relationships or combinatorial methods. These are "forward design" strategies that start with a molecule, evaluate its properties, and suggest strategies for improvement.

Inverse methods begin with the target properties and seek structures that optimize these properties.[3] "Inverse design" can be implemented with continuous or discrete methods. Integer-based exploration of discrete space can be inefficient (e.g., branch and bound methods[4]). Although continuous optimization has been explored previously,[5,6] it can be challenging to associate optimization results with specific molecules. We recently described a linear combination of atomic potentials (LCAP) approach to transform molecular optimization, a challenge of discrete optimization, into a continuous optimization problem.[7] The LCAP approach has since been used with a plane wave based density functional theory and a gradient directed Monte Carlo approach[8] to optimize molecular hyperpolarizabilities. A Hückel-based continuous optimization of atom types has also been implemented on molecular scaffolds that produce extremely large libraries.[9] Others have applied related ideas to drug design[10,11] and protein folding.[12]

The LCAP method expands the external electrostatic potentials as a linear combinations of atomic (or chemical group) potentials

$$v(r) = \sum_{R=1}^{M} \sum_{i=1}^{n_R} \lambda_i^R \, v_i^R(r) \qquad \sum_{i=1}^{n_R} \lambda_i^R = 1 \qquad 0 \le \lambda_i^R \le 1 \tag{1}$$

Equation 1 describes the external electrostatic potential $v(r)$ of a molecule consisting of M sites, each with n_R possible groups. The λ_i^R coefficients define the admixture of an atom or group at position R. When $\lambda_i^R = 0$ for all choices of i except a single one and $\lambda_i^R = 1$ for exactly one choice of i for all R sites, $v(r)$ corresponds to a real chemical species; such potentials are called chemical representable (CR).[7] Otherwise, $v(r)$ corresponds to an "intermediate" or alchemical species. Since λ_i^R are continuous variables, the property optimization is a search for λ_i^R coefficients that produce molecules with the most favorable property values.

In our initial tests of the LCAP strategy, we have found that property surfaces need not be smooth, especially when changes of λ_i^R are accompanied by changes in the total number of electrons or of the molecular geometry. The roughness of the surface increased with the molecular diversity of the library. Following property derivatives continuously on the surface may therefore lead to two main problems. First, the search may stall in one of many local extrema corresponding to alchemical species. Second, most of the optimization cost will be spent exploring alchemical species, rather than real molecules. Trying to overcome the first problem by interpolation of the alchemical atoms to the nearest real atoms does not guarantee retaining the maximal property values. In order to overcome these challenges, a different optimization method is used here, namely a "gradient directed jump".[8] In this approach, the property derivatives near a real molecule are calculated using the LCAP method. Then a "jump" is made (based on derivative information) to the next real molecule. The "gradient-directed jump" retains knowledge of the property surface but avoids some of the difficulties associated with surface roughness. The optimization visits a series of discrete molecules, with jumps between molecules directed by property gradients calculated only near each molecule.

Here we describe the development of a gradient-directed LCAP jump method in an atomic-orbital AM1 semiempirical SCF framework[13] and apply it to (1) maximizing the first electronic hyperpolarizability of organic molecules and (2) lowering the LUMO energy for structures based on an unsaturated scaffold, with the aim of designing n-type organic semiconductors.

II. Computational Details

II.1. LCAP in AM1. Implementing the LCAP optimization with an AM1 semiempirical Hamiltonian requires modifying the AM1 Hamiltonian[13] and the wavefunctions to describe

* To whom correspondence should be addressed.

10.1021/jp0646168 CCC: $37.00
Published on Web 12/14/2006

alchemical atoms. The essential changes include introducing alchemical atomic masses, defining an alchemical valence orbital basis set, and modifying the one- and two-center integrals appropriately. The approach described here can be applied to first-principles LCAO-MO frameworks as well. The calculations described here were preformed with the DYNAMO implementation of AM1.[14]

For an alchemical atom X, the atomic charge is a linear combination of the atomic charges of the atoms included in the chemical library for site R

$$Z_X^R = \sum_{i=1}^{n_R} \lambda_i^R Z_i \tag{2}$$

Z_i is the integer atomic charge of atom i. AM1 calculations use a minimal valence (Slater) basis set.[15] Since orbitals depend on atom types, it is also necessary to change the atomic orbitals based on λ. For alchemical atoms in our LCAO-LCAP approach, each basis function is defined as a fixed linear combination of the atomic orbitals associated with the limiting atoms

$$\phi_\mu^X = \sum_{i=1}^{n_R} \lambda_i^R \phi_\mu^i \tag{3}$$

For atoms considered here, μ can be s, p_x, p_y, or p_z. For example, an alchemical atom X, intermediate between F and I, is a combination of F atom basis functions with strength λ_1 and I atom basis functions with strength λ_2

$$\begin{aligned}
\phi_{\mathrm{s}}^X &= \lambda_1 \phi_{\mathrm{s}}^{\mathrm{F}} + \lambda_2 \phi_{\mathrm{s}}^{\mathrm{I}} \\
\phi_{\mathrm{p}_x}^X &= \lambda_1 \phi_{\mathrm{p}_x}^{\mathrm{F}} + \lambda_2 \phi_{\mathrm{p}_x}^{\mathrm{I}} \\
\phi_{\mathrm{p}_y}^{\mathrm{X}} &= \lambda_1 \phi_{\mathrm{p}_y}^{\mathrm{F}} + \lambda_2 \phi_{\mathrm{p}_y}^{\mathrm{I}} \\
\phi_{\mathrm{p}_z}^{\mathrm{X}} &= \lambda_1 \phi_{\mathrm{p}_z}^{\mathrm{F}} + \lambda_2 \phi_{\mathrm{p}_z}^{\mathrm{I}}
\end{aligned} \tag{4}$$

This scheme matches the AOs of limiting atoms and constructs a new AO from the linear combination of eq 3. The variational LCAO-MO alchemical wavefunction is

$$\psi_i = \sum_\mu c_{\mu i} \phi_\mu^X \tag{5}$$

The one-center Hamiltonian integrals, such as $U_{\mu\mu}$ (the one-electron one-center integrals) are[13]

$$U_{\mu\mu}^A = \left\langle \phi_\mu^A \left| \frac{-\nabla^2}{2} - \frac{Z_A}{r} \right| \phi_\mu^A \right\rangle \tag{6}$$

which becomes, using eq 3

$$U_{\mu\mu}^X = \left\langle \sum_{i=1}^{n_R} \lambda_i^R \phi_\mu^i \left| \frac{-\sum_{i=1}^{n_R} \lambda_i^R \nabla_i^2}{2} - \frac{\sum_{i=1}^{n_R} \lambda_i^R Z_i}{r} \right| \sum_{i=1}^{n_R} \lambda_i^R \phi_\mu^i \right\rangle \tag{7}$$

for the alchemical species. There are three sums in eq 7. The cross terms between different atom types are approximated as averages

$$\left\langle \lambda_i^R \phi_\mu^i \left| \frac{-\lambda_i^R \nabla_i^2}{2} - \frac{\lambda_i^R Z_i}{r} \right| \lambda_i^R \phi_\mu^i \right\rangle = \lambda_i^{R^3} U_{\mu\mu}^i \tag{8}$$

$$\left\langle \lambda_i^R \phi_\mu^i \left| \frac{-\lambda_j^R \nabla_j^2}{2} - \frac{\lambda_j^R Z_j}{r} \right| \lambda_k^R \phi_\mu^k \right\rangle = \lambda_i^R \lambda_j^R \lambda_k^R \left(\frac{1}{3} U_{\mu\mu}^i + \frac{1}{3} U_{\mu\mu}^j + \frac{1}{3} U_{\mu\mu}^k \right) \tag{9}$$

So

$$U_{\mu\mu}^X = \sum_i \sum_j \sum_k \lambda_i^R \lambda_j^R \lambda_k^R \left(\frac{1}{3} U_{\mu\mu}^i + \frac{1}{3} U_{\mu\mu}^j + \frac{1}{3} U_{\mu\mu}^k \right) \tag{10}$$

The same approach is used to calculate β_μ.

Two-center integrals $(\mu\nu|\eta\sigma)$ are calculated using an analytical expression based on the NDDO (neglect of diatomic differential overlap) approximation.[13] Two-center integrals for alchemical atoms are calculated using the real atoms' two-center integrals, as implemented in the standard AM1 approach. The two-center integrals are

$$(\mu\nu|\eta\sigma) = \left\langle \phi_\mu^A \phi_\nu^A \left| \frac{1}{r} \right| \phi_\eta^B \phi_\sigma^B \right\rangle \tag{11}$$

Using eq 3 to define the alchemical atomic orbitals

$$(\mu\nu|\eta\sigma)^X = \left\langle \sum_{i=1}^{n_{R1}} \lambda_i^{R_1} \phi_\mu^i \sum_{j=1}^{n_{R2}} \lambda_j^{R_2} \phi_\nu^j \left| \frac{1}{r} \right| \sum_{k=1}^{n_{R3}} \lambda_k^{R_3} \phi_\eta^k \sum_{l=1}^{n_{R4}} \lambda_l^{R_4} \phi_\sigma^l \right\rangle \tag{12}$$

The assumptions of eqs 8 and 9 give

$$(\mu\nu|\eta\sigma)^X = \sum_{i=1}^{n_{R_1}} \sum_{j=1}^{n_{R_2}} \sum_{k=1}^{n_{R_3}} \sum_{l=1}^{n_{R_4}} \lambda_i^{R_1} \lambda_j^{R_2} \lambda_k^{R_3} \lambda_l^{R_4} \left[\frac{1}{4}(\mu\mu|\eta\eta) + \frac{1}{4}(\mu\mu|\sigma\sigma) + \frac{1}{4}(\nu\nu|\eta\eta) + \frac{1}{4}(\nu\nu|\sigma\sigma) \right] \tag{13}$$

Core−core repulsion terms are calculated using the same assumptions (Supporting Information).

II.2. LCAP Gradient Directed Jumping Search. In the molecular optimization, we use the gradient-directed jumping search developed in our laboratory.[8,9] The optimization in the discrete molecular space is carried out by following the LCAP gradients that improve the molecular property. In the earlier studies,[8,9] the gradients were obtained analytically. Here we use numerical derivatives (although analytical derivatives could be developed). Each structure optimization begins with a random molecule (random choice of binary values for λ_i^R), uses the LCAP to calculate the gradients of the property surface near that structure, finds the steepest gradients, and "jumps" to the next candidate molecule pointed to by those gradients. The specific scheme is:

1. Begin with a random molecule **A** (λ_i^R coefficients 0 or 1)
2. Calculate property $\mathbf{P_A}$ of molecule **A**
3. Compute the property gradients (numerically) with respect to coefficients. There are $j = 1...M$ positions, and n_R possible groups for each R_j position. For each position R_j compute the gradient of all possible atoms or groups X_i, to find the largest one:

• Build LCAP molecule **B** in the geometry of molecule **A** by changing only group X_i at position R_j to have small positive ($\lambda_i^R = +0.01$) LCAP contribution, keeping all other groups as in **A**

• Calculate property $\mathbf{P_B}$ of molecule **B**

• Build LCAP molecule **C** in the geometry of molecule **A** by changing group X_i at position R_j to have small negative ($\lambda_i^R = -0.01$) LCAP coefficient, keeping all other groups as in **A**

• Calculate property $\mathbf{P_C}$ of molecule **C**

• The property derivative associated with changes to group X_i at position R_j is $\Delta = \mathbf{P_B} - \mathbf{P_C}$

When looking for maxima, find the group X_i that has the largest property derivative for each position R_j. When looking for minima, find the group X_i that has the lowest property derivative for each position R_j.

4. Build the next molecule $\mathbf{A_{new}}$, containing all the X_i groups (one per R_j position) that had the largest (or lowest) property derivative.

5. Test to see if the new molecule, $\mathbf{A_{new}}$, was previously visited

If **no** - go to step 2 for another cycle

If **yes** - end optimization

II.3. Calculations of Properties. Both the electric dipole moments (μ) and the static hyperpolarizabilities (β) are calculated using the finite-field method.[16] The electric field F dependent dipole is

$$\mu_i = \left[-\frac{3}{2}\{E(F_i) - E(-F_i)\} + \frac{1}{12}\{E(2F_i) - E(-2F_i)\}\right]/F_i$$

$$\mu_{tot} = \sqrt{\sum_{i=x,y,z} \mu_i^2} \qquad (14)$$

The total hyperpolarizability is

$$\beta_{iii} = \left[-\frac{1}{2}\{E(2F_i) - E(-2F_i)\} + \{E(F_i) - E(-F_i)\}\right]/(F_i)^3$$

$$\beta_i = \frac{1}{3}\sum_{k=x,y,z} (\beta_{ikk} + \beta_{kik} + \beta_{kki})$$

$$\beta_{tot} = \sqrt{\beta_x^2 + \beta_y^2 + \beta_z^2} \qquad (15)$$

E is the electronic energy. Typical F_i values used are 0.1 au.

III. Applications

Electronic Energy and Dipole-Moment Surfaces. We used the AM1-LCAO-LCAP approach to explore electronic energy (Figure 1) and dipole moment (Figure 2) surfaces for the continuous change $\mathbf{H_3C-CH_3} \leftrightarrow \mathbf{NC-CN}$. This is a two-site system where **X** and **Y** correspond to $-CH_3$ and $-CN$, respectively. $\lambda = 0$ denotes CH_3 and $\lambda = 1$ denotes CN. There are three molecules in this family: $\mathbf{H_3C-CH_3}$, $\mathbf{NC-CH_3}$, and $\mathbf{NC-CN}$.

The initial geometry is the AM1 optimized geometry of $\mathbf{H_3C-CH_3}$. The nitrogen and carbon atoms of $\mathbf{NC-CN}$ and $\mathbf{H_3C-CH_3}$ occupy the same positions in space, and ethane has an eclipsed structure. One hydrogen atom in $\mathbf{H_3C-CH_3}$ is changed to nitrogen in $\mathbf{NC-CN}$, and the other two hydrogen atoms on each carbon in $\mathbf{NC-CN}$ disappear in the $\mathbf{NC-CN}$ structure (λ values go from one to zero). The number of valence electrons is changed from 14 to 18. Figures 1 and 2 show that the electronic energy and the dipole moments change smoothly with the LCAP coefficients.

First Hyperpolarizability Optimization of PNA Derivatives. Organic molecules with large electronic hyperpolarizabilities typically have unsaturated bridges linking donor and acceptor substituents. *p*-Nitroaniline (PNA), **1a**,[17,18] and its derivatives have been studied intensively in this context. The electronic structure and hyperpolarizability are sensitive to the chemistry of the donor, bridge, and acceptor, as well as to the molecular geometry.[19,20]

Here, as an example, we optimized the chemical structure of PNA derivatives by changing the donor and acceptor units, as well as by changing atom types in the bridge. Chart **1b** shows all of the possible groups and their locations on a PNA-like framework. The donor group **Z** library includes $-N(CH_3)_2$, $-OH$, $-NH_2$ (planar), $-NH_2$* (* indicates that the NH_2 plane is perpendicular to the ring), $-I$, $-Br$, $-F$, or $-H$. The acceptor

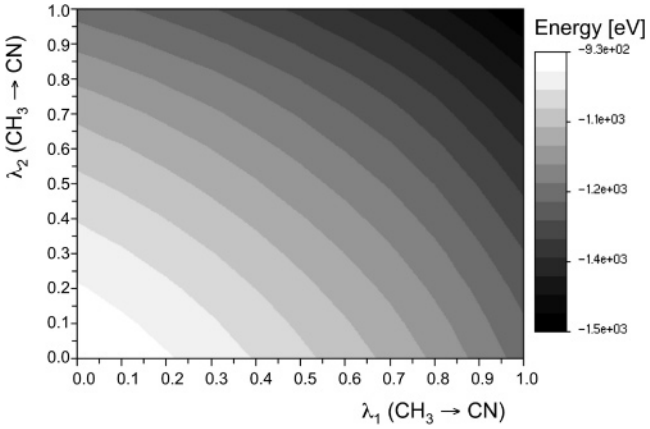


Figure 1. Electronic energy surface for $\mathbf{H_3C-CH_3} \leftrightarrow \mathbf{NC-CN}$. The contours are drawn as a function of the two CH_3 weighing coefficients (λ_1 and λ_2). The lower left corner corresponds to $\mathbf{H_3C-CH_3}$, the upper right corner corresponds to **NC**−**CN**, and the other corners correspond to $\mathbf{NC-CH_3}$.

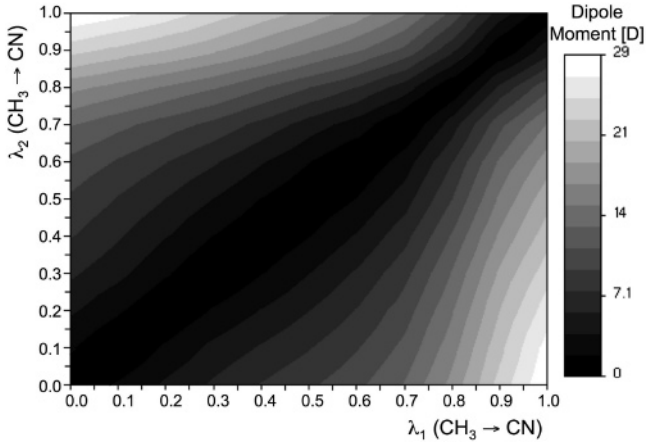
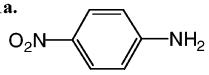


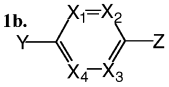
Figure 2. Dipole moment surface of $\mathbf{H_3C-CH_3} \leftrightarrow \mathbf{NC-CN}$. The contours are drawn as a function of the two CH_3 weighing coefficients (λ_1 and λ_2). The lower left corner corresponds to $\mathbf{H_3C-CH_3}$, the top right corner corresponds to **NC**−**CN**, and the two other corners correspond to $\mathbf{NC-CH_3}$.

CHART 1

1a.

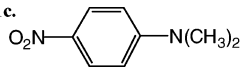


1b.



$\mathbf{X_{1-4}}$ = CH or N
Y (acceptor) = -CN, -H, -NO_2, -NO_2* (perpendicular)
Z (donor) = -$N(CH_3)_2$, -OH, -NH_2, -NH_2* (perpendicular), -I, -Br, -F, -H

1c.



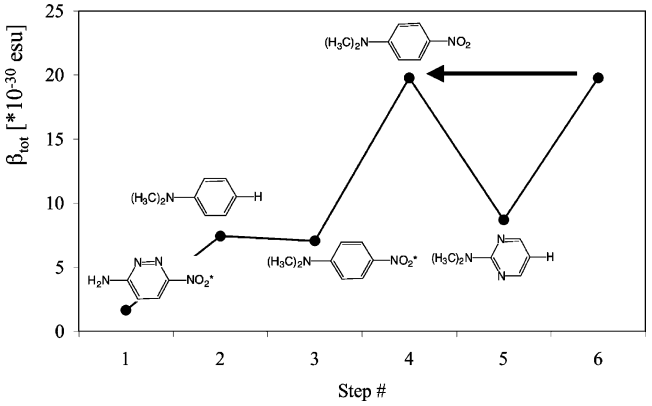


Figure 3. Example of an optimization profile that searches for the highest β_{tot} (in units of 10^{-30} esu) of PNA analogues. *N*,*N*-Dimethyl *p*-nitroaniline, **1c**, is found to be the optimum.

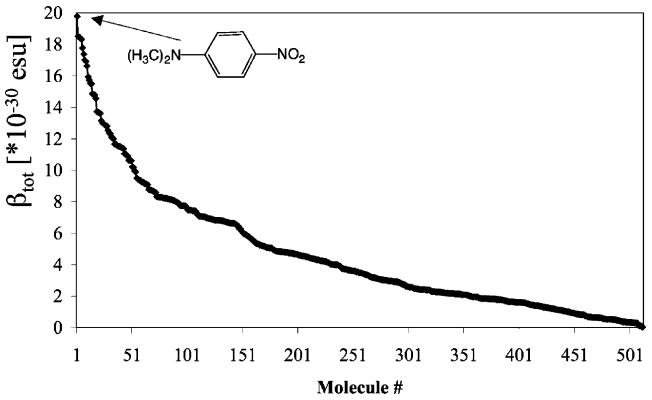


Figure 4. Calculated β_{tot} values for the 512 molecules in the library, assuming a frozen geometry.

group **Y** may be $-NO_2$ (planar), $-NO_2$* (perpendicular), $-CN$, or $-H$, and each atom ring (**X**) can be CH or N. Perpendicular geometries are twisted by 90° to the ring plane, in order to include several geometries of the same chemical group in the optimization. This is a small library, allowing enumeration of all 512 structures and their properties to explore the outcome of the LCAP optimization.

The strategy described in section II.2 was applied to optimize β_{tot}, and each β_{tot} value was calculated using the finite-field method of section II.3. Molecular geometries were frozen and were based on the AM1-optimized geometries of the largest molecule ($\mathbf{X_{1-4}}$ = CH, **Y** = NO_2, **Z** = $N(CH_3)_2$), which is planar. For all chemical groups, we assume that atoms are located in the same positions as in the AM1-optimized geometry. For example, O (of OH), N (of NH_2), Br, F, and H in the donor group **Z** are all located at the position of the nitrogen of $N(CH_3)_2$. For iodine, the original distance of 1.39 Å between I and C was too small for SCF convergence, so it was extended to 1.85 Å. Twelve optimization runs were conducted, each beginning with a randomly chosen molecule. On average, five steps were required to complete the optimization. Figure 3 shows the progress of an optimization resulting in the *N*,*N*-dimethyl *p*-nitroaniline structure, **1c**. *N*,*N*-Dimethyl *p*-nitroaniline was found to be the optimum in all 12 searches, independent of the initial random molecule selected.

The molecule with the largest β_{tot}, found by enumerating and analyzing all molecules in the set (Figure 4), is *N*,*N*-dimethyl *para*-nitroaniline, **1c**. The same molecule was found to be the maximum in the LCAP optimization. To test the viability of the frozen geometry assumption, we preformed AM1 geometry optimization for each of the 512 molecules in the library and calculated β_{tot} for each (Supporting Information). Comparing the frozen geometry with the optimized geometry for several molecules, we found that the two geometries were similar, justifying the frozen geometry assumption. This analysis confirms that *N*,*N*-dimethyl *p*-nitroaniline is indeed the species with the largest hyperpolarizability in this library.

Organic n-Type Semiconductors. Organic semiconductors are of great interest for applications in thin-film transistors, light-emitting diodes, and other electronic devices.[21] n-type (electron accepting) organic semiconductors are relatively uncommon,[22] and there is great interest in discovering more candidates. In 2003, the Marks group showed that a perfluoroarene-substituted polythiophene, **2a**, can be used as an n-type organic semiconductor.[23] The investigators attributed this behavior, among other things, to a low LUMO energy. Using **2a** as the lead structure for our LCAP optimization, a large library was created and optimization was conducted in an attempt to discover structures with even lower LUMO energies.

Heteroatom species were varied in the LCAP optimization, as were electronegative substituents on the terminal rings (see Chart 2b). The chemistry at 14 sites was changed, with three possibilities at each site, leading to 4.8 million structures (without reducing the count for symmetry equivalent structures).

The PNA-based library assumed planar frozen geometries. In the library of n-type semiconductors, the geometry changes dramatically with chemical changes. Figure 5 compares the AM1 optimized geometry of **2a** with that of **2d**. When the 5-member-ring heteroatoms are changed from sulfur to oxygen, the geometry changes from twisted to planar.

We changed the LCAP optimization protocol to allow geometry evolution at each step by introducing geometry

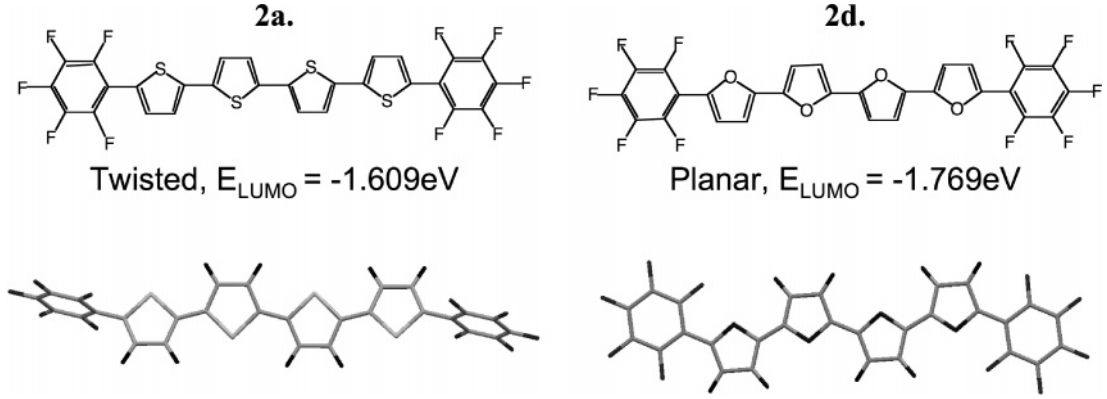
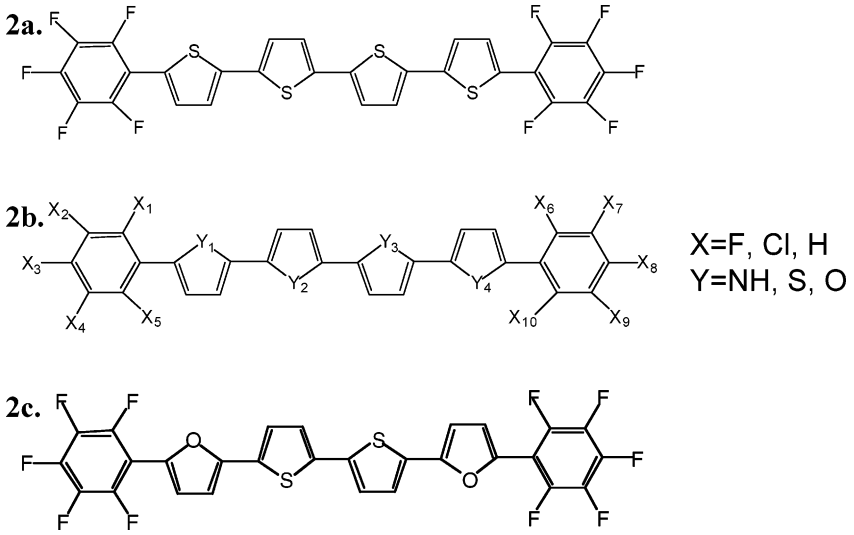


Figure 5. Comparison of the AM1 optimized geometry of **2a** perfluoroarene-modified polythiophene with **2d** perfluoroarene-modified polyfuran. When the 5-member-ring heteroatoms are changed from sulfur to oxygen, the geometry is changed from twisted to planar.

CHART 2



optimization for each lead molecule **A** in the second LCAP optimization step. In each cycle of the optimization, lead molecule $\mathbf{A_i}$ is geometry optimized (without LCAP), the E_{LUMO} value is calculated, and the property derivatives are calculated for that specific geometry (with LCAP). Following the steepest property derivatives, a new lead molecule $\mathbf{A_{i+1}}$ is chosen and a new optimization cycle begins (details appear in the Supporting Information).

Nineteen optimization runs were conducted, each begins with a randomly chosen molecule that has the prescribed covalent framework. An average of five steps was required to reach an optimized structure. Figure 6 shows an optimization profile, with the lowest LUMO energy found for perfluoroarene-furan-thiophene−thiophene-furan-perfluoroarene, **2c**. Perfluoroarene-furan-thiophene−thiophene-furan-perfluoroarene, with planar geometry and $E_{LUMO} = -1.844$ eV, was found in all nineteen runs as the optimum structure, independent of the randomly chosen initial molecule.

It is difficult to prove that **2c** is the global optimum in the molecular library, since the library contains 4.8 million structures. Figure 7 shows calculated E_{LUMO} values for 200 randomly chosen structures and Table 1 shows calculated E_{LUMO} values for several key molecules in the **2b** library. All calculated E_{LUMO} values are above that designed using the LCAP approach, which yields structure **2c**, with $E_{LUMO} = -1.844$ eV.

Table 1 shows the chemical, structural, and E_{LUMO} data for several candidates with low E_{LUMO}'s. Three important factors

TABLE 1: AM1 Calculated E_{LUMO} Values for Several Key Molecules in the 2b Library

	X_{1-10}	Y_1	Y_2	Y_3	Y_4	geometry	E_{LUMO} [eV]
2a	F	S	S	S	S	twisted	−1.609
2e	Cl	S	S	S	S	twisted	−1.495
2f	H	S	S	S	S	planar	−1.397
2d	F	O	O	O	O	planar	−1.769
2g	F	NH	NH	NH	NH	twisted	−1.494
2h	F	S	O	O	S	twisted	−1.398
2c	**F**	**O**	**S**	**S**	**O**	**planar**	**−1.844**

can be deduced from this table as controlling the LUMO energy. The first is the electronegativity of the phenyl substituents: more electronegative substituents are more favorable.[24] The second factor is the heteroatom: the stronger the π-donor strength, the lower the LUMO energy.[25] The third factor is the coplanarity of the rings. Steric interactions cause chain twisting that decreases the effective conjugation and raise E_{LUMO}. For example, the geometry of **2a** is twisted and the LUMO is located only on the four heteroatom-containing rings. In contrast, **2c** is a planar molecule, the LUMO is more delocalized, and the LUMO is localized on the phenyl rings and on the heteroatom-containing rings (Supporting Information). Based on these observations, a molecule with a low LUMO energy in this structure class should be planar, with as many fluorines on the phenyl rings as possible, and a maximum number of sulfur atoms in the 5-member-ring heterocycles. Even knowing these general structure−function rules, discovering the optimum

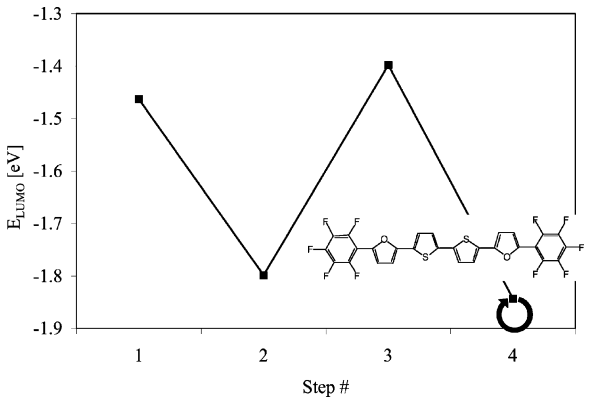


Figure 6. Typical optimization profile for the lowest E_{LUMO} (in eV) in the library based on **2b**. Perfluoroarene-furan-thiophene−thiophene-furan-perfluoroarene, **2c**, is found to be the optimum structure.

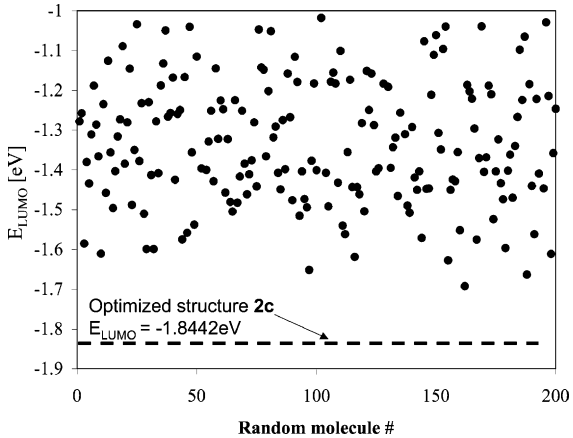


Figure 7. AM1 calculated E_{LUMO} values for 200 randomly chosen molecules in the **2b** library.

structure from a library of 4.8 million possibilities is difficult without the LCAP strategy. These three structural factors have been described previously,[24,25] and they have significant implications for the design of new organic semiconductors.

IV. Conclusions

The LCAP method was implemented in a semiempirical SCF-LCAO electronic structure theory framework. This LCAO-LCAP approach was used with the AM1 Hamiltonian to optimize the properties and structures in large molecular spaces using a gradient directed jumping strategy. We explored the optimization of first molecular hyperpolarizabilities and LUMO energies. In the latter case, the LCAP approach was combined with geometry optimization. The structural library was constructed based on known lead molecules (**2a**), and the optimized structure was predicted to have a LUMO energy about ~0.2 eV below known structures.

This study shows that the LCAO-LCAP inverse design strategy can explore large molecular spaces using semiempirical SCF methods. We have also demonstrated that the approach is applicable to important molecular design challenges, including hyperpolarizabilities and orbital energy engineering.

Acknowledgment. We are grateful to Prof. Martin Field for providing the DYNAMO program. Support of the DARPA Predicting Real Optimized Materials project through ARO is gratefully acknowledged (W911NF-04-1-0243).

Supporting Information Available: A total of two figures (β_{tot} for the 512 geometry-optimized molecules in the PNA library and AM1 HOMO and LUMO for **2a** and **2c**) and two text pages (how to calculate the core−core repulsion term in LCAP and LCAP combined with geometry optimization). This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Dobson, C. M. *Nature* **2004**, *432*, 824.

(2) Lipinski, C.; Hopkins, A. *Nature* **2004**, *432*, 855.

(3) Edgar, T. F.; Dixon, D. A.; Reklaitis, G. V. Vision 2020: Computational needs of the chemical industry. In *Impact of advances in computing and communications technologies on chemical science and technology*; National Academy Press: Washington, DC, 1999; pp 74.

(4) Ostrovsky, G. M.; Achenie, L. E. K.; Sinha, M. *Comput. Chem.* **2002**, *26*, 645.

(5) Kuhn, C.; Beratan, D. N. *J. Phys. Chem.* **1996**, *100*, 10595.

(6) Franceschetti, A.; Zunger, A. *Nature* **1999**, *402* (6757), 60.

(7) Wang, M.; Hu, X.; Beratan, D. N.; Yang, W. *J. Am. Chem. Soc.* **2006**, *128*, 3228.

(8) Hu, X.; Yang, W.; Beratan, D. N. A gradient directed Monte-Carlo approach to inverse molecular design. In preparation, **2006**.

(9) Xiao, D.; Beratan, D. N.; Yang, W. Inverse molecular design in a molecular orbital framework. Submitted, **2006**.

(10) von Lilienfeld, O. A.; Lins, R. D.; Rothlisberger, U. *Phys. Rev. Lett.* **2005**, *95*, 153002.

(11) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *J. Chem. Phys.* **2005**, *122*, 14113.

(12) Koh, S. K.; Ananthasuresh, G. K.; Vishveshwara, S. *Int. J. Rob. Res.* **2005**, *24*, 109.

(13) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(14) Field, M. J.; Albe, M.; Bret, C.; Proust-De, Martin, F.; Thomas, A. *J. Comp. Chem.* **2000**, *21*, 1088.

(15) Pople, J. A.; Beveridge, D. L. *Approximate molecular orbital theory*; McGraw-Hill Book Company: New York, 1970.

(16) Kurtz, H. A.; Dudis, D. S. Quantum mechanical methods for predicting nonlinear optical properties. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New-York, 1998; Vol. 12, p 241.

(17) Beratan, D. N. Electronic Hyperpolarizability and Chemical Structure. In *Materials for Nonlinear Optics: Chemical Perspective*; Marder, S. R., Sohn, J. E., Stucky, G. D., Eds.; American Chemical Society: Washington, DC, 1991; Vol. 455, p 89.

(18) *Nonlinear Optical Materials: Theory and Modeling*; Karna, S. P., Yeates, A. T., Eds.; American Chemical Society: Washington, DC, 1996; Vol. 628.

(19) Albert, I. D. L.; Marks, T. J.; Ratner, M. A. First hyperpolarizability of molecular chromophores: practical computational approach. In *Progress in Organic Nonlinear Optics*; Kuzyk, M., Dirk, C. R., Eds.; Marcel-Dekker: New-York, 1998; p 37.

(20) Keinan, S.; Zojer, E.; Brédas, J. L.; Ratner, M. A.; Marks, T. J. *THEOCHEM* **2003**, *633*, 227.

(21) Law, K. Y. *Chem. Rev.* **1993**, *93*, 449.

(22) Chua, L.-L.; Zamseil, J.; Chang, J.-F.; Ou, E. C.-W.; Ho, P. K.-H.; Sirringhaus, H.; Friend, R. H. *Nature* **2005**, *434*, 194.

(23) Facchetti, A.; Yoon, M.-H.; Stern, C. L.; Katz, H. E.; Marks, T. J. *Angew. Chem., Int. Ed.* **2003**, *42*, 3900.

(24) Hutchison, G. R.; Ratner, M. A.; Marks, T. J. *J. Phys. Chem. B* **2005**, *109*, 3126.

(25) Salzner, U.; Lagowski, J. B.; Pickup, P. G.; Poirier, R. A. *Synth. Met.* **1998**, *96*, 177.