articles

molecular pharmaceutics

Rapid Quantum Mechanical Models for the Computational Estimation of C−H Bond Dissociation Energies as a Measure of Metabolic Stability

John L. Lewin and Christopher J. Cramer*

Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received January 8, 2004

Abstract: Several relatively inexpensive levels of theory are surveyed together with alternative algorithmic methods for the estimation of C−H bond dissociation energies (BDEs), such energies being useful for the prediction of metabolic stability in drug-like molecules. In particular, bond stretching potentials of several C−H bonds are computed using the AM1, PM3, HF/MIDI!, and B3LYP/MIDI! levels of electronic structure theory, and selected points are fit to Morse and parabolic potentials. BDEs computed by an AM1 fit to the Morse function show the smallest mean unsigned error in prediction (±3−4 kcal/mol) over 32 diverse C−H bonds. An alternative method for correlating the AM1 parabolic force constant from a two-point unrelaxed potential provides only a slightly decreased accuracy and is computationally particularly inexpensive. Both methods should prove to be useful for the rapid *in silico* screening of drug-like molecules for metabolic stability to C−H bond oxidizing enzymes.

Keywords: C−H homolysis; bond energy; oxidation; bond stretching; metabolic stability

Introduction

Recent advances in synthetic chemistry, especially combinatorial methodology, together with increased market pressure associated with aging populations have led to increasing numbers of compounds entering drug discovery programs.[1] The cost of screening new compounds, however, remains high and may be increasing;[2] the cost of a late failure is particularly large, as clinical trials represent the most expensive phase of pharmaceutical research and development[2] (the average cost of introducing a drug to market has been estimated to be upward of $500 million with time frames on the order of a decade or more[3,4]). Coupled with such costs is the further estimation that "nearly 50% of drug candidates fail because of unacceptable efficacy", including "undesirable metabolic stability".[5] These exigencies have led pharmaceutical researchers to utilize ADME/Tox (absorption, distribution, metabolism, excretion, and toxicity) screening early in the drug discovery process to minimize undesirable properties and allow only candidates that pass such tests to be developed further.[2,3,5−7] For metabolic screening, pharmaceutical companies have developed high-throughput *in*

* To whom correspondence should be addressed. Phone: (612) 624-0859. Fax: (612) 626-2006. E-mail: cramer@chem.umn.edu.

(1) Wring, S. A.; Silver, I. S.; Serabjit-Singh, C. J. Automated Quantitative and Qualitative Analysis of Metabolic Stability: A Process for Compound Selection during Drug Discovery. *Methods Enzymol.* **2002**, *357*, 285−295.

(2) Stouch, T. R.; Kenyon, J. R.; Johnson, S. R.; Chen, X.-Q.; Doweyko, A.; Li, Y. *In silico* ADME/Tox: why models fail. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 83−92.

(3) Dearden, J. C. *In silico* prediction of drug toxicity. *J. Comput.-Aided Mol. Des.* **2003**, *17*, 119−127.

(4) Yevich, J. P. Drug Development: From Discovery to Marketing. In *A Textbook of Drug Design and Development*, 2nd ed.; Krogsgaard-Larsen, P., Liljefors, T., Madsen, U., Eds.; Harwood: Amsterdam, 1996; Chapter 18, pp 508−523.

(5) Li, A. P. Screening for human ADME/Tox drug properties in drug discovery. *Drug Discovery Today* **2001**, *6*, 357−366.

(6) Ekins, S.; Rose, J. In silicio ADME/Tox: the state of the art. *J. Mol. Graphics Modell.* **2002**, *20*, 305−309.

(7) Masimirembwa, C. M.; Ridderström, M.; Zamora, I.; Andersson, T. B. Combining Pharmacophore and Protein Modeling to Predict CYP450 Inhibitors and Substrates. *Methods Enzymol.* **2002**, *357*, 133−144.

10.1021/mp049977r CCC: $27.50
Published on Web 02/05/2004

vitro techniques to assay large numbers of compounds for stability, often using liver microsomes (homogenized liver fractions of smooth endoplasmic reticulum containing metabolizing enzymes such as cytochrome P450).[1,4,5,8] However, such methods can be time-consuming[9] and give false measurements, for example, the recent observation that some drugs can bind to the non-enzymatic portion of microsomal proteins under *in vitro* conditions, where non-specific binding leads to an inverse relationship between metabolic rate and microsomal protein concentration.[5] Alternative *in vivo* animal models also pose challenges, simply because they are inherently different from human systems and thus have the potential to give different metabolic descriptions for a compound of interest.[9]

In addition to high-throughput *in vitro* methods, there is growing acknowledgment of the potential utility of computational models for the prediction of ADME/Tox properties before a compound is ever synthesized, *in silico* techniques being particularly economical high-throughput assays.[2,5−7] Theoretical data may then guide synthetic efforts, such as the design of a combinatorial library or the prioritization of the preparation of compounds within one.[2,6]

One interesting computational goal is the estimation of homolytic carbon−hydrogen bond dissociation energies (BDEs),[10] as these play a role in dictating the susceptibility to cytochrome P450 oxidation in Phase I (oxidative) metabolism.[11−15] Cytochrome P450 is the most important participant in Phase I metabolism reactions, and is responsible for the majority of known drug biotransformations.[3,16−21] Cytochrome P450 acts on roughly 90% of all drugs in clinical use,[22] and ∼60% of the commonly prescribed drugs in the United States.[19] With the goal of finding a computationally inexpensive, rapid, and straightforward method, we survey a range of theoretical models for a molecular test set incorporating a variety of local chemical functionality. We identify two models that on average estimate homolytic C−H BDEs to within 3−4 kcal/mol with a minimum of computational investment and that should prove to be straightforward in automating and employing.

Computational Considerations and Models

Homolytic carbon−hydrogen BDEs are usually defined as the enthalpy change for the reaction

$$R{-}H \rightarrow R^{\bullet} + H^{\bullet} \quad (1)$$

which is the difference in the heats of formation between products and reactant at 298 K and 1 atm.[23,24] Thus

$$D_{RH} \equiv \Delta H^{\circ}_{f,298}[R^{\bullet}_{(g)}] + \Delta H^{\circ}_{f,298}[H^{\bullet}_{(g)}] - \Delta H^{\circ}_{f,298}[RH_{(g)}] \quad (2)$$

where D_{RH} is the bond dissociation energy. Equation 2 reflects how D_{RH} is determined in practice experimentally and sometimes computationally; i.e., the heat of formation at 298 K is measured or computed for the reactant and the radical products. However, the computation (or measurement for that matter) of accurate heats of formation, particularly for radicals, can be decidedly nontrivial. Accurate computation typically requires either expensive and possibly multideterminantal methods that extensively account for electron correlation or multilevel methods involving several sequential high-level calculations and corrections to the energy, such as the popular G3 or CBS theories.[25−29]

Such theoretical rigor is inconsistent with our goal of developing a *rapid and inexpensive* technique of estimating

(8) Kronbach, T. Hepatic Microsomes and Heterologous Expression Systems as In Vitro Models for Human Drug Metabolism. In *Pharmacokinetics: regulatory, industrial, academic perspectives*, 2nd ed.; Welling, P. G., Tse, F. L. S., Eds.; Dekker: New York, 1995; Chapter 10, pp 235−236.

(9) Propst, C. L.; Perun, T. J. Introduction to Computer-Aided Drug Design. In *Computer Aided Drug Design: methods and applications*; Perun, T. J., Propst, C. L., Eds.; Dekker: New York, 1989; Chapter 1, pp 1−16.

(10) Benson, S. W. *Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters*, 2nd ed.; Wiley: New York, 1976.

(11) Jones, J. P.; Korzekwa, K. R. Predicting the Rates and Regioselectivity of Reactions Mediated by the P450 Superfamily. *Methods Enzymol*. **1996**, *272*, 326−335.

(12) Ortiz de Montellano, P. R. The Cytochrome P450 Oxidative System. In *Handbook of Drug Metabolism*; Woolf, T. F., Ed.; Dekker: New York, 1999; Chapter 4, pp 109−130.

(13) Omura, T. Forty Years of Cytochrome P450. *Biochem. Biophys. Res. Commun.* **1999**, *266*, 690−698.

(14) Ewing, T. J. A.; Patel, P. I.; Tieu, H.; Korzekawa, K. R. Predicting metabolic stability of drug molecules. U.S. Patent Appl. 2001-811283 20010315, 2001 (continuation of U.S. Patent 613,875).

(15) Ewing, T. J. A.; Kocher, J.-P.; Tieu, H.; Korzekawa, K. R. Predicting susceptibility of reactive sites on molecules to metabolism and accessibility correction factors for electronic models of cytochrome P450 metabolism. U.S. Patent Appl. 2001-902470 20010709, 2002 (continuation of U.S. Patent 613,875).

(16) Gibson, G. G.; Skett, P. *Introduction to Drug Metabolism*, 2nd ed.; Chapman and Hall: London, 1994; pp 1−49.

(17) Mabic, S.; Castagnoli, K.; Castagnoli, N., Jr. Oxidative Metabolic Bioactivation of Xenobiotics. In *Handbook of Drug Metabolism*; Woolf, T. F., Ed.; Dekker: New York, 1999; Chapter 2, pp 49−79.

(18) Smith, D. A.; van de Waterbeemd, H.; Walker, D. K. *Pharmacokinetics and Metabolism in Drug Design*; Wiley-VCH: Weinheim, Germany, 2001; Chapter 7, pp 75−97.

(19) Venkatakrishnan, K.; von Moltke, L. L.; Greenblatt, D. J. Human Drug Metabolism and the Cytochromes P450: Application and Relevance of In Vitro Models. *J. Clin. Pharmacol.* **2001**, *41*, 1149−1179.

(20) McManus, M. E.; McKinnon, R. A. Measurement of Cytochrome P450 Activation of Xenobiotics Using the Ames *Salmonella* Test. *Methods Enzymol.* **1991**, *206*, 501−509.

(21) Anzenbacher, P.; Anzenbacherová, E. Cytochromes P450 and metabolism of xenobiotics. *Cell. Mol. Life Sci.* **2001**, *28*, 737−747.

(22) Lewis, D. F. V. *Cytochromes P450: structure, function and mechanism*; Taylor and Francis: London, 1996; Chapters 3 and 4, pp 79−167.

(23) McMillen, D. F.; Golden, D. M. Hydrocarbon Bond Dissociation Energies. *Annu. Rev. Phys. Chem.* **1982**, *33*, 493−532.

(24) Blanksby, S. J.; Ellison, G. B. Bond Dissociation Energies of Organic Molecules. *Acc. Chem. Res.* **2003**, *36*, 255−263.

C−H bond dissociation energies in drug-like molecules. For our purposes, we would prefer the speed inherent in restricted single-determinant theories, such as semiempirical and *ab initio* Hartree−Fock (HF) theory as well as density functional theory (DFT). Of these various levels, semiempirical calculations can be up to several orders of magnitude faster than DFT and *ab initio* HF theory.[30,31]

Restricted HF theory (and its DFT Kohn−Sham equivalent) expresses the wave function as a single Slater determinant composed of spin-paired electrons in self-consistently determined doubly occupied molecular orbitals.[31−34] While this adequately describes closed-shell, ground-state molecules, the restriction to doubly occupied orbitals becomes problematic in homolytic bond breaking, when the interatomic distance is stretched far beyond the equilibrium value. Ionic terms in the RHF wave function cause the stretching potential to be unphysically steep.[33,35] However, as this is an error that develops systematically along a bond dissociation coordinate, it is possible to correct for it in a systematic fashion. To that end, we here compute points on the bond dissociation curve relatively near the equilibrium structure for a variety of molecules and use these points to fit more global potentials from which D_{RH} can be estimated. This approach avoids the requirement for high levels of theory associated with the explicit modeling of open-shell species and the direct computation of accurate bond dissociation energies.

In this work, we employ semiempirical Austin Model 1[36] (AM1) and Parametric Method 3[37] (PM3). Both of these effective Hamiltonians are based on the neglect of differential diatomic overlap (NDDO) formalism in which various elements of the Fock matrix are either set to zero or computed using analytical functions that contain parameters optimized to reproduce various properties of ground-state, closed-shell neutral molecules, such as heats of formation, molecular geometries, ionization potentials, and dipole moments. These approximations greatly increase the speed of semiempirical HF calculations.

We also examine the *ab initio* restricted HF methodology in which all elements of the Fock matrix are computed in full. The restricted DFT method we employ is the hybrid B3LYP functional, comprised of Becke's B3 three-parameter gradient-corrected exchange functional[38,39] with the LYP correlation functional of Lee, Yang, and Parr[40] as originally described by Stevens et al.[41] Consistent with our goal of inexpensive calculations, both HF and B3LYP made use of the MIDI! basis set.[42] MIDI! was specifically designed to render calculations on large molecules more efficient. Although it is smaller than the popular 6-31G(d) basis set, it has been reported to yield more accurate geometries and charges as judged by comparison to MP2/cc-pVDZ calculations.[42]

(25) Lynch, B. J.; Truhlar, D. G. Robust and Affordable Multicoefficient Methods for Thermochemistry and Thermochemical Kinetics: The MCCM/3 Suite and SAC/3. *J. Phys. Chem. A* **2003**, *107*, 3898−3906.

(26) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. Gaussian-3 (G3) theory for molecules containing first and second-row atoms. *J. Chem. Phys.* **1998**, *109*, 7764−7776.

(27) Ochterski, J. W.; Petersson, J. A.; Montgomery, J. A. A complete basis set model chemistry. V. Extensions to six or more heavy atoms. *J. Chem. Phys.* **1996**, *104*, 2598−2619.

(28) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A fifth-order perturbation comparison of electron correlation theories. *Chem. Phys. Lett.* **1989**, *157*, 479−483.

(29) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **1987**, *87*, 5968−5975.

(30) Thiel, W. Thermochemistry from Semiempirical Molecular Orbital Theory. In *Computational Thermochemistry*; Irikura, K. K., Frurip, D. J., Eds.; ACS Symposium Series 677; American Chemical Society: Washington, DC, 1998; Chapter 8, pp 142−161.

(31) Cramer, C. J. *Essentials of Computational Chemistry*; Wiley: New York, 2002.

(32) Roothaan, C. C. J. New Developments in Molecular Orbital Theory. *Rev. Mod. Phys.* **1951**, *23*, 69−89.

(33) Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry*; Macmillan: New York, 1982.

(34) Koch, W.; Holthausen, M. C. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2000.

(35) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986; pp 29−32, 270−285.

(36) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(37) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method. *J. Comput. Chem.* **1989**, *10*, 209−220.

(38) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.

(39) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(40) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785−789.

(41) Stevens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *Ab Initio* Calculations of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623−11627.

(42) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges. *Theor. Chim. Acta* **1996**, *93*, 281−301.

(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A., Jr.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, revision A11; Gaussian, Inc.: Pittsburgh, PA, 2001.

Table 1. Experimental C−H BDEs for the Initial Hydrocarbon Test Set

molecule	BDE (kcal/mol)	ref
methane	104.99 ± 0.03	24
ethane	101.1 ± 0.4	24
propane (2°)	98.6 ± 0.4	24
tert-butane (3°)	96.5 ± 0.4	24
acetylene	133.32 ± 0.07	24
propene (allylic)	88.8 ± 0.4	24
ethene	110.7 ± 0.6	24
cyclopentane	97.6	44
cyclohexane	98.6	44
benzene	112.9 ± 0.5	24
toluene (benzylic)	89.8 ± 0.6	24
naphthalene (α)	112.2 ± 1.3	45
tetrahydronaphthalene (benzylic)	82.9 ± 1.2	44

All electronic structure calculations were performed using the *Gaussian98* (revision A11) suite of electronic structure programs.[43]

Protocols for Predicting BDEs

We began with the assembly of a small test set of 13 hydrocarbons for which high-quality experimental BDEs were available in an effort to test various computational protocols for their ability to generate useful bond stretching potentials. These hydrocarbons, together with their experimental BDEs, are listed in Table 1. The test set includes saturated, unsaturated, aromatic, allylic, and cyclic molecules having H atoms bound to formally sp, sp^2, aromatic, and sp^3 carbon types. The set moreover provides BDE values ranging over ∼50 kcal/mol and is thus reasonably robust in providing coverage of different bond strengths.

To begin, all structures in Table 1 were fully optimized at the AM1, PM3, HF/MIDI!, and B3LYP/MIDI! levels, and all structures were verified as local minima by subsequent computation of analytic frequencies at the same levels. Subsequently, additional energies were computed for each molecule at each level of theory with the target C−H bond either shortened or lengthened by increments of 0.05 Å so that pointwise stretching coordinates were generated over the range from $r = r_{eq} - 0.4$ Å to $r = r_{eq} + 0.4$ Å. These coordinates were generated both with (relaxed) and without (unrelaxed) reoptimization of the remaining geometric degrees of freedom.

Morse Potential Fits. Various points along these coordinates were then fit to a Morse potential, which has the form

$$E(r) = D_{RH}[1 - e^{-\alpha_{RH}(r - r_{eq})}]^2 \qquad (3)$$

where D_{RH} is the zero-point-excluding BDE and α_{RH} is a fitting constant.[31] We considered various fitting protocols. If three points are chosen, the D_{RH} and α_{RH} parameters may in general be determined to fit the chosen energies exactly. If more points are chosen, best fit values may be determined from standard statistical approaches. We examined nine different protocols for accomplishing the fits as outlined in Table 2. In eight of the protocols, three points were chosen, consisting of the equilibrium bond length (corresponding to a relative energy of zero) and two points symmetrically displaced to either side of the equilibrium bond length by differing multiples of 0.05 Å. We also considered one five-point fit.

Table 2. Points Used in Potential Fitting Protocols [listed by displacement (Å) from r_{eq}]

protocol	point set
3.1	−0.05, 0.00, 0.05
3.2	−0.10, 0.00, 0.10
3.3	−0.15, 0.00, 0.15
3.4	−0.20, 0.00, 0.20
3.5	−0.25, 0.00, 0.25
3.6	−0.30, 0.00, 0.30
3.7	−0.35, 0.00, 0.35
3.8	−0.40, 0.00, 0.40
5.1	−0.10, −0.05, 0.00, 0.05, 0.10

D_{RH} and α_{RH} were optimized by nonlinear fitting using standard NAG Fortran 90 routines;[47a] multiple solutions from different initial guesses were generated to ensure that the global minimum solution was reached. The Solver program[47b] in Microsoft Excel was also employed for this parameter

(44) Laarhoven, L. J. J.; Mulder, P. α-C−H Bond Strengths in Tetralin and THF: Application of Competition Experiments in Photoacoustic Calorimetry. *J. Phys. Chem. B* **1997**, *101*, 73−77.

(45) Reed, D. R.; Kass, S. R. Experimental determination of the α and β C−H bond dissociation energies in naphthalene. *J. Mass Spectrom.* **2000**, *35*, 534−539.

(46) (a) Morse, P. M. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Phys. Rev.* **1929**, *34*, 57−63. (b) Warshel, A. Calculation of the Anharmonicity in the Vibrational Frequencies of Alkane Molecules by the CFF Functions. *J. Chem. Phys*. **1971**, *55*, 3327−3335. (c) Warshel, A. *Computer Modeling of Chemical Reactions in Enzymes and Solutions*; Wiley-Interscience: New York, 1991; pp 18−22.

(47) (a) *NAG Fortran 90 Library fl90*, release 3; The Numerical Algorithms Group, Ltd.: Oxford, U.K., 1999. (b) *Solver*; Frontline Systems, Inc.: Incline Village, NV, 1990−1995. *Microsoft Excel 98*; Microsoft Corp.: Seattle, 1985−1998.

(48) Wenthold, P. G.; Squires, R. R. Gas-Phase Properties and Reactivity of the Acetate Radical Anion. Determination of the C−H Bond Strengths in Acetic Acid and Acetate Ion. *J. Am. Chem. Soc.* **1994**, *116*, 11890−11897.

(49) Holmes, J. L.; Lossing, F. P.; Mayer, P. M. Heats of Formation of Oxygen-Containing Organic Free Radicals from Appearance Energy Measurements. *J. Am. Chem. Soc.* **1991**, *113*, 9723−9728.

(50) Zhu, X.-Q.; Li, H.-R.; Li, Q.; Ai, T.; Lu, J.-Y.; Yang, Y.; Cheng, J.-P. Determination of the C4−H Bond Dissociation Energies of NADH Models and Their Radical Cations in Acetonitrile. *Chem. Eur. J.* **2003**, *9*, 871−880.

(51) Wayner, D. D. M.; Clark, K. B.; Rauk, A.; Yu, D.; Armstrong, D. A. C−H Bond Dissociation Energies of Alkyl Amines: Radical Structures and Stabilization Energies. *J. Am. Chem. Soc.* **1997**, *119*, 8925−8932.

(52) Griller, D.; Lossing, F. P. On the Thermochemistry of α-Aminoalkyl Radicals. *J. Am. Chem. Soc.* **1981**, *103*, 1586−1587.

(53) Lalevée, J.; Allonas, X.; Fouassier, J.-P. N−H and α(C−H) Bond Dissociation Enthalpies of Aliphatic Amines. *J. Am. Chem. Soc.* **2002**, *124*, 9613−9621.

fitting and provided essentially identical results. The fitted D_{RH} values were then linearly regressed on the experimental bond dissociation energies to assess the quality of the relationship. This linear regression corrects for systematic errors associated with individual levels of theory and also corrects for the zero-point-excluding nature of the Morse potential. In the latter case, the correction may be considered essentially a constant: the total variation in zero-point energies is $^1/_2hc\Delta\omega$,[31] where h is Planck's constant, c is the speed of light, and $\Delta\omega$ is the range of C−H stretching frequencies in wavenumber units; as this range is conservatively no more than 600 cm^{-1} for typical C−H bonds, the total variation in zero-point vibrational energies is less than 1 kcal/mol, on the order of the experimental error in the measurements in many instances.

Parabolic Fits. If we represent the bond stretching potential in the region of the equilibrium structure as a Taylor expansion in the bond length truncated at the second derivative, we have

$$E(r) = E(r_{eq}) + \left.\frac{dE}{dr}\right|_{r_{eq}} (r - r_{eq}) + \frac{1}{2!}\left.\frac{d^2E}{dr^2}\right|_{r_{eq}} (r - r_{eq})^2 \quad (4)$$

Noting that the first term on the right-hand side of eq 4 defines the relative energy of zero on the stretching potential and that the first derivative is zero as it is evaluated at the minimum on the potential coordinate, we can simplify eq 4 to the well-known harmonic oscillator result

$$E(r) = {}^1/_2k(r - r_{eq})^2 \quad (5)$$

where k is the harmonic force constant defined by reference to eq 4 as the second derivative of the energy with respect to motion along the stretching coordinate evaluated at the equilibrium position. If we consider E to have the form of the Morse potential in eq 3, we may also write

$$\left.\frac{d^2E_{Morse}}{dr^2}\right|_{r_{eq}} = 2D_{RH}\alpha_{RH}{}^2 \quad (6)$$

Thus, providing that the parameter α is not particularly sensitive to the nature of the R−H bond, we may expect a linear relationship between harmonic force constants k and BDEs.

To explore this relationship further, we fit the points used in protocols 3.1−3.8 to parabolas (analytic geometry permits an exact fit for any three non-collinear points) for both relaxed and unrelaxed stretching coordinates. We also considered the simpler protocol made possible if the equilibrium bond length is assumed to be the critical point of the parabola. In that case, only a single additional point is required to uniquely define the harmonic potential. We chose to use the forward point from each three-point protocol (i.e., the stretched point) for this purpose. Subsequent linear regression of the force constants on the experimental BDEs completes the definition of a fitting protocol.

Results and Discussion

A complete analysis of all of the statistical data associated with the various fitting protocols applied to the hydrocarbon test set is provided as Supporting Information. In the interest of brevity, we note here simply that protocol 3.6 proved to be robust for essentially every choice of theory, fitting potential, and relaxed versus unrelaxed stretching coordinate. The sensitivity to choice of protocol was rather low over the range from approximately protocol 3.4 to 3.7, particularly for the parabolic fits, but in general, protocol 3.6 tended to provide very slightly reduced errors compared to the others. For protocol 3.8, the rapid increase in energy associated with the repulsive side of the potential curve appeared to reduce the quality of BDE estimates, while for protocols 3.1−3.3, the small step sizes away from the equilibrium position failed to adequately reflect differences in BDEs for different C−H bonds. For the Morse potential fits, protocol 5.1 performed in a manner essentially equivalent to that of protocol 3.2, implying that the two additional points present in the former protocol were not particularly useful. As five-point protocols are obviously nearly twice as costly as three-point protocols in terms of computational resources, we did not consider them further. Note that for the two-point parabolic force constants, we refer to the protocol as protocol 3.6 even

Table 3. Performance of Protocol 3.6 for Correlation of Relaxed and Unrelaxed Morse Potential BDEs with Experiment for the Hydrocarbon Test Set[a]

statistic	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
pre-regression				
mse	74.5 (89.2)	24.3 (28.0)	3.3 (8.5)	10.4 (−5.1)
mue	74.5 (89.2)	24.3 (28.0)	3.8 (8.5)	10.4 (6.3)
rmse	74.8 (89.3)	25.1 (28.6)	5.1 (9.9)	10.9 (7.1)
post-regression				
slope	0.66 (1.04)	0.77 (0.90)	0.93 (1.10)	1.14 (1.31)
intercept	−15.2 (−96.4)	4.4 (−14.8)	3.6 (19.7)	−2.7 (−25.1)
R^2	0.91 (0.88)	0.79 (0.80)	0.91 (0.85)	0.94 (0.90)
mue	3.0 (3.9)	4.9 (4.8)	2.7 (4.1)	2.3 (3.5)
rmse	3.8 (4.4)	5.7 (5.6)	3.8 (4.8)	3.1 (3.9)

[a] mse, mean signed error (kcal/mol); mue, mean unsigned error (kcal/mol); rmse, root-mean-square error (kcal/mol); slope (unitless); intercept (kcal/mol); R, Pearson correlation coefficient (unitless). Results presented as relaxed (unrelaxed).

Table 4. Performance of Protocol 3.6 for Correlation of Relaxed and Unrelaxed Three-Point Parabolic Force Constants with Experiment for the Hydrocarbon Test Set[a]

statistic	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
slope	0.19 (0.22)	0.17 (0.15)	0.17 (0.17)	0.17 (0.18)
intercept	−62.3 (−90.0)	−38.6 (−23.2)	−73.1 (−80.5)	−63.8 (−72.2)
R^2	0.86 (0.84)	0.50 (0.37)	0.88 (0.86)	0.88 (0.86)
mue	4.0 (4.4)	7.8 (8.5)	3.7 (4.0)	3.7 (4.1)
rmse	4.7 (5.1)	8.9 (10.0)	4.4 (4.7)	4.3 (4.7)

[a] slope (Å²); intercept (kcal/mol); *R*, Pearson correlation coefficient (unitless); mue, mean unsigned error (kcal/mol); rmse, root-mean-square error (kcal/mol). Results presented as relaxed (unrelaxed).

Table 5. Performance of Protocol 3.6 for Correlation of Relaxed and Unrelaxed Two-Point Parabolic Force Constants with Experiment for the Hydrocarbon Test Set[a]

statistic	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
slope	0.30 (0.38)	0.30 (0.32)	0.28 (0.31)	0.32 (0.34)
intercept	−47.6 (−92.9)	−31.1 (−43.3)	−34.1 (−51.9)	−34.8 (−48.9)
R^2	0.89 (0.85)	0.66 (0.61)	0.90 (0.86)	0.93 (0.86)
mue	3.4 (4.2)	6.5 (6.9)	3.1 (4.1)	2.8 (4.1)
rmse	4.2 (4.8)	7.3 (7.9)	3.9 (4.7)	3.4 (4.7)

[a] slope (Å²); intercept (kcal/mol); *R*, Pearson correlation coefficient (unitless); mue, mean unsigned error (kcal/mol); rmse, root-mean-square error (kcal/mol). Results presented as relaxed (unrelaxed).

though only a single calculation of the energy for $r_{eq} + 0.3$ Å is required.

Tables 3−5 provide details about the quality of the protocol 3.6 fits for the hydrocarbon test set data for all levels of theory. In the case of the Morse potential (Table 3), errors in BDEs both before and after linear regression are presented. For the parabolic fits, the linear regression may be thought to change the units of the force constant from kilocalories per mole per angstrom squared to kilocalories per mole, so any preregression analysis is not meaningful (note that the SI unit for the force constant is newtons per centimeter, which is equal to the commonly tabulated cgs unit millidyne per angstrom; the conversion factor is 1 kcal mol^{-1} Å^{-2} = 0.00695 N cm^{-1}).

It is noteworthy in the Morse potential analysis that AM1 and PM3 do very badly in raw prediction of D_{RH}, AM1 particularly so. HF/MIDI!, on the other hand, provides surprisingly good estimates of the BDEs directly from D_{RH}, while B3LYP/MIDI! is less successful for a relaxed stretching coordinate and somewhat more successful with an unrelaxed coordinate. The large errors for AM1 appear to be highly systematic since after linear regression the model provides an accuracy approximately equal to that of HF/MIDI!. PM3 is markedly less improved after regression. The postregression B3LYP/MIDI! results are the best, with a relaxed rmse of only 3.1 kcal/mol (we note that the dispersion in the data is 12.6 kcal/mol). Of course, the DFT model is much more expensive (*vide infra*).

The three-point parabolic predictions are less accurate than the Morse potential predictions in every case, with the PM3 results being almost statistically meaningless. The difference between the quality of the relaxed and unrelaxed fits is smaller for the parabolic predictions than for the Morse potential predictions. Interestingly, the two-point parabolic predictions are in almost every case *more* accurate than the three-point predictions.

Having chosen a protocol on the basis of high-quality hydrocarbon data, we next expanded our test set by addition of 23 BDE data for more functionalized organic molecules

Table 6. Experimental C−H BDEs for Additional Organic Molecules

molecule	BDE (kcal/mol)	ref
acetaldehyde (methyl)	94 ± 2	24
acetic acid	94.6 ± 3.0	48, 49
1-benzyl-1,4-dihydronicotinamide[a]	67.9 ± <1	50
dimethylamine	87 ± 2.4	51
dimethyl ether	93 ± 2	52
ethanolamine (*N*-α)	90.7 ± 2	53
formaldehyde	88.144 ± 0.008	24
methanol	96.1 ± 0.2	24
methylamine	94.2	52, 54
methanethiol	94.2	24
morpholine (*N*-α)	93 ± 2.4	51
N,*N*-dimethylaniline (methyl)	91.7 ± 1.3	55
piperidine (*N*-α)	92 ± 2.4	51
propylamine (*N*-α)	93.1 ± 2	53
pyridine (2 position)	105 ± 2	56
pyridine (3 position)	112 ± 2	56
pyridine (4 position)	112 ± 2	56
pyrimidine (2 position)	98 ± 2	56
pyrimidine (4 position)	103 ± 2	56
pyrimidine (5 position)	112 ± 2	56
pyrrolidine (*N*-α)	90.1 ± 2.4	51
tetrahydrofuran (*O*-α)	92.1 ± 1.6	44
trimethylamine	86.5 ± 2.2	51, 52

[a] C(4)−H.

(54) Burkey, T. J.; Castelhano, A. L.; Griller, D.; Lossing, F. P. Heats of Formation and Ionization Potentials of Some α-Aminoalkyl Radicals. *J. Am. Chem. Soc.* **1983**, *105*, 4701−4703.

(55) Dombrowski, G. W.; Dinnocenzo, J. P.; Farid, S.; Goodman, J. L.; Gould, I. R. α-C−H Bond Dissociation Energies of Some Tertiary Amines. *J. Org. Chem.* **1999**, *64*, 427−431.

Table 7. Performance of Protocol 3.6 for Correlation of Relaxed and Unrelaxed Morse Potential BDEs with Experiment for Expanded Test Set[a]

statistic	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
slope	0.61 (0.78)	0.73 (0.75)	0.68 (0.89)	0.70 (0.90)
intercept	−6.9 (−46.1)	10.3 (4.0)	28.7 (1.8)	37.3 (14.5)
R^2	0.80 (0.69)	0.77 (0.64)	0.72 (0.67)	0.62 (0.66)
mue	3.4 (4.2)	4.5 (5.0)	4.9 (5.3)	5.6 (5.6)
rmse	5.0 (6.2)	5.3 (6.7)	5.9 (6.4)	6.9 (6.6)

[a] slope (unitless); intercept (kcal/mol); *R*, Pearson correlation coefficient (unitless); mue, mean unsigned error (kcal/mol); rmse, root-mean-square error (kcal/mol). Results presented as relaxed (unrelaxed).

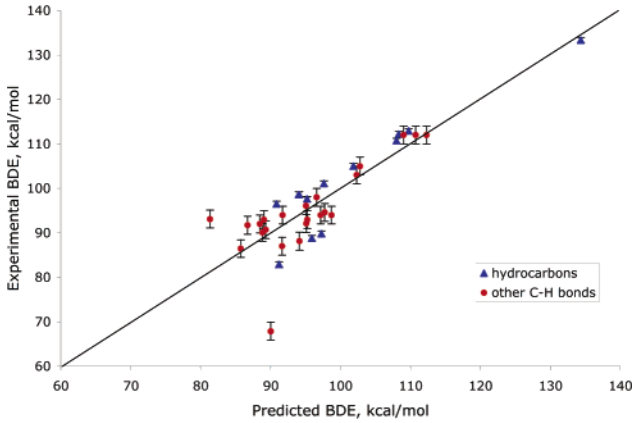

Figure 1. Predicted vs experimental BDE using the AM1 unrelaxed two-point parabola protocol.

(Table 6), for which experimental error bars were generally larger. We then employed exclusively protocol 3.6 to examine the extent to which the Morse and parabolic potential fits continued to be robust predictors of BDEs within the context of a more diverse test set.

As a technical point, we note that in constructing the stretching potentials for the compounds examined here, levels of theory occasionally disagreed on the global minimum energy conformer. For example, AM1 and PM3 predict the lowest-energy conformation of piperidine to have the N−H bond in the axial position while HF/MIDI! and B3LYP/MIDI! predict the same bond to prefer to be equatorially disposed. In such instances, the lowest-energy structure from a given level of theory was used irrespective of conflicting results from higher levels of theory (this being in the spirit of how one would expect to use a given model in practice, once defined).

Although we restricted our consideration of protocols to protocol 3.6 on the basis of our experience with the hydrocarbon test set, we carried out new linear regressions for the combined 38 data to evaluate the accuracy of our fits. The results are presented in Tables 7−9. Inclusion of the new data tended to degrade the predictive accuracy of the models based on fitting to the Morse potential, particularly for B3LYP/MIDI!. Part of this increase in error is associated with 1-benzyl-1,4-dihydronicotinamide, the BDE for which tended to be overestimated by all models by as much as 20 kcal/mol. Smaller systematic errors tended to be associated with underestimation of BDEs for allylic and benzylic C−H bonds. The relaxed and unrelaxed three-point parabolic fits also showed a loss of accuracy with the inclusion of the new test data, with the more expensive levels of theory showing the greatest loss in predictive accuracy. While some of this decrease in accuracy may be associated with greater experimental uncertainty in the new test data, it also appears that errors in the different protocols may be dependent on the nature of functionality present near the breaking C−H bond. That is, if there were to be sufficient data, it might prove to be a significant advantage to have separate regression equations for C−H bonds in hydrocarbons, C−H bonds α to nitrogens, etc. (Of course, in the limit

Table 8. Performance of Protocol 3.6 for Correlation of Relaxed and Unrelaxed Three-Point Parabolic Force Constants with Experiment for the Expanded Test Set[a]

statistic	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
slope	0.20 (0.23)	0.19 (0.15)	0.14 (0.15)	0.14 (0.15)
intercept	−70.4 (−100.1)	−57.5 (−27.2)	−53.0 (−60.9)	−31.8 (−41.8)
R^2	0.76 (0.73)	0.41 (0.25)	0.70 (0.67)	0.68 (0.68)
mue	4.0 (4.4)	7.4 (8.2)	5.1 (5.3)	5.4 (5.5)
rmse	5.5 (5.8)	8.6 (9.7)	6.2 (6.5)	6.4 (6.4)

[a] slope (Å^2); intercept (kcal/mol); *R*, Pearson correlation coefficient (unitless); mue, mean unsigned error (kcal/mol); rmse, root-mean-square error (kcal/mol). Results presented as relaxed (unrelaxed).

Table 9. Performance of Protocol 3.6 for Correlation of Relaxed and Unrelaxed Two-Point Parabolic Force Constants with Experiment for the Expanded Test Set[a]

statistic	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
slope	0.30 (0.37)	0.33 (0.33)	0.23 (0.26)	0.23 (0.26)
intercept	−48.1 (−89.3)	−43.6 (−48.4)	−11.1 (−29.9)	3.3 (−12.4)
R^2	0.79 (0.75)	0.66 (0.52)	0.74 (0.68)	0.68 (0.65)
mue	3.6 (4.1)	5.6 (6.4)	4.8 (5.3)	5.4 (5.8)
rmse	5.2 (5.6)	6.5 (7.8)	5.6 (6.4)	6.4 (6.7)

[a] slope (Å^2); intercept (kcal/mol); *R*, Pearson correlation coefficient (unitless); mue, mean unsigned error (kcal/mol); rmse, root-mean-square error (kcal/mol). Results presented as relaxed (unrelaxed).

Table 10. Computational Times (central processing unit seconds) Required To Compute Energies for Points on the *N*-α C−H Stretching Coordinate for Piperidine[a]

no. of points	AM1	PM3	HF/MIDI!	B3LYP/MIDI!
2	12 (5)	13 (6)	276 (33)	1783 (211)
3	19 (6)	16 (6)	376 (48)	2217 (312)

[a] Single IBM SP 375 MHz Power3 processor. Results presented as relaxed (unrelaxed).

of a *very* large set of experimental data, one might be able to avoid quantum chemical calculations altogether and find sufficiently accurate predictive models based on structure alone, e.g., graph-theoretical models, but the total number of accurately measured BDEs does not yet permit an evaluation of this point.)

Interestingly, the computationally simplest model, the two-point parabolic correlation, has predictive power over the full test set approximately as good as, if not better than, the Morse and three-point parabolic potential correlations. Indeed, with AM1 the performance of the two-point parabolic correlation is approximately as good over the full test set as over just the hydrocarbons. When unrelaxed geometries are used, the price/performance ratio of the AM1 model is quite impressive (Table 10 provides computational times for the two-point and three-point calculations required for the various models when applied to piperidine).

(56) Barckholtz, C.; Barckholtz, T. A.; Hadad, C. M. C−H and N−H Bond Dissociation Energies of Small Aromatic Hydrocarbons. *J. Am. Chem. Soc.* **1999**, *121*, 491−500.

(57) Beck, B.; Horn, A.; Carpenter, J. E.; Clark, T. Enhanced 3D-Databases: A Fully Electrostatic Database of AM1-Optimized Structures. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 1214−1217.

We conclude, then, that a particularly efficient method for estimating C−H BDEs is first to optimize the molecular structure with AM1 (we note that in 1998 Beck et al.[57] reported having carried out the AM1 optimization of a database of 53 000 compounds in 14 h on a 128-processor Origin 2000 computer). Next, every C−H bond of interest should be stretched by 0.3 Å and the energy computed without reoptimization (a fairly simple procedure to automate). From analytic geometry, the unrelaxed force constant for the bond is then 2/0.09 times the energy increase in units of kilocalories per mole per angstrom squared. Multiplication of this value by 0.37 $Å^2$ and subtraction of 89.3 will then provide an estimate of the BDE in kilocalories per mole with a mean unsigned error that, based on our 38-molecule test set, should be on the order of $\leq$4 kcal/mol. The performance of this model for the test set is depicted graphically in Figure 1. Such predicted BDEs may then prove to be useful in models correlating BDE with susceptibility to metabolic oxidation.

Acknowledgment. We thank Mr. Jason Thompson for technical assistance with some of the calculations and Prof. Tim Clark for stimulating discussion. Partial support for this work was provided by the National Science Foundation (Grant CHE-0203346).

Supporting Information Available: Statistical data for performance of various fitting protocols over the hydrocarbon test set. This material is available free of charge via the Internet at http://pubs.acs.org.

MP049977R