Toward a computed peptide structure database: The role of a universal atomic numbering system of amino acids in peptides and internal hierarchy of database

ARTICLE *in* INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY · DECEMBER 2001

Impact Factor: 1.43 · DOI: 10.1002/qua.947

8 AUTHORS, INCLUDING:

Gregory Chass
Queen Mary, University of London
105 PUBLICATIONS **1,325** CITATIONS
SEE PROFILE

Sandor Lovas
Creighton University
123 PUBLICATIONS **2,047** CITATIONS
SEE PROFILE

Ödön Farkas
Eötvös Loránd University
74 PUBLICATIONS **5,269** CITATIONS
SEE PROFILE

Jean-Louis. Rivail
University of Lorraine
124 PUBLICATIONS **3,779** CITATIONS
SEE PROFILE

Toward a Computed Peptide Structure Database: The Role of a Universal Atomic Numbering System of Amino Acids in Peptides and Internal Hierarchy of Database

GREGORY A. CHASS,[1,2,3,5] MICHELLE A. SAHAI,[1,2] JACQUELINE M. S. LAW,[1,2] SANDOR LOVAS,[3] ÖDÖN FARKAS,[4] ANDRAS PERCZEL,[4] JEAN-LOUIS RIVAIL,[5] IMRE G. CSIZMADIA[1,2]

[1]*Velocet Research and Development, 210 Dundas St. W., Suite 810, Toronto, ON, M5G 2E8 Canada*
[2]*Department of Chemistry, University of Toronto, 80 St. George St., Toronto, ON, M5S 3H6 Canada*
[3]*Department of Biomedical Sciences, Creighton University, 2500 California Plaza, Omaha, Nebraska 68178*
[4]*Department of Organic Chemistry , Eötvös Loránd University, Pázmány Péter sétány 1/A, H-1117, Budapest, Hungary*
[5]*Laboratoire de Chimie Théorique, Université Henri Poincaré BP239, 54506 Vandoeuvre-les-Nancy, France*

Received 15 September 2001; accepted 5 March 2002

DOI 10.1002/qua.947

ABSTRACT: With the construction and implementation of a logical and standardized numbering of atomic nuclei, to define mono-, di-, and oligo-peptide systems, automation of input file generation and data extraction could greatly improve the efficiency of the search for the structural energy minima on the potential energy hypersurface of these systems. The internal hierarchy of the database covering constitutional structures, protective groups, levels of theory, and basis sets used, as well as the variety of possible conformations, is also discussed. Int J Quantum Chem 90: 933–968, 2002

Key words: peptide computations; atomic numbering; protein folding; internal coordinates; database internal hierarchy; ab initio

Correspondence to: G. A. Chass; e-mail: gchass@fixy.org.

Dedication

This article is written in the "teaching spirit" of our immortal teacher Professor P. O. Löwdin. We hope that our article and the methodology and ideals within will improve the understanding of young researchers in the whole field of molecular science, from physics and mathematics through chemistry and biology, all the way to medicine and even computational sciences, specifically for those who wish to enter the field of study comprising peptide and protein folding. In this spirit, we wish to dedicate this article to the legacy and living memory of Professor P. O. Löwdin.

Preamble

We are now in a new age where we are able to produce vast amounts of data but do not capitalize on their value. We cannot thoroughly analyze all of the data, due to limitations in extraction of relevant variables and values, coupled with a serious lack of cross-referencing/pattern-matching. Consequently, one must develop "data-engines" and "data-mechanics" for multivariable analyses of systems commonly having hundreds or thousands of degrees of freedom. The process is particularly difficult when intersystem assemblies are made of known intrasystem building blocks, such as with modularly constructed oligopeptides and glycopeptides interacting with substrate molecules. In fact, the direction of approach, environmental conditions, and many other interactive relations become very important, especially in the development of kinetic models. Although intersystem dynamics and mechanics are very complex and yield a wealth of understanding, they are more of a holistic approach [1] to modern challenges in theoretical areas. The internal thermodynamics of oligopeptidic secondary and tertiary structural elements is perhaps one of the most challenging aspects of chemistry as a whole, in all of its related subdisciplines.

Using this reductionist approach [1] and breaking the problem of secondary and tertiary peptide/protein structure up into its constituent elements, one is able to isolate each, creating and refining them, until they are understood in a form that allows them to be reassembled into a more accurate and holistic picture [2]. This would include the formulation of a set of analytic rules to account for the coupling behavior of all parameters in the system.

However, at this point there is little true understanding of the internal conformational behavior of the constituent peptides. These geometries are governed by the spatial orientations of all composite nuclei, which can also be described in numerous different ways through the use of internal dihedral or torsional angles. Of course the four-atom definition of these torsions can be formulated in a logical, ordered, and related manner, making data management also routine and automatable.

As all peptides have the same backbone nuclei, differing and not to great extent—only in their side-chains—a reproducible and consistent numbering system may be defined. Standardizing this system would provide an opportunity for the simplification of the extraction of all internal structural variables. This, in turn, would allow for the construction of a much more complete computational structure data base. The standardized structural definitions and numbering would then be used as the infrastructure of data base searches for complex patterns, trends, and deviations from ideal or expected structures, in folded proteins.

Backbone dihedral angles ϕ and ψ are the primary contributors to overall secondary and tertiary structure, but folding is dependent on many more variables that sometimes sway the approximations "the other way," away from the ideal or expected state. Some ab initio calculations carried out on N- and C-protected single amino acid residues have yielded the inverse γ-turn (γ_L) conformation as the global minimum, due to internal hydrogen bonding (C=O$\cdots$H–N). However, γ_L may not be the lowest energy conformer anymore when other interactions dominate. This can happen particularly when one residue makes a hydrogen bond with another residue. Because of such interactions structural distortion can occur, for example the peptide bond can be strained to 150° from 180°, etc., which is never expected in nor worked into calibrations for classical models. Computations can help to advance understanding of preferred conformers, so as to aid in a better understanding of experimental results and macroscale observables. Full comprehension of peptide and protein folding could lead to an ability to understand many diseases, which is a prerequisite to finding a cure.

HISTORY OF PEPTIDE CALCULATIONS

It is reasonable to relate the start of peptide computations with respect to the date when Ramachandran first proposed the concept of conformational potential energy surface (PES) of a peptide residue [2]. It seems in retrospect that this date, 1963, really represents the origin of our time coordinate. The history is illustrated schematically in Figure 1.

Ramachandran's idea came from mechanical considerations of hard ball molecular models. It took over a quarter of a century for ab initio surfaces to be published as two-dimensional (2D) contour representations by Pople [3] and as 3D landscape representations by Perczel [4].

The sequence of selected events in peptide computations is shown on four time scales in Figure 1. These are shown from top to bottom: basic concepts, force fields, semiempirical molecular orbital (MO), and ab initio quantum chemical computations.

Soon after Ramachandran's epoch-making paper [2], Nemethy and Scheraga [5] published in 1965 their force field method (ECEPP) parametized for peptides. A paper by Liquori [6] followed in 1969. The well-established force field CHARMM arrived on the scene in 1983 from Karplus [7] and the popular AMBER by Kollman [8] arrived in 1984. These force fields are used nowadays in molecular mechanics and molecular dynamics packages.

Ronald Hoffman [9] in 1969 was the first to use the semiempirical MO method to calculate peptides and he was followed by many [10–18] including Hopfinger [10] and Maigret [15–17].

Ab initio quantum chemistry was in its infancy in the early 1960s. Thus, the smallest molecule that had a peptide bond, formamide, could not even be considered. Since in the early 1960s all mainframe computers were vacuum tube machines, only the smaller formyl fluoride (HCOF), which was isoelectronic to formamide ($HCONH_2$), could be computed in 1963 [18, 19]. A few years later mainframe computers were transistorized and by 1967 Basch [20] and 1968 Robb [21] were reporting ab initio computations on formamide. About a decade had to go by for small peptide calculations. Several people [22–29], including Peters [22] in 1979 and Boyd [26] in 1981, carried out ab initio peptide computations without geometry optimization. It was Schafer [29] in 1982 who pointed out first that geometry optimization is necessary. Gradually, ab initio computations became standard and after Perczel's entry in the field in 1990 [30] and 1991 [31], numerous studies have been reported. The conformations of dipeptide [32–34], tripeptide [35], and tetrapeptide conformations [36–39]

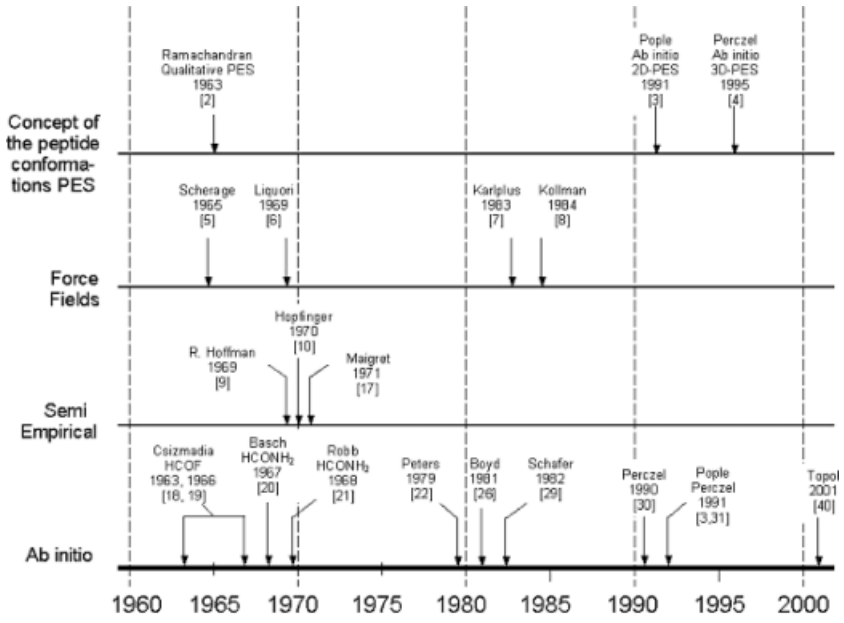


FIGURE 1. A schematic illustration of the history of peptide computations.

were studied using alanine as the only amino acid. Oligopeptides containing eight amino acids [40] or more are now within the realm of possibility.

FUTURE OF PEPTIDE CALCULATIONS

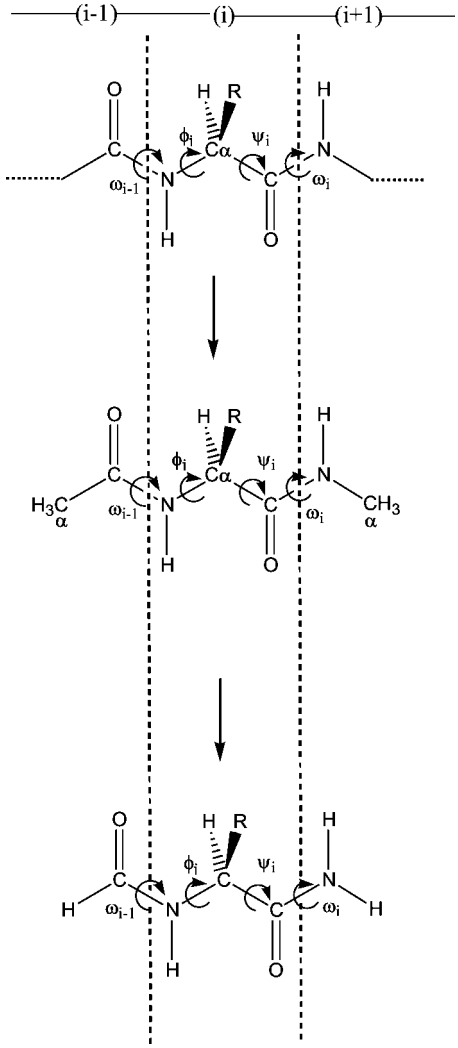
It should be noted that with the new age of technology comes a new age of students. The present generation (termed Generation X by the media) has been raised in an environment saturated by electronics and digital devices, including computers, since birth. Each of these operates on a numerical logistic and pervades daily living and thought. It comes therefore as no surprise that a number of students of this present generation are experiencing and pioneering a new age of thought, one that is steeped in numerical logic. This new thought may be one of the tools to be used in the search for the key to understanding the mechanics and forces behind the folding of polypeptides into functional proteins.

Of the 24 mono-peptide systems computed in this work, not one of the 96 computations (four levels of theory) was prepared using a visualization software tool. All conformational inputs were prepared numerically and examined in a like manner after convergence. This methodology could also be used in molecular mechanics and dynamics simulations of tens or hundreds of thousands of atoms, in order to track absolutely every variable in the system. Perhaps this best explains the terminology used when announcing the possibility the authors introduce in this work that we are no longer limited to relying on 3D slices of these highly dimensional polypeptide systems. Rather, we now have the means to perform an n-dimensional examination at any one time, leading to a new set of questions in this quest and epic battle.

> Computers are useless, they can only give you answers.
>
> Pablo Picasso

Introduction

An amino acid residue in the protein chain may be modeled in a number of ways as shown below in Scheme 1. Within IUPAC [41–49] convention using the central structure of Scheme 1, the torsional angles may be defined as follows: Dihedral ω_{i-1} is defined as $C_{\alpha i-1}$–C–N–$C_{\alpha i}$, ϕ_i is defined as C–N–$C_{\alpha i}$–C, ψ_i is defined N–$C_{\alpha i}$–C–N, and ω_i is $C_{\alpha i}$–C–N–$C_{\alpha i+1}$.



SCHEME 1.

Since an amino acid is principally a double rotor, involving two dihedral angles (ϕ and ψ), a total of $3 \times 3 = 9$ conformers are expected. The conformational assignments ($g^+g^+, ag^+, \ldots$, etc.) shown in Figure 2(a) are a bit cumbersome to use. Very often some names are invented for sake of convenience. While names may be arbitrary, such as A, B, C, etc., a more rational name might, however, be more appropriate. Such nomenclature is shown in Figure 2(b).

The names [Fig. 2(b)] are subscripted Greek letters. The Greek letters originate from earlier nomenclature (involving α, β, γ, etc. for α-helix, β-sheet,

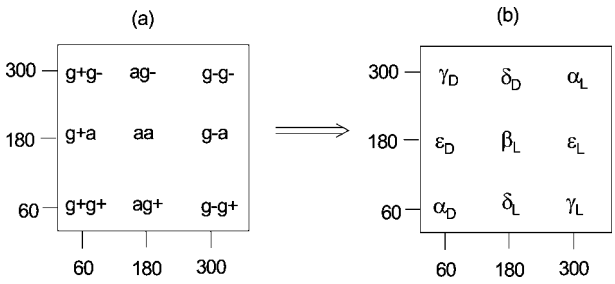


FIGURE 2. Conformational assignments (a) and names (b) of peptide conformers on the conformational (PES) of a peptide (P_N–CONH–CHR–CONH–P_C).

FIGURE 4. A schematic representation of the conformational PEHS of a peptide (P_N–CONH–CHR–CONH–P_C). The approximate locations of the conformations are symbolized by subscripted Greek letters. This is presented in the $-360° \leq \phi \leq +360° \leq$ and $-360° \leq \psi \leq +360° \leq$ range of independent variables, repetition of the traditional cut, with the dashed lines showing the IUPAC conventional cut.

γ-turn, etc.) while the L and D subscripts originate from the observation that the L-amino acids favor L subscripted conformations while D amino acids favor D-subscripted conformations. The names also suggest that we are experiencing the combination of the chirality of the constitutional structure (R or S configuration) and the chirality originating from the conformational twist or folding. This is summarized in Figure 3.

Figure 4 shows the symbolic pattern of the conformational PES for two full cycles of rotation ($-360° \rightarrow 0° \rightarrow +360°$). The four quadrants give the traditional cuts and the central square marked by broken lines specifies a cut according to IUPAC convention (from $-180°$ to $+180°$). The negative sign corresponds to counterclockwise rotation while the positive sign denotes clockwise rotation. An energy

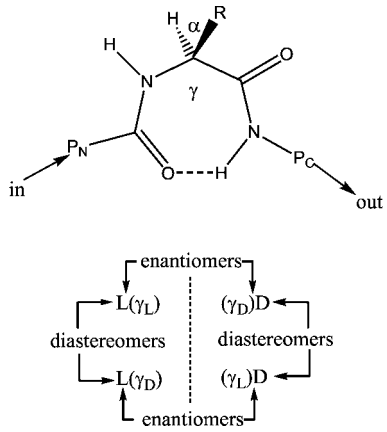


FIGURE 3. Stereochemical relationships of γ-turns. Note that only the α carbon has point chirality but there is also chirality in the twisting of the backbone. The combination of these two types of chiralities leads to enantiomeric and diastereoisomeric structures. D and L denote the chirality of the configuration C_α, while γ_L and γ_D denote the chirality of the conformation.

contour diagram of a peptide conformational potential energy surface is normally referred to as a Ramachandran map, in honor of the Indian chemist who first outlined their utility in peptide conformational analysis as seen as in Figure 5 for alanine and valine residues. As noted in the figure, the sidechain dihedral of valine was not restrained and remained in the g^+ conformation throughout the scanning of the backbone dihedrals. One may conclude that for this system, under these isolated conditions, the backbone and sidechain torsions, ϕ and ψ and χ respectively, are not coupled to an extent that would change the system to another sidechain conformer.

There are 20 naturally occurring amino acids and 18 of them have the same structural feature of their backbone folding as shown in Figure 2(b) (i.e., nine discrete conformations). Proline has its nitrogen locked in a five-membered ring as shown in Scheme 2.

For the proline residue ϕ can only be around $-60°$ (i.e., $+300°$) and therefore only three backbone conformations are possible: α_L, ε_L, and γ_L. The other unique amino acid is glycine (Scheme 3), which is achiral.

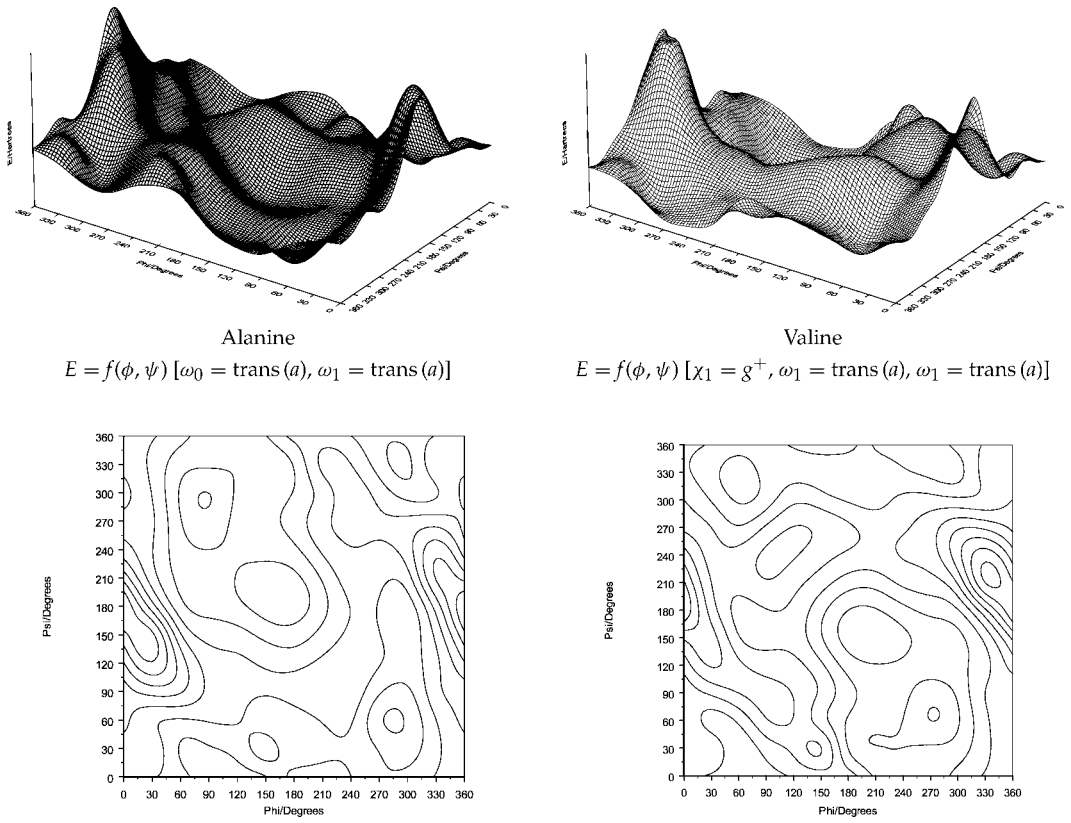


FIGURE 5. Potential energy surface diagrams of the N- and C-protected amino acids N-acetyl–alanine–N-methylamide and N-acetyl–valine–N-methylamide computed at the RHF/3-21G level of theory. Top: Landscape representation. Bottom: Contour representation. It should be noted that the valine sidechain remained fully relaxed during the scans.

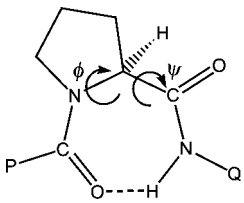
Finally, it should be mentioned that molecular computations have determined, in the case of single [31] and double [32–34] amino acid diamides, the actual location of the nine minima established above. The ϕ and ψ values deviate somewhat from the ideal values but nevertheless the resemblance is astonishing. Table I specifies these numerical values. The information tabulated in Table I is also presented graphically in Figure 6.

Method

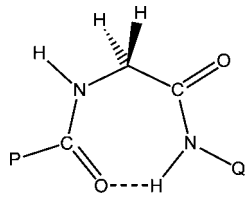
TORSIONAL/DIHEDRAL ANGLE DEFINITION AND CONSERVATION OF GEOMETRY

Examining the definitions of a dihedral angle, also known as a torsional angle, the four-atom geometric relation we find is shown in Scheme 4.

This particular dihedral angle η is shown in a number of different conformational states. It is numerically defined by the atomic nuclei that comprise



SCHEME 2.



SCHEME 3.

TABLE I

Optimized ϕ, ψ torsional angle pairs for alanine diamide ($HCONH–CHMe–CONH_2$) done at the restricted Hartree–Fock method level of theory. The idealized torsional angle pairs, together with their conformational classification, are also shown for the sake of comparison.

Conformer	Optimized values		Idealized values		Conformational classification
	ϕ	ψ	ϕ	ψ	
α_L	−66.6	−17.5	−60	−60	g^-g^+
α_D	+61.8	+31.9	+60	+60	g^+g^+
β_L	−167.6	+169.9	+180	+180	aa
γ_L	−84.5	+68.7	−60	+60	g^-g^+
γ_D	+74.3	−59.5	+60	−60	g^+g^-
δ_L	−126.2	+26.5	+180	+60	ag^+
δ_D	−179.6	−43.7	+180	−60	ag^-
ε_L	−74.7	+167.8	−60	+180	g^-a
ε_D	+64.7	−178.6	+60	−180	g^+a

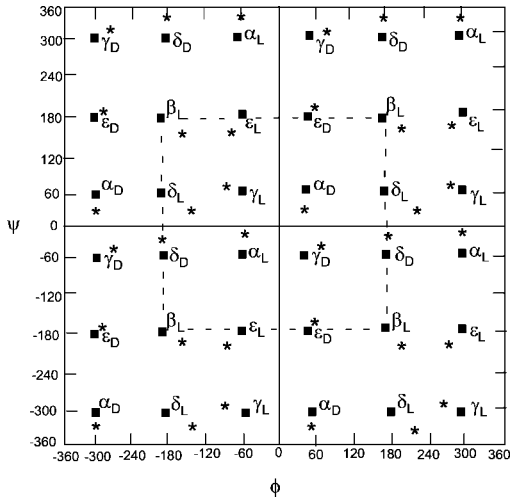


FIGURE 6. A schematic illustration of the PEHS of an amino acid diamide. The idealized positions are marked by the solid squares and the computationally determined (optimized) positions are shown as stars. These points were all computed at the restricted Hartree Fock (RHF/3-21G) level of theory. The names of the conformers are given as subscripted Greek letters. This is presented in the $-360° \leq \phi \leq +360$ and $-360° \leq \psi \leq +360$ range of independent variables, repetition of the traditional cut, with the dashed lines showing the IUPAC conventional cut.

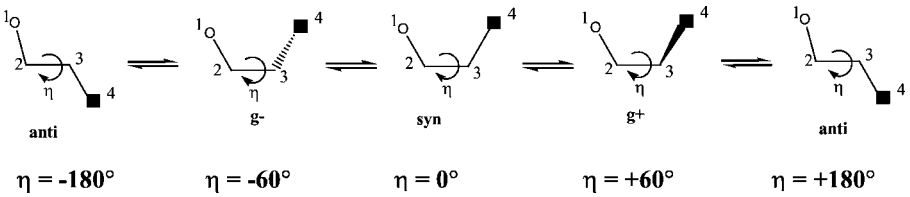
it, mainly atomic centers 1, 2, 3, and 4. Thus, the dihedral angle for atom 4 (D4) is defined by the sequence of atoms 4, 3, 2, 1. The order of the atomic nuclei may be reversed but not rearranged. This can be likened to "viewing" the dihedral angle from the other side.

Therefore,

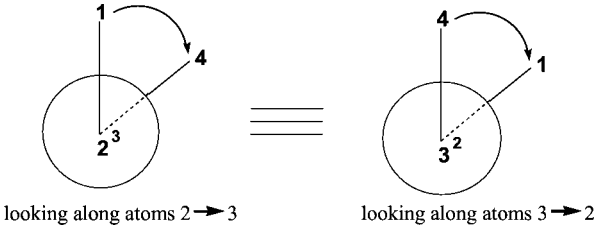
$$(4, 3, 2, 1) \equiv (1, 2, 3, 4).$$

Using the *gauche*$^+$ (g^+) example from above, represented with a Newman projection, we see the same conformation, regardless of whether it is viewed from the front or the back as shown in Scheme 5.

The spatial orientation of atom 4 is +60° (clockwise) away from atom 1 for the left portion of Scheme 5. The opposite view can be seen at the right hand side of Scheme 5. For a tetrahedral center, the same conditions exist whereby geometry is conserved as illustrated by Scheme 6. Note that in the first structure the atomic center 5, in the back of the Newman projection, is +60° (clockwise) away while the atomic center 6, in the back, is −60° (counterclockwise) away, both in terms of their spatial relation to atom 1 located at the front.



SCHEME 4.



SCHEME 5.

CONSTRUCTION OF PEPTIDE MODULES (SINGLE AMINO ACID DIAMIDES)

Moving on to a system of particular interest here, one finds that peptides can be easily constructed and manipulated both quickly and accurately with all stereochemical concerns being accounted for. In a peptide chain, each segment or component is numbered in its entirety (cf. Table II) prior to moving on to the next. This modular approach allows for quick additions, removals, rearrangements, or substitutions of any or all of the modules without having to redefine any other portions. Thus, any protecting groups of constituent amino acids would be numbered completely, one at a time, prior to progressing to the next. The numbering begins at the N-terminus and progresses to the C-terminus. The protecting groups are relatively simple and the numbering procedure will not be elaborated on; rather their numbering will be self-explanatory and shown in the examples that follow.

The peptide modules are numbered identically, through to the C_β (H in the case of glycine). Note that in Scheme 7, z is the number of atoms from N-acetyl-end through the amino acid in the sequence; i.e., $z_{Gly} = 13$, $z_{Ala} = 16$, $z_{Val} = 24$, etc. Dihedral angle χ_i^1 represents the first sidechain dihedral of the ith amino acid residue and subsequent sidechain dihedrals increasing with superscript indices. Also, note that the backbone atoms of an amino acid, N_7, C_8 (i.e., C_α), C_9, O_{10}, H_{11}, H_{12}, and $C_{\beta 13}$ (H_{13} in the case of glycine), are always numbered in this order.

By IUPAC definition, the following backbone dihedrals can be numerically described:

N-Protective group

$$\psi_{i-1} = (7,3,2,1) \equiv (1,2,3,7) \rightarrow D_7$$

$$\omega_{i-1} = (8,7,3,2) \equiv (2,3,7,8) \rightarrow D_8$$

Amino acid residue

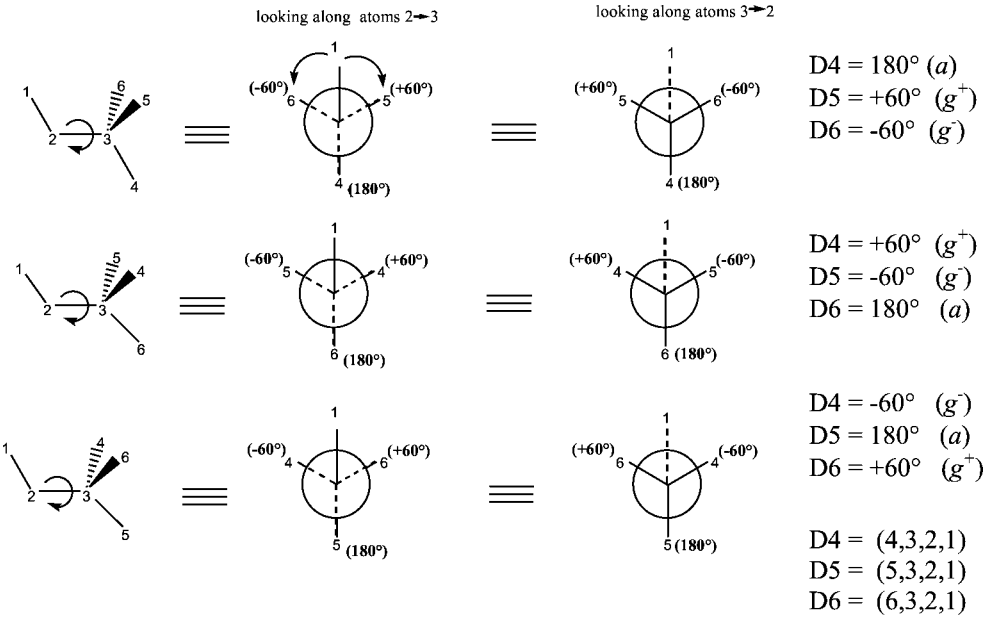
$$\phi_i = (9,8,7,3) \equiv (3,7,8,9) \rightarrow D_9$$

$$\psi_i = (z+1,9,8,7) \equiv (7,8,9,z+1) \rightarrow D_{z+1}$$

C-Protective group

$$\omega_i = (z+2,z+1,9,8) \equiv (8,9,z+1,z+2) \rightarrow D_{z+2}$$

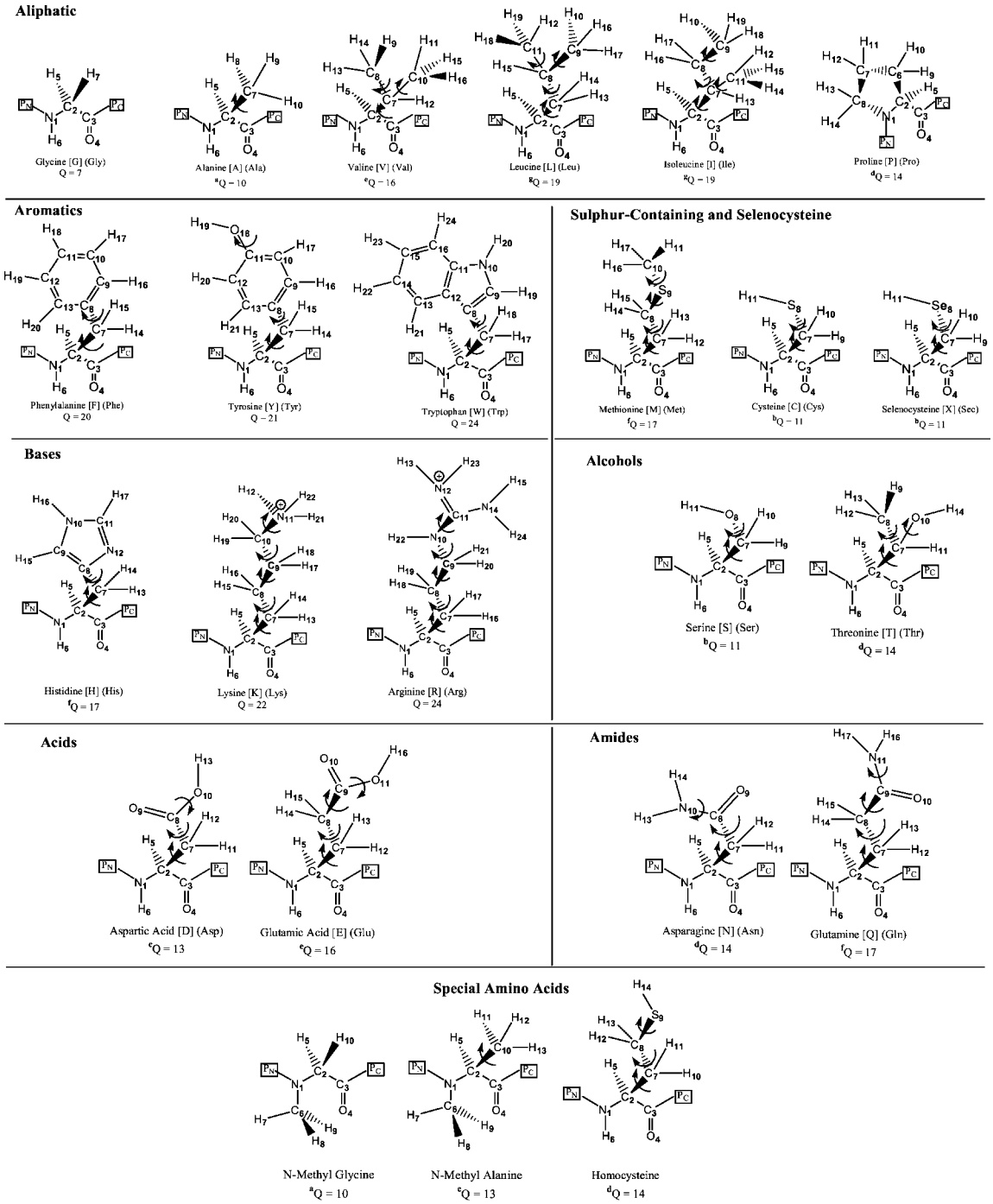
$$\phi_{i+1} = (z+3,z+2,z+1,9) \equiv (9,z+1,z+2,z+3) \rightarrow D_{z+3}.$$



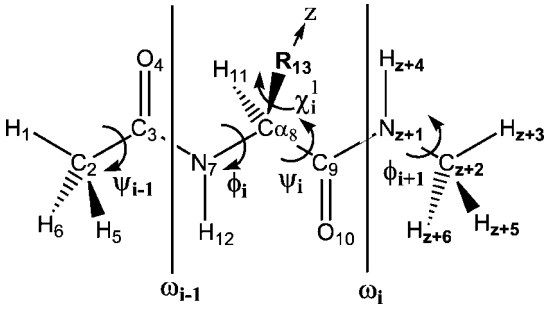
SCHEME 6.

TABLE II

Structures of amino acid residues. Note that in addition to the 20 naturally occurring amino acids, which have both DNA and RNA codons, the 21st amino acid, selenocysteine (Sec), which has only an RNA codon, is also included. Three special amino acids, N-methyl glycine (sarcosine), N-methyl alamine, and S-demethylated methionine (homocysteine) are also included.

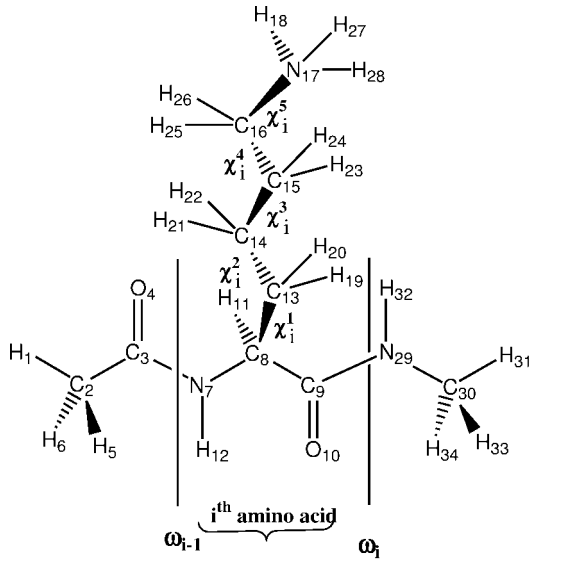


Peptides with similar *Q* numbers (a, b, c, d, e, and f) are differentiated through the examination of their sidechain atomic nuclei (see Table VIII).



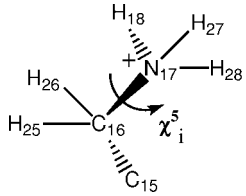
SCHEME 7.

The sidechain dihedrals labeled χ_i^j are defined by IUPAC (C_{14}–C_{13}–$C_{\alpha 8}$–C_7 for the first χ), as the sidechain to central backbones "heavy atom" spatial orientation, relative to the peptide backbone. Thus, the atomic numbering and resultant dihedral numeric definitions progress along the sidechain backbone, toward and into the peptide backbone. Note that for lysine (Scheme 8) $z_{\mathrm{Lys}} = 22 + 6 = 28$, where 22 is the number of atoms in the lysine amino acid and 6 is the number of atoms in the N-acetyl terminus.



$\chi_i^1 = (14,13,8,7) \equiv (7,8,13,14) \longrightarrow D_{14}$
$\chi_i^2 = (15,14,13,8) \equiv (8,13,14,15) \longrightarrow D_{15}$
$\chi_i^3 = (16,15,14,13) \equiv (13,14,15,16) \longrightarrow D_{16}$
$\chi_i^4 = (17,16,15,14) \equiv (14,15,16,17) \longrightarrow D_{17}$
$\chi_i^5 = (18,17,16,15) \equiv (15,16,17,18) \longrightarrow D_{18}$

SCHEME 8.



χ_i^5 = D18 (18,17,16,15)
or D27 (27,17,16,15)
or D28 (28,17,16,15)

SCHEME 9.

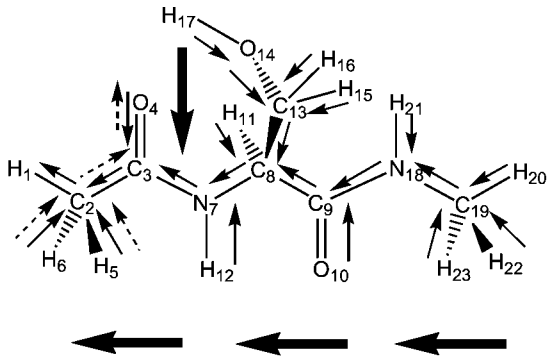
This lysine example brings up an interesting case whereby one could define χ_i^5 using any one of the three seemingly equivalent hydrogen atoms (18, 27, and 28) to be the fourth atom in the dihedral angle definition of χ_i^5 (i.e., D18, D27, or D28), as illustrated in Scheme 9.

RELATED DIHEDRAL ANGLES

This may bring about a new question or phenomenon of interest due to the fact that torsions or dihedrals may be defined in a number of different ways. In this method the numbering follows a pattern, which seeks to satisfy all IUPAC definitions [41–49], by reading along the peptide backbone from the largest to the smallest atom number. The example shown below is a serine monopeptide where the convention follows with the thick arrows, from the largest to the smallest number, with only two exceptions at the N-terminus. Note that in Scheme 10 the dashed lines ·····➤ going from H_6 and H_5 toward O_4 are the exceptions to the descending numbering of the dihedral angles [i.e., D6 = (6, 2, 3, 4), D5 = (5, 2, 3, 4)]. Here we follow the backbone in this ascending order so as not to be defining a dihedral through space [i.e., D6 = (6, 2, 1, 3) and D5 = (5, 2, 1, 3)] but rather through the covalently bonded structure.

However, by consequence, a number of dihedrals can be defined in a unique manner. Specifically, two dihedrals may be considered equivalent when their "lead" atoms (highest atom number) are being defined adjacent to a trigonal planar center. Alternatively, when the adjacent atom is a tetrahedral one, there will be observed three dihedrals with the same three-number terminus. This is more clearly understood by making use of Schemes 11 and 16.

As one can see in the example on the right side of Scheme 10 and at the left of Scheme 11, there are three dihedral angles, or three related atoms defining the ϕ rotational torsional potential. This requires a numerical assurance/value to prevent the superposition of these atomic nuclei. Taking the example of serine (Scheme 11) the bolded bonds show the



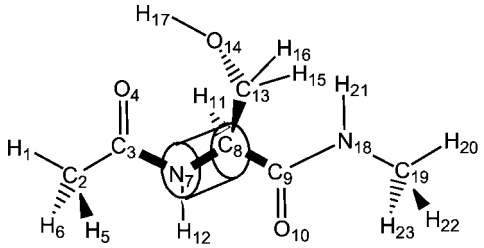
D (of atom #8) = D8 = (8,7,3,2) = ω_0
D (of atom # 12) = D12 = (12,7,3,2) = ω_0'

D (of atom # 9) = D9 = (9,8,7,3) = ϕ_1
D (of atom #11) = D11 = (11, 8,7,3) = ϕ_1'
D (of atom # 13) = D13 = (13, 8,7,3) = ϕ_1"

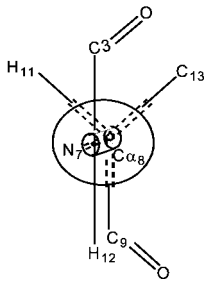
SCHEME 10.

interatomic bonds comprising these "related" dihedral angles.

D_9, D_{11}, and D_{13} could be termed as the "related" dihedral angles associated with the spatial orientations of atom 9, 11 and 13. This also implies that none of them could ever be equivalent or for that matter even proximate to one another. If two dihedral angles are equivalent, i.e., $(9,\underline{\mathbf{8}},7,3) \equiv (11,\underline{\mathbf{8}},7,3) \equiv (13,\underline{\mathbf{8}},7,3)$, then the atoms will be nearly superimposed, which is energetically disallowed.



D9 = (9,<u>**8**</u>,7,3) = ϕ
D11 = (11,<u>**8**</u>,7,3) = ϕ'
D13 = (13,<u>**8**</u>,7,3) = ϕ"

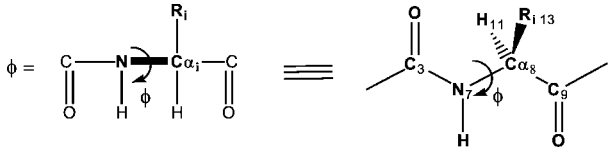


SCHEME 11.

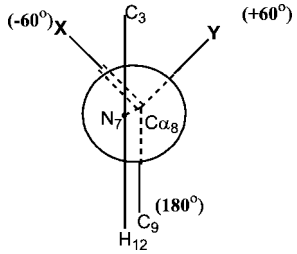
For this set of "related" dihedrals, angles (ϕ, ϕ', ϕ''), defined through a tetrahedral center (C_α #8), one could be confident that these three torsions must be approximately 120° different, where 360° must be divided equally (360/3 = 120°). Consequently, when D9 = 180°, the other two angels must be −60° and +60°. The question of stereochemistry does indeed arise when assigning the particular numerical values (−60° and 60° in this case) with the atomic nuclei in question (#11 and #13 here). This will be explained in the next section as well. The amount of change in any one would have to be reflected in the other two. So if D9 = 0°, the other two angles (previously −60° and 60°) would become +120° and −120°, respectively.

STEREOCHEMISTRY

Let us represent the four atoms defining ϕ (cf. Scheme 12) using a Fischer projection for $\phi = 180°$. In accordance with Schemes 11 and 12, using this standardized numbering we obtain for $\phi = (9,8,7,3) = D9 = 180°$; yet, as we have previously stated, D9 = D11 +/− 120° = D13 −/+ 120°. More specifically, when D11 = D9 + 120°, then D13 = D9 − 120° and when D11 = D9 − 120° then D9 = D13 + 120°. This clearly shows the presence of a stereochemical event. How can one differentiate



SCHEME 12.



SCHEME 13.

numerically between the stereochemical identities of each state?

Let us now represent this, using a Newman projection for $\phi = 180°$; therefore we can place our C (#9) with confidence at the 180° position in Scheme 13. What values do ϕ' ($C_\alpha - H$) and ϕ'' ($C_\alpha - R$) now have? Where do we now place our remaining two atoms, attached to the $C_{\alpha 8}$?

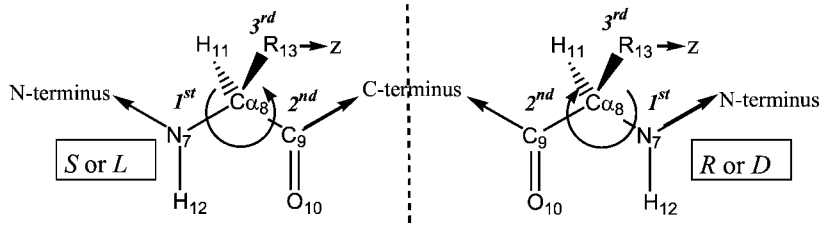
We must assign the X and Y positions to the H and R atomic nuclei of the C_α–H and C_α–R bonds and with the L or D enantiomeric arrangement desired. Note that with the exception of cysteine and selenocysteine, all L-amino acids have *S*-absolute configuration at C_α. Conversely all *D*-amino acids have *R*-absolute configuration at C_α. An examination of the *L* and *D* enantiomers of a peptide "module" represented with the C_α–H pointing away from the viewer (cf. Scheme 14) and behind the plane of the page helps with this determination.

Note the order of preference, whereby the $C_{\alpha 8}$–$\overset{\overset{O_{10}}{\|}}{C_9}$ is second in preference [2^{nd}] after the nitrogen [1^{st}], since no amino acid sidechain has a larger group of atoms than C–C or C–O (with the exception of cysteine and selenocysteine). Therefore the R-group has the third [3^{rd}] priority.

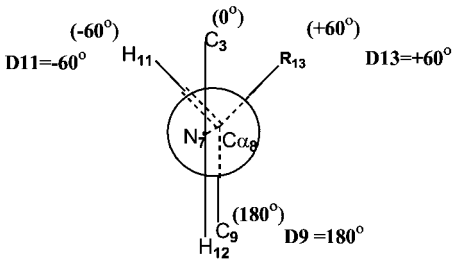
One can clearly see (Scheme 14) that in order to achieve the L-enantiomer for the peptide module using this specific numbering approach, the C_α–H must therefore occupy the $-60°$ position at $\phi_i = 180°$ (β_L); therefore, $\phi_i' = -60°$. Consequently, the C_α–R_i must occupy the $+60°$ position, making $\phi_i'' = +60°$ (therefore D11 = $-60°$ and D13 = $+60°$). For the D enantiomer, the inverse is true; whereby with $\phi_i = 180°$ (β_D), the C_α–H and C_α–R atomic nuclei would occupy the spatial orientations of C_α–H = $+60°$ ($\phi_i' = +60°$) and C_α–R = $-60°$ ($\phi_i'' = -60°$), respectively (therefore D11 = $+60°$ and D13 = $-60°$). These results are summarized in Scheme 15.

As trigonal planar centers have no stereochemical potential (i.e., can never be stereocenters), there is no concern of the numeric ordering of their dihedral angles (Scheme 16). However, all centers have the issue of "related angles" as mentioned above. One torsional definition may not be independently varied without the same change (magnitude and sign/direction of rotation) being applied to all other related centers dihedral angles in the set, about the atomic center in question. As the only existing backbone stereocenter of any and all α-peptides is the C_α, therefore only these centers have to be scrutinized for the numerical ordering of the related dihderal angles about them in terms of their stereochemistry. In the case of the first peptide, the stereochemically dependent dihedrals of the backbone are D9, D11, and D13. Additional general examples follow along with generalized trends and ideas for automation.

Upon examination of the sidechains, two amino acids may have stereocenters along their sidechains, specifically isoleucine and threonine, which have

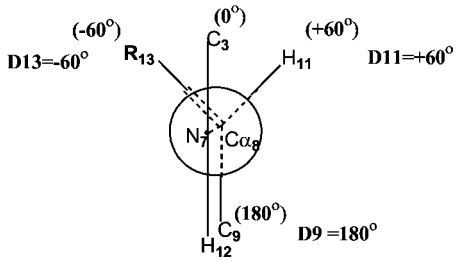


SCHEME 14.



L-Enantiomer

D(C3-**N7-C8**-C9) = D9 = 180 = ϕ_1
D(C3-**N7-C8**-C11) = D11 = -60 = ϕ_1'
D(C3-**N7-C8**-C13) = D13 = +60 = ϕ_1"

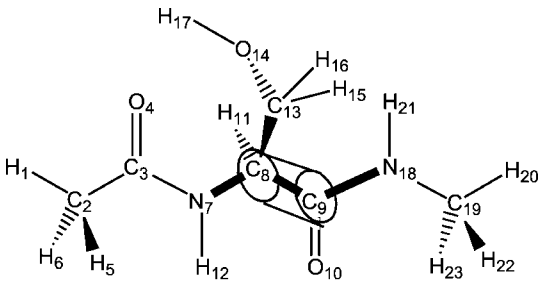


D-Enantiomer

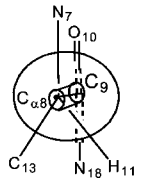
D(C3-**N7-C8**-C9) = D9 = 180 = ϕ_1
D(C3-**N7-C8**-C11) = D11 = +60 = ϕ_1'
D(C3-**N7-C8**-C13) = D13 = -60 = ϕ_1"

SCHEME 15.

stereocenters at their β-carbons. Other amino acids (e.g., proline and lysine) may be oxidized after they have been incorporated in the protein chain. In general all sidechain stereocenters are of *R*-configuration but allowance in the standardized numbering should be made for enantiomeric sidechain configurations as well. It is advisable therefore to discriminate between H_R and H_S in all prochiral CH_2 groups.



D18 = (18,**9**,8,7) = ψ
D10 = (10,**9**,8,7) = ψ'



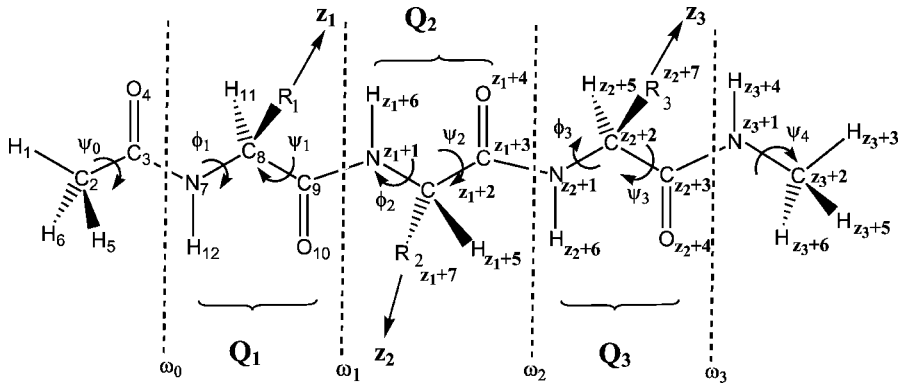
SCHEME 16.

COMPLEX TRENDS IN OLIGOPEPTIDES

Using a tripeptide, as a general example, we can easily see that this atomic numbering method starts at the N-acetyl-end, working its way from left to right, until the N-methyl amide or C-terminus, with priority given to the backbone. R_1, R_2, and R_3 define any sidechain attached to the general structure for any amino acid backbone. This will always be true when the numbering starts from the N-acetyl end, moving from left to right, as it has its own generalized numbering pattern. Using an N-acetyl protecting group, there will always be six atoms on the N-terminus of the peptide chain.

When the numbering has reached the *R*-sidechain, one immediately encounters a problem; it must be taken into consideration that the *R*-sidechain can be different for all 21 naturally occurring amino acids. Knowing that there can be different *R*-sidechains for any of the 21 amino acids, the letter designations z_1, z_2, z_3 account for the number of atoms in all preceding amino acid portions, given the designation Q_1, Q_2, Q_3 for the number of atoms in the first, second, and third amino acid, respectively, while also including the six atoms in the N-acetyl protecting group (Scheme 17).

Note that R_1 tends to z_1 as the amino acid size changes depending on $Q_1 + 6$; also R_2 tends to z_2 because of $Q_2 + Q_1 + 6$, also R_3 tends to z_3 because of



SCHEME 17.

$Q_3+Q_2+Q_1+6$. So in general $z_3 > z_2 > z_1$. Of course many different protecting groups may be used at either the N- or C-termini; therefore the labels P_N and P_C are given to them, respectively, as can be seen in Table II. The dependence of backbone torsional angles on P_N and Q_i is summarized in Table III and those of the sidechain torsional angles are specified in Table IV.

The example in Scheme 18 shows the tripeptide Pro–Cys–Lys, where Q is different for each of the peptide modules; specifically $Q_{Pro} = 14$, $Q_{Cys} = 11$, $Q_{Lys} = 22$ for proline, cysteine, and lysine respectively. If a peptide chain were to have more than these three constituent amino acids, the Q-values of those peptides would also be added in, until the chain reached the C-terminus. Finally, 6 (the number of atomic nuclei in the protecting group) in the N-methyl protecting group is added in to complete the numbering of the whole peptide chain, as seen in Scheme 17 in a general way and in Scheme 18 for a specific case.

TABLE III

The dependence of the "lead" atom number of the backbone torsional angles on P_N and Q_i for various forms of amino acids. The resultant number becomes the dihedral number of that particular torsion (e.g., ϕ_1 = D9 for N-Ac–Ser–NHMe).

		Free amino acid		Protected amino acid	
Dihedral	General formula	Neutral	Zwitterion	N-formyl–ser–NH_2	N-acetyl–ser–NHMe
ψ_0	P_N+1	2	3	4	7
ω_0	P_N+2	3	4	5	8
ϕ_1	P_N+3	4	5	6	9
ψ_1	P_N+1+Q_1	$2+Q_1$	$3+Q_1$	$4+Q_1$	$7+Q_1$
ω_1	P_N+2+Q_1	$3+Q_1$	$4+Q_1$	$5+Q_1$	$8+Q_1$
ϕ_2	P_N+3+Q_1	$4+Q_1$	$5+Q_1$	$6+Q_1$	$9+Q_1$
ψ_2	$P_N+1+Q_1+Q_2$	$2+Q_1+Q_2$	$3+Q_1+Q_2$	$4+Q_1+Q_2$	$7+Q_1+Q_2$
ω_2	$P_N+2+Q_1+Q_2$	$3+Q_1+Q_2$	$4+Q_1+Q_2$	$5+Q_1+Q_2$	$8+Q_1+Q_2$
ϕ_3	$P_N+3+Q_1+Q_2$	$4+Q_1+Q_2$	$5+Q_1+Q_2$	$6+Q_1+Q_2$	$9+Q_1+Q_2$
ψ_3	$P_N+1+Q_1+Q_2+Q_3$	$2+Q_1+Q_2+Q_3$	$3+Q_1+Q_2+Q_3$	$4+Q_1+Q_2+Q_3$	$7+Q_1+Q_2+Q_3$
ω_3	$P_N+2+Q_1+Q_2+Q_3$	$3+Q_1+Q_2+Q_3$	$4+Q_1+Q_2+Q_3$	$5+Q_1+Q_2+Q_3$	$8+Q_1+Q_2+Q_3$
ϕ_3	$P_N+3+Q_1+Q_2+Q_3$	$4+Q_1+Q_2+Q_3$	$5+Q_1+Q_2+Q_3$	$6+Q_1+Q_2+Q_3$	$9+Q_1+Q_2+Q_3$
⋮	⋮	⋮	⋮	⋮	⋮
ϕ_N	$P_N+1\sum_{i=1}^{N} Qi$	$2+\sum_{i=1}^{N} Q_i$	$3+\sum_{i=1}^{N} Q_i$	$4+\sum_{i=1}^{N} Q_i$	$7+\sum_{i=1}^{N} Q_i$
ψ_N	$P_N+2\sum_{i=1}^{N} Qi$	$3+\sum_{i=1}^{N} Q_i$	$4+\sum_{i=1}^{N} Q_i$	$5+\sum_{i=1}^{N} Q_i$	$8+\sum_{i=1}^{N} Q_i$
ω_N	$P_N+3\sum_{i=1}^{N} Qi$	$4+\sum_{i=1}^{N} Q_i$	$5+\sum_{i=1}^{N} Q_i$	$6+\sum_{i=1}^{N} Q_i$	$9+\sum_{i=1}^{N} Q_i$

TABLE IV

The dependence of the "lead" atom number of the sidechain torsional angle on P_N and Q_i for various forms of amino acid.

Peptide	Free amine				N-Formyl				N-Acetyl			
	χ_1	χ_2	χ_3	χ_4	χ_1	χ_2	χ_3	χ_4	χ_1	χ_2	χ_3	χ_4
Gly												
Ala	9				11				14			
Val*	9	10*			11	12*			14	15*		
Leu*	9	10	11*		11	12	13*		14	15	16*	
Ile*	9	10*	11		11	12*	13		14	15*	16	
Pro	8	9	14		10	11	16		13	14	19	
Phe	9	10			11	12			14	15		
Tyr	9	10	20		11	12	22		14	15	25	
Trp	9	10			11	12			14	15		
Met	9	10	11	12	11	12	13	14	14	15	16	17
Cys	9	12			11	14			14	15		
Sel	9	12			11	14			14	15		
His	9	10			11	12			14	15		
Lys	9	10	11	12	11	12	13	14	14	15	16	17
Arg	9	10	11	12	11	12	13	14	14	15	16	17
Ser	9	12			11	14			14	15		
Thr*	9	10*			11	12*			14	15		
Asp	9	11	14		11	13	16		14	15	19	
Glu	9	10	12	17	11	12	14	19	14	15	17	22
Asn	9	11	14		11	13	16		14	15	19	
Gln	9	10	12	17	11	12	14	19	14	15	17	22
Sar									14	15		
Ala–N-Me	9				11							
Hom	9	10	15		11	12	17		14	15	20	

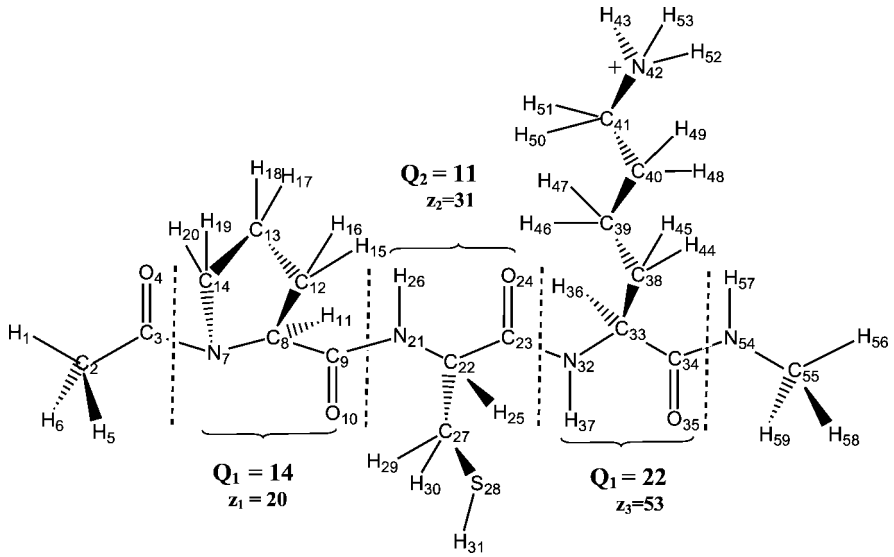
Peptides with more than 4 sidechain torsions

Peptide	Free amine		N-Formyl		N-Acetyl	
	χ_5	χ_6	χ_5	χ_6	χ_5	χ_6
Lys	13		15		18	
Arg	13	14	15	16	18	19

Peptides with branched sidechains, providing the opportunity to define torsions in an alternative fashion

Peptide	Free amine			N-Formyl			N-Acetyl		
	$\chi_{2'}$	$\chi_{3'}$	$\chi_{6'}$	$\chi_{2'}$	$\chi_{3'}$	$\chi_{6'}$	$\chi_{2'}$	$\chi_{3'}$	$\chi_{6'}$
Val	12			14			17		
Leu		13			15			18	
Ile	13			15			18		
Thr	15			17			20		
Arg			16			18			21

Of course there are more peptides that have this duality in sidechain torsion definition. However, only those with equivalent groups (i.e., two methyl groups) are considered here. Groups such as the carboxylate groups (i.e., asp, glu) having two equivalent *atomic* nuclei make use of the torsion—chosen at the researchers' discretion—that fits the trend desired.



SCHEME 18.

DATA BASE HIERARCHY

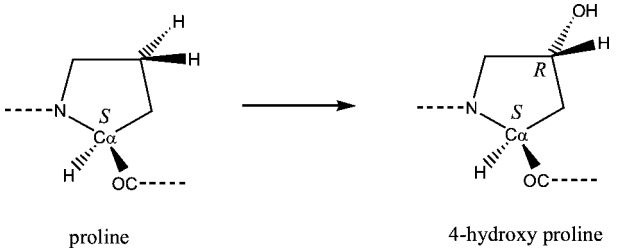
There are 20 DNA encoded amino acids; selenocysteine (Sec) has only an RNA codon but not a DNA codon. Also, amino acid derivatives may be obtained by methylation or demethylation (see Scheme 19). These 21 amino acids and three other derivatives (homocysteine, N-methyl glycine, and N-methyl alanine) have been included in Table II. However, there are other important amino acids, which are produced by degradation or by oxidation of the sidechain, which will be dealt with in subsequent works.

Also, the sidechain oxidation of proline may result in hydroxy-proline (Scheme 20). Therefore a comprehensive amino acid/peptide database must have more than 20 or 21 entries. In fact it should be rather open ended as far as molecular constitution is concerned.

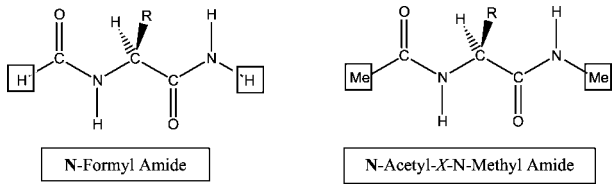
Stereochemistry is also important. Eukaryotes build their bodies in terms of L-amino acids but prokaryotes sometimes use D-amino acids (e.g., in bacterial cell-wall). This chirality is for the backbone, associated with the α-carbon. However, sidechain chirality at the β-, γ-, and δ-carbon may be *R* or *S*. It is predominantly *R* but a database must cater to both enantiomeric forms.

The size of a peptide is an important consideration, whereby we may have mono-, di-, tri-, oligo-, and polypeptides. The term "monopeptides" is used for simple amino acid diamides, which of course have two peptide groups. A recent publication by Topal [50] on the octapeptide HCO–$(Ala)_8$–NH_2 is expected to be superseded soon due to technical developments.

Conformations represent a major challenge: a total of $(3 \times 3)^n = 9^n$ backbone conformations may be expected which for monopeptides ($n = 1$), dipeptides ($n = 2$), and tripeptides ($n = 3$) correspond to 9, 81, and 729 backbone conformations. Each of these is multiplied by the number of sidechain conformations.



SCHEME 19.



SCHEME 20.

TABLE V

The increase in the number of possible backbone conformers (N_0) with the number of amino acid residues (n) in the polypeptide chain.

Degree of polymerization (n)	Number of possible conformers (N_0)
1	$9^1 = 10^{0.954242} = 9.00000 \times 10^0$
10	$9^{10} = 10^{9.542425} = 3.48678 \times 10^9$
100	$9^{100} = 10^{95.42425} = 2.65614 \times 10^{95}$
1000	$9^{1000} = 10^{954.2425} = 1.74787 \times 10^{954}$

For the tripeptide Arg-Gly-Asp (RGD) this leads to a total of

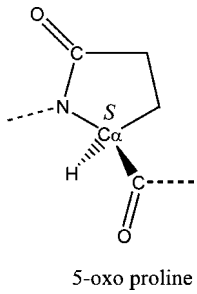
$$\underbrace{3^6}_{\substack{729\\ \text{Backbone}}} \times \underbrace{3^6}_{\substack{729\\ \text{Side chain}}} = 729^2 = 531,\ 441 \text{ conformers.}$$

Clearly, when not a tripeptide ($n = 3$) but a decapeptide ($n = 10$) is considered the number of backbone conformers will increase from $9^3 = 729$ to $9^{10} = 3.48674 \times 10^9$. Table V summarizes the dramatic increase of backbone conformation (N_0) with the increasing degree of polymerization (n):

$$N_0 = 9^n = 10^{n \log 9}.$$

Protecting groups add another dimension. We may have unprotected neutral or zwitterionic amino acids. However, protective groups are often used. The N-formyl group was popular in the 20th century in ab initio computation because of its smaller size. However, the N-acetyl group has becomes manageable; these two possibilities are shown below in Scheme 21, which is also illustrated in Scheme 1. In Scheme 21 the "X" in the N-acetyl-X-N-methyl amide represents the amino acid within the N-acetyl protecting group.

However, esters are also needed in this database and sometimes protections are made in the case of one terminus. For example, in minimicing peptide



5-oxo proline

SCHEME 22.

biosynthesis [41] we may need the following (free amino group but the carboxyl is esterified) as shown in Scheme 22.

Level of theory is the most important in computation. These may include a variety of semiempirical MO, ab initio Hartree–Fock (HF), Møeller Plesset second order (MP2), various density functional theory (DFT), and coupled cluster (i.e., CCSD, CCSDT) methods.

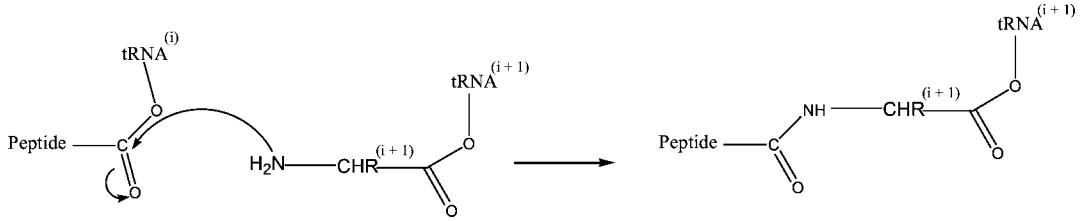
The *basis sets* used may also contain a large variety starting with 3-21G through 6-311++G(d, p) all the way to what facilities will permit. In the long run the sky is no limit.

Figure 7 summarizes the hierarchy outlined above.

Results and Discussion

In Table VI the symbolic z-matrix for serine is shown as an example with various N- and C-terminal protecting groups. The first two columns show the free amino acid in neutral and zwitterionic forms. The third and fourth columns show N- and C-protected forms of serine. In the third column, shorter protective groups are used: HCO–Ser–NH_2 while in the fourth column longer protective groups are used: MeCO–Ser–NHMe.

In Table VII the symbolic z-matrices of the 21 regular amino acids plus 3 special cases, the S-demethylated methionine, called homocysteine, the



SCHEME 21.

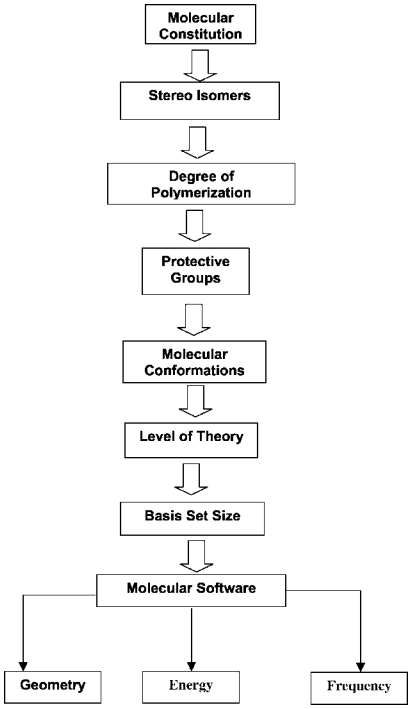


FIGURE 7. A schematic illustration of the data base hierarchy.

N-methylated glycine called sarcosine (Sar), and N-methylated alanine, are summarized.

Of the 24 examples included in Table VII, all have the same bonding pattern except proline, in which the sidechain is folded back and connected to the amide nitrogen, to form a five-member ring.

In the case of proline, if the sidechain is logically managed (to a reduced space), it can be made to bind with the N–C_α. This space is in the vicinity of -30, $+30$, -30 for D12, D13, and D14, respectively. Alternatively, one may use the $+30$, -30, $+30$, for D12, D13, and D14, respectively, if the ring flip parameter is to be used with any other results (i.e., full stereochemical considerations). These values will give the inverse ring state (exo- and endo-nomenclatures are dependent on L or D configuration of proline residue. For the L-enantiomer, the $+, -, +$ series will give the *exo* and the $-, +, -$ result in the *endo*). The D-enantiomer has the inverse relationship (*exo* $= -, +, -$ vs. *endo* $= +, -, +$).

Sidechains may have the same number of atoms such as leucine (Leu) and isoleucine (Ile), which are isomers, or serine (Ser) and cysteine (Cys), which are congeners. These pairs have identical Q numbers yet their sidechains need to be differentiated. Such differentiation is presented in Table VIII.

TABLE VI
Sample input *Z*-matrix for various forms of serine.

Free amino acid		Protected amino acid		
Neutral	Zwitterion	N-Formyl–Ser–NH_2	N-Acetyl–Ser–NHMe	
			H C *1* R2 C *2* R3 *1* A3	P_N (N-terminal protecting group)
	H	H	O *3* R4 *2* A4 *1* D4	
H	N *1* R2	C *1* R2 O *2* R3 *1* A3	H *2* R5 *3* A5 *4* D5 H *2* R6 *3* A6 *4* D6	
N *1* R2 C *2* R3 *1* A3 C *3* R4 *2* A4 *1* D4 O *4* R5 *3* A5 *2* D5 H *3* R6 *2* A6 *1* D6 H *2* R7 *3* A7 *4* D7 C *3* R8 *2* A8 *1* D8	H *2* R3 *1* A3 C *2* R4 *1* A4 *3* D4 C *4* R5 *2* A5 *1* D5 O *5* R6 *4* A6 *2* D6 H *4* R7 *2* A7 *1* D7 H *2* R8 *1* A8 *3* D8 C *4* R9 2 A9 *1* D9	N *2* R4 *1* A4 *3* D4 C *4* R5 *2* A5 *1* D5 C *5* R6 *4* A6 *2* D6 O *6* R7 *5* A7 *4* D7 H *5* R8 *4* A8 *2* D8 H *4* R9 *2* A9 *1* D9 C *5* R10 *4* A10 *2* D10	N *3* R7 *2* A7 *1* D7 C *7* R8 *3* A8 *2* D8 C *8* R9 *7* A9 *3* D9 O *9* R10 *8* A10 *7* D10 H *8* R11 *7* A11 *3* D11 H *7* R12 *3* A12 *2* D12 C *8* R13 *7* A13 *3* D13	Peptide backbone and C_α–C_β (C_α–H for Gly)
O *8* R9 *3* A9 *2* D9 H *8* R10 *3* A10 *2* D10 H *8* R11 *3* A11 *2* D11 H *9* R12 *8* A12 *3* D12	O *9* R10 *4* A10 *2* D10 H *9* R11 *4* A11 *2* D11 H *9* R12 *4* A12 *2* D12 H *10* R13 *9* A13 *4* D13	O *10* R11 *5* A11 *4* D11 H *11* R12 *10* A12 *8* D12 H *10* R13 *5* A13 *4* D13 H *10* R14 *5* A14 *4* D14	O *13* R14 *8* A14 *7* D14 H *14* R15 *13* A15 *8* D15 H *13* R16 *8* A16 *7* D16 H *13* R17 *8* A17 *7* D17	Sidechain
O *4* R13 *3* A13 *2* D13 H *13* R14 *4* A14 *3* D14	O *5* R14 *4* A14 *2* D14	N *9* R18 *8* A18 *7* D18 C *18* R19 *9* A19 *8* D19 H *19* R20 *18* A20 *9* D20 H *18* R21 *9* A21 *8* D21	N *9* R18 *8* A18 *7* D18 C *18* R19 *9* A19 *8* D19 H *19* R20 *18* A20 *9* D20 H *18* R21 *9* A21 *8* D21 H *19* R22 *18* A22 *9* D22 H *19* R23 *18* A23 *9* D23	P_C (C-terminal protecting group)

TABLE VII

Symbolic *Z*-matrix for all the natural and a few special amino acids in their N-acteyl and N-methylated protected forms.

Glycine	Alanine	Valine	Leucine
H	H	H	H
C *1* R2	C *1* R2	C *1* R2	C *1* R2
C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3
O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4
H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5
H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6
N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7
C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8
C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9
O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10
H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11
H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12
H *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13
	H *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14
	H *13* R15 *8* A15 *7* D15	H *14* R15 *13* A15 *8* D15	C *14* R15 *13* A15 *8* D15
	H *13* R16 *8* A16 *7* D16	C *13* R16 *8* A16 *7* D16	H *15* R16 *14* A16 *13* D16
		H *16* R17 *13* A17 *8* D17	C *14* R17 *13* A17 *8* D17
		H *13* R18 *8* A18 *7* D18	H *17* R18 *14* A18 *13* D18
		H *14* R19 *13* A19 *8* D19	H *13* R19 *8* A19 *7* D19
		H *14* R20 *13* A20 *8* D20	H *13* R20 *8* A20 *7* D20
		H *16* R21 *13* A21 *8* D21	H *14* R21 *13* A21 *8* D21
		H *16* R22 *13* A23 *8* D22	H *15* R22 *14* A22 *13* D22
			H *15* R23 *14* A23 *13* D23
			H *17* R24 *14* A24 *13* D24
			H *17* R25 *14* A25 *13* D25
N *9* R14 *8* A14 *7* D14	N *9* R17 *8* A17 *7* D17	N *9* R23 *8* A23 *7* D23	N *9* R26 *8* A26 *7* D26
C *14* R15 *9* A15 *8* D15	C *17* R18 *9* A18 *8* D18	C *23* R24 *9* A24 *8* D24	C *26* R27 *9* A27 *8* D27
H *15* R16 *14* A16 *9* D16	H *18* R19 *17* A19 *9* D19	H *24* R25 *23* A25 *9* D25	H *27* R28 *26* A28 *9* D28
H *14* R17 *9* A17 *8* D17	H *17* R20 *9* A20 *8* D20	H *23* R26 *9* A26 *8* D26	H *26* R29 *9* A29 *8* D29
H *15* R18 *14* A18 *9* D18	H *18* R21 *17* A21 *9* D21	H *24* R27 *23* A27 *9* D27	H *27* R30 *26* A30 *9* D30
H *15* R19 *14* A19 *9* D19	H *18* R22 *17* A22 *9* D22	H *24* R28 *23* A28 *9* D28	H *27* R31 *26* A31 *9* D31

Isoleucine	Proline	Phenylalanine	Tyrosine
H	H	H	H
C *1* R2	C *1* R2	C *1* R2	C *1* R2
C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3
O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4
H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5
H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6
N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7
C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8
C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9
O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10
H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11
H *7* R12 *3* A12 *2* D12		H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12
C *8* R13 *7* A13 *3* D13	C *8* R12 *7* A12 *3* D12	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13

(Continued)

TABLE VII
(Continued)

Isoleucine	Proline	Phenylalanine	Tyrosine
C *13* R14 *8* A14 *7* D14	C *12* R13 *8* A13 *7* D13	C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14
C *14* R15 *13* A15 *8* D15	C *13* R14 *12* A14 *8* D14	C *14* R15 *13* A15 *8* D15	C *14* R15 *13* A15 *8* D15
H *15* R16 *14* A16 *13* D16	H *12* R15 *8* A15 *7* D15	C *15* R16 *14* A16 *13* D16	C *15* R16 *14* A16 *13* D16
C *13* R17 *8* A17 *7* D17	H *12* R16 *8* A16 *7* D16	C *16* R17 *15* A17 *14* D17	C *16* R17 *15* A17 *14* D17
H *17* R18 *13* A18 *8* D18	H *13* R17 *12* A17 *8* D17	C *17* R18 *16* A18 *15* D18	C *17* R18 *16* A18 *15* D18
H *17* R19 *13* A19 *8* D19	H *13* R18 *12* A18 *8* D18	C *18* R19 *17* A19 *16* D19	C *18* R19 *17* A19 *16* D19
H *17* R20 *13* A20 *8* D20	H *14* R19 *13* A19 *12* D19	H *13* R20 *8* A20 *7* D20	H *13* R20 *8* A20 *7* D20
H *13* R21 *8* A21 *7* D21	H *14* R20 *13* A20 *12* D20	H *13* R21 *8* A21 *7* D21	H *13* R21 *8* A21 *7* D21
H *14* R22 *13* A22 *8* D22		H *14* R22 *13* A22 *8* D22	H *15* R22 *14* A22 *13* D22
H *14* R23 *13* A23 *8* D23		H *15* R23 *14* A23 *13* D23	H *16* R23 *15* A23 *14* D23
H *15* R24 *14* A24 *13* D24		H *16* R24 *15* A24 *14* D24	O *17* R24 *16* A24 *15* D24
H *15* R25 *14* A25 *13* D25		H *17* R25 *16* A25 *15* D25	H *25* R25 *17* A25 *16* D25
		H *18* R26 *17* A26 *16* D26	H *18* R26 *17* A26 *16* D26
		H *19* R27 *18* A27 *17* D27	H *19* R27 *18* A27 *17* D27
N *9* R26 *8* A26 *7* D26	N *9* R21 *8* A21 *7* D21	N *9* R28 *8* A28 *7* D28	N *9* R28 *8* A28 *7* D28
C *26* R27 *9* A27 *8* D27	C *21* R22 *9* A22 *8* D22	C *28* R29 *9* A29 *8* D29	C *26* R29 *9* A29 *8* D29
H *27* R28 *26* A28 *9* D28	H *22* R23 *21* A23 *9* D23	H *29* R30 *28* A30 *9* D30	H *27* R30 *26* A30 *9* D30
H *26* R29 *9* A29 *8* D29	H *21* R24 *9* A24 *8* D24	H *28* R31 *9* A31 *8* D31	H *26* R31 *9* A31 *8* D31
H *27* R30 *26* A30 *9* D30	H *22* R25 *21* A25 *9* D25	H *29* R32 *28* A32 *9* D32	H *27* R32 *26* A32 *9* D32
H *27* R31 *26* A31 *9* D31	H *22* R26 *21* A26 *9* D26	H *29* R33 *28* A33 *9* D33	H *27* R33 *26* A33 *9* D33

Tryptophan	Methionine	Cysteine	Selenocysteine
H	H	H	H
C *1* R2	C *1* R2	C *1* R2	C *1* R2
C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3
O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4
H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5
H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6
N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7
C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8
C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9
O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10
H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11
H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12
C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13
C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14	S *13* R14 *8* A14 *7* D14	Se *13* R14 *8* A14 *7* D14
C *14* R15 *13* A15 *8* D15	S *14* R15 *13* A15 *8* D15	H *13* R15 *8* A15 *7* D15	H *13* R15 *8* A15 *7* D15
C *15* R16 *14* A16 *13* D16	C *15* R16 *14* A16 *13* D16	H *13* R16 *8* A16 *7* D16	H *13* R16 *8* A16 *7* D16
C *16* R17 *15* A17 *14* D17	H *16* R17 *15* A17 *14* D17	H *14* R17 *13* A17 *8* D17	H *14* R17 *13* A17 *8* D17
C *17* R18 *16* A18 *15* D18	H *13* R18 *8* A18 *7* D18		
C *18* R19 *17* A19 *16* D19	H *13* R19 *8* A19 *7* D19		
C *19* R20 *18* A20 *17* D20	H *14* R20 *13* A20 *8* D20		
C *20* R21 *19* A21 *18* D21	H *14* R21 *13* A21 *8* D21		
C *21* R22 *20* A22 *19* D22	H *16* R22 *15* A22 *14* D22		
H *13* R23 *8* A23 *7* D23	H *16* R23 *15* A23 *14* D23		
H *13* R24 *8* A24 *7* D24			
H *15* R25 *14* A25 *13* D25			
H *16* R26 *15* A26 *14* D26			
H *19* R27 *18* A27 *17* D27			

(Continued)

TABLE VII
(Continued)

Tryptophan	Methionine	Cysteine	Selenocysteine
H 20 R28 19 A28 18 D28			
H 21 R29 20 A29 19 D29			
H 22 R30 21 A30 20 D30			
N 9 R31 8 A31 7 D31	N 9 R24 8 A24 7 D24	N 9 R18 8 A18 7 D18	N 9 R18 8 A18 7 D18
C 26 R32 9 A32 8 D32	C 24 R25 9 A25 8 D25	C 18 R19 9 A19 8 D19	C 18 R19 9 A19 8 D19
H 27 R33 26 A33 9 D33	H 25 R26 24 A26 9 D26	H 19 R20 18 A20 9 D20	H 19 R20 18 A20 9 D20
H 26 R34 9 A34 8 D34	H 24 R27 9 A27 8 D27	H 18 R21 9 A21 8 D21	H 18 R21 9 A21 8 D21
H 27 R35 26 A35 9 D35	H 25 R28 24 A28 9 D28	H 19 R22 18 A22 9 D22	H 19 R22 18 A22 9 D22
H 27 R36 26 A36 9 D36	H 25 R29 24 A29 9 D29	H 19 R23 18 A23 9 D23	H 19 R23 18 A23 9 D23
Histidine	Lysine	Arginine	Serine
H	H	H	H
C 1 R2	C 1 R2	C 1 R2	C 1 R2
C 2 R3 1 A3	C 2 R3 1 A3	C 2 R3 1 A3	C 2 R3 1 A3
O 3 R4 2 A4 1 D4	O 3 R4 2 A4 1 D4	O 3 R4 2 A4 1 D4	O 3 R4 2 A4 1 D4
H 2 R5 3 A5 4 D5	H 2 R5 3 A5 4 D5	H 2 R5 3 A5 4 D5	H 2 R5 3 A5 4 D5
H 2 R6 3 A6 4 D6	H 2 R6 3 A6 4 D6	H 2 R6 3 A6 4 D6	H 2 R6 3 A6 4 D6
N 3 R7 2 A7 1 D7	N 3 R7 2 A7 1 D7	N 3 R7 2 A7 1 D7	N 3 R7 2 A7 1 D7
C 7 R8 3 A8 2 D8	C 7 R8 3 A8 2 D8	C 7 R8 3 A8 2 D8	C 7 R8 3 A8 2 D8
C 8 R9 7 A9 3 D9	C 8 R9 7 A9 3 D9	C 8 R9 7 A9 3 D9	C 8 R9 7 A9 3 D9
O 9 R10 8 A10 7 D10	O 9 R10 8 A10 7 D10	O 9 R10 8 A10 7 D10	O 9 R10 8 A10 7 D10
H 8 R11 7 A11 3 D11	H 8 R11 7 A11 3 D11	H 8 R11 7 A11 3 D11	H 8 R11 7 A11 3 D11
7 R12 3 A12 2 D12	H 7 R12 3 A12 2 D12	H 7 R12 3 A12 2 D12	H 7 R12 3 A12 2 D12
C 8 R13 7 A13 3 D13	C 8 R13 7 A13 3 D13	C 8 R13 7 A13 3 D13	C 8 R13 7 A13 3 D13
C 13 R14 8 A14 7 D14	C 13 R14 8 A14 7 D14	C 13 R14 8 A14 7 D14	O 13 R14 8 A14 7 D14
C 14 R15 13 A15 8 D15	C 14 R15 13 A15 8 D15	C 14 R15 13 A15 8 D15	H 13 R15 8 A15 7 D15
N 15 R16 14 A16 13 D16	C 15 R16 14 A16 13 D16	N 15 R16 14 A16 13 D16	H 13 R16 8 A16 7 D16
C 16 R17 15 A17 14 D17	N 16 R17 15 A17 14 D17	C 16 R17 15 A17 14 D17	H 14 R17 13 A17 8 D17
N 17 R18 16 A18 15 D18	H 17 R18 16 A18 15 D18	N 17 R18 16 A18 15 D18	
H 13 R19 8 A19 7 D19	H 13 R19 8 A19 7 D19	H 18 R19 17 A19 16 D19	
H 13 R20 8 A20 7 D20	H 13 R20 8 A20 7 D20	N 17 R20 16 A20 15 D20	
H 18 R21 17 A21 16 D21	H 14 R21 13 A21 8 D21	H 20 R21 17 A21 16 D21	
	H 14 R22 13 A22 8 D22	H 13 R22 8 A22 7 D22	
	H 15 R23 14 A23 13 D23	H 13 R23 8 A23 7 D23	
	H 15 R24 14 A24 13 D24	H 14 R24 13 A24 8 D24	
	H 16 R25 15 A25 14 D25	H 14 R25 13 A25 8 D25	
	H 16 R26 15 A26 14 D26	H 15 R26 14 A26 13 D26	
	H 17 R27 16 A27 15 D27	H 15 R27 14 A27 13 D27	
	H 17 R28 16 A28 15 D28	H 16 R28 15 A28 14 D28	
		H 18 R29 17 A29 16 D29	
		H 20 R30 17 A30 16 D30	
N 9 R22 8 A22 7 D22	N 9 R29 8 A29 7 D29	N 9 R31 8 A31 7 D31	N 9 R18 8 A18 7 D18
C 22 R23 9 A23 8 D23	C 29 R30 9 A30 8 D30	C 31 R32 9 A32 8 D32	C 18 R19 9 A19 8 D19
H 23 R24 22 A24 9 D24	H 30 R31 29 A31 9 D31	H 32 R33 31 A33 9 D33	H 19 R20 18 A20 9 D20
H 22 R25 9 A25 8 D25	H 29 R32 9 A32 8 D32	H 31 R34 9 A34 8 D34	H 18 R21 9 A21 8 D21
H 23 R26 22 A26 9 D26	H 30 R33 29 A33 9 D33	H 32 R35 31 A35 9 D35	H 19 R22 18 A22 9 D22
H 23 R27 22 A27 9 D27	H 30 R34 29 A34 9 D34	H 32 R36 31 A36 9 D36	H 19 R23 18 A23 9 D23

(Continued)

TABLE VII
(Continued)

Threonine	Aspartic Acid	Glutamic Acid	Asparagine
H	H	H	H
C *1* R2	C *1* R2	C *1* R2	C *1* R2
C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3
O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4
H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5
H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6
N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7
C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8
C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9
O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10
H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11
H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12
C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13
C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14
H *14* R15 *13* A15 *8* D15	O *14* R15 *13* A15 *8* D15	C *14* R15 *13* A15 *8* D15	O *14* R15 *13* A15 *8* D15
O *13* R16 *8* A16 *7* D16	O *14* R16 *13* A16 *8* D16	O *15* R16 *14* A16 *13* D16	N *14* R16 *13* A16 *8* D16
H *13* R17 *8* A17 *7* D17	H *13* R17 *8* A17 *7* D17	O *15* R17 *14* A17 *13* D17	H *16* R17 *14* A17 *13* D17
H *14* R18 *13* A18 *8* D18	H *13* R18 *8* A18 *7* D18	H *13* R18 *8* A18 *7* D18	H *13* R18 *8* A18 *7* D18
H *14* R19 *13* A19 *8* D19	H *16* R19 *14* A19 *13* D19	H *13* R19 *8* A19 *7* D19	H *13* R19 *8* A19 *7* D19
H *16* R20 *13* A20 *8* D20		H *14* R20 *13* A20 *8* D20	H *16* R20 *14* A20 *13* D20
		H *14* R21 *13* A21 *8* D21	
		H *17* R22 *15* A22 *14* D22	
N *9* R21 *8* A21 *7* D21	N *9* R20 *8* A20 *7* D20	N *9* R23 *8* A23 *7* D23	N *9* R21 *8* A21 *7* D21
C *21* R22 *9* A22 *8* D22	C *20* R21 *9* A21 *8* D21	C *23* R24 *9* A24 *8* D24	C *21* R22 *9* A22 *8* D22
H *22* R23 *21* A23 *9* D23	H *21* R22 *20* A22 *9* D22	H *24* R25 *23* A25 *9* D25	H *22* R23 *21* A23 *9* D23
H *21* R24 *9* A24 *8* D24	H *20* R23 *9* A23 *8* D23	H *23* R26 *9* A26 *8* D26	H *21* R24 *9* A24 *8* D24
H *22* R25 *21* A25 *9* D25	H *21* R24 *20* A24 *9* D24	H *24* R27 *23* A27 *9* D27	H *22* R25 *21* A25 *9* D25
H *22* R26 *21* A26 *9* D26	H *21* R25 *20* A25 *9* D25	H *24* R28 *23* A28 *9* D28	H *22* R26 *21* A26 *9* D26

Glutamine	Homocysteine	N-Methyl-Glycine	N-Methyl-Alanine
H	H	H	H
C *1* R2	C *1* R2	C *1* R2	C *1* R2
C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3	C *2* R3 *1* A3
O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4	O *3* R4 *2* A4 *1* D4
H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5	H *2* R5 *3* A5 *4* D5
H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6	H *2* R6 *3* A6 *4* D6
N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7	N *3* R7 *2* A7 *1* D7
C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8	C *7* R8 *3* A8 *2* D8
C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9	C *8* R9 *7* A9 *3* D9
O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10	O *9* R10 *8* A10 *7* D10
H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11	H *8* R11 *7* A11 *3* D11
H *7* R12 *3* A12 *2* D12	H *7* R12 *3* A12 *2* D12	C *7* R12 *3* A12 *2* D12	C *7* R12 *3* A12 *2* D12
C *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13	H *8* R13 *7* A13 *3* D13	C *8* R13 *7* A13 *3* D13
C *13* R14 *8* A14 *7* D14	C *13* R14 *8* A14 *7* D14		H *13* R14 *8* A14 *7* D14
C *14* R15 *13* A15 *8* D15	S *14* R15 *13* A15 *8* D15		H *13* R15 *8* A15 *7* D15
O *15* R16 *14* A16 *13* D16	H *13* R16 *8* A16 *7* D16		H *13* R16 *8* A16 *7* D16
N *15* R17 *14* A17 *13* D17	H *13* R17 *8* A17 *7* D17		
H *17* R18 *15* A18 *14* D18	H *14* R18 *13* A18 *8* D18		
H *13* R19 *8* A19 *7* D19	H *14* R19 *13* A19 *8* D19		
H *13* R20 *8* A20 *7* D20	H *15* R20 *14* A20 *13* D20		
H *14* R21 *13* A21 *8* D21			
H *14* R22 *13* A22 *8* D22			
H *17* R23 *15* A23 *14* D23			
		H *12* R14 *7* A14 *3* D14	H *12* R17 *7* A17 *3* D17
		H *12* R15 *7* A15 *3* D15	H *12* R18 *7* A18 *3* D18
		H *12* R16 *7* A16 *3* D16	H *12* R19 *7* A19 *3* D19

(Continued)

TABLE VII
(Continued)

Glutamine	Homocysteine	N-Methyl-Glycine	N-Methyl-Alanine
N *9* R24 *8* A24 *7* D24	N *9* R21 *8* A21 *7* D21	N *9* R17 *8* A17 *7* D17	N *9* R20 *8* A20 *7* D20
C *21* R25 *9* A25 *8* D25	C *21* R22 *9* A22 *8* D22	C *17* R18 *9* A18 *8* D18	C *20* R21 *9* A21 *8* D21
H *22* R26 *21* A26 *9* D26	H *22* R23 *21* A23 *9* D23	H *18* R19 *17* A19 *9* D19	H *21* R22 *20* A22 *9* D22
H *21* R27 *9* A27 *8* D27	H *21* R24 *9* A24 *8* D24	H *17* R20 *9* A20 *8* D20	H *20* R23 *9* A23 *8* D23
H *22* R28 *21* A28 *9* D28	H *22* R25 *21* A25 *9* D25	H *18* R21 *17* A21 *9* D21	H *21* R24 *20* A24 *9* D24
H *22* R29 *21* A29 *9* D29	H *22* R26 *21* A26 *9* D26	H *18* R22 *17* A22 *9* D22	H *21* R25 *20* A25 *9* D25

Structures are usually preoptimized at the AM1 level of theory. These resulting dihedral angles are shown in Table IX. Such preoptimizations are usually helpful in making a good "educated guess" for ab initio input geometries, particularly where bond lengths and bond angles are concerned. The ab initio calculations carried out at the RHF/3-21G level of theory represent usually the first

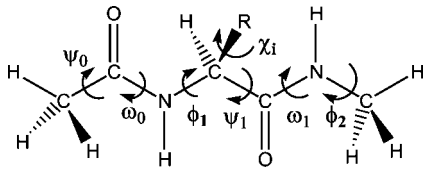
TABLE VIII
Differentiation of amino acids with identical *Q* numbers, according to their sidechain atomic nuclei ordering.

Amino acid	*Q*#	β	γ	δ	ε	ϕ	Number of χ's	Neglected sidechain dihedrals[b]
Aliphatic								
Gly	7	H						H
Ala	10	C	H					$1 \times$ Me
Val	16	C	C	C(δ'=C)[a]	H		1	$2 \times$ Me
Leu	19	C	C	C	C(ε'=C)[a]		2	$2 \times$ Me
Ile	19	C	C	C(δ'=C)[a]	C	H	2	$2 \times$ Me
Pro	14	C	C	C	C	C	3	
Aromatic								
Phe	20	C	C	C	C	C	2	
Tyr	21	C	C	C	C	C	3	
Trp	24	C	C	C	C	N	2	
Alcohols								
Ser	11	C	C	O	H		2	
Thr	12	C	C(γ'=O)[a]	H	H(ε'=C)[a]		2	$1 \times$ Me
Sulphur/Selenium Containing								
Met	17	C	C	S	C	H	3	$1 \times$ Me
Cys	11	C	C	S	H		2	
Sel	11	C	C	Se	H		2	
Acids								
Asp	11	C	C	C	O(ε'=O)[a]		2	
Glu	15	C	C	C	C	O(ϕ'=O)[a]	3	
Bases								
His	17	C	C	C	N	C	2	
Lys	22	C	C	C	C	N	5	
Arg	24	C	C	C	N	N	5	$2 \times NH_2$
Amides								
Asp	14	C	C	C	O(ε'=N)[a]	H	2	$1 \times NH_2$
Gln	17	C	C	C	C	O(ϕ'=N)[a]	3	$1NH_2$

[a] At these positions, along the sidechain, moving away from the peptide backbone (i.e., going from C_α to X_β, X_γ, X_δ, etc., with X being the atomic nuclei specific to that amino acid), we find that branching brings about a "duality" in atomic nuclei nomenclature.
[b] These amino acids do not have any sidechain torsions; methyl-hydrogen spatial orientations are not considered as "true" torsions here; however, the sidechain group is still identified.

TABLE IX

Monopeptides (N-acetyl-amino acid-NHMe) in their β_L or γ_L backbone conformations optimized at the AM1 level of theory.



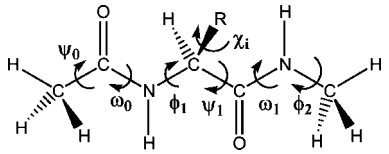
Amino acid	Dihedral angles										
	Energy	ψ_0	ω_0	ϕ_1	ψ_1	ω_1	ϕ_2	χ_1	χ_2	χ_3	χ_4
Glycine	−0.1323794	180.00	179.99	−179.96	179.98	180.00	180.00				
Alanine	−0.1403346	177.06	171.80	−106.29	144.64	−178.81	−11.06	−178.15			
Valine	−0.1587100	178.96	−179.34	−83.83	77.29	−178.47	−7.61	−169.83	−177.10		
Leucine	−0.1680543	178.94	−179.94	−83.77	75.40	−178.09	−5.12	−144.69	170.41	−178.09	
Isoleucine	−0.1678695	179.10	−179.43	−83.68	78.57	−177.97	−9.34	−168.25	−160.98	178.38	
Proline	0.0560454	−171.17	−169.95	−76.57	51.30	150.80	16.37	3.68	−2.48	0.38	2.06
Phenylalanine	−0.0977912	179.06	173.61	−142.36	170.93	178.86	11.40	−143.12	−117.49		
Tyrosine	−0.1686142	178.96	174.08	−144.42	171.06	179.02	10.45	−141.97	−117.36	−178.92	
Tryptophan	−0.0473614	178.99	173.57	−144.28	174.68	178.39	14.82	−136.86	63.62		
Methionine	−0.1491683	179.41	−178.50	−84.80	69.10	179.17	2.21	−163.89	−163.67	82.95	−150.51
Cysteine	−0.1250601	177.60	171.39	−132.24	166.80	176.63	17.18	−150.19	72.27		
Selenocysteine	−0.0497119	178.80	−179.49	−83.69	68.02	178.09	5.14	61.17	14.18		
Histidine	0.0614690	178.68	174.81	−162.01	165.45	174.55	26.15	−167.32	161.13		
Lysine	0.0487048	−178.64	−174.25	−86.90	61.47	−179.48	13.21	−159.91	−159.73	106.44	−52.10
Arginine	0.0845921	173.83	166.32	−132.91	145.92	−179.34	−8.90	−93.79	83.93	−179.27	77.43
Serine	−0.2165257	178.00	170.50	−133.78	174.99	176.28	19.39	−155.60	72.35		
Threonine	−0.2220033	−177.42	−168.96	−91.17	71.28	−178.09	−6.97	−166.43	−167.03		
Aspartic acid	−0.2862105	177.72	171.01	−122.01	149.21	177.69	11.59	−155.45	171.97	−179.52	
Glutamic acid	−0.2975508	174.43	−178.45	−85.15	68.59	179.40	1.93	−167.11	172.87	172.66	−179.67
Asparagine	−0.2072116	179.30	176.26	−167.22	175.15	177.63	18.52	−134.04	84.93	−177.25	
Glutamine	−0.2175854	−179.39	−177.66	−85.23	68.10	179.34	2.01	−165.85	−145.28	70.18	−179.94
N-Methyl glycine	−0.1273086	−177.82	−179.49	−84.32	62.80	176.15	11.71				
N-Methyl alanine	−0.1305043	167.87	−174.84	−90.22	71.07	178.82	1.39	−172.69			
Homocysteine	−0.1393681	179.31	−178.84	−84.65	70.09	179.28	1.29	−164.53	−177.50	68.39	

Amino acids with sidechains longer than χ^4

	χ_5	χ_6
Arginine	−179.16	−176.34
Lysine	−158.53	

TABLE X

Monopeptides (N-acetyl-aminoacid-NHMe) in their β_L, γ_L or ε_D backbone conformations optimized at the RHF/3-21G level of theory.



Amino acid	Dihedral angles										
	Energy	ψ_0	ω_0	ϕ_1	ψ_1	ω_1	ϕ_2	χ_1	χ_2	χ_3	χ_4
Glycine	−451.2927807	179.91	180.00	−179.97	−179.98	180.00	180.00				
Alanine	−490.1181535	179.53	178.77	−167.03	169.46	177.93	−0.77	179.54			
Valine	−567.7552433	−179.49	176.78	−136.44	141.46	179.02	−1.99	−174.58	−178.68		
Leucine	−606.5711127	−179.35	−179.87	−161.09	149.63	178.88	1.26	−176.42	144.61	179.22	
Isoleucine	−606.5724303	−179.30	177.38	−134.99	136.40	179.77	−3.22	−174.23	−167.46	176.94	
Proline[a]	−566.4027904	172.15	−176.19	−77.22	57.41	158.86	16.75	37.92	−42.53	30.04	−6.29
Phenylalanine	−718.3880988	178.52	177.60	−165.98	173.09	179.16	1.20	−146.85	−111.65		
Tyrosine	−792.8287435	178.56	177.68	−165.94	172.96	179.07	0.87	−145.80	−111.79	−177.94	
Tryptophan	−848.4237310	178.76	177.43	−165.86	172.14	178.18	−2.51	−161.93	76.21		
Methionine	−963.3094976	−179.05	−179.94	−161.73	153.76	178.54	−3.80	−174.10	177.74	179.93	−179.77
Cysteine	−885.6727478	178.42	176.78	−168.18	173.42	178.86	3.56	−162.59	79.05		
Selenocysteine	−2878.5966717	178.73	177.66	−168.17	168.55	178.48	3.12	−164.58	83.20		
Histidine	−712.5299564	177.95	177.37	−165.88	174.09	179.48	0.47	−126.07	−117.74		
Lysine	−661.6690128	177.37	178.22	−166.87	147.61	−179.45	−11.13	−173.82	177.06	−178.36	179.13
Arginine	−770.0338086	169.24	172.15	−160.46	166.73	178.59	−2.99	−83.55	102.29	178.72	85.24
Serine	−564.5586773	176.46	174.02	−169.55	−173.50	−178.81	8.50	−172.88	158.03		
Threonine	−603.3659325	−172.99	−172.64	−137.18	131.87	−179.89	−4.84	−174.94	178.08	−54.91	
Aspartic acid	−676.6879695	177.83	176.74	−170.03	171.55	179.87	−179.50	−162.41	174.53	−178.36	
Glutamic acid	−715.4975830	−178.57	−179.56	−161.92	153.79	178.56	−6.38	−174.21	177.72	−178.84	179.86
Asparagine	−656.9738844	176.62	176.02	−172.54	174.40	−179.75	178.50	−159.29	164.96	−179.82	
Glutamine	−695.7782314	−177.78	−179.02	−160.26	152.38	178.79	−5.90	−173.63	179.81	−174.74	−179.62
N-Methyl glycine[b]	−490.0994309	−178.03	−172.50	87.17	−174.02	−179.88	3.21				
N-Methyl alanine[a]	−528.9273762	168.17	−174.80	−93.73	77.46	−176.71	−8.67	−170.72			
Homocysteine	−924.4885309	−179.33	179.90	−161.77	152.88	178.56	−5.22	−174.27	176.90	66.47	

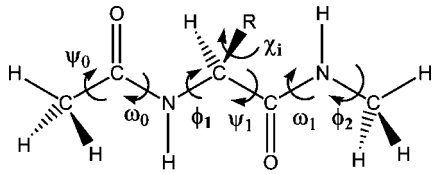
Amino acids with sidechains longer than χ^4

	χ_5	χ_6
Arginine	178.52	177.52
Lysine	−179.96	

[a] γ_L Backbone conformation.
[b] ε_D Backbone conformation.

TABLE XI

Monopeptides (N-acetyl-aminoacid-NHMe) in their β_L, γ_L or γ_D backbone conformations optimized at the RHF/6-31G(d) level of theory.



Amino acid	Dihedral angles										
	Energy	ψ_0	ω_0	ϕ_1	ψ_1	ω_1	ϕ_2	χ_1	χ_2	χ_3	χ_4
Glycine	−453.8239710	180.00	180.00	−179.99	179.97	180.00	180.00				
Alanine	−492.8608896	−135.98	−179.97	−157.32	158.86	179.38	−169.24	−178.50			
Valine[a]	−570.9281584	157.39	168.64	−85.57	159.74	−179.18	−169.19	−64.98	−175.03		
Leucine	−609.9591406	−143.94	−176.17	−150.92	133.94	−178.94	−170.95	−176.56	145.99	179.38	
Isoleucine[a]	−609.9612693	141.76	176.18	−85.80	89.22	−171.76	−169.45	−177.31	−170.22	175.21	
Proline[a]	−569.5607586	172.85	−174.62	−79.44	54.35	155.07	19.10	36.43	−40.80	28.56	−5.67
Phenylalanine	−722.4080463	124.77	178.81	−156.86	149.68	172.83	166.99	−169.75	−107.26		
Tyrosine	−797.2626363	127.71	177.88	−157.43	153.19	173.53	8.09	−165.61	−106.78	−177.79	
Tryptophan	−853.1742948	−14.12	−178.44	−156.78	143.40	172.63	163.35	−174.49	87.87		
Methionine	−968.4341365	−144.54	−176.31	−151.14	137.12	−178.46	−167.10	−175.47	176.19	179.00	−179.94
Cysteine	−890.3679926	144.40	173.97	−159.23	164.06	175.58	172.80	−160.83	76.41		
Selenocysteine	−2890.4359045	145.76	173.39	−158.65	166.15	176.11	172.74	−161.50	67.71		
Histidine	−716.5285614	4.36	175.95	−163.21	170.61	177.25	173.44	−130.68	−114.77		
Lysine	−665.3571496	−156.71	−172.10	−155.85	37.36	174.35	169.87	−158.24	−176.29	179.44	−178.47
Arginine	−774.3144690	14.27	−170.96	−144.80	155.41	179.97	−166.68	−75.15	88.30	178.50	86.45
Serine	−567.7119188	153.54	169.62	−155.62	178.05	177.32	171.59	−170.35	168.76		
Threonine[a]	−606.7476366	−154.20	−165.61	−88.40	76.38	−174.44	−172.42	−168.11	179.08	−46.71	
Aspartic acid	−680.4811329	143.30	173.99	−160.53	163.13	178.19	178.55	−160.60	177.88	−178.49	
Glutamic acid	−719.5104346	−146.89	−175.55	−152.27	137.23	−178.23	−167.34	−174.43	176.99	−179.17	−179.97
Asparagine	−660.6519753	−111.11	173.10	−166.31	−177.21	179.96	175.86	−137.69	92.92	−169.45	
Glutamine	−699.6754531	−148.13	−174.82	−148.66	133.68	−177.09	−167.30	−175.46	176.58	175.56	173.16
N-Methyl glycine[b]	−492.8482044	−159.94	174.95	91.15	−81.59	174.12	169.03				
N-Methyl alanine[a]	−531.8827346	165.60	−175.03	−94.30	83.29	−173.28	−168.85	−173.18			
Homocysteine	−929.4001868	−145.35	−176.21	−151.22	135.26	−178.58	−168.62	−174.97	176.45	65.85	

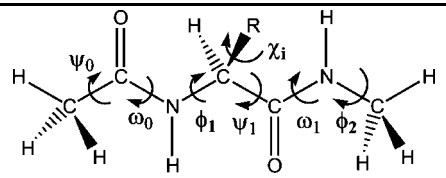
Amino acids with sidechains longer than χ^4

	χ_5	χ_6
Arginine	179.83	179.78
Lysine	180.00	

[a] γ_L Backbone conformation.
[b] γ_D Backbone conformation.

TABLE XII

Monopeptide (N-acetyl-aminoacid-NHMe) in their β_L, γ_L or γ_D backbone conformations optimized at the B3LYP/6-31G(d) level of theory.



Amino acid	Dihedral angles										
	Energy	ψ_0	ω_0	ϕ_1	ψ_1	ω_1	ϕ_2	χ_1	χ_2	χ_3	χ_4
Glycine	−456.5352684	180.00	180.00	−179.99	−179.99	180.00	180.00				
Alanine	−495.8528898	−121.65	176.74	−158.40	164.22	177.64	−0.83	−177.94			
Valine	−574.4787405	121.03	173.45	−126.70	132.89	179.60	−1.75	179.07	179.36		
Leucine	−613.7896911	−124.16	−178.20	−154.06	138.00	178.07	2.50	−177.32	146.24	179.62	
Isoleucine[c]	−613.7904969	170.36	176.46	−129.04	130.96	179.71	−1.38	−179.65	−173.93	175.38	
Proline[a]	−573.0831617	170.36	−174.31	−75.52	54.50	154.36	18.61	35.52	−41.24	29.81	−7.51
Phenylalanine	−726.9049625	128.44	174.49	−158.19	170.73	177.34	8.36	−155.17	−111.23		
Tyrosine[d]	−802.1210157	141.61	172.49	−157.95	173.47	178.23	−92.20	−153.55	−111.56	−179.18	
Tryptophan	−858.4741567	2.22	176.39	−159.58	166.64	173.42	11.80	−162.63	76.57		
Methionine	−972.6616179	−127.91	−179.19	−153.61	144.61	179.24	−7.34	−173.33	176.59	178.87	−179.96
Cysteine	−894.038196	130.09	173.44	−159.94	172.73	177.00	14.93	−163.13	72.86		
Selenocysteine[c]	−2895.230453	124.23	173.15	−159.41	172.84	177.12	24.75	−164.55	63.25		
Histidine[c]	−720.8807901	2.82	175.30	−162.15	173.11	177.76	10.80	−127.39	−115.23		
Lysine[c]	−669.5067466	6.02	−174.30	−151.57	35.62	175.83	6.93	−160.03	−177.21	179.08	−178.15
Arginine[d]	−779.065637	−3.61	169.08	−149.10	162.54	180.61	−160.82	−75.91	89.90	178.57	92.88
Serine	−571.0620126	123.28	171.17	−159.62	−173.99	178.77	22.52	−170.03	178.27		
Threonine[a]	−610.3788778	118.41	−166.45	−83.78	69.28	−176.75	−16.83	−162.58	−179.95		
Aspartic acid	−684.4240259	−115.40	175.09	−163.33	167.54	178.54	2.83	−161.65	173.57	−179.18	
Glutamic acid	−723.7299128	−60.04	−179.46	−154.06	144.06	179.66	−8.70	−172.51	177.74	−178.64	−179.96
Asparagine	−664.5621495	−115.95	173.63	−167.28	−175.14	−179.28	5.82	−137.70	−167.37	−167.37	
Glutamine	−703.8607095	4.97	179.40	−155.21	148.39	−179.60	−22.21	−169.36	−176.11	−137.61	174.85
N-Methyl glycine[b]	−495.8444405	−170.61	176.77	87.15	−79.54	174.47	18.64				
N-Methyl alanine[a]	−535.1593261	166.95	−176.80	−89.45	81.32	−173.75	−18.38	−171.58			
Homocysteine	−933.3475112	−121.58	179.78	−152.91	143.16	179.71	−7.56	−172.79	176.07	63.85	

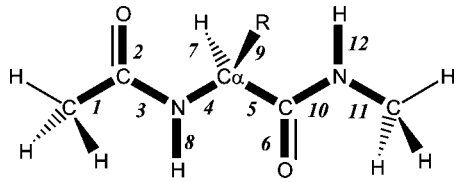
Amino acids with sidechains longer than χ^4

	χ_5	χ_6
Arginine	177.44	176.63
Lysine	179.86	

[a] γ_L Backbone conformation.
[b] γ_D Backbone conformation.
[c] Energy fully converged, geometry only partially converged.
[d] Energy and geometry partially converged.

TABLE XIII

Bond lengths in angstroms optimized at the AM1 level of theory.

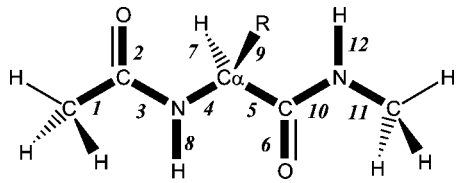


Amino acids	Bond lengths											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Glycine	1.5077	1.2479	1.3802	1.4329	1.5342	1.2477	1.1299	0.9936	1.1299	1.3769	1.4293	0.9899
Alanine	1.5084	1.2471	1.3877	1.4413	1.5523	1.2480	1.1332	0.9950	1.5300	1.3754	1.4284	0.9914
Valine	1.5081	1.2498	1.3804	1.4408	1.5558	1.2471	1.1342	0.9911	1.5442	1.3733	1.4276	0.9967
Leucine	1.5081	1.2497	1.3809	1.4430	1.5561	1.2470	1.1349	0.9912	1.5371	1.3736	1.4276	0.9966
Isoleucine	1.5080	1.2498	1.3805	1.4408	1.5560	1.2469	1.1342	0.9911	1.5458	1.3737	1.4277	0.9968
Proline	1.5090	1.2557	1.3739	1.4644	1.5191	1.4501	1.1320	1.4499[a]	1.5481	1.4961	1.4612	1.0244
Phenylalanine	1.5081	1.2485	1.3834	1.4417	1.5486	1.2487	1.1367	0.9948	1.5464	1.3748	1.4282	0.9945
Tyrosine	1.5081	1.2488	1.3826	1.4417	1.5484	1.2486	1.1366	0.9947	1.5468	1.3748	1.4282	0.9944
Tryptophan	1.5083	1.2495	1.3823	1.4422	1.5473	1.2491	1.1370	0.9949	1.5495	1.3751	1.4283	0.9961
Methionine	1.5078	1.2494	1.3811	1.4408	1.5554	1.2484	1.1344	0.9912	1.5371	1.3715	1.4281	0.9968
Cysteine	1.5075	1.2474	1.3868	1.4420	1.5524	1.2485	1.1362	0.9961	1.5389	1.3740	1.4294	0.9961
Selenocysteine	1.5079	1.2497	1.3802	1.4404	1.5520	1.2470	1.1326	0.9915	1.5389	1.3730	1.4277	0.9964
Histidine	1.5077	1.2487	1.3806	1.4414	1.5491	1.2497	1.1343	0.9950	1.5475	1.3745	1.4305	0.9994
Lysine	1.5041	1.2474	1.3891	1.4372	1.5565	1.2623	1.1366	0.9909	1.5374	1.3572	1.4331	1.0011
Arginine	1.5060	1.2565	1.3775	1.4448	1.5572	1.2465	1.1326	0.9997	1.5423	1.3710	1.4323	0.9910
Serine	1.5075	1.2473	1.3885	1.4410	1.5484	1.2489	1.1363	0.9967	1.5523	1.3749	1.4293	0.9964
Threonine	1.5075	1.2498	1.3848	1.4435	1.5541	1.2479	0.9947	1.1338	1.5504	1.3715	1.4281	0.9970
Aspartic acid	1.5080	1.2458	1.3894	1.4403	1.5539	1.2498	1.1355	0.9962	1.5415	1.3700	1.4292	0.9721
Glutamic acid	1.5075	1.2494	1.3813	1.4412	1.5547	1.2476	1.1347	0.9914	1.5356	1.3717	1.4282	0.9966
Asparagine	1.5075	1.2512	1.3787	1.4418	1.5501	1.2491	1.1363	0.9951	1.5432	1.3727	1.4301	0.9985
Glutamine	1.5074	1.2492	1.3817	1.4405	1.5558	1.2500	1.1346	0.9912	1.5374	1.3699	1.4287	0.9972
N-Methyl glycine	1.5062	1.2496	1.3919	1.4379	1.5407	1.2476	1.1290	1.4353[a]	1.1272	1.3730	1.4280	0.9964
N-Methyl alanine	1.5070	1.2509	1.3886	1.4513	1.5528	1.2469	1.1358	1.4337[a]	1.5258	1.3729	1.4273	0.9964
Homocysteine	1.5077	1.2493	1.3815	1.4410	1.5548	1.2479	1.1346	0.9911	1.5371	1.3717	1.4281	0.9967

[a] Although these peptides do not have a hydrogen (H) attached to this nitrogen (N–C_α) but are instead attached to a carbon (C), these bond lengths are nonetheless included.

TABLE XIV

Bond lengths in angstroms optimized at the RHF/3-21G level of theory.

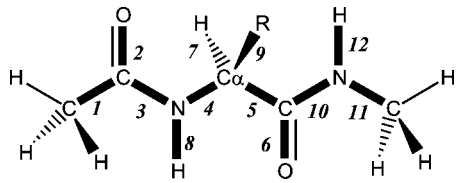


	Bond lengths											
Amino acids	1	2	3	4	5	6	7	8	9	10	11	12
Glycine	1.5147	1.2202	1.3511	1.4445	1.5211	1.2218	1.0843	0.9988	1.0843	1.3451	1.4646	0.9959
Alanine	1.5152	1.2220	1.3505	1.4498	1.5237	1.2238	1.0824	0.9993	1.5433	1.3463	1.4638	0.9962
Valine	1.5153	1.2205	1.3556	1.4559	1.5249	1.2244	1.0789	0.9988	1.5506	1.3462	1.4643	0.9960
Leucine	1.5155	1.2219	1.3511	1.4565	1.5217	1.2244	1.0806	0.9989	1.5498	1.3464	1.4643	0.9957
Isoleucine	1.5153	1.2204	1.3559	1.4572	1.5250	1.2244	1.0786	0.9987	1.5395	1.3465	1.4643	0.9961
Proline	1.5138	1.2411	1.3335	1.4870	1.5471	1.5598	1.0822	1.4779[a]	1.5395	1.5438	1.4627	1.0204
Phenylalanine	1.5141	1.2252	1.3489	1.4482	1.5234	1.2248	1.0808	0.9994	1.5618	1.3436	1.4634	0.9973
Tyrosine	1.5140	1.2256	1.3485	1.4482	1.5228	1.2247	1.0807	0.9995	1.5623	1.3438	1.4635	0.9973
Tryptophan	1.5151	1.2236	1.3495	1.4502	1.5264	1.2267	1.0815	0.9997	1.5575	1.3404	1.4634	0.9967
Methionine	1.5145	1.2212	1.3528	1.4527	1.5231	1.2237	1.0806	0.9990	1.5501	1.3467	1.4651	0.9960
Cysteine	1.5136	1.2229	1.3514	1.4492	1.5327	1.2257	1.0818	1.0001	1.5427	1.3391	1.4647	0.9974
Selenocysteine	1.5141	1.2228	1.3508	1.4500	1.5308	1.2258	1.0811	0.9999	1.5440	1.3730	1.4277	0.9964
Histidine	1.5124	1.2299	1.3436	1.4489	1.5227	1.2234	1.0809	1.0006	1.5696	1.3745	1.4305	0.9994
Lysine	1.5112	1.2227	1.3567	1.4519	1.5223	1.2215	1.0807	0.9991	1.5470	1.3435	1.4645	0.9987
Arginine	1.5106	1.2408	1.3275	1.4575	1.5382	1.2232	1.0822	1.0051	1.5528	1.3390	1.4708	0.9969
Serine	1.5148	1.2228	1.3514	1.4486	1.5334	1.2273	1.0832	1.0008	1.5325	1.3358	1.4611	1.0021
Threonine	1.5148	1.2176	1.3585	1.4486	1.5201	1.2238	0.9975	1.0768	1.5348	1.3450	1.4639	0.9962
Aspartic acid	1.5142	1.2231	1.3504	1.4498	1.5332	1.2262	1.0816	1.0004	1.5410	1.3352	1.4626	0.9997
Glutamic acid	1.5143	1.2214	1.3528	1.4525	1.5229	1.2234	1.0810	0.9989	1.5460	1.3470	1.4660	0.9960
Asparagine	1.5154	1.2242	1.3482	1.4515	1.5362	1.2282	1.0816	1.0010	1.5442	1.3314	1.4610	1.0040
Glutamine	1.5146	1.2214	1.3525	1.4535	1.5221	1.2232	1.0807	0.9988	1.5461	1.3488	1.4650	0.9960
N-Methyl glycine	1.5140	1.2206	1.3677	1.4477	1.5272	1.2193	1.0838	1.4649[a]	1.0790	1.3521	1.4626	0.9965
N-Methyl alanine	1.5156	1.2303	1.3556	1.4779	1.5288	1.2236	1.0764	1.4671[a]	1.5261	1.3458	1.4603	1.0010
Homocysteine	1.5143	1.2210	1.3533	1.4526	1.5227	1.2239	1.0806	0.9989	1.5508	1.3464	1.4656	0.9962

[a] Although these peptides do not have a hydrogen (H) attached to this nitrogen (N–C_α) but are instead attached to a carbon (C), these bond lengths are nonetheless included.

TABLE XV

Bond lengths in angstroms optimized at the RHF/6-31G(d) level of theory.

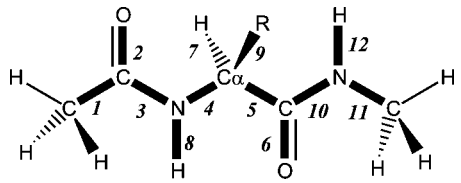


	Bond lengths											
Amino acids	1	2	3	4	5	6	7	8	9	10	11	12
Glycine	1.5136	1.2022	1.3497	1.4336	1.5205	1.2029	1.0856	0.9956	1.0856	1.3444	1.4485	0.9936
Alanine	1.5133	1.2040	1.3479	1.4421	1.5260	1.2041	1.0830	0.9952	1.5348	1.3450	1.4484	0.9934
Valine	1.5136	1.2014	1.3567	1.4443	1.5293	1.2043	1.0821	0.9965	1.5499	1.3463	1.4484	0.9935
Leucine	1.5135	1.2032	1.3502	1.4479	1.5245	1.2049	1.0813	0.9953	1.5406	1.3439	1.4482	0.9929
Isoleucine	1.5122	1.2055	1.3513	1.4567	1.5347	1.2035	1.0809	0.9952	1.5376	1.3469	1.4470	0.9969
Proline	1.5134	1.2184	1.3370	1.4812	1.5341	1.4705	1.0839	1.4632[a]	1.5326	1.4775	1.4499	1.0086
Phenylalanine	1.5130	1.2047	1.3476	1.4427	1.5268	1.2055	1.0813	0.9951	1.5478	1.3431	1.4480	0.9943
Tyrosine	1.5130	1.2052	1.3470	1.4423	1.5271	1.2056	1.0816	0.9952	1.5492	1.3429	1.4481	0.9941
Tryptophan	1.5134	1.2045	1.3478	1.4445	1.5258	1.2065	1.0811	0.9952	1.5462	1.3424	1.4486	0.9943
Methionine	1.5130	1.2027	1.3514	1.4451	1.5259	1.2040	1.0813	0.9953	1.5398	1.3449	1.4489	0.9933
Cysteine	1.5122	1.2041	1.3498	1.4408	1.5325	1.2053	1.0824	0.9959	1.5445	1.3402	1.4489	0.9944
Selenocysteine	1.5122	1.2039	1.3501	1.4412	1.5344	1.2053	1.0822	0.9960	1.5415	1.3407	1.4486	0.9951
Histidine	1.5104	1.2109	1.3417	1.4406	1.5258	1.2041	1.0825	0.9961	1.5604	1.3420	1.4489	0.9953
Lysine	1.5107	1.1985	1.3677	1.4498	1.5337	1.2050	1.0836	0.9966	1.5359	1.3402	1.4511	0.9938
Arginine	1.5082	1.2196	1.3340	1.4512	1.5352	1.2028	1.0830	0.9989	1.5427	1.3394	1.4543	0.9940
Serine	1.5130	1.2024	1.3536	1.4416	1.5341	1.2062	1.0842	0.9968	1.5315	1.3398	1.4472	0.9954
Threonine	1.5116	1.2062	1.3517	1.4572	1.5378	1.2048	0.9950	1.0835	1.5310	1.3411	1.4462	0.9977
Aspartic acid	1.5125	1.2040	1.3490	1.4419	1.5339	1.2065	1.0816	0.9958	1.5399	1.3365	1.4477	0.9956
Glutamic acid	1.5128	1.2027	1.3517	1.4450	1.5261	1.2038	1.0816	0.9954	1.5382	1.3455	1.4492	0.9935
Asparagine	1.5108	1.2098	1.3439	1.4408	1.5324	1.2060	1.0857	0.9967	1.5535	1.3366	1.4482	0.9995
Glutamine	1.5129	1.2026	1.3521	1.4460	1.5256	1.2035	1.0812	0.9954	1.5371	1.3472	1.4489	0.9935
N-Methyl glycine	1.5143	1.2085	1.3539	1.4562	1.5258	1.2039	1.0805	1.4530	1.0790	1.3445	1.4457	0.9969
N-Methyl alanine	1.5153	1.2085	1.3568	1.4677	1.5324	1.2043	1.0774	1.4555	1.5240	1.3456	1.4460	0.9969
Homocysteine	1.5129	1.2024	1.3520	1.4455	1.5258	1.2044	1.0814	0.9953	1.5403	1.3441	1.4492	0.9933

[a] Although these peptides do not have a hydrogen (H) attached to this nitrogen (N–C_α) but are instead attached to a carbon (C), these bond lengths are nonetheless included.

TABLE XVI

Bond lengths in angstroms optimized at the B3LYP/6-31G(d) level of theory.



	Bond lengths											
Amino acids	1	2	3	4	5	6	7	8	9	10	11	12
Glycine	1.5228	1.2266	1.3630	1.4401	1.5321	1.2280	1.0994	1.0138	1.0994	1.3570	1.4555	1.0097
Alanine	1.5217	1.2284	1.3621	1.4490	1.5377	1.2291	1.0977	1.0135	1.5419	1.3598	1.4532	1.0102
Valine	1.5217	1.2265	1.3680	1.4564	1.5366	1.2295	1.0953	1.0125	1.5542	1.3602	1.4535	1.0099
Leucine	1.5217	1.2279	1.3633	1.4551	1.5333	1.2294	1.0958	1.0122	1.5500	1.3604	1.4534	1.0098
Isoleucine	1.5217	1.2266	1.3676	1.4565	1.5368	1.2294	1.0952	1.0123	1.5572	1.3600	1.4530	1.0100
Proline	1.5188	1.2474	1.3483	1.5009	1.5451	1.5194	1.0975	1.4725[a]	1.5407	1.4958	1.4595	1.0416
Phenylalanine	1.5205	1.2304	1.3616	1.4471	1.5398	1.2304	1.0968	1.0142	1.5631	1.3573	1.4535	1.0117
Tyrosine	1.5204	1.2307	1.3612	1.4476	1.5412	1.2308	1.0971	1.0147	1.5622	1.3560	1.4541	1.0117
Tryptophan	1.5210	1.2301	1.3608	1.4474	1.5387	1.2311	1.0964	1.0140	1.5632	1.3570	1.4537	1.0116
Methionine	1.5210	1.2275	1.3641	1.4515	1.5354	1.2286	1.0961	1.0124	1.5508	1.3607	1.4543	1.0101
Cysteine	1.5201	1.2289	1.3637	1.4488	1.5457	1.2313	1.0973	1.0150	1.5510	1.3541	1.4542	1.0143
Selenocysteine	1.5201	1.2287	1.3639	1.4488	1.5479	1.2313	1.0973	1.0152	1.5467	1.3540	1.4540	1.0170
Histidine	1.5172	1.2365	1.3554	1.4467	1.5361	1.2292	1.0963	1.0150	1.5733	1.3570	1.4540	1.0130
Lysine	1.5174	1.2228	1.3829	1.4588	1.5440	1.2302	1.0972	1.0123	1.5427	1.3550	1.4550	1.0110
Arginine	1.5146	1.2352	1.3453	1.4569	1.5506	1.2265	1.0967	1.0189	1.5530	1.3530	1.4610	1.0100
Serine	1.5215	1.2277	1.3658	1.4486	1.5529	1.2318	1.0994	1.0159	1.5380	1.3530	1.4530	1.0160
Threonine	1.5186	1.2347	1.3607	1.4682	1.5563	1.2299	1.0128	1.0984	1.5389	1.3537	1.4514	1.0204
Aspartic acid	1.5208	1.2294	1.3615	1.4486	1.5479	1.2320	1.0964	1.0147	1.5474	1.3510	1.4526	1.0160
Glutamic acid	1.5224	1.2271	1.3655	1.4517	1.5357	1.2285	1.0964	1.0128	1.5477	1.3614	1.4545	1.0965
Asparagine	1.5178	1.2361	1.3567	1.4468	1.5424	1.2316	1.0993	1.0156	1.5600	1.3518	1.4534	1.0214
Glutamine	1.5213	1.2286	1.3624	1.4518	1.5362	1.2282	1.0968	1.0126	1.5491	1.3614	1.4550	1.0100
N-Methyl glycine	1.5213	1.2359	1.3672	1.4665	1.5391	1.2279	1.0921	1.4583	1.0924	1.3601	1.4512	1.0169
N-Methyl alanine	1.5217	1.2362	1.3699	1.4805	1.5464	1.2288	1.0915	1.4611	1.5263	1.3605	1.4516	1.0168
Homocysteine	1.5209	1.2274	1.3645	1.4515	1.5353	1.2289	1.0961	1.0122	1.5519	1.3600	1.4550	1.0100

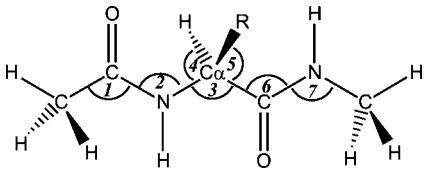
[a] Although these peptides do not have a hydrogen (H) attached to this nitrogen (N–C_α) but are instead attached to a carbon (C), these bond lengths are nonetheless included.

phase. Due to fortuitous cancellation of errors, the RHF/3-21G results are very good representations of the conformational intricacies of peptides. For the 24 chosen amino acids (summarized in Table II) the torsional angles obtained at the RHF/3-21G level of theory are shown in Table X. Optimized torsional angles obtained at RHF/6-31G(d) and B3LYP/6-31G(d) levels of theory are given in Tables XI and XII. Tables XIII–XVI show bond lengths computed at the four levels of theory discussed above. Finally Tables XVII–XX summarize bond angles obtained at the chosen four levels of theory.

A general effort was made to include as many peptide structures as possible but many more to be included in the developmental stage. Most of these additional structures are modifications of naturally occurring amino acids such as 5-oxo proline shown in Scheme 22.

As noted, a very large number of combinations or procedures for numbering atomic nuclei exist. This article seeks only to describe one that is functional, reproducible, and automatable, as well as useful in the teaching and education of younger colleagues in the area of quantum chemical computations. Any number of modifications to this methodology are possible and one cannot yet weigh the advantages and disadvantages against the standing system until the development of the fully functional form is completed. It is hoped that the rela-

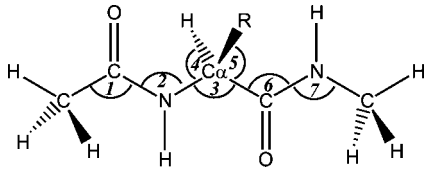
TABLE XVII
Bond angles optimized at the AMI level of theory.



Amino acids	Bond angles						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Glycine	117.39	120.91	112.83	109.26	109.26	116.12	121.56
Alanine	117.01	121.71	111.09	107.50	112.90	117.10	122.85
Valine	117.50	122.27	110.04	107.75	113.61	115.99	122.92
Leucine	117.47	122.27	110.25	107.38	112.69	115.97	122.97
Isoleucine	117.50	122.22	109.84	107.73	113.75	115.97	122.87
Proline	118.94	124.05	119.05	108.17	106.62	118.87	118.91
Phenylalanine	117.24	121.19	111.37	107.47	113.07	116.06	122.78
Tyrosine	117.27	121.22	111.37	107.40	113.17	116.01	122.82
Tryptophan	117.28	121.07	111.63	107.38	113.25	115.96	122.60
Methionine	117.48	122.44	111.33	107.86	112.43	116.35	122.99
Cysteine	117.14	121.25	111.30	107.67	112.62	116.28	122.43
Selenocysteine	117.54	122.68	111.97	108.06	112.25	116.45	122.95
Histidine	117.37	121.63	111.16	107.81	112.38	116.23	121.56
Lysine	117.73	121.81	111.83	107.98	112.63	117.77	123.43
Arginine	118.07	122.83	109.56	107.74	114.23	117.09	122.60
Serine	117.46	121.24	110.35	118.50	113.45	116.09	122.28
Threonine	117.16	120.77	111.90	108.44	112.38	116.15	122.81
Aspartic acid	116.90	121.65	110.85	108.26	111.94	116.88	122.67
Glutamic acid	117.48	122.40	111.36	107.77	111.99	116.33	122.93
Asparagine	117.53	121.31	111.57	106.93	113.28	116.05	122.07
Glutamine	117.50	122.45	111.41	107.90	112.36	116.49	123.05
N-Methyl glycine	118.73	119.99	114.87	108.83	109.28	117.29	122.73
N-Methyl alanine	119.42	119.45	111.14	107.29	114.20	116.19	123.02
Homocysteine	117.47	122.36	111.24	107.78	112.37	116.29	122.97

TABLE XVIII

Bond angles optimized at the RHF/3-21G level of theory.



Amino acids	Bond angles						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Glycine	114.48	120.73	107.71	111.19	111.19	115.16	120.42
Alanine	114.44	121.45	106.47	109.72	111.23	114.95	121.88
Valine	114.11	121.91	105.61	107.65	111.86	116.37	121.81
Leucine	114.45	121.51	105.37	109.02	109.57	115.81	121.85
Isoleucine	114.10	121.98	105.36	107.43	101.60	116.56	121.79
Proline	117.22	122.96	117.49	109.59	102.12	107.96	119.09
Phenylalanine	114.82	120.42	106.73	109.30	111.96	114.81	121.81
Tyrosine	114.85	120.36	106.81	109.25	112.04	114.73	121.84
Tryptophan	114.58	121.26	105.99	109.77	110.98	115.31	122.02
Methionine	114.50	121.21	105.84	109.33	109.65	115.79	121.75
Cysteine	114.61	121.32	105.75	110.11	109.59	115.80	121.71
Selenocysteine	114.56	121.49	105.56	110.11	109.41	115.87	121.74
Histidine	115.14	121.07	107.03	109.10	112.94	114.11	121.98
Lysine	114.94	120.42	106.39	109.48	109.36	115.84	121.42
Arginine	115.77	124.30	104.29	109.72	114.34	115.61	121.84
Serine	114.33	122.42	106.07	111.22	110.93	114.90	121.76
Threonine	113.86	121.46	106.95	121.96	110.35	116.40	121.76
Aspartic acid	114.53	121.97	105.41	110.55	110.23	115.76	120.28
Glutamic acid	114.52	121.10	105.93	109.23	109.69	115.75	121.72
Asparagine	114.39	122.74	105.04	110.79	110.54	115.67	120.37
Glutamine	114.49	121.10	105.87	109.01	109.65	115.77	121.69
N-Methyl glycine	117.15	116.42	111.12	109.73	108.65	113.79	121.84
N-Methyl alanine	117.79	117.59	106.95	106.29	112.42	113.45	121.79
Homocysteine	114.53	121.12	105.91	109.34	109.62	115.82	121.77

tively flexible, open-ended, and modular nature of the system described herein will allow for future progress.

Concluding Remarks

About a century ago, the co-discoverer of saccharine, Ira Remsen (1846–1927), expressed his fear of the unknown concerning peptide and protein chemistry [42]:

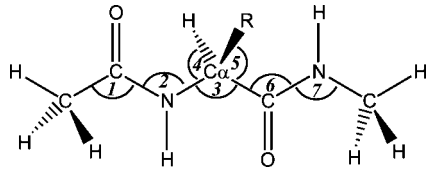
> *I always feel like running away when any one begins to talk about proteids in my presence. In my youth I had a desire to attack these dragons, but now I am afraid of them. They are unsolved problems of chemistry; and let me add, they are likely to remain such for generations to come. Yet every one who knows anything about chemistry and physiology, knows that these proteids must be understood, before we can hope to have a clear conception of the chemical processes of the human body.*

The problem at hand, which is the lack of understanding of the rules of peptide and protein folding, takes the role of the proverbial dragon. The solution to that problem is the Saint Georgian killing of that proverbial dragon.

We hope that the construction of the peptide structure database represents a proverbial sword or spear for a major stab or hack at the heart

TABLE XIX

Bond angles optimized at the RHF/6-31G(d) level of theory.



Amino acids	Bond angles						
	1	2	3	4	5	6	7
Glycine	115.16	121.43	109.19	110.97	110.97	115.03	121.67
Alanine	115.86	122.24	107.43	108.21	112.19	115.65	121.72
Valine	114.88	122.25	107.92	107.29	113.37	115.69	121.69
Leucine	115.57	122.57	106.11	107.28	111.05	116.54	121.69
Isoleucine	115.97	122.42	107.21	106.33	111.81	114.63	121.08
Proline	117.67	123.15	118.77	108.92	102.94	110.13	120.79
Phenylalanine	116.07	122.00	107.01	108.25	111.24	116.11	121.54
Tyrosine	116.08	121.86	107.02	108.29	111.32	116.09	121.58
Tryptophan	115.86	122.32	106.74	108.18	111.13	116.20	121.76
Methionine	115.58	122.37	106.62	107.47	111.15	116.51	121.64
Cysteine	115.82	121.62	106.94	108.96	110.31	116.27	121.29
Selenocysteine	115.76	121.53	106.94	108.90	110.25	116.24	121.15
Histidine	116.67	120.91	107.96	108.21	112.84	114.89	121.72
Lysine	115.54	120.33	110.72	105.83	112.97	115.98	121.70
Arginine	116.80	124.11	105.74	108.24	114.81	116.24	121.63
Serine	115.27	121.85	107.36	110.04	110.41	115.86	121.18
Threonine	115.59	121.82	109.90	117.14	109.73	114.83	121.10
Aspartic acid	115.78	122.05	106.71	109.27	110.48	116.08	121.37
Glutamic acid	115.54	122.21	106.66	107.40	111.19	116.51	121.60
Asparagine	116.43	121.18	107.83	108.45	112.20	115.44	121.20
Glutamine	115.44	122.37	106.67	107.12	111.21	116.51	121.46
N-Methyl glycine	118.28	118.40	111.32	108.64	108.99	114.99	121.45
N-Methyl alanine	118.18	118.00	108.79	105.99	112.72	114.44	121.47
Homocysteine	115.57	122.30	106.61	107.44	111.18	116.62	121.63

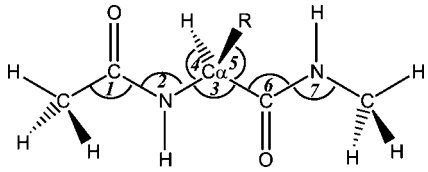
of that proverbial dragon. This would be in the fighting spirit of our warrior teacher Professor Löwdin.

ACKNOWLEDGMENTS

The authors thank Christopher N. J. Marai, Tania A. Pecora, David H. Setiadi, and Michael D. Hendrickson for their helpful input and discussions, as well as for the preparation of this manuscript, tables, and figures. The work of Graydon Hoare in the areas of database management, network support, software, and distributive processing development is also recognized. A special thanks is extended to Andrew M. Chasse for his continuing and ongoing development of novel scripting and coding techniques, indirectly bringing about a reduction in the necessary number of CPU cycles for each computation. Last but not least, the authors wish to acknowledge Kenneth P. Chasse (Jr.) of Velocet Communications Inc. for his pioneering of great advances in all composite computer-cluster software and hardware architectures. He and the other directors of Velocet Communications Inc., specifically David C. L. Gilbert and Adam A. Heaney, are also acknowledged for providing spare CPU time for the computations needed in the preparation of this article.

TABLE XX

Bond angles optimized at the B3LYP/6-31G(d) level of theory.



Amino acids	Bond angles						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Glycine	114.92	121.65	108.43	110.98	110.98	115.31	121.45
Alanine	115.53	122.05	106.89	108.22	112.04	114.99	123.04
Valine	115.20	122.64	106.24	106.33	112.33	116.07	122.96
Leucine	115.43	122.39	106.02	107.26	110.85	115.76	122.93
Isoleucine	115.18	122.67	106.02	106.19	112.57	116.12	122.95
Proline	117.46	123.58	119.40	108.30	103.15	108.70	120.17
Phenylalanine	115.81	121.09	106.85	108.36	111.69	115.17	120.92
Tyrosine	115.55	121.09	107.12	108.53	111.90	115.37	122.25
Tryptophan	115.81	121.62	106.68	108.55	111.36	115.10	122.77
Methionine	115.49	122.11	106.46	107.50	110.91	115.77	122.82
Cysteine	115.63	121.50	106.45	109.00	109.97	115.53	122.50
Selenocysteine	115.66	121.42	106.35	108.80	109.92	115.73	122.06
Histidine	116.26	121.19	107.60	108.13	113.11	114.05	122.96
Lysine	115.78	119.98	111.27	104.76	112.81	114.97	123.09
Arginine	116.90	124.34	104.70	108.20	114.92	115.98	121.64
Serine	115.43	121.89	106.62	109.92	110.94	115.16	122.23
Threonine	115.95	122.36	111.83	117.51	109.16	113.81	122.19
Aspartic acid	115.63	122.21	106.07	109.32	110.39	115.18	122.71
Glutamic acid	114.82	122.16	106.49	107.44	111.06	115.82	122.77
Asparagine	116.14	120.96	107.83	108.08	112.47	114.41	122.43
Glutamine	115.53	122.05	106.30	107.53	110.97	115.88	122.38
N-Methyl glycine	118.36	118.00	111.80	108.52	108.15	113.62	122.25
N-Methyl alanine	118.21	117.51	108.77	105.16	112.76	113.37	122.32
Homocysteine	115.55	122.02	106.50	107.44	110.89	115.86	122.84

One of the authors, G. A. Chass, extends his deepest thanks and respects to Ollie and Kenneth L. Chasse (Sr.), for their teachings about the battle between the "plus guys" and the "minus guys" and just how many would be left over if they met, this from the earliest of childhood memories of Saturday morning arithmetic lessons and pancakes. This battle continues today and in this work.

References

1. Chasse, G. A.; Rodriguez, A. M.; Mak, M. L.; Deretey, E.; Perczel, A.; Sosa, C. P.; Enriz, R. D.; Csizmadia, I. G. J Mol Struct (Theochem) 2001, 537, 319–361.
2. Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. J Mol Biol 1963, 7, 95.
3. Head-Gordon, T.; Head-Gordon, M.; Frish, M. J.; Brooks, C.; Pople, J. J Am Chem Soc 1991, 113, 5989.
4. Perczel, A.; Farkas, Ö.; Csizmadia, I. G. J Am Chem Soc 1995, 117, 1653–1654.
5. Nemethy, G.; Scheraga, H. A. Biopolymers 1965, 3, 155.
6. Liquori, A. M. Q Rev Biophys 1969, 2, 65.
7. Brooks, B. B.; Bruccelori, R. E.; Olafson, B. D.; States, D. J.; Swaminatham, S.; Karplus, M. J Comput Chem 1983, 4, 187.
8. Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, Jr., S.; Weiner, P. J Am Chem Soc 1984, 106, 765.
9. Hoffman, R.; Imamura, A. Biopolymers 1969, 7, 207.
10. Hopfinger, A. J.; Walton, A. G. Biopolymers 1970, 9, 29.

11. Damm, W.; Van Gunsteren, W. F. J Comput Chem 2000, 31, 774.
12. Yan, J. F.; Momany, F. A.; Hoffman, R.; Scheraga, H. A. J Phys Chem 1970, 74, 420.
13. Momany, F. A.; McGuire, R. F.; Yan, J. F.; Scheraga, H. A. J Phys Chem 1971, 75, 2286.
14. Kleier, D. A.; Lipscomb, W. N. Int J Quantum Chem Quantum Biol Symp 1977, 4, 73.
15. Maigret, B.; Pullman, B.; Dreyfus, M. J Theoret Biol 1970, 26, 321.
16. Maigret, B.; Perania, D.; Pullman, B. J Theoret Biol 1970, 29, 275.
17. Maigret, B.; Perania, D.; Pullman, B. Biopolymers 1971, 10, 491.
18. Csizmadia, I. G.; Harrison, M. C.; Sutcliffe, B. T. Quart Progr Rept (MIT SSMTG) 1963, 50, 1.
19. Csizmadia, I. G.; Harrison, M. C.; Sutcliffe, B. T. Theoret Chim Acta 1966, 6, 217.
20. Basch, H. M.; Robin, B.; Kuebler, N. A. J Chem Phys 1967, 47, 1201.
21. Robb, M. A.; Csizmadia, I. G. Theoret Chim Acta 1968, 10, 269.
22. Peters, D.; Peters, J. J Mol Struct (Theochem) 1979, 53, 103.
23. Peters, D.; Peters, J. J Mol Struct (Theochem) 1984, 109, 137.
24. Peters, D.; Peters, J. J Mol Struct (Theochem) 1980, 68, 249.
25. Peters, D.; Peters, J. J Mol Struct (Theochem) 1980, 69, 249.
26. Ramani, R.; Boyd, R. J. Int J Quantum Chem Quantum Biol Symp 1981, 8, 117.
27. Ramani, R.; Boyd, R. J. Can J Chem 1981, 59, 3232.
28. Boyd, R. J.; Perkyns, J. S.; Ramani, R. Can J Chem 1983, 61, 1082.
29. Scardale, J. N.; van Alsenoy, C.; Klimkowski, V. J.; Schafer, L.; Momany, F. A. J Am Chem Soc 1983, 105, 3438.
30. Perczel, A.; Daudel, R.; Angyan, J. G.; Csizmadia, I. G. Can J Chem 1990, 68, 1882–1888.
31. Perczel, A.; Angyan, J. G.; Kajtar, M.; Viviani, W.; Rivail, J. L.; Marcoccia, J. F.; Csizmadia, I. G. J Am Chem Soc 1991, 113, 6256.
32. McAllister, M. A.; Perczel, A.; Császár, P.; Csizmadia, I. G. J Mol Struct (Theochem) 1993, 288, 181–198.
33. Perczel, A.; McAllister, M. A.; Császár, P.; Csizmadia, I. G. J Am Chem Soc 1993, 115, 4849.
34. Perczel, A.; McAllister, M. A.; Császár, P.; Csizmadia, I. G. Can J Chem 1994, 72, 2050–2070.
35. Cheung, M.; McGovern, M. E.; Jin, T.; Zhao, D. C.; McAllister, M. A.; Perczel, A.; Császár, P.; Csizmadia, I. G. J Mol Struct (Theochem) 1994, 309, 151–224.
36. Liegener, C.-M.; Endrédi, G.; McAllister, M. A.; Perczel, A.; Ladik, J.; Csizmadia, I. G. J Am Chem Soc 1993, 115, 8275–8277.
37. Endrédi, G.; Liegner, C.-M.; McAllister, M. A.; Perczel, A.; Ladik, J.; Csizmadia, I. G. J Mol Struct (Theochem) 1994, 306, 1–7.
38. Endrédi, G.; McAllister, M. A.; Perczel, A.; Császár, P.; Ladik, J.; Csizmadia, I. G. J Mol Struct (Theochem) 1995, 331, 5–10.
39. Endrédi, G.; McAllister, M. A.; Farkas, Ö.; Perczel, A.; Ladik, J.; Csizmadia, I. G. J Mol Struct (Theochem) 1995, 331, 11–26.
40. Fang, D. C.; Fu, X. Y.; Tang, T. H.; Csizmadia, I. G. J Mol Struct (Theochem) 1998, 427, 243–252.
41. Arch Biochem Biophys 1971, 145, 405–421.
42. Biochem J 1971, 121, 577–585.
43. Biochemistry 1971, 9, 3471–3479.
44. Biochim Biophys Acta 1971, 229, 1–17.
45. Eur J Biochem 1969, 17, 193–201.
46. J Biol Chem 1970, 245, 6489–6497.
47. J Mol Biol 1970, 52, 1–17.
48. Pure Appl Chem 1974, 40, 291–308.
49. Biochemical Nomenclature and Related Documents, 2nd ed.; Portland Press: 1992; pp. 73–81.
50. Topol, I. A.; Burt, S. K.; Deretey, E.; Tang, T. H.; Perczel, A.; Rashin, A.; Csizmadia, I. G. Math Chem 2001, 123, 6054.
51. Remsen, I. In Modern Inventions and Discoveries; J. A. Hill: New York, 1904; also quoted in Neurath, H. Protein Sci 1992, 1, 2.