Szeged Matrix Property Indices as Descriptors to Characterize Fullerenes

Lorentz Jäntschi and Sorana D. Bolboacă

Supplementary material

Schematic flowchart of the applied experimental design is presented in Figure 1.

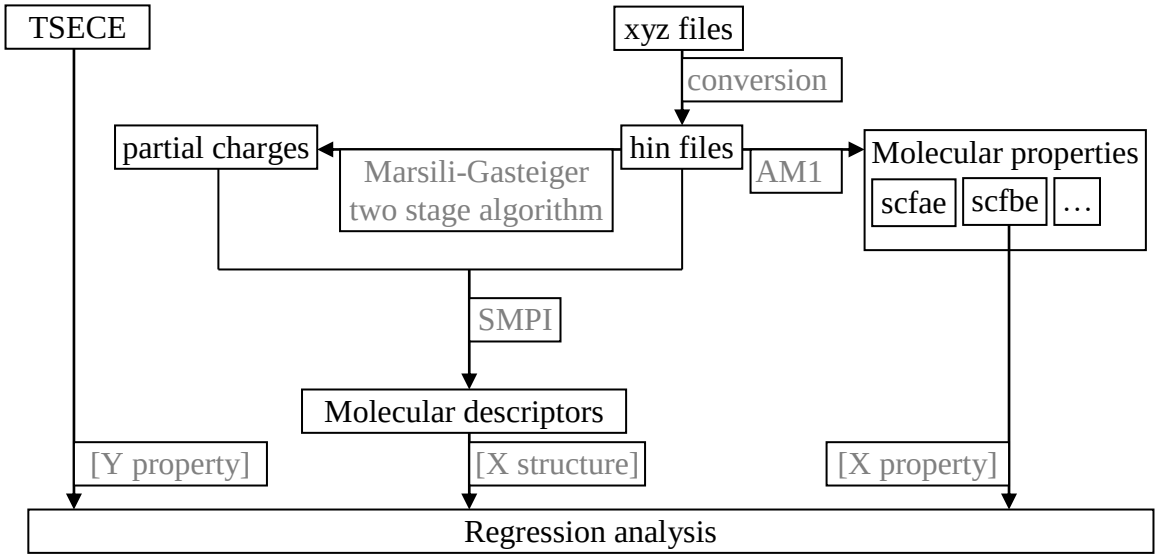

Figure 1. Summary of the applied design (TSECE = total strain energy (continuum elasticity), AM1 = Austin Model 1; scfbe = scf-binding-energy)

The results of each scenario are listed in Table 1 and Table 2, r^2 being the determination coefficient, and Q^2 being the determination coefficient in leave-one-out analysis, both expressed as percentages. The analysis of the results presented in Table 1 and 2 revealed that the total strain energy derived in the context of continuum elasticity theory (our Y) correlates high with a AM1 (Austin Model 1 [1], an accurate semi-empirical SCF method) calculated property, namely 'scfbe' (scf-binding-energy in the HyperChem formalism [2]) - over 94% of the variance explained (model with one variable in scenario 2). Identification of a model with a high goodness-of-fit as the first model in the second scenario could discourage further analysis, which it will be a mistake because the contribution of this property is discharge on model 17 in the second scenario.

[1] Dewar M. J. S., E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart. Development and use of quantum mechanical molecular models. 76. AM1: A new general purpose quantum mechanical molecular model. J. Am. Chem. Soc **107**, 3902 (1985).
[2] HyperChem(TM) Professional 7.51, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.

Table 1. First scenario regressions: structure-property analysis

Model ($\hat{Y}$ =)	r^2 (%)	Q^2 (%)
1v model		
Y=b0(t0)+b1(t1)*IJUGB	77.02	75.06
2v models		
4.70e+2(t=9.48e+0)+IEUUG*5.35e-5(t=3.09e+0)+IIUTG*-1.25e-4(t=1.12e+1)	77.62	71.84
4.96e+3(t=1.20e+1)+IEUUG* 1.62e-4(t=8.31e+0)+LIUTC*-6.14e+2(t=1.21e+1)	80.19	75.35
4.23e+2(t=1.06e+1)+IEUUG* 1.59e-4(t=8.75e+0)+IIUTC*-1.39e-1(t=1.30e+1)	82.41	78.69
-1.08e+2(t=4.85e+0)+IEUUG* 5.21e-5(t=3.40e+0)+IJUGB*5.26e+0(t=1.30e+1)	82.49	78.39
4.45e+2(t=1.13e+1)+IEUGG* 3.59e-4(t=8.97e+0)+IIUTC*-1.48e-1(t=1.34e+1)	82.97	79.25
-2.86e+3(t=3.93e+0)+LJUGG* 1.18e+2(t=3.89e+0)+IJUGB*7.37e+0(t=1.06e+1)	83.68	81.28
-4.42e+3(t=7.31e+0)+IIUGG*-7.23e-5(t=7.74e+0)+LJUGE*7.42e+2(t=8.00e+0)	85.04	77.84
-4.82e+2(t=4.89e+0)+IIUGG*-7.50e-5(t=9.18e+0)+IJUGE*1.37e+0(t=9.47e+0)	88.07	81.26
-5.29e+3(t=1.10e+1)+LJEGG* 3.88e+1(t=9.71e+0)+LJUGE*6.60e+2(t=8.01e+0)	88.97	85.31
-1.82e+3(t=1.95e+1)+LJEGG* 3.96e+1(t=1.17e+1)+IJUGE*1.22e+0(t=9.84e+0)	91.66	88.13
-3.39e+2(t=4.10e+0)+IIUGF*-2.60e-3(t=1.26e+1)+IJUGE*1.80e+0(t=1.70e+1)	92.59	84.58
-1.81e+3(t=1.87e+1)+RFUGE*-3.77e+5(t=1.15e+1)+IFEGE*4.91e+0(t=1.90e+1)	92.62	90.42
5.68e+3(t=1.63e+1)+RFUGE*-3.91e+5(t=1.17e+1)+RFDGA*-4.31e+5(t=1.91e+1)	92.67	90.86
5.27e+3(t=1.61e+1)+RJUGE*-6.06e+5(t=1.76e+1)+LIUGE*-6.54e+2(t=1.29e+1)	93.17	88.87
1.77e+3(t=2.66e+1)+RJUGE*-6.07e+5(t=2.04e+1)+IIUGE*-1.15e+0(t=1.53e+1)	94.88	92.04
3v models		
2.29e+3(t=2.50e+1)+IEUUG*-2.98e-5(t=3.38e+0)+RJUGE*-4.27e+5(t=1.37e+1)+IJUTE*-1.79e+0(t=1.57e+1)	95.58	93.20
2.04e+3(t=2.63e+1)+IEUUG*-3.81e-5(t=4.25e+0)+RJUGE*-4.51e+5(t=1.50e+1)+IIUTE*-1.76e+0(t=1.60e+1)	95.70	93.44
2.07e+3(t=1.77e+1)+LMUUG*-7.66e+0(t=2.98e+0)+RJUGE*-6.62e+5(t=2.01e+1)+IIUGE*-1.35e+0(t=1.40e+1)	95.89	90.94
2.08e+3(t=1.77e+1)+LMUUG*-7.89e+0(t=3.05e+0)+RJUGE*-6.74e+5(t=2.03e+1)+IIUGD*-1.50e-3(t=1.40e+1)	95.90	93.50
2.49e+3(t=1.69e+1)+LFUGG*-2.54e+1(t=4.66e+0)+RJUGE*-4.42e+5(t=1.50e+1)+IIUTE*-1.67e+0(t=1.72e+1)	95.97	94.08
1.49e+3(t=2.29e+1)+IIUGG*-1.95e-4(t=1.22e+1)+RJUGE*-5.80e+5(t=1.85e+1)+IJPTC*1.17e-3(t=7.85e+0)	96.11	94.71
-2.37e+3(t=6.26e+0)+LJEGG*5.98e+1(t=1.35e+1)+RJUGE*-3.74e+5(t=1.23e+1)+IEUGD*1.35e-2(t=5.48e+0)	96.13	94.42
-1.45e+3(t=7.65e+0)+IEUGF*2.95e-2(t=6.16e+0)+IIUGF*-4.01e-3(t=1.48e+1)+IJUGE*1.95e+0(t=2.47e+1)	96.39	84.96
1.04e+3(t=7.85e+0)+IEUGF*2.27e-2(t=5.91e+0)+RJUGE*-6.40e+5(t=2.88e+1)+IIUGE*-1.61e+0(t=1.69e+1)	97.40	95.73
1.02e+3(t=7.81e+0)+IEUGF*2.37e-2(t=6.24e+0)+RJUGE*-6.53e+5(t=2.98e+1)+IIUGD*-1.80e-3(t=1.73e+1)	97.52	96.04
-9.29e+2(t=4.59e+0)+IEUGF*3.20e-2(t=8.26e+0)+RJUGE*-4.17e+5(t=1.94e+1)+IIEGA*2.83e-4(t=1.89e+1)	97.87	96.93
3.01e+3(t=1.15e+1)+IEEGF*-2.57e-2(t=8.82e+0)+RJUGE*-4.96e+5(t=2.43e+1)+IIEGA*3.24e-4(t=1.78e+1)	98.05	97.01
2.50e+3(t=1.28e+1)+RJUGE*-5.99e+5(t=2.46e+1)+IFEGE*-2.22e+0(t=9.31e+0)+IIEGA*3.73e-4(t=1.68e+1)	98.19	97.48
4v models		
-1.49e+3(t=1.84e+1)+LMUUG*-4.96e+0(t=2.59e+0)+IFUGE*3.34e+0(t=8.58e+0)+IJUGE*1.23e+0(t=1.82e+1)+IIEGA*2.92e-4(t=2.22e+1)	98.21	97.69
2.43e+3(t=1.28e+1)+LMUUG*-3.81e+0(t=2.16e+0)+RJUGE*-6.03e+5(t=0.59e+1)+IFEGE*-2.06e+0(t=8.60e+0)+IIEGA*3.75e-4(t=1.77e+1)	98.40	97.63
3.96e+3(t=1.41e+1)+IIUGG*-5.57e-5(t=9.22e+0)+IEEGF*-1.80e-2(t=7.74e+0)+RJUGE*-4.60e+5(t=2.47e+1)+IJUTE*-1.58e+0(t=1.41e+1)	98.42	97.40
2.58e+3(t=2.33e+1)+IIUGG*-4.87e-5(t=9.13e+0)+RFUGE*-1.02e+5(t=8.21e+0)+RJUGE*-3.95e+5(t=2.12e+1)+IJUTE*-1.44e+0(t=1.43e+1)	98.54	97.81
3.73e+3(t=1.58e+1)+IIUGG*-6.10e-5(t=9.95e+0)+RJUGE*-5.25e+5(t=2.56e+1)+IFEGE*-1.49e+0(t=8.26e+0)+IJUTE*-1.75e+0(t=1.48e+1)	98.55	97.85
2.50e+3(t=1.41e+1)+ImUGG*1.00e-3(t=2.99e+0)+RJUGE*-6.23e+5(t=2.66e+1)+IFEGE*-2.18e+0(t=1.01e+1)+IIEGA*3.91e-4(t=1.86e+1)	98.56	98.10
2.44e+3(t=1.45e+1)+LMEGG*-3.10e+0(t=3.72e+0)+RJUGE*-6.20e+5(t=2.86e+1)+IFEGE*-2.05e+0(t=9.78e+0)+IIEGA*3.90e-4(t=1.99e+1)	98.70	98.33
IFUGE*2.71e+0(t=1.04e+1)+IJUGE*2.58e+0(t=3.04e+1)+IIUGE*-3.08e+0(t=2.48e+1)+IIPTB*-5.32e-1(t=1.46e+1)	98.76	98.30

The most important point related to the results presented in the Supplementary Material is that the SMPI performs better in the estimation of the total strain energy compared with scf-binding-energy on MRL models with three or four variables. Of course, the expected results, which profs the reproducibility, is that the 'best to moment' equations from one scenario to another are changed (different) only at the beginning when the supplementary descriptors - calculated properties - make their room for describing the association with the total strain energy. Adding the calculated properties in structure-property-property analysis decrease the number of equations with improvement in r^2 and converged to identical models for the last four models with three variables and all equations with four variables.

Table 2. Second scenario: structure-property-property analysis

Model ($\hat{Y}$ =)	r^2 (%)	Q^2 (%)
1v model		
b0(t0)+b1(t1)*scfbe	94.94	94.12
2v models		
2.09e+2(t=7.06e+0)+LMUUG*-5.28e+0(t=2.16e+0)+scfbe*2.33e-2(t=2.06e+1)	95.51	94.42
1.95e+2(t=9.33e+0)+LMEGG*-3.01e+0(t=2.39e+0)+scfbe*2.40e-2(t=1.84e+1)	95.62	94.66
1.64e+2(t=1.97e+1)+IMEGG*-2.91e-5(t=2.52e+0)+scfbe*2.42e-2(t=1.83e+1)	95.68	94.74
-3.84e+2(t=2.54e+0)+IEUGF*1.15e-2(t=3.50e+0)+scfbe*2.35e-2(t=2.59e+1)	96.20	94.65
4.06e+2(t=5.66e+0)+RFUGE*-5.06e+4(t=3.63e+0)+scfbe*2.32e-2(t=2.74e+1)	96.27	94.80
3v models		
LJUGG*-2.28e+1(t=3.44e+0)+IEUGF*1.48e-2(t=4.38e+0)+scfbe*2.29e-2(t=3.19e+1)	96.62	95.39
1.03e+3(t=3.53e+0)+LJUGG*-2.35e+1(t=2.20e+0)+RFUGE*-6.52e+4(t=4.40e+0)+scfbe*2.25e-2(t=2.59e+1)	96.71	95.03
-6.10e+2(t=3.79e+0)+IIUGG*-2.26e-5(t=2.80e+0)+IEUGF*1.79e-2(t=4.73e+0)+scfbe*2.16e-2(t=2.04e+1)	96.88	95.49
6.37e+2(t=6.44e+0)+IIUGG*-2.47e-5(t=3.09e+0)+RFUGE*-8.13e+4(t=5.07e+0)+scfbe*2.10e-2(t=2.03e+1)	97.05	95.78
LJEGG*2.17e+1(t=7.24e+0)+RFUGE*-1.01e+5(t=5.65e+0)+scfbe*1.84e-2(t=2.86e+1)	97.12	96.53
-1.55e+3(t=4.70e+0)+LIEGG*4.30e+1(t=3.84e+0)+IEUGF*1.98e-2(t=5.59e+0)+scfbe*2.12e-2(t=2.17e+1)	97.30	96.42
6.88e+2(t=3.53e+0)+LIEGG*5.31e+1(t=4.33e+0)+IEEGF*-1.63e-2(t=5.67e+0)+scfbe*2.30e-2(t=2.24e+1)	97.34	96.67
LIEGG*3.20e+1(t=7.90e+0)+RFUGE*-8.48e+4(t=5.92e+0)+scfbe*2.16e-2(t=3.69e+1)	97.41	96.85
3.32e+2(t=2.03e+0)+LIEGG*4.69e+1(t=4.20e+0)+REUGE*-1.12e+5(t=5.94e+0)+scfbe*2.12e-2(t=2.24e+1)	97.46	93.80
-1.69e+3(t=5.09e+0)+LIEGG*4.69e+1(t=4.22e+0)+IEUGD*1.03e-2(t=5.97e+0)+scfbe*2.11e-2(t=2.23e+1)	97.47	96.80
1.02e+3(t=7.81e+0)+IEUGF*2.37e-2(t=6.24e+0)+RJUGE*-6.53e+5(t=2.98e+1)+IIUGD*-1.80e-3(t=1.73e+1)	97.52	96.06
-9.29e+2(t=4.59e+0)+IEUGF*3.20e-2(t=8.26e+0)+RJUGE*-4.17e+5(t=1.94e+1)+IIEGA*2.83e-4(t=1.89e+1)	97.87	96.93
3.01e+3(t=1.15e+1)+IEEGF*-2.57e-2(t=8.82e+0)+RJUGE*-4.96e+5(t=2.43e+1)+IIEGA*3.24e-4(t=1.78e+1)	98.05	97.03
2.50e+3(t=1.28e+1)+RJUGE*-5.99e+5(t=2.46e+1)+IFEGE*-2.22e+0(t=9.31e+0)+IIEGA*3.73e-4(t=1.68e+1)	98.19	97.49
4v models		
-1.49e+3(t=1.84e+1)+LMUUG*-4.96e+0(t=2.59e+0)+IFUGE*3.34e+0(t=8.58e+0)+IJUGE*1.23e+0(t=1.82e+1)+IIEGA*2.92e-4(t=2.22e+1)	98.21	97.69
2.43e+3(t=1.28e+1)+LMUUG*-3.81e+0(t=2.16e+0)+RJUGE*-6.03e+5(t=2.59e+1)+IFEGE*-2.06e+0(t=8.60e+0)+IIEGA*3.75e-4(t=1.77e+1)	98.40	97.64
3.96e+3(t=1.41e+1)+IIUGG*-5.57e-5(t=9.22e+0)+IEEGF*-1.80e-2(t=7.74e+0)+RJUGE*-4.60e+5(t=2.47e+1)+IJUTE*-1.58e+0(t=1.41e+1)	98.42	97.42
2.58e+3(t=2.33e+1)+IIUGG*-4.87e-5(t=9.13e+0)+RFUGE*-1.02e+5(t=8.21e+0)+RJUGE*-3.95e+5(t=2.12e+1)+IJUTE*-1.44e+0(t=1.43e+1)	98.54	97.82
3.73e+3(t=1.58e+1)+IIUGG*-6.10e-5(t=9.95e+0)+RJUGE*-5.25e+5(t=2.56e+1)+IFEGE*-1.49e+0(t=8.26e+0)+IJUTE*-1.75e+0(t= .48e+1)	98.55	97.86
2.50e+3(t=1.41e+1)+ImUGG* 1.00e-3(t=2.99e+0)+RJUGE*-6.23e+5(t=2.66e+1)IFEGE*-2.18e+0(t=1.01e+1)+IIEGA*3.91e-4(t=1.86e+1)	98.56	98.10
2.44e+3(t=1.45e+1)+LMEGG*-3.10e+0(t=3.72e+0)+RJUGE*-6.20e+5(t=2.86e+1)+IFEGE*-2.05e+0(t=9.78e+0)+IIEGA*3.90e-4(t=1.99e+1)	98.70	98.33
IFUGE*2.71e+0(t=1.04e+1)+IJUGE*2.58e+0(t=3.04e+1)+IIUGE*-3.08e+0(t=2.48e+1)+IIPTB*-5.32e-1(t=1.46e+1)	98.76	98.30