

## Chemistry 1001

### The Periodic Table for Macromolecular Crystallographers and Structural Bioinformatics

In actual practice, structural biology needs to consider only a restricted set of elements:

Hydrogen	Carbon	Nitrogen	Oxygen
		Phosphorus	Sulfur

and recognize the existence of ionic metals, as well as a few special cases of other elements.

Chemists often consider an atom to be a ball, with standard radius and color (radii are listed in Appendix 3).

- C Carbon, black or green
- N Nitrogen, blue
- O Oxygen, red
- P Phosphorus, pale yellow
- S Sulfur, bright yellow
- H Hydrogen, white or gray

Crystallography is an imaging process, and as such, sees existence (of electrons) and shapes. So geometry of interactions is the main aspect that we will be able to observe. Fortunately, local geometry is dominated by the nature of the individual atoms and the connections between them.

Carbon has two important geometries: tetrahedral with 4 surrounding close neighbors, and planar with 3 surrounding neighbors (trigonal). (Those of you with some chemistry background might recognize  $sp^3$  and  $sp^2$  hybridization.)

Nitrogen and Oxygen also have both tetrahedral and planar aspects, but sometimes, some directions seem unoccupied (unshared pairs of electrons occupy binding directions, but are difficult to observe directly.)

Phosphorus and Sulfur are dominantly tetrahedral; Phosphorus often has four obvious neighbors, but for Sulfur either 4 or only 2 are actually atoms. However, the arrangement of neighbors is close to the corners of a tetrahedron.

Hydrogen can only make one strong bond, so is usually an end twig on the branching tree framework of a covalently connected molecule. Of course, it is these hydrogen "twigs" that stick out from the bulk of a multi-atom molecule and so it is the hydrogens that make most of the contacts between molecules and molecular units. In particular, "polar" hydrogens (those bonded to O, N, or S) can interact with certain other "acceptor" near neighbors (O, N, S, aromatic-ring faces, etc). These "Hydrogen-bonds" are weaker and less restricted in their geometry than covalent bonds, but are the architects of secondary structure (helices and sheets in proteins, base pairs in nucleic acids).

So with some imagination for the seemingly empty directions and hydrogens sticking out like twigs, macromolecules can be described as combinations of tetrahedral bits and planar bits with "covalent" connections between them. Even the mutual arrangement of further covalent neighbors is dominated by the nature of the two central atoms, thus "dihedral" angles (the mutual angle of two other atoms connected to each of a central pair) have preferred values determined by those central two atoms. Macromolecules are surprisingly well relaxed into preferred dihedral positions, so that for instance dihedral angles around two atoms that are both tetrahedral will nearly always be near  $-60^\circ$ ,  $+60^\circ$ , or  $180^\circ$  (minus, plus, or trans and "staggered" rather than "eclipsed").

X-rays interact with the electrons of the atoms, so x-ray crystallography gives a picture of the electron density of the molecule. Atoms are seen in direct proportion to their number of electrons, so C,N,O are close to the same density, while H is markedly weaker. Indeed, even that one electron of Hydrogen does not diffract x-rays as efficiently as other electrons, so Hydrogen contributes relatively much less to the image of electron density. Hence, crystallographers have traditionally ignored Hydrogen in their constructed models, even though Hydrogens, by their chemical nature of being at the edges of connected sets of atoms, are the dominant atom type making contact in packing and molecular interactions.

Geometric values:

Bond lengths (approx; vary some with context):

- C-C 1.5
- C=C (arom) 1.3
- C=O 1.2
- C-H 1.1
- N-H,O-H 1.0
- S-H 1.3

Van der Waals radii:

- H (non-polar) 1.15
- H (polar) 0.8-1.0
- C 1.75
- C (carbonyl)1.65
- N 1.55
- O 1.4
- P 1.8
- S 1.8

Bond angle examples (from Engh & Huber 1991, Acta Cryst A47:392):

- (exact tetrahedron:  $109.6^\circ$ )
- Cbeta-Calpha-C:  $110.1^\circ$
- N-Calpha-C:  $111.2^\circ$  "tau" (somewhat wider variability)
- Cbeta-Calpha-N:  $110.5^\circ$
- (exact trigonal:  $120^\circ$ )
- Calpha-N-C:  $116.2^\circ$
- N-C-O:  $123.0^\circ$