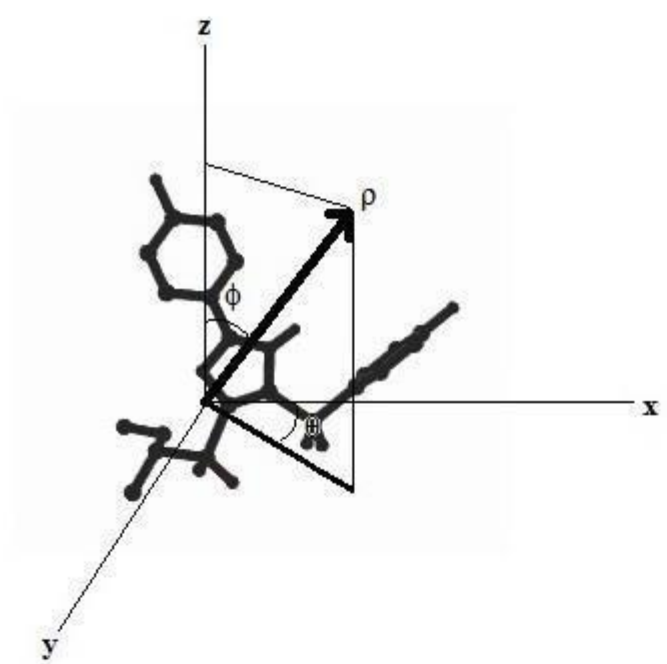
GEE Descriptors

Robert L. Jackson, Stilwell, Kansas

Abstract:

This paper introduces a new set of descriptors for usage in cheminformatics.

<u>A gradient of Electrostatic [Potential] Energy, or GEE, Descriptors.</u> Consider a molecule so either Cartesian or spherical coordinates can be used to describe the atoms present in the compound:



.

Electrostatic potential for an atom directly correlates with or is proportional to the atom's electronegativity:

$$V_{E,i} \propto \chi_i,$$

where $V_{E,i}$ is the electrostatic potential of atom E and χ is the electronegativity of atom i. Thus, the electrostatic potential energy for atom i, or E_ι, can be expressed as the following:

$$E_i \propto \chi_i \rho_i,$$

where ρ_i is the magnitude of the distance of atom i energy from an origin, or the center of the molecule. The above would suggest the total electrostatic energy of a molecule about its center would be defined as:

$$E = \sum_i \chi_i \rho_i.$$

The gradient of electrostatic potential energy about the center of a molecule would be as follows:

$$\nabla E = \nabla \sum_i \chi_i \rho_i,$$

where the ∇ is the del operator. After the application of the chain rule of differentiation, one obtains:

$$\nabla E = \nabla \sum_i \chi_i \rho_i$$

$$= \sum_i \rho_i \nabla \chi_i + \sum_i \chi_i \nabla \rho_i$$

Since the electronegativity of an atom is relatively constant, the left term of the above expression drops out leaving:

$$. \quad \nabla E = \sum_i \chi_i \nabla \rho_i.$$

Assuming one is working with spherical coordinates, the gradient of ρ_i will reduce to:

$$\nabla \rho_i = \frac{1}{\rho_i} \left\langle x_i, y_i, z_i \right\rangle,$$

where x_i, y_i, and z_i are the Cartesian coordinates of atom i from the center of the molecule.

Substituting in the spherical coordinates associated with the Cartesian system:

$$x = \rho \cos\theta \sin\phi,$$

$$y = \rho \sin\theta \sin\phi,$$

$$z = \rho \cos\phi.$$

The gradient of ρ_i simplifies to:

$$\nabla \rho_i = \left\langle \cos\theta_i \sin\phi_i, \sin\theta_i \sin\phi_i, \cos\phi_i \right\rangle.$$

Knowing θ and ϕ are defined as:

$$\theta = arc\tan\left(\frac{y}{x}\right),$$
$$\phi = arc\cos\left(\frac{z}{\rho}\right).$$

then the gradient of ρ_i becomes:

$$\nabla \rho_i = \left\langle \cos\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \sin\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \cos\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right)\right\rangle,$$
$$= \left\langle \cos\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \sin\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \frac{z_i}{\rho_i}\right\rangle.$$

Thus, the gradient of the electrostatic potential energy about the center of the molecule is given as the following expression:

$$\nabla E = \sum_i \chi_i \left\langle \cos\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \sin\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \cos\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right)\right\rangle$$

Next, one can define a set of descriptors based upon the distance from the center of a compound:

$$GEE_j(\rho) = \sum_i \sum_j \left\{u_{\rho_j}(\rho) - u_{\rho_j+40}(\rho)\right\}\chi_i \left\langle \cos\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \sin\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right), \frac{z_i}{\rho_i}\right\rangle$$
,

which can be reduced to Cartesian components:

$$GEE_{j,x}(\rho) = \sum_i \sum_j \left\{u_{\rho_j}(\rho) - u_{\rho_j+40}(\rho)\right\}\chi_i \cos\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right),$$
$$GEE_{j,y}(\rho) = \sum_i \sum_j \left\{u_{\rho_j}(\rho) - u_{\rho_j+40}(\rho)\right\}\chi_i \sin\left(arc\tan\left(\frac{y_i}{x_i}\right)\right)\sin\left(arc\cos\left(\frac{z_i}{\rho_i}\right)\right),$$
$$GEE_{j,z}(\rho) = \sum_i \sum_j \left\{u_{\rho_j}(\rho) - u_{\rho_j+40}(\rho)\right\}\chi_i \frac{z_i}{\rho_i}.$$

Finally, the net GEE descriptors for ρ at j would be defined as the following expression:

$$GEE_j(\rho) = \sqrt{GEE_{j,x}^2(\rho) + GEE_{j,y}^2(\rho) + GEE_{j,z}^2(\rho)}\ .$$