PowerMVDescriptorsV01

PowerMV will compute a number of molecular descriptors.

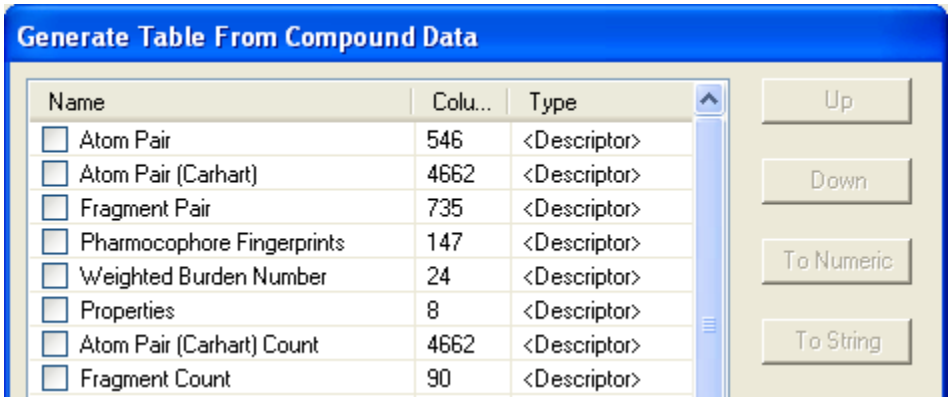
For a description of PowerMV and many of the descriptors that PowerVM can compute, see

J. Chem. Inf. Model. **2005,** *45,* 515−522 **515**

PowerMV: A Software Environment for Molecular Viewing, Descriptor Generation, Data Analysis and Hit Evaluation

Kejun Liu,† Jun Feng,† and S. Stanley Young*

Eight molecular descriptor sets can be computed using PowerMV. Four are bit string, one is continuous, one is a collection useful for judging the drug-like nature of a molecule and one gives simple counts of 90 atom types.



For the bit string descriptors, each bit is set to "1" when a certain feature is presented and "0" when it is not. For **Atom Pair** and **Atom Pair(Carhart)** we adopt the Carhart strategy where a feature refers to two chemical groups or atoms

Table 1. List of the Carhart Atom Types Used in PowerMV

C(1,0)	C(2,0)	C(3,0)
C(4,0)	C(1,1)	C(2,1)
C(3,1)	C(1,2)	C(2,2)
O(1,0)	O(2,0)	O(1,1)
O(2,1)	N(1,0)	N(2,0)
N(3,0)	N(4,0)	N(1,1)
N(2,1)	N(3,1)	S(1,0)
S(2,0)	S(2,1)	S(1,1)
S(3,1)	S(4,2)	F
Cl	Br	I
P(4,1)	Si	B
Se	As	Y

separated by a certain 2D path length, the bond count of the shortest path between the two atom types. The atoms are typed in the following way: The atom symbol is give, e.g. C for carbon, O for oxygen, N for nitrogen, etc.; next is given the number of non-

hydrogen connections of the atom; finally the number of pi electrons. So C(1,0) refers to a carbon, connected to one non-hydrogen, having no pi electrons. C(1,0) stands for -CH_3. Halogen atoms only have one possibility, (1,0), in organic molecules, so their extended notation is ignored. All undefined atom features are assigned to feature Y. If longer paths are counted then we go from 546 (paths up to seven bonds) features to 4662 features (**paths up to xx bonds**). For **Atom Pair** and **Atom Pair(Carhart)**, **Atom Pair**, we simply note the absence or presence, 0/1, of the feature. Note that we deviate from the original Carhart in noting presence/absence rather than using counts of features. For **Atom Pair (count)** we give the counts of the features in a molecule. For **Fragment Fingerprints** we replace atom types with groups of atoms, Table 2,

Table 2. Fragment-Based Descriptors

5 or 6-member aromatic rings	primary, secondary, and tertiary amine, with positive charge or not
carboxylic acid, sulfinate, sulfone	trifluromethyl, nitrile, nitro, sulfonamide,
halogen	double bond center
triple bond center	disulfide, ester
hydroxyl, thiol, primary amine	aliphatic ring centers
amide, carboxylic ester	ketone, imine, thione
ethyl	methyl

and again count the shortest through-bond distance between the pharmacophores. For fragment-based descriptors, 14 classes are defined. For example, two phenyl rings, which are separated by two bonds, are expressed as AR_02_AR.

Pharmacophore Fingerprint descriptors were built based on bioisosteric principles (Two atoms or groups that are expected to have roughly the same biological effect are called bioisosteres.). For example, the disulfide (-S-) is often used to replace ester group (-O-), so we assign these two groups to the same type. This type of thinking leads to our pharmacophore-based descriptors, giving six classes; see Table 3.

Table 3. Pharmacophore-Based Descriptors

An atom bearing a formal negative charge or groups such as carboxylic, sulfinic, tetrazole, and phosphinic acids
An atom bearing a formal positive charge or groups such as nitrogen in primary, secondary and tertiary amines
Hydrogen bond donor, oxygen or nitrogen atom with hydrogen attached
Hydrogen bond acceptor, oxygen or nitrogen atom with a lone pair electron
Aromatic center, any five- or six-member aromatic ring system
Hydrophobic center, a fragment in which most atoms are hydrophobic atoms, like aliphatic carbon ring systems or aliphatic carbon chains with few heteroatom substitutions

The continuous descriptors we implemented are a variation on the Burden number. We place one of three properties on the diagonal of the Burden connectivity matrix: electro negativity, Gasteiger partial charge or atomic lipophilicity, XLogP. It is common to scale the off diagonal elements of the connectivity matrix before computing eigen values. The off-diagonal elements were weighted by one of the following values: 2.5, 5.0, 7.5 or 10.0. We use the largest and smallest eigen values. This procedure gives us a total of 24 numerical descriptors. Our procedure is similar to the method used by Dr. Pearlman calculating his BCUT descriptors. Dragon software also has Burden Number inspired eigen value descriptors. All three methods are computed somewhat differently, but all are inspired by Burden.

The **Properties** set of descriptors include eight descriptors useful for judging the drug-like nature of a molecule, XlogP (a measure of the propensity of a molecule to partition into water or oil), polar surface area, PSA, number of rotatable bonds, H-bond donors, H-bond acceptors, molecular weight, blood-brain indicator (0 does not go into the brain, 1, goes into the brain) and bad group indicator (the molecule contains a chemically reactive or toxic group). These properties are useful for judging the drug-like nature of a molecule.

The **Fragment Count** gives the counts of 90 different typed atoms. These counts were very useful for a regression model for water solubility and are expected to be useful for other molecular physical properties.