

22. MOLECULAR GEOMETRY AND FEATURES

Table 22.1.4.1

Standard atomic radii (Å)

For '*' see following notes on specific sets of values. *Bondi*: Values assigned on the basis of observed packing in condensed phases (Bondi, 1968). *Lee & Richards*: Values adapted from Bondi (1964) and used in Lee & Richards (1971). *Shrake & Rupley*: Values taken from Pauling (1960) and used in Shrake & Rupley (1973). >C= value can be either 1.5 or 1.85. *Richards*: Minor modification of the original Bondi set in Richards (1974). (Rationale not given.) See original paper for discussion of aromatic carbon value. *Chothia*: From packing in amino-acid crystal structures. Used in Chothia (1975). *Richmond & Richards*: No rationale given for values used in Richmond & Richards (1978). *Gelin & Karplus*: Origin of values not specified. Used in Gelin & Karplus (1979). *Dunfield et al.*: Detailed description of deconvolution of molecular crystal energies. Values represent one-half of the heavy-atom separation at the minimum of the Lennard–Jones 6–12 potential functions for symmetrical interactions. Used in Nemethy *et al.* (1983) and Dunfield *et al.* (1979). *ENCAD*: A set of radii, derived in Gerstein *et al.* (1995), based solely on the *ENCAD* molecular dynamics potential function in Levitt *et al.* (1995). To determine these radii, the separation at which the 6–12 Lennard–Jones interaction energy between equivalent atoms was $0.25 k_B T$ was determined ($0.15 \text{ kcal mol}^{-1}$; $1 \text{ kcal} = 4.184 \text{ kJ}$). *CHARMM*: Determined in the same way as the *ENCAD* set, but for the *CHARMM* potential (Brooks *et al.*, 1983) (parameter set 19). *Tsai et al.*: Values derived from a new analysis (Tsai *et al.*, 1999) of the most common distances of approach of atoms in the Cambridge Structural Database.

Atom type and symbol	Bondi (1968)	Lee & Richards (1971)	Shrake & Rupley (1973)	Richards (1974)	Chothia (1975)	Richmond & Richards (1978)	Gelin & Karplus (1979)	Dunfield <i>et al.</i> (1979)	ENCAD derived (1995)	CHARMM derived (1995)	Tsai <i>et al.</i> (1999)	
–CH ₃	Aliphatic, methyl	2.00	1.80	2.00	2.00	1.87	1.90	1.95	2.13	1.82	1.88	1.88
–CH ₂ –	Aliphatic, methyl	2.00	1.80	2.00	2.00	1.87	1.90	1.90	2.23	1.82	1.88	1.88
>CH–	Aliphatic, CH	–	1.70	2.00	2.00	1.87	1.90	1.85	2.38	1.82	1.88	1.88
≥CH=	Aromatic, CH	–	1.80	1.85	*	1.76	1.70	1.90	2.10	1.74	1.80	1.76
>C=	Trigonal, aromatic	1.74	1.80	*	1.70	1.76	1.70	1.80	1.85	1.74	1.80	1.61
–NH ₃ ⁺	Amino, protonated	–	1.80	1.50	2.00	1.50	0.70	1.75	–	1.68	1.40	1.64
–NH ₂	Amino or amide	1.75	1.80	1.50	–	1.65	1.70	1.70	–	1.68	1.40	1.64
>NH	Peptide, NH or N	1.65	1.52	1.40	1.70	1.65	1.70	1.65	1.75	1.68	1.40	1.64
=O	Carbonyl oxygen	1.50	1.80	1.40	1.40	1.40	1.40	1.60	1.56	1.34	1.38	1.42
–OH	Alcoholic hydroxyl	–	1.80	1.40	1.60	1.40	1.40	1.70	–	1.54	1.53	1.46
–OM	Carboxyl oxygen	–	1.80	1.89	1.50	1.40	1.40	1.60	1.62	1.34	1.41	1.42
–SH	Sulfhydryl	–	1.80	1.85	–	1.85	1.80	1.90	–	1.82	1.56	1.77
–S–	Thioether or –S–S–	1.80	–	–	1.80	1.85	1.80	1.90	2.08	1.82	1.56	1.77

accessible surface of the same atoms. In complementary fashion, the re-entrant surface is also not closed but is a series of concave patches that is part of the probe surface where it contacts two or three atoms simultaneously. At infinite probe radius, the re-entrant areas are plane surfaces, at which point the molecular surface becomes a convex surface. The re-entrant surface cannot be divided up and assigned unambiguously to individual atoms. Note that the molecular surface is simply the union of the contact and re-entrant surfaces, so in terms of area $MS = CS + RS$.

22.1.3.3.4. Further points

The detail provided by these surfaces will depend on the radius of the probe used for their construction.

One may argue that the behaviour of the rolling probe sphere does not accurately model real hydrogen-bonded water. Instead, its 'rolling' more closely mimics the behaviour of a nonpolar solvent. An attempt has been made to incorporate more realistic hydrogen-bonding behaviour into the probe sphere, allowing for the definition of a hydration surface more closely linked to the behaviour of real water (Gerstein & Lynden-Bell, 1993c).

The definitions of accessible surface and molecular surface can be related back to the Voronoi construction. The molecular surface is similar to 'time-averaging' the surface formed from the faces of Voronoi polyhedra (the Voronoi surface) over many water configurations, and the accessible surface is similar to averaging the Delaunay triangulation of the first layer of water molecules over many configurations.

There are a number of other definitions of protein surfaces that are unrelated to either the probe-sphere method or Voronoi polyhedra and provide complementary information (Kuhn *et al.*, 1992; Leicester *et al.*, 1988; Pattabiraman *et al.*, 1995).

22.1.4. Definitions of atomic radii

The definition of protein surfaces and volumes depends greatly on the values chosen for various parameters of linear dimension – in particular, van der Waals and probe-sphere radii.

22.1.4.1. van der Waals radii

For all the calculations outlined above, the hard-sphere approximation is used for the atoms. (One must remember that in reality atoms are neither hard nor spherical, but this approximation has a long history of demonstrated utility.) There are many lists of the radii of such spheres prepared by different laboratories, both for single atoms and for unified atoms, where the radii are adjusted to approximate the joint size of the heavy atom and its bonded hydrogen atoms (clearly not an actual spherical unit).

Some of these lists are reproduced in Table 22.1.4.1. They are derived from a variety of approaches, *e.g.* looking for the distances of closest approach between atoms (the Bondi set) and energy calculations (the *CHARMM* set). The differences between the sets often come down to how one decides to truncate the Lennard–Jones potential function. Further differences arise

22.1. PROTEIN VOLUMES, AREAS AND DISTANCES

Table 22.1.4.2

Probe radii and their relation to surface definition

The values of 1.4 and, especially, 10 Å are only approximate. One could, of course, use 1.5 Å for a water radius or 15 Å for a ligand radius, depending on the specific application.

Probe radius (Å)	Part of probe sphere	Type of surface
0	Centre (or tangent)	van der Waals surface (VDWS)
1.4	Centre	Solvent-accessible surface (SAS)
1.4	Tangent (one atom)	Contact surface (CS, from parts of atoms)
1.4	Tangent (two or three atoms)	Re-entrant surface (RS, from parts of probe)
1.4	Tangent (one, two, or three atoms)	Molecular surface (MS = CS + RS)
10	Centre	A ligand- or reagent-accessible surface
∞	Tangent	Minimum limit of MS (related to convex hull)
∞	Centre	Undefined

from the parameterization of water and other hydrogen-bonding molecules, as these substances really should be represented with two radii, one for their hydrogen-bonding interactions and one for their VDW interactions.

Perhaps because of the complexities in defining VDW parameters, there are some great differences in Table 22.1.4.1. For instance, the radius for an aliphatic CH (>CH=) ranges from 1.7 to 2.38 Å, and the radius for carboxyl oxygen ranges from 1.34 to 1.89 Å. Both of these represent at least a 40% variation. Moreover, such differences are practically quite significant, since many geometrical and energetic calculations are very sensitive to the choice of VDW parameters, particularly the relative values within a single list. (Repulsive core interactions, in fact, vary almost exponentially.) Consequently, proper volume and surface comparisons can only be based on numbers derived through use of the same list of radii.

In the last column of the table we give a recent set of VDW radii that has been carefully optimized for use in volume and packing calculations. It is derived from analysis of the most

common distances between atoms in small-molecule crystal structures in the Cambridge Structural Database (Rowland & Taylor, 1996; Tsai *et al.*, 1999).

22.1.4.2. The probe radius

A series of surfaces can be described by using a probe sphere with a specified radius. Since this is to be a convenient mathematical construct in calculation, any numerical value may be chosen with no necessary relation to physical reality. Some commonly used examples are listed in Table 22.1.4.2.

The solvent-accessible surface is intended to be a close approximation to what a water molecule as a probe might 'see' (Lee & Richards, 1971). However, there is no uniform agreement on what the proper water radius should be. Usually it is chosen to be about 1.4 Å.

22.1.5. Application of geometry calculations: the measurement of packing

22.1.5.1. Using volume to measure packing efficiency

Volume calculations are principally applied in measuring packing. This is because the packing efficiency of a given atom is simply the ratio of the space it could minimally occupy to the space that it actually does occupy. As shown in Fig. 22.1.5.1, this ratio can be expressed as the VDW volume of an atom divided by its Voronoi volume (Richards, 1974, 1985; Richards & Lim, 1994). (Packing efficiency also sometimes goes by the equivalent terms 'packing density' or 'packing coefficient'.) This simple definition masks considerable complexities – in particular, how does one determine the volume of the VDW envelope (Petitjean, 1994)? This requires knowledge of what the VDW radii of atoms are, a

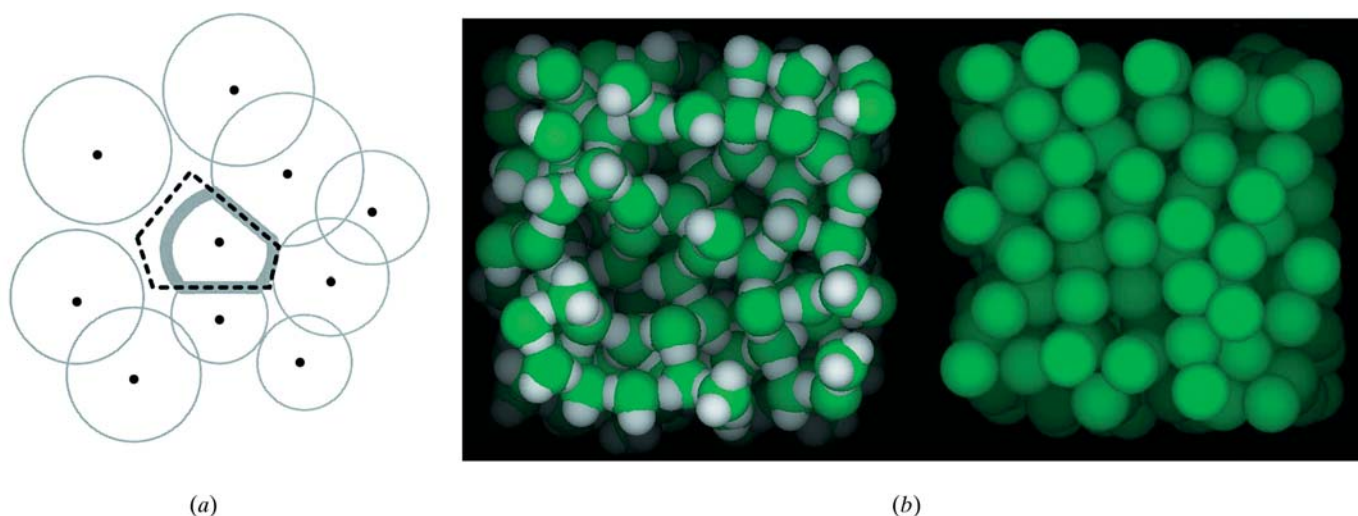


Figure 22.1.5.1

Packing efficiency. (a) The relationship between Voronoi polyhedra and packing efficiency. Packing efficiency is defined as the volume of an object as a fraction of the space that it occupies. (It is also known as the 'packing coefficient' or 'packing density'.) In the context of molecular structure, it is measured by the ratio of the VDW volume (V_{VDW} , shown by a light grey line) and Voronoi volume (V_{Vor} , shown by a dotted line). This calculation gives absolute packing efficiencies. In practice, one usually measures a relative efficiency, relative to the atom in a reference state: $(V_{VDW}/V_{Vor})/[V_{VDW}/V_{Vor}(\text{ref})]$. Note that in this ratio the unchanging VDW volume of an atom cancels out, leaving one with just a ratio of two Voronoi volumes. Perhaps more usefully, when one is trying to evaluate the packing efficiency P at an interface, one computes $P = p \sum V_i / \sum v_i$, where p is packing efficiency of the reference data set (usually 0.74), V_i is the actual measured volume of each atom i at the interface and v_i is the reference volume corresponding to the type of atom i . (b) A graphical illustration of the difference between tight packing and loose packing. Frames from a simulation are shown for liquid water (left) and for liquid argon, a simple liquid (right). Owing to its hydrogen bonds, water is much less tightly packed than argon (packing efficiency of 0.35 versus ~0.7). Each water molecule has only four to five nearest neighbours while each argon atom has about ten.