

II. APPENDIX

II.A1. Conformations. Models and Definitions

a. Sets of coordinates for backbone conformations

Let us consider a macromolecule composed of n structural *units* along the backbone. These units may be atoms or unified groups of atoms (*). The simplest set of coordinates for defining the instantaneous *configuration* of the molecule is the set of *position vectors* $\{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \dots, \mathbf{R}_{n-1}, \mathbf{R}_n\}$ with respect to a laboratory-fixed frame OXYZ, as shown in *Figure II.A1.1*. In statistical treatments of chain conformational mechanics, a chain-embedded frame $O_1X_1Y_1Z_1$ affixed to the first bond is generally used as a reference. The position of the i^{th} unit with respect to the frame $O_1X_1Y_1Z_1$ is represented by the position vector $\mathbf{r}_i = (x_i \ y_i \ z_i)^T$, where T denotes the transpose. This choice of reference frame removes the three *translational* degrees of freedom of the molecule. Furthermore, the overall *rotational* degrees of freedom can be eliminated upon fixing the absolute orientation of the molecule. For example, to this aim, the first bond of the chain can be chosen along the X_1 axis, and the third can be constrained to lie in the plane spanned by the axes X_1 and Y_1 . Clearly, $\mathbf{r}_1 = \mathbf{0}$, and $y_2 = z_2 = z_3 = 0$ in this representation. The remaining $3n - 6$ coordinates $\{x_2, x_3, y_3, x_4, y_4, z_4, \dots, x_n, y_n, z_n\}$ define the *internal* configuration of the macromolecule.

The position of the i th unit with respect to OXYZ can be expressed in terms of the internal position vectors \mathbf{r}_i as

$$\mathbf{R}_i = \mathbf{R}_1 + \mathbf{T}_1 \mathbf{r}_i = \mathbf{R}_1 + \mathbf{T}_1 \sum_{k=2}^i \mathbf{l}_k \quad (\text{II.A1.1})$$

where $\mathbf{l}_i = \mathbf{r}_i - \mathbf{r}_{i-1}$ is the *bond vector* connecting the units $i-1$ and i , pointing from $i-1$ to i (*Figure II.A1.2*), and \mathbf{T}_1 is the transformation matrix for the passage from the frame $O_1X_1Y_1Z_1$ into the laboratory-fixed frame OXYZ. We note that in this notation the first bond of the chain is denoted as \mathbf{l}_2 . See § II.A2 for the definition of transformation matrices in general, and for the conventional bond-based frames and their associated transformation matrices used in polymer statistics.

(*) -----Footnote-----

In most theoretical treatments the hydrogen atoms are assumed to be included implicitly together with the heavy atoms to which they are attached. Furthermore, groups of atoms such as the carbonyl C and O, or members of a ring, are regarded as forming a unified group rigidly moving during conformational motions. The backbone of the macromolecule can thus include all atoms other than bulky side chains. We therefore choose to refer to *structural units*, rather than atoms, for describing the main chain. It should be understood that this term includes isolated or unified atoms.

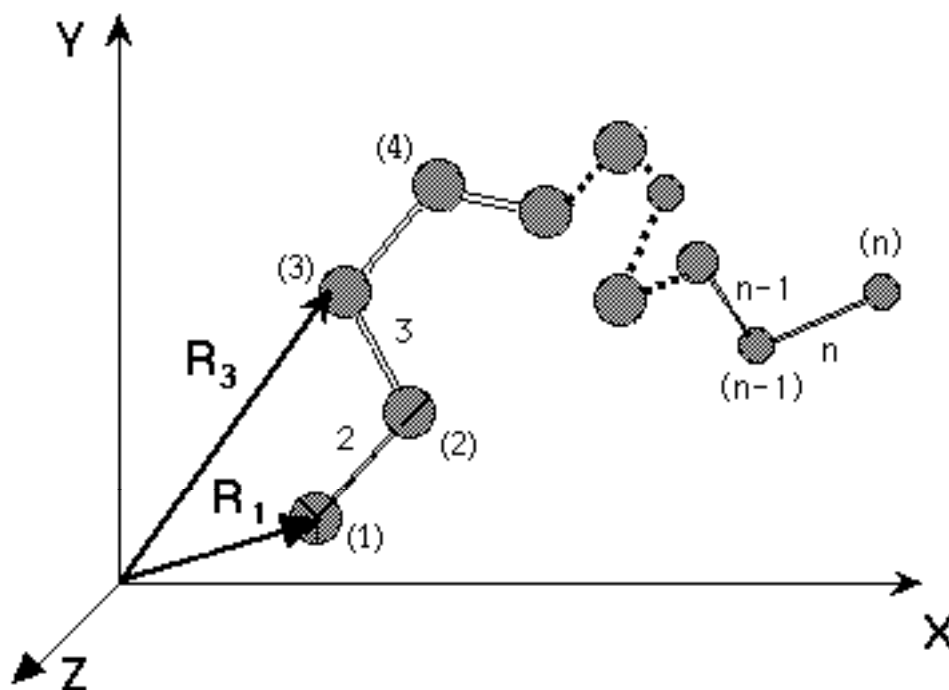


Figure II.A1.1. Schematic representation of a chain of n backbone units. Bonds are labeled from 2 to n , and structural units from 1 to n . The location of the i th unit with respect to the laboratory-fixed frame $OXYZ$ is indicated by the position vector \mathbf{R}_i . \mathbf{R}_1 and \mathbf{R}_3 are explicitly shown.

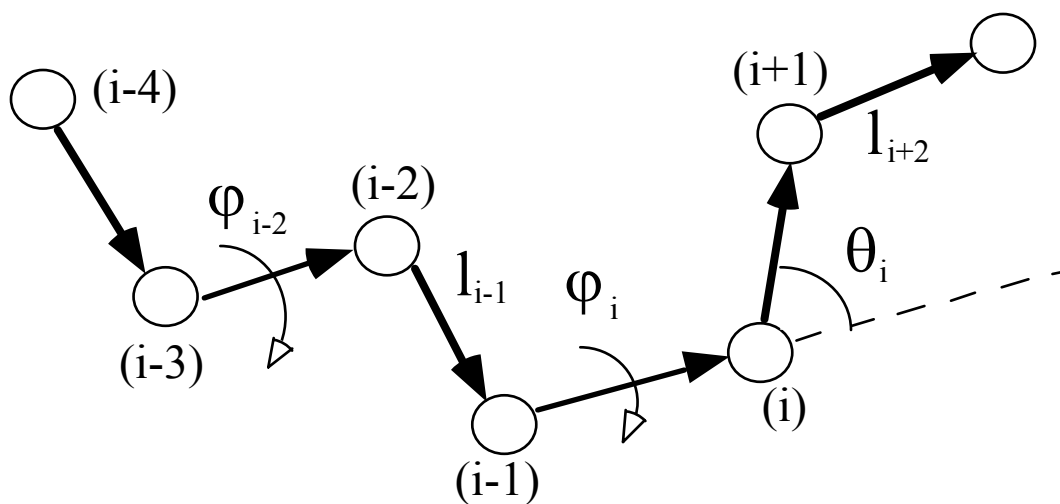


Figure II.A1.2. Schematic representation of a portion of the main chain of a macromolecule. \mathbf{l}_i is the bond vector between units $i-1$ and i , as shown. φ_i denotes the torsional angle about bond i .

The set of coordinates $\{\mathbf{R}_1, l_2, l_3, \dots, l_{n-1}, l_n\}$ can be conveniently used, instead of $\{\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3, \dots, \mathbf{R}_{n-1}, \mathbf{R}_n\}$, for describing the three-dimensional structure of the chain.

Alternatively, the set of $3n$ variables consisting of

- (i) the position vector \mathbf{R}_1 of the first unit with respect to OXYZ,
- (ii) three Euler angles describing the absolute orientation of the protein in space,
- (iii) $n-1$ bond lengths l_i , i varying in the range $2 \leq i \leq n$,
- (iv) $n-2$ supplemental bond angles θ_i between bonds l_i and l_{i+1} , $2 \leq i \leq n-1$ (Figure II.A1.2),
- (v) $n-3$ rotational or dihedral angles φ_i describing the torsional rotations of all internal bonds ($3 \leq i \leq n-1$) about their own axes, as illustrated in Figure II.A1.3,

may be adopted for describing the conformation of a protein backbone. This choice of *generalized coordinates* offers the advantage of separating the *internal* and *external* degrees of freedom. Items (i) and (ii) represent the external state. They refer to the rigid-body translation and rotation of the molecule, respectively. The remaining $3n - 6$ variables in items (iii)-(v) describe the internal conformational state.

In analytical treatments, it will often be necessary to switch between *rectilinear* or *rectangular* coordinates $\{x_2, x_3, y_3, \dots, x_n, y_n, z_n\}$ and *curvilinear* or *polar* coordinates $\{l_2, l_3, \dots, l_n, \theta_2, \theta_3, \dots, \theta_{n-1}, \varphi_3, \varphi_4, \dots, \varphi_{n-1}\}$ or vice versa. See Appendix II.A3 for the set of equations commonly used for the passage between these two sets of coordinates.

Depending upon the investigated property, one or the other set may be advantageously adopted for computation. For instance, the former readily yields the interatomic distances. These are needed when calculating non-bonded intramolecular interactions in MD simulations. Rectangular coordinates are also used for finding the position of the *center of mass* from the summation over all units $1 \leq i \leq n$,

$$\mathbf{R}_G = \sum_i m_i \mathbf{R}_i / \sum_i m_i \quad (\text{II.A1.2})$$

and the *radius of gyration* S from

$$S^2 = \sum_i m_i |\mathbf{R}_i - \mathbf{R}_G|^2 / \sum_i m_i \quad (\text{II.A1.3})$$

where m_i denotes the mass of the i th structural unit. The vector $\mathbf{R}_i - \mathbf{R}_G$ represents the position of the i th unit with respect to the center of mass, and the vertical bars designate the magnitude of the enclosed vector. S^2 can conveniently be expressed in terms of internal distances along as {Flory 1969 ID: 460}

$$S^2 = \sum_{1 \leq i < j \leq n} (\mathbf{R}_j - \mathbf{R}_i)^2 / n^2 \quad (\text{II.A1.4})$$

The polar variables, on the other hand, are used in the harmonic potentials governing bond stretching and bending motions, in serial expansions describing the bond rotational energy distributions as a function of torsion angles, or in simulations done by holding the bond lengths and bond angles fixed.

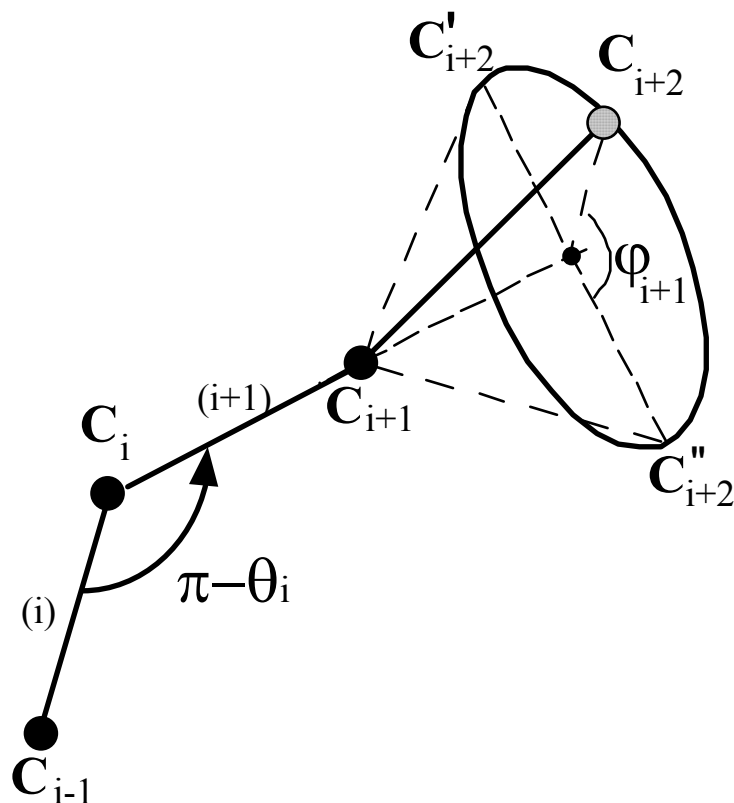


Figure II.A1.3. Spatial representation of the torsional mobility around the bond $i+1$. The torsional angle φ_{i+1} of bond $i+1$ determines the position of the atom C_{i+2} relative to C_{i-1} . C'_{i+2} and C''_{i+2} represent the locations of the atom $i+2$, when φ_i assumes the respective values 180° and 0° , characteristic of the *trans* and *cis* rotameric states. (see Fig II.A1.5).

b. Torsional angles of backbone bonds are a major determinant of macromolecular conformation

The internal conformation, rather than the absolute location and orientation in space, is of interest in most applications, and the set of internal variable $\{l_2, l_3, \dots, l_n, \theta_2, \theta_3, \dots, \theta_{n-1}, \varphi_3, \varphi_4, \dots, \varphi_{n-1}\}$ is commonly resorted to. For fixed bond lengths and bond angles, which may be regarded as a reasonable first approximation for treating chain statistics, specification of the ensemble of rotational angles $\{\varphi_3, \varphi_4, \dots, \varphi_{n-1}\}$ is sufficient for defining a given conformation. This approximation relies on the fact that the rotational

angles are the softest degrees of freedom, usually responsible for conformational changes occurring in macromolecules, while bond angles, and bond lengths in particular, are more severely constrained in their fluctuations about mean values.

Not all torsional angles are equally probable, in general. Some torsional angles, referred to as rotational isomeric states (RIS) {Flory 1969 ID: 460} {Mattice & Suter 1994 ID: 493}, are more frequent than others, these being favored by the intrinsic torsional potentials of the particular bonds. For example, tetrahedral bonds show three equally probable isomeric states, separated each by $\Delta\varphi = 120^\circ$. Their intrinsic torsional potential is shown in *Figure II.A1.4* (dashed curve). In polyethylene chains, $-(\text{CH}_2)_n-$, two of these states, called the *gauche*⁺ and *gauche*⁻ isomers, become less probable than the third, *trans* isomer (*Figure II.A1.5*). This is due to the constraints on rotational states imposed by the interactions between bonded near neighbors, for example those between carbon atoms $i-2$ and $i+1$ when rotating the i th bond. The steric clash between the same pair of atoms is also responsible for the large hindrances in the *cis* isomer. This distribution over the dihedral angle is illustrated by the solid curve in *Figure II.A1.4*.

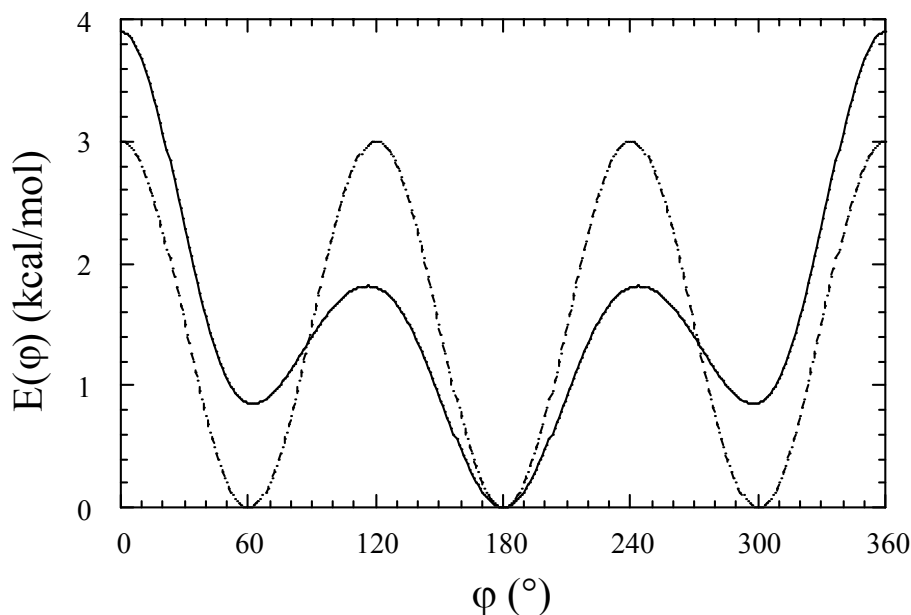
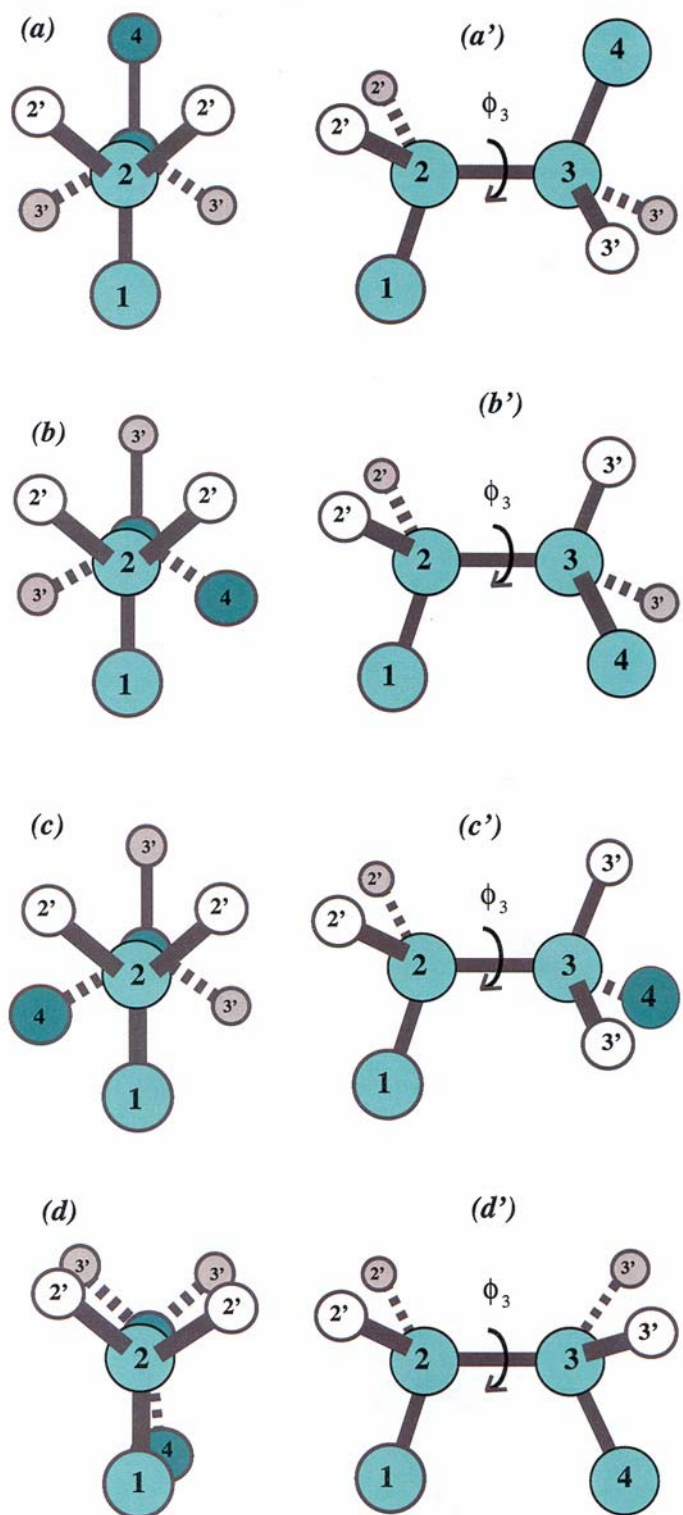


Figure II.A1.4. Rotational energy as a function of dihedral angle for a threefold symmetric torsional potential (dashed curve) and a three-state potential with a preference for the *trans* isomer ($\varphi = 180^\circ$) over the *gauche* isomers (60° and 300°) (solid curve), and the *cis* (0°) state being most unfavorable. See *Figure II.A1.5*.

Figure II.A1.5. (next page) Rotational isomeric states for the central bond in a segment of four backbone atoms. Large blue spheres show backbone atoms. They are indexed from 1 to 4. The small spheres show side groups; they are labeled by the indices of the backbone atoms to which they are affixed (with a prime sign). Side groups on atoms 1 and 4 are omitted for clarity. Four isomeric conformations, characterized by the rotational (or dihedral) angle φ_3 of the bond connecting atoms 2 and 3, are illustrated. Two views are shown for each conformation: the Newmann diagrams **(a)-(d)** where atom 3 is eclipsed by atom 2, and the side views **(a')-(d')** where the first two bonds lie in the plane of the both sides lie on a common plane. Atoms 1 and 4 are located at the largest separation permitted by bond angles and bond lengths; **(b)** *gauche*⁺ (*g*⁺). The middle bond undergoes right-handed rotation of 120° with respect to the *trans* state, and becomes $\varphi_3 = 300^\circ$ (or -60° due to periodicity); **(c)** *gauche*⁻ (*g*⁻). The middle bond is rotated by -120° with respect to the *trans* state, and becomes $\varphi_3 = 60^\circ$; **(d)** *cis* (*c*), with $\varphi_2 = 0^\circ$. The three successive bonds are planar, as in the *trans* state, but atoms 1 and 4 are at their closest separation.



II.A2. Bond-based coordinates for describing polymeric structures

In the conformational analysis of macromolecules, it proves convenient to *define bond-based* frames, or *local* frames embedded in the chain. These move with the backbone as the latter undergoes conformational changes. Vectorial quantities rigidly affixed in these frames are first evaluated in their local bond-based frames, and then transformed by standard matrix multiplication methods into their representation in the laboratory-fixed frame OXYZ. *Figure II.A2.1* illustrates the conventional definition for the bond-based frame $OX_{i+1}Y_{i+1}Z_{i+1}$ appended to the i th bond along the chain.

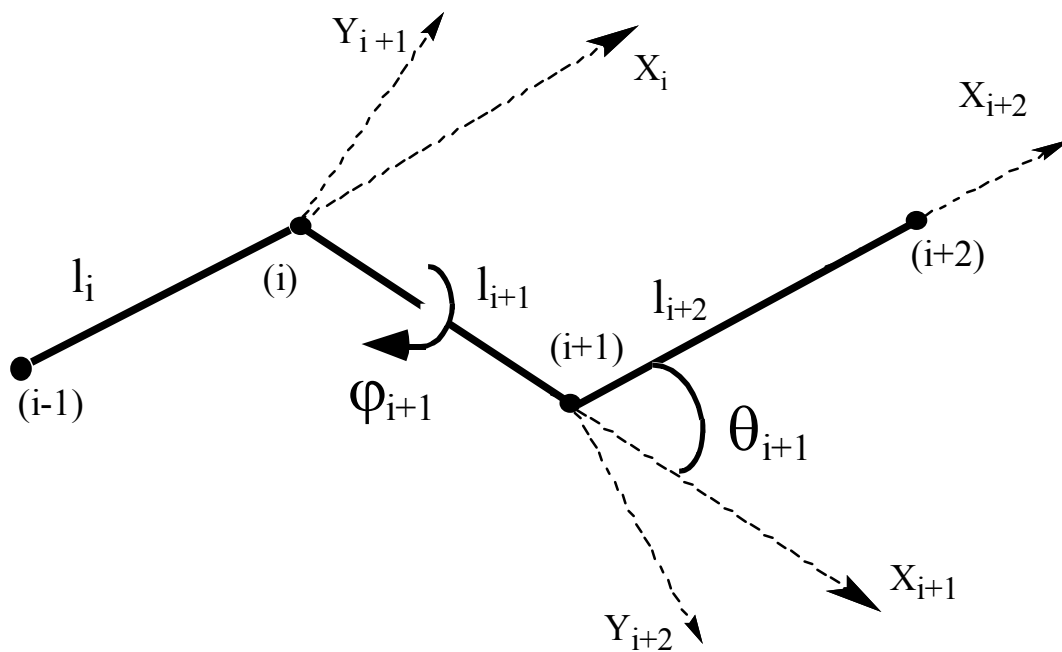


Figure II.A2.1. Schematic representation of a chain segment of four bonds. Atomic serial indices are indicated in parentheses. The i th bond connects atoms $i-1$ and i along the main chain, and its torsion angle is denoted as φ_i . θ_i is the supplemental bond angle defined by bonds i and $i+1$. The X_{i+1} and Y_{i+1} axes of the bond-based coordinate system $X_{i+1}Y_{i+1}Z_{i+1}$ appended to the bond $i+1$ are shown. Y_{i+1} lies in the plane defined by bonds i and $i+1$, and makes an acute angle with X_i . The Z_{i+1} axis, not shown, completes a right-handed coordinate system.

II.A3. Transformations between Cartesian and generalized coordinates

The transformation from the Cartesian coordinates $\{x_2, x_3, y_3, \dots, x_n, y_n, z_n\}$ into the generalized coordinates $\{l_2, l_3, \dots, l_n, \theta_2, \theta_3, \dots, \theta_{n-1}, \varphi_3, \varphi_4, \dots, \varphi_n\}$ is achieved by invoking the relationships

$$l_k = |\mathbf{r}_k - \mathbf{r}_{k-1}| \quad (\text{II.A3.1})$$

$$\theta_k = \theta_k(\mathbf{r}_{k-1}, \mathbf{r}_k, \mathbf{r}_{k+1}) = \cos^{-1} \left[\frac{\mathbf{l}_k \cdot \mathbf{l}_{k+1}}{|\mathbf{l}_k| |\mathbf{l}_{k+1}|} \right] \quad (\text{II.A3.2})$$

$$\varphi_k = \varphi_k(r_{k-2}, r_{k-1}, r_k, r_{k+1}) = \text{sign}[\sin(\varphi_k)] \cos^{-1}(\mathbf{n}_k \cdot \mathbf{n}_{k-1}) \quad (\text{II.A3.3})$$

where the symbols \mathbf{x} and \cdot refer to vector and scalar products, respectively, \mathbf{n}_k is the unit normal vector, perpendicular to the plane spanned by \mathbf{l}_k and \mathbf{l}_{k+1} , found from

$$\mathbf{n}_k = [(\mathbf{l}_k \times \mathbf{l}_{k+1}) / |\mathbf{l}_k \times \mathbf{l}_{k+1}|] \quad (\text{II.A3.4})$$

and $\text{sign}[\sin(\varphi_k)]$ represents the sign (+ or -) of $\sin(\varphi_k) = (\mathbf{n}_k \times \mathbf{n}_{k-1}) / |\mathbf{n}_k| |\mathbf{n}_{k-1}|$. We note that the inverse cosines in equations II.A3.2 and II.A3.3 yield angles in the range $[0^\circ, 180^\circ]$. The bond angles vary in the range $[0^\circ, 180^\circ]$, and can be directly found from eq II.A3.2. The torsional angles, on the other hand, vary in the range $[-180^\circ, 180^\circ]$, and one needs to choose between two values, either positive or negative rotations with respect to the *trans* state (180°), both having the same cosine value. The term $\text{sign}[\sin(\varphi_k)]$ selects for the appropriate torsional angle.

Equations II.A3.1-4 are useful for calculating the bond angles θ_k and rotational angles φ_k in structures whose rectangular coordinates are available. A typical application is the evaluation of the dihedral angle distributions for protein databank structures. The passage from polar coordinates to the Cartesian ones, on the other hand, is usually performed with the help of the transformation matrices. The matrix $\mathbf{T}_i(\theta_i, \varphi_i)$ transforms the vectorial quantities of the bond-based frame $i+1$ into their representation in the i th frame. For example, the bond connecting atoms i and $i+1$ is expressed as $\mathbf{l}_{i+1}^* = (l_{i+1} \ 0 \ 0)^T$ in the local frame $i+1$, the asterisk indicating that the vector refers to the *local* frame $i+1$. The same vector's representation in the preceding bond-based frame is $\mathbf{T}_i(\theta_i, \varphi_i) \mathbf{l}_{i+1}^*$. Successive applications of such transformations leads, for the vector between unit 1 and i , to

$$\mathbf{r}_i = \mathbf{l}_2^* + \mathbf{T}_2(\theta_2, \varphi_2) \mathbf{l}_3^* + \mathbf{T}_2(\theta_2, \varphi_2) \mathbf{T}_3(\theta_3, \varphi_3) \mathbf{l}_4^* + \dots + \prod_{k=2}^{i-1} [\mathbf{T}_k(\theta_k, \varphi_k)] \mathbf{l}_i^* \quad (\text{II.A3.5})$$

where φ_2 is set equal to 180° , and the transformation matrix $\mathbf{T}_i(\theta_i, \varphi_i)$ is defined as

$$\mathbf{T}_i(\theta_i, \varphi_i) = \begin{bmatrix} \cos \theta_i & \sin \theta_i & 0 \\ -\sin \theta_i \cos \varphi_i & \cos \theta_i \cos \varphi_i & -\sin \varphi_i \\ -\sin \theta_i \sin \varphi_i & \cos \theta_i \sin \varphi_i & \cos \varphi_i \end{bmatrix} \quad (\text{II.A3.6})$$

A typical application of equations (II.A3.5) and (II.A3.6) is the computer generation of chains whose dihedral angles, bond lengths and bond angles are known. MC simulations using the generalized coordinates is a typical application. See Chapter VII.

II.A4. Virtual bond representation of inter-residue coordination angles

Two spherical angles have been defined in order to describe the coordination geometry of residues (see *Figure II.3.8*) at a coarse-grained level {Bahar & Jernigan 1996 ID: 83}. In this coarse-grained description, each residue (i) is represented by two sites, its α -carbon (C^α_i), and a sidechain site S_i . The site S_i is located either at the centroid of sidechain, at the functional/charged group. The position of a neighbor S_j with respect to S_i is defined by two coordination angles, θ_{ij} and φ_{ij} , referred to as the polar and azimuthal angles, respectively. See *Figure II.A4.1*.

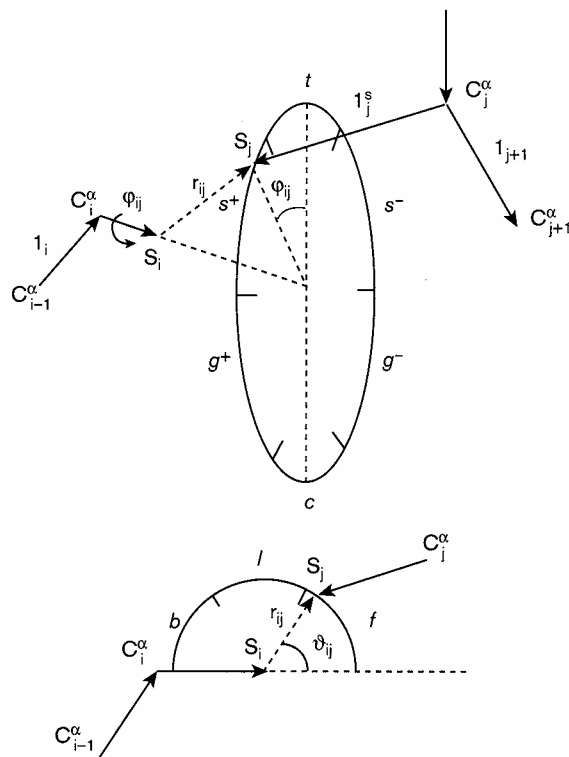


Figure II.A4.1. Coarse-grained representation of the angular coordination between side chain S_i and S_j . r_{ij} , (dashed line with an arrow at the end) is the vector pointing from S_i to S_j . The polar angle ϑ_{ij} is the angle between r_{ij} and the extension of side chain bond vector l_i^s . The azimuthal angle ϕ_{ij} is the torsional angle about bond l_i^s . It is defined as 0° for the trans (t) position with respect to atoms C_{i-1}^α , C_i^α and S_i . Other dihedral angle positions are indicated as s^+ (skew⁺, 60°), g^+ (120°), c (cis, 180°), g^- (gauche⁻, 240°) and s^- (skew⁻, 300°); whereas polar angles are classified as front (f), lateral (l) and back(b) position. The geometric variables r_{ij} , ϑ_{ij} and ϕ_{ij} define the coordination of S_i by S_j . (from {Bahar & Jernigan 1996 ID: 83})