J|A|C|S
COMMUNICATIONS

Published on Web 03/18/2006

Orbital Energetics and Molecular Recognition

Aaron George,[†] Yonas Abraham,[†] Carlo Sbraccia,[§] Vishali Mogallapu,[‡] Rebecca Harris,[†] Roberto Car,[§] and Jeffrey D. Schmitt*,[†]

Targacept, Inc., 200 East First Street, Suite 300, Winston-Salem, North Carolina 27101, Department of Electrical Engineering, University of North Carolina at Charlotte, Charlotte, North Carolina 28223, and Chemistry Department, Princeton University, Princeton, New Jersey 08544

Received December 5, 2005; E-mail: jeff.schmitt@targacept.com

The field of computational chemistry has established that molecular orbital theory is a useful and fundamental way of representing and simulating molecular structure. Energetics of molecular orbitals has provided a direct window to experimentally measured quantities, mainly spectroscopic observables. In this communication, we first show that orbital energy fluctuations, based on canonical ab initio molecular dynamics simulations, exhibit significant asymmetry. We hypothesize that the asymmetry of orbital energy fluctuation will reflect how the molecule interacts with the environment. To test this hypothesis, we developed a new class of QSAR/QSPR descriptors, termed DYNEVA (DYNamic EigenVAlues). This work posts an improvement upon the predictive power of quantum mechanically derived descriptors by including the temporal dimension of atomic orbital energetics, further support that capturing the evolution of fundamental quantities can lead to more accurate prediction of biomolecular interaction.

This investigation has been inspired by an earlier and successful QSAR approach, EEVA, which uses quantum mechanically derived electronic structure information (electronic eigenvalues) as the basis for descriptors. In EEVA, a Gaussian spread (σ) is applied to each eigenvalue—the underlying assumption being that the energy distribution is symmetric—to create a pseudospectrum which is then divided into intervals. The integral of each interval leads to a single EEVA descriptor.[1] Our investigation was further inspired by a significant body of work striving to account for conformational flexibility in the context of QSAR. Hopfinger has developed descriptors as measures of conformational flexibility and/or entropy.[2] Dobler and Vedani allow the representation of molecules by an ensemble of conformations, orientations, and protonation states.[3] Likewise, Mekenyan and co-workers have developed a probabilistic descriptor weighting approach that takes conformational flexibility into account to improve model quality.[4] The ALPHA descriptor is derived from Gaussian smoothing of power spectra and, like EEVA, requires a spreading parameter.[5]

We posit that asymmetric behavior in molecular orbital energetics is due to individual orbital response to local intramolecular fields that fluctuate through dynamical evolution of the molecule. To test our hypothesis, we use ab initio Density Functional Theory (DFT)-based Car−Parrinello molecular dynamics (MD).[6] Canonical finite temperature MD simulations are performed at 800 K, accomplished by using a Nose−Hoover thermostat (a single mass, oscillating at 30.5 THz). Orbital eigenvalues, total energy, and energy constituents are recorded at regular intervals of 7.25 fs.

To illustrate the power of this approach, we use DYNEVA descriptors to build QSAR/QSPR models for two structurally diverse, evenly distributed datasets constructed from published data of pharmacological interest: (1) $\alpha 4\beta 2$ neuronal nicotinic receptor (NNR) affinity (K_i),[7] and (2) Topliss oral bioavailability.[8] The asymmetric nature of the eigenvalue data is depicted in a histogram of the absolute difference between mean and median values for each dataset (Figure 1A,B); the magnitude of |mean−median| is an illustration of the distributional asymmetry. We point out that this difference can be as high as 0.045 eV. Digging deeper, we plot example histograms of the HOMO for molecules TC-2216 (Figure 2A), terbutaline (Figure 2B), and ethosuximide (Figure 2C), illustrating the striking asymmetry of individual orbital distributions. Interestingly, there is a marked difference in the bioavailability between 2B and 2C. Finally, we chart the |mean−median| values of the highest 20 orbitals of all three example molecules (Figure 3). We posit that the asymmetric distribution of these data (presumably due to the interaction of local, intramolecular fields) contains more information than the symmetric approximation and may provide further insight into molecular recognition phenomena. Additionally, DYNEVA may be useful for differentiation of broad-based biological properties.

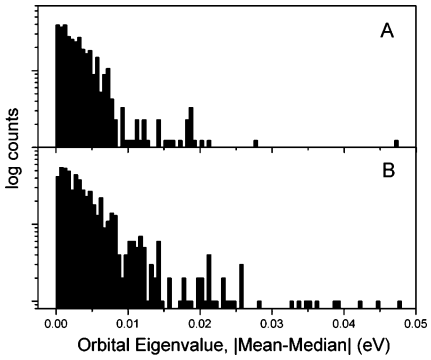

Figure 1. The asymmetric nature of the dynamic eigenvalue data is illustrated for Topliss (A) and $\alpha 4\beta 2$ (B) datasets. Histograms represent the |mean−median| of all the electronic eigenvalues for all the molecules in the datasets. |mean−median| = 0 implies a symmetric distribution. Data were collected using 30 ps of Car−Parrinello molecular dynamics.

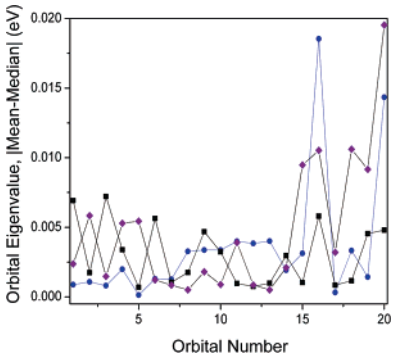

Figure 2. Eigenvalue distribution data by orbital for molecules TC-2216 (diamonds), terbutaline (circles), and ethosuximide (squares). The 20 highest energy orbitals |mean − median| are displayed.

[†] Targacept, Inc.
[‡] University of North Carolina at Charlotte.
[§] Princeton University.

10.1021/ja057826r CCC: $33.50

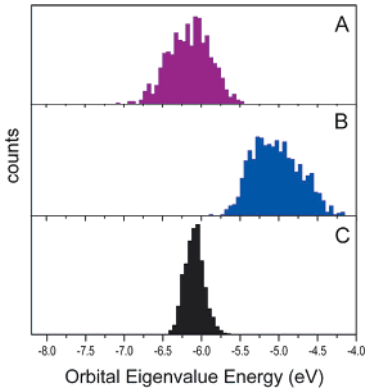

Figure 3. The asymmetric nature of the eigenvalue data is further illustrated for the Highest Occupied Molecular Orbital (HOMO): TC-2216 (A), terbutaline (B), and ethosuximide (c).

Table 1. Comparison of DYNEVA, EEVA, and MOE Descriptor Models Generated with GPLS and PLS[a]

		PLS		GPLS	
		r^2	q^2	r^2	q^2
α4β2	DYNEVA	0.82	0.06	0.95	0.67
	EEVA ($T = 800$ K)	0.43	0.00	0.44	0.26
	EEVA	0.41	0.13	0.56	0.33
	MOE	0.31	0.10	0.38	0.12
Topliss	DYNEVA	0.88	0.13	0.95	0.56
	EEVA ($T = 800$ K)	0.50	0.09	0.72	0.38
	EEVA	0.84	0.16	0.80	0.42
	MOE	0.20	0.41	0.89	0.51

[a] Correlation coefficient (r^2), leave-one-out cross-validation (q^2) scores for the α4β2 ($n = 30$) and Topliss ($n = 20$) datasets. See Supporting Information for additional modeling parameters.

One can assume that the harmonic vibrations of a molecule at finite temperature induce an oscillating intramolecular potential that acts as a perturbation on the electronic states. The probability distribution for each eigenvalue can thus be estimated by assuming equipartition of energy and by using perturbation theory. In the first order of the perturbing potential, the probability distribution of each eigenvalue is given by a sum of Gaussians (one for each normal mode) centered at the unperturbed eigenvalue, so that the probability distribution remains symmetric. The asymmetry in the distribution (as seen in Figure 3) enters via second-order perturbation. These terms are quadratic in the perturbing potential, and their strength depends on the extent of coupling between different states.[9] For the HOMO and LUMO states, which are most important in forecasting molecular reactivity and binding, the asymmetry in the eigenvalue distribution is related to the HOMO−LUMO gap. This measure is, however, more sophisticated than just the HOMO−LUMO gap because it includes the effect of the molecular vibrations on the eigenstates. The extent to which conformational dynamics contributes to asymmetry in the eigenvalue signature is a matter of further investigation. Next, we build descriptors to map this fluctuation in a form highly applicable to QSAR and QSPR studies. Before describing DYNEVA, we remind the reader that the number of eigenvalues varies between molecules, thus direct comparison is precluded. As a solution, eigenvalue trajectories (histograms) are registered on a common energy scale, which is then divided into 250 bins of equal width. The total count in each bin becomes a particular DYNEVA descriptor, thus creating a standardized numerical vector, which can be used to compare two or more molecules.

Next we utilize Cerius2's Genetic-PLS[10] (GPLS) to evaluate the descriptors. Partial Least Squares[11] (PLS) results are included for reference only. A key advantage of GPLS is that its genetic algorithm significantly reduces the number of descriptors used in the resulting models relative to PLS. To illustrate the increase in information content due to the inclusion of dynamics, we provide basic correlation measures (r^2, squared correlation coefficient; and q^2, the leave-one-out cross-validation score) for DYNEVA, EEVA, and MOE[12] descriptors (Table 1). Also included in this table is the EEVA approximation at finite ($T = 800$ K) temperature. The first section shows results for the α4β2 affinity dataset, and the second section, the Topliss dataset. Using GPLS as the basis for comparison, we observe a significant improvement in the DYNEVA descriptors over the static descriptor data (EEVA and MOE). The PLS models are likely over-fit. Establishing the physical basis for the robustness of DYNEVA descriptors will require further investigation, but we believe that ligand behavior is more broadly captured in the signature of local field interactions since the descriptors described herein show improvement over those employing Gaussian kernelling, such as the case with EEVA.

We show preliminary evidence for asymmetry in the energy distribution of molecular orbitals. On the basis of this finding, we have outlined a method of creating descriptors based upon molecular dynamics at the electronic structure level. Preliminary results demonstrate the promise of the DYNEVA approach, although further statistical analysis will be required to established its true utility. Our expectation is that this method can be applied to dynamics derived from any level of quantum molecular theory. We have observed the eigenvalue trajectory asymmetry using semiempirical Hartree−Fock dynamics and have obtained promising preliminary modeling results on the aforementioned datasets (data not shown). A key advantage of this semiempirical implementation is that we have observed a 100-fold speed increase with far less memory usage. A study of eigenvalue distribution as a function of finite temperature may also be useful. Rigorous statistical analysis of DYNEVA descriptors using other datasets is currently underway.

Acknowledgment. This material is based upon work supported by the Advanced Technology Program of the National Institute of Standards and Technology (NIST), Award No. 70NANB3H3065.

Supporting Information Available: Car−Parrinello simulation and QSAR/QSPR modeling parameters. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Tuppurainen, K.; Viisas, M.; Laatikainen, R.; Peräkylä, M. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 607−613.
(2) Hopfinger, A. J.; Wang, S.; Tokarski, J. S.; Jin, B.; Albuquerque, M.; Madhav, P. J.; Duraiswami, C. *J. Am. Chem. Soc.* **1997**, *119*, 10509−10524.
(3) Dobler, A.; Vedani, M. *J. Med. Chem.* **2002**, *45*, 2139−2149.
(4) (a) Mekenyan, O.; Nikolova, N.; Schmieder, P.; Veith, G. *QSAR Comb. Sci.* **2004**, *23*, 5−18. (b) Mekenyan, O.; Nikolova, N.; Schmieder, P. *J. Mol. Struct. (THEOCHEM)* **2003**, *622*, 147−165. (c) Bradbury, S. P.; Mekenyan, O.; Ankley G. T. *Environ. Sci. Technol.* **1983**, *17*, 15−25.
(5) Lopez de Compadre, R. L.; Pearlstein, R. A.; Hopfinger, A. J.; Seydel, J. K. *Pharmacochemistry Library* **1987**, *10* (QSAR Drug Des. Toxicol.), 149−156.
(6) Additional parameters provided in Supporting Information. (a) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864−871. (b) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133−1138. (c) Vanderbilt D. *Phys. Rev. B* **1990**, *41*, 7892−7895. (d) Car, R.; Parrinello, M. *Phys. Rev. Lett.* **1985**, *55*, 2471−2474. (e) Giannozzi, P.; De Angelis, F.; Car, R. *J. Chem. Phys.* **2004**, *13*, 5903−5915. (f) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865−3868. (g) Baroni, S.; Dal Corso, A.; De Gironcoli, S.; Giannozzi, P. Plane-wave self-consistent field modeling package, http://www.pwscf.org.
(7) Schmitt, J. D. *Curr. Med. Chem.* **2000**, *7*, 749−800.
(8) Yoshida, F.; Topliss, J. G. *J. Med. Chem.* **2000**, *43*, 2575−2585.
(9) The coupling decreases as the inverse of the difference of the unperturbed eigenvalues.
(10) *Cerius*^{2}, release 4.9; Accelrys Inc.: San Diego, 2003.
(11) (a) R Development Core Team. *R, A language and enironment for statistical computing*; R Foundation for Statistical Computing: Vienna, Austria, 2004; http://www.R-project.org. (b) Lindberg, W.; Persson, J. A.; Wold, S. *Anal. Chem.* **1983**, *55*, 643−648.
(12) Chemical Computing Group, Inc. *MOE modeling environment*, release 2005.08.

JA057826R