

Appendix 1.

Selection of a basis set

Models of a chemical system generally consist of the combination of a theoretical method such as restricted HF and B3LYP with a basis set. Each such unique pairing of method with basis set represents a different approximation to the Schrödinger equation. *Choosing a model of a chemical system, almost always involves a trade-off between accuracy and computational cost.* More accurate methods and larger basis sets make jobs run longer.

A basis set is a mathematical description of the orbitals within a system used to perform the theoretical calculation. Larger basis sets approximate more accurately the orbitals by imposing fewer restrictions on the locations of the electrons in space.

Standard basis sets for electronic structure calculations use linear combinations of gaussian functions to form the orbitals. Basis sets assign a group of basis functions to each atom within a molecule to approximate its orbitals. These basis functions themselves are composed of a linear combination of gaussian functions; such basis functions are referred to as *contracted functions*, and the component gaussian functions are referred to as *primitives*. A basis function consisting of a single gaussian function is termed *uncontracted*.

Cartesian Gaussian functions are of the form

$$g_{ijk} = Nx^i y^j z^k e^{-\alpha r^2}$$

where i, j and k are nonnegative integrals, α is a positive orbital exponent and N is a normalization constant.

- **Minimal basis sets**

Minimal basis sets contain the minimum number of basis functions that are needed for each atom. For example,

H: 1s

C: 1s, 2s, 2p_x, 2p_y, 2p_z

Minimal basis sets use fixed size atomic type orbitals. The **STO-3G** basis set is a minimal basis set (through it is not the smallest possible basis set). It uses three

gaussian primitives per basis function (“3G”). “STO” stands for “Slater type orbitals”, and the STO-3G basis set approximates Slater orbitals with gaussian functions. STO has the form

$$Nr^{n-1}e^{-\xi r}Y_l^m(\theta, \phi)$$

where $Y_l^m(\theta, \phi)$ are the spherical harmonic.

- **Split valence basis sets**

One way to increase the size of a basis set is to take more basis functions per atom. ***Split valence basis sets***, such as 3-21G and 6-31G basis sets, have two (or more) sizes of basis function for each valence orbital. For example,

H: 1s, 1s'

C: 1s, 2s, 2s', 2p_x, 2p_y, 2p_z, 2p_x', 2p_y', 2p_z'

where the primed and unprimed orbitals differ in size. The ***double zeta basis sets*** form all molecular orbitals from linear combinations of two sizes of functions for each atomic orbital. Similarly, ***triple split valence basis sets*** such as 6-311G, use three sizes of contracted functions for each orbital type.

- **Polarized basis sets**

Split valence basis sets allow orbitals to change size, but not to change shape. ***Polarized basis sets*** remove this limitation by adding orbitals with angular momentum beyond what is required for the ground state to the description of each atom. For example, polarized basis sets add *d*-functions to carbon atoms and some of them add *p*-functions to hydrogen atoms. Examples for polarized basis sets are the 6-31G(d) and the 6-311G(d, p) basis sets.

- **Diffused functions**

Diffuse functions are large size versions of *s*- and *p*-type functions. They allow orbitals to occupy a larger region of space. Basis sets with diffuse functions are important for systems where electrons may be far from the nucleus. One example for diffuse basis function is the 6-311+G(d, p) basis set (see table).

If the name of the basis set is of the form:

$$n_1 - n_2 n_3 n_4 G$$

where n_1, n_2, n_3 and n_4 natural numbers.

For atoms with an atomic number larger than two, the $1s$ function is constructed from n_1 gaussian functions. Each of the $2s$ and $2p$ functions are constructed from n_2 gaussian functions. Each of the $3s$ and $3p$ functions are constructed from n_3 gaussian functions and each of the $4s$ and $4p$ functions are constructed from n_4 gaussian functions. However, for hydrogen and helium, the $1s$ functions are constructed from n_2 gaussian functions. The $2s$ functions are constructed from n_3 gaussian functions. The $3s$ functions are constructed from n_4 gaussian functions.

If the basis function is of the form

$$n_1 - n_2 n_3 n_4 G^* \text{ or } n_1 - n_2 n_3 n_4 G(d)$$

For atoms with atomic number larger than two, we add d functions (each constructed from one gaussian function).

If the basis function is of the form

$$n_1 - n_2 n_3 n_4 G^{**} \text{ or } n_1 - n_2 n_3 n_4 G(d, p)$$

we add also p function to the hydrogen and helium atoms.

Examples:

STO-3G – each occupied atomic orbital of each atom is constructed from three gaussian functions.

3-21G

For atoms with atomic number larger than two,

$1s$ – 3 gaussians

$2s$ and $2p$ – 2 gaussians

$3s$ and $3p$ – 1 gaussian

For hydrogen and helium,

$1s$ – 2 gaussians

$2s$ – 1 gaussian

6-311G*

For atoms with atomic number larger than two,

$1s$ – 6 gaussians

$2s$ and $2p$ - 3 gaussians

$3s$ and $3p$ – 1 gaussian

$4s$ and $4p$ – 1 gaussian

d – 1 gaussian

For hydrogen and helium,

$1s$ – 3 gaussians

$2s$ – 1 gaussian

$3s$ – 1 gaussian

6-311G**

As above, in addition we add p function for hydrogen and helium atoms.

The table below describes some common basis sets

| Basis set [Applicable atoms] | Description | # basis functions for first row atoms | # basis functions for hydrogen atoms | Default function types |
|------------------------------------|--|---|--|------------------------------|
| STO-3G [H – Xe] | Minimal basis set: use for more qualitative results on very large systems, when you cannot afford even 3-21G. | 5 | 1 | 6D |
| 3-21G [H – Xe] | Split valence: two sets of functions in the valence region provide a more accurate representation of orbitals. Use for very large molecules for which 6-31G(d) is too expensive. | 9 | 2 | 6D |
| 6-31G(d) [H – Cl] | Adds polarization functions to heavy atoms: use for most jobs on up to medium/large sized systems. (This basis set uses the 6-component type <i>d</i> functions) | 15 | 5 | 6D, 7F |
| 6-311+G(d, p) [H – Br] | Triple zeta: adds extra valence functions (three sizes of <i>s</i> - and <i>p</i> -functions). | 22 | 6 | 5D, 7F |

6D denotes Cartesian, 6-component *d*-functions, 5D and 7F denote 5-component *d*-functions and 7-component *f*-functions, respectively.