Maximum allowed solvent accessibilites of residues in proteins

Supporting Information

Matthew A. Tien, Austin G. Meyer, Stephanie J. Spielman, Claus O. Wilke

Supporting Figures

Figure S1. Distribution of RSA values for alanine. RSA was calculated using our theoretically determined normalization values. This distribution is highly non-normal with a strong right skew. Therefore, mean RSA is a poor measure of center for this distribution. Similarly skewed distributions are found for most amino acids.

$1 - 5$

-5

Supporting Tables

Table S1. Backbone conformation of maximally exposed trimer structures. Multiple rows per residue indicate alternative conformations with comparable solvent exposure.

Residue	Empirical		Theoretical	
	ϕ	ψ	ϕ	ψ
Alanine	$-66.7°$	$-13.1°$	$-60°$	$-15°$
	$-52.1°$	$-33.6°$		
	$-51.9°$	$-37.9°$		
Arginine	$-79.2°$	$-20.2°$	$-70°$	$-5°$
			$-70°$	$-25°$
			$-60°$	$-15°$ to $-10°$
			$-55°$	$-30°$
			$-40°$	$-50°$
Asparagine	$-94.8°$	$-3.4°$	$-50°$	$-40°$
			$-50°$	$-30°$
			$70°$	$-5°$
Aspartate	$-79.4°$	$83.0°$	$70°$	$-5°$
Cysteine	$-87.7°$	$-45.1°$	$60°$	$-40°$
Glutamine	$-65.5°$	$-24.0°$	$70°$	$-5°$
Glutamate	$-55.0°$	$-49.9°$	$-60°$	$-15°$
Glycine	$80.2°$	$7.2°$	$-75°$	$20°$
			$-75°$	$50°$
			$-70°$	$0°$
			$-65°$	$-10°$
			$-60°$	$-15°$
			$-50°$	$-25°$
			$70°$	$-15°$
			$75°$	$-30°$
Histidine	$51.2°$	$32.4°$	$-180°$	$155°$
			$-80°$	$170°$ to $175°$
			$-80°$ to $-75°$	$130°$
Isoleucine	$-64.1°$	$-21.9°$	$-55°$	$-25°$
			$-50°$	$-40°$
Leucine	$-81.6°$	$-13.5°$	$-70°$	$-5°$
	$-63.5°$	$-44.5°$		
	$-60.9°$	$-29.3°$		
	$-55.8°$	$-36.4°$		
Lysine	$63.1°$	$23.1°$	$-45°$	$-45°$ to $-40°$
Methionine	$-67.5°$	$-27.5°$	$50°$	$-40°$
Phenylalanine	$-50.3°$	$135.2°$	$-45°$	$-40°$
			$70°$	$-10°$
Proline	$-63.8°$	$-21.6°$	$-55°$	$-20°$
Serine	$-58.1°$	$-27.3°$	$65°$	$-45°$
	$-103.4°$	$1.1°$		
Threonine	$-57.4°$	$-17.3°$	$-45°$	$-45°$
Tryptophan	$-68.0°$	$-62.3°$	$65°$	$-40°$ to $-35°$
			$70°$	$-55°$
Tyrosine	$-67.6°$	$-9.8°$	$-50°$	$-40°$ to $-35°$
			$-45°$	$-45°$
Valine	$-56.2°$	$-31.7°$	$-55°$	$-25°$

Table S2. Hydrophobicity scales derived in this work.

Amino Acid	Mean RSA (theor)[a]	Mean RSA (emp)[b]	100% buried[c]	95% buried[d]
Alanine	0.796	0.782	0.228	0.399
Arginine	0.651	0.639	0.0121	0.0749
Asparagine	0.672	0.658	0.0451	0.146
Asparate	0.646	0.634	0.0276	0.104
Cysteine	0.911	0.899	0.287	0.576
Glutamine	0.654	0.636	0.0289	0.109
Glutamate	0.605	0.589	0.0183	0.0717
Glycine	0.749	0.731	0.166	0.291
Histidine	0.723	0.731	0.0532	0.188
Isoleucine	0.876	0.875	0.247	0.516
Leucine	0.861	0.853	0.213	0.486
Lysine	0.565	0.554	0.00597	0.0283
Methionine	0.856	0.841	0.217	0.484
Phenylalanine	0.87	0.864	0.186	0.483
Proline	0.669	0.658	0.0607	0.162
Serine	0.744	0.722	0.105	0.241
Threonine	0.742	0.728	0.0987	0.237
Tryptophan	0.849	0.837	0.0979	0.368
Tyrosine	0.818	0.813	0.0797	0.306
Valine	0.864	0.857	0.25	0.494

[a]Scale based on mean RSA, as calculated using the theoretially derived SA normalization values. The actual scale is defined as $1-$(mean RSA), to yield increasingly larger values for more hydrophobic residues.

[b]Same scale as in (a), but calculated using the empirically derived SA normalization values.

[c]Fraction of 100% buried residues, with RSA $= 0$ (corresponding to SA < 1Å).

[d]Fraction of 95% buried residues, with RSA < 0.05.

Supporting Text

Exhaustive surveying of model tripeptides

To find the theoretical maximum solvent accessibility (SA) for each amino acid X, we computationally constructed Gly-X-Gly tripeptides. Each tripeptide was modeled by specifying coordinates of each constituent atom using bond lengths and angles from our empirically mined protein structures. Once constructed, we exhaustively rotated ϕ and ψ dihedral backbone angles in discrete 1° increments, holding ω constant at 180°. For each (ϕ, ψ) combination, we additionally rotated through all possible χ rotamer angles, as found in the Dunbruck Rotamer Database [?]. Rotamer angles were grouped into three 120° sectors (60°, -60°, and 180°) and averaged within each sector. For amino acids where the side chain could assume more than ten distinct rotamer conformations (e.g. for L, I, M, K, N), we selected ten rotamer conformations at random instead of exhaustively enumerating all rotamer conformations. A different set of randomly chosen rotamer conformations was generated for each combination of (ϕ, ψ) angles.

For each tripeptide conformation examined, a corresponding PDB file was created and inputted into the program DSSP [?] to compute the SA of amino acid X. For each (ϕ, ψ) combination, we recorded the largest SA value from all rotamer variations examined.

Tripeptide construction

We construct tripeptides by placing atoms one-by-one at the correct location in 3D space. We always begin with the N-terminus residue, which we place at the origin: The α carbon is placed at coordinates $(0, 0, 0)$, and the carbonyl carbon is placed at $(1.52, 0, 0)$, reflecting the 1.52Å bond length between a carbonyl group and a carbon atom. Next, the nitrogen atom is placed at $(\ell \cos\theta, \ell \sin\theta, 0)$, where ℓ is the bond length and θ is the bond angle. To ensure that the constructed residue is in the L-conformation, the nitrogen atom is rotated positively from the x-axis.

All subsequent atoms are placed using the following procedure: We identify three reference atoms, which we refer to as A, B, and C. We translate these points such that C lies at the origin. We then calculate the coordinates (x, y, z) of a fourth atom X at a bond length ℓ from C, a bond-angle θ relative to the CB vector (which we refer to as $\mathbf{b}$), and a dihedral angle τ relative to the ABC plane (see figure on next page). In this calculation, we proceed in two steps: We initially place the atom at the correct distance from C and correct bond angle relative to the CB vector, but with an arbitrary dihedral angle. We then rotate the atom around the CB vector to place it at the correct dihedral angle.

To achieve the initial placement, we solve a system of three equations, corresponding to the three conditions that the distance from C to X should be ℓ, the angle between CB and CX be θ, and that X be in a plane perpendicular to ABC:

$$x^2 + y^2 + z^2 = \ell^2, \tag{1}$$

$$\mathbf{b} \cdot (x, y, z) = \ell ||\mathbf{b}|| \cos\theta, \tag{2}$$

$$Ix + Jy + Kz = 0. \tag{3}$$

Schematic drawing of the placement of an atom relative to three reference atoms A, B, and C. Atom C is assumed to be at the origin. The vector from C to B is denoted by $\mathbf{b}$, with coordinates

$(x_{\mathbf{b}}, y_{\mathbf{b}}, z_{\mathbf{b}})$. Similarly, the vector from C to A is denoted by $\mathbf{a}$, with coordinates $(x_{\mathbf{a}}, y_{\mathbf{a}}, z_{\mathbf{a}})$. The distance from C to X is ℓ, the angle between CB and CX is θ, and the dihedral angle between ABC and CX is τ.

Here, the constants I, J, and K are coefficients of the target plane, and are obtained from the cross product of vectors $\mathbf{a}$ and $\mathbf{b}$:

$$
\begin{align}
I &= (y_{\mathbf{a}}z_{\mathbf{b}}) - (z_{\mathbf{a}}y_{\mathbf{b}}), \tag{4} \\
J &= (z_{\mathbf{a}}x_{\mathbf{b}}) - (x_{\mathbf{a}}z_{\mathbf{b}}), \tag{5} \\
K &= (x_{\mathbf{a}}y_{\mathbf{b}}) - (y_{\mathbf{a}}x_{\mathbf{b}}). \tag{6}
\end{align}
$$

Solving for x, y, and z yields the following expressions:

$$
x = \frac{R - IJP y_{\mathbf{b}} + PK(Kx_{\mathbf{b}} - Iz_{\mathbf{b}}) + J^2 Px_{\mathbf{b}}}{Q}, \tag{7}
$$

$$
y = \frac{-I(Rz_{\mathbf{b}} + J^2 Px_{\mathbf{b}}z_{\mathbf{b}} - JPKx_{\mathbf{b}}y_{\mathbf{b}}) + K[Rx_{\mathbf{b}} - P(Jz_{\mathbf{b}} - Ky_{\mathbf{b}})^2] + I^2 Py_{\mathbf{b}}(Jz_{\mathbf{b}} - Ky_{\mathbf{b}})}{Q(Jz_{\mathbf{b}} - Ky_{\mathbf{b}})}, \tag{8}
$$

$$
z = \frac{IRy_{\mathbf{b}} - JRx_{\mathbf{b}} + I^2 Pz_{\mathbf{b}}(Jz_{\mathbf{b}} - Ky_{\mathbf{b}}) + IPKx_{\mathbf{b}}(Ky_{\mathbf{b}} - Jz_{\mathbf{b}}) + JP(Jz_{\mathbf{b}} - Ky_{\mathbf{b}})^2}{Q(Jz_{\mathbf{b}} - Ky_{\mathbf{b}})}, \tag{9}
$$

where constants Q and R are defined as

$$
Q = I^2(y_{\mathbf{b}}^2 + z_{\mathbf{b}}^2) + J^2(x_{\mathbf{b}}^2 + z_{\mathbf{b}}^2) + K^2(x_{\mathbf{b}}^2 + y_{\mathbf{b}}^2) - 2Jy_{\mathbf{b}}(Ix_{\mathbf{b}} + Kz_{\mathbf{b}}) - 2IKx_{\mathbf{b}}z_{\mathbf{b}}, \tag{10}
$$

$$
R = \sqrt{(Jz_{\mathbf{b}} - Ky_{\mathbf{b}})^2 [Q\ell^2 - P^2(I^2 + J^2 + K^2)]}. \tag{11}
$$

As we can see from the denominator in Equations (8) and (9), the expressions for y and z are undefined if both $y_{\mathbf{b}}$ and $z_{\mathbf{b}}$ are zero. In this case, the appropriate solution is

$$
y = \frac{-IJx + S}{J^2 + K^2}, \tag{12}
$$

$$
z = \frac{-IK^2 x + JS}{K(J^2 + K^2)}, \tag{13}
$$

with x given by equation (7) and S defined as

$$
S = \sqrt{K^2[-I^2 x^2 + J^2 K^2(\ell - x)(\ell + x)]}. \tag{14}
$$

These equations for x, y, and z yield two possible solutions; the first corresponds to a dihedral angle of $0°$ from ABC, and the second corresponds to a dihedral angle of $180°$ from ABC. We arbitrarily selected one of these solutions. We then rotated the point X around $\mathbf{b}$ until it was located at the appropriate dihedral angle τ relative to ABC.

Once atom X is placed at position (x, y, z), we translate these coordinates back to the original coordinate system in which C is not at the origin.