Reprinted from

THEO CHEM

Journal of Molecular Structure (Theochem) 500 (2000) 5–58

Prospects in computational molecular medicine: a millennial mega-project on peptide folding

M.A. Berg[a], G.A. Chasse[a], E. Deretey[a], A.K. Füzéry[a], B.M. Fung[a], D.Y.K. Fung[a], H. Henry-Riyad[a], A.C. Lin[a], M.L. Mak[a], A. Mantas[a], M. Patel[a], I.V. Repyakh[a], M. Staikova[a], S.J. Salpietro[a], Ting-Hua Tang[a], J.C. Vank[a], A. Perczel[b,‡], G.I. Csonka[c,‡], Ö. Farkas[b,‡], L.L. Torday[d,‡], Z. Székely[a,‡], I.G. Csizmadia[a,*]

[a]*Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6*
[b]*Department of Organic Chemistry, Loránd Eötvös University, H-1117 Budapest, Hungary*
[c]*Department of Inorganic Chemistry, Technical University, H-1111 Budapest, Hungary*
[d]*Department of Pharmacology and Pharmacotherapy, Albert Szent-Györgyi Medical University, H-6720 Szeged, Dóm tér 12, P.O. Box 417, Hungary*

ELSEVIER

ELSEVIER

Journal of Molecular Structure (Theochem) 500 (2000) 5–58

THEO CHEM

www.elsevier.nl/locate/theochem

Prospects in computational molecular medicine: a millennial mega-project on peptide folding

M.A. Berg[a], G.A. Chasse[a], E. Deretey[a], A.K. Füzéry[a], B.M. Fung[a], D.Y.K. Fung[a], H. Henry-Riyad[a], A.C. Lin[a], M.L. Mak[a], A. Mantas[a], M. Patel[a], I.V. Repyakh[a], M. Staikova[a], S.J. Salpietro[a], Ting-Hua Tang[a], J.C. Vank[a], A. Perczel[b,‡], G.I. Csonka[c,‡], Ö. Farkas[b,‡], L.L. Torday[d,‡], Z. Székely[a,‡], I.G. Csizmadia[a,*]

[a]*Department of Chemistry, University of Toronto, Toronto, Ontario, Canada M5S 3H6*
[b]*Department of Organic Chemistry, Loránd Eötvös University, H-1117 Budapest, Hungary*
[c]*Department of Inorganic Chemistry, Technical University, H-1111 Budapest, Hungary*
[d]*Department of Pharmacology and Pharmacotherapy, Albert Szent-Györgyi Medical University, H-6720 Szeged, Dóm tér 12, P.O. Box 417, Hungary*

Abstract

During the second half of the 20th century, *Molecular Computations* have reached to a level that can revolutionize chemistry. The next target will be structural biology, which will be followed soon by *Molecular Medicine*. The present paper outlines where we are at, in this field, at the end of the 20th century, and in what direction the development may take in the new millennium. In view of the gigantic nature of the problem, it is suggested that a suitably designed cooperative *Millennial Mega-project* might accelerate our schedule. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Biomolecular computations; Peptide models; Peptide folding; Molecular medicine

1. Preamble

Every disease starts at the molecular level. Thus, ultimately every cure has to be achieved at the molecular level. This recognition leads us inevitably to the direction of Molecular Medicine.[1]

By the end of the 20th century, a powerful network of new disciplines has been developed which can in fact support molecular medicine. This interdisciplinary approach emerges as a new *mega discipline*. Its translation to practice, i.e. to the drug discovery process, involves at least five major fields, as illustrated in Scheme 1.

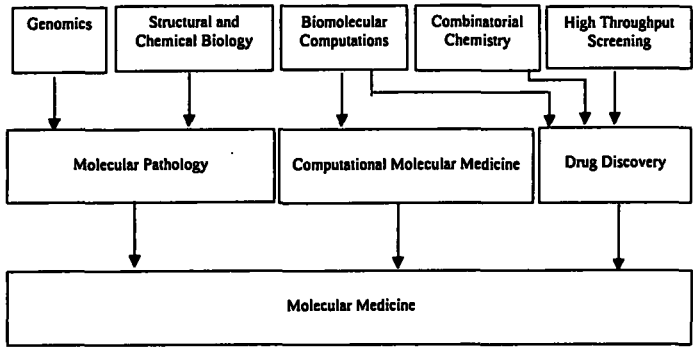
Genomics (human genome project).
Structural biochemistry and biology.
Biomolecular computation.
Solid phase supported combinatorial chemistry.
High throughput screening.

Based on the soon-to-be completed human genome project, a massive number of protein sequences are to

* Corresponding author. Tel.: +1-416-978-3598; fax: +1-416-978-3598.
‡ Senior authors.
E-mail address: icsizmad@alchemy.chem.utoronto.ca (I.G. Csizmadia).

[1] Gene Therapy, which is also part of Molecular Medicine, will not be discussed in this paper.

PII: S0166-1280(00)00448-6



Scheme 1.

be released; however, the biological functions for some of these proteins are yet to be identified. Using easily available biotechnological methods, a highly diverse protein collection will soon be available for structural studies.

X-ray diffraction and NMR spectroscopy, the two determinative arms of the structural *proteomics*, are developing rapidly for high throughput data collection. For example, synchrotron sources for X-ray studies are currently in use. Furthermore, multi-magnet NMR facilities will become functional in the near future. These NMR installations which contain 10–20 magnets, each in the vicinity of 1 GHz magnitude, are equipped with semi-automatic co-ordinate determination. This development will lead to a high-throughput NMR set-up. Such a complex facility would be able to handle the booming proteomics-depot.

The re-organized Protein Data Bank, supported by all other data banks derived from the high throughput screening, is serving a reliable platform to establish one of the biggest challenges in molecular medicine, the interface between chemistry and biology.

Since the late 1980s, a very important paradigm shift has taken place. There was a move from individually synthesized (and tested) compounds to combinatorially synthesized (and tested) mixtures. These collections of molecules are called *libraries* by the new terminology. This approach has been initiated in peptide chemistry due to its simplicity and the availability of automated procedures developed on solid phase. The success achieved with peptides [1–5] immediately triggered the development of solid phase organic chemistry [6]. Finally this led to small, pharmacologically relevant compound libraries.

Using robotics, the problem of screening of the emerging number of libraries, containing synthetic compounds, as well as natural products, is now being solved. It has to be emphasized that for successful handling of the above mentioned multi-source data collections, up-to-date data processing software should be used.

In spite of our outstanding and ever developing technologies, as outlined above, we do have conceptual shortcomings. The paramount obstacle to sophisticated drug discovery is our ignorance of the law of protein folding [7,8]. If we could pronounce the law by which proteins fold, we would be able to proceed with unprecedented speed and efficiency. However, this is precisely the basic knowledge that is missing, and we may not find the answer to this burning question for centuries, unless we change our method of investigation.

Until this point, in scientific practice, scientists have always started with folded proteins, of which there are literally billion types in the human body, and tried to find the rules that govern their folding patterns. Although some success has been made, the problem is far from being solved. For example, we cannot even understand what forces determine the

way a short polypeptide, containing for instance 10 amino acids, is folded. How can we then expect to understand why a long polypeptide chain, containing, for instance, 1000 amino acids, is folded in the way it is? Clearly, in order to gain a deeper understanding, we need to launch a project in deciphering the rules of *peptide folding*, yet even this smaller problem will prove to be a *mega-project*.

2. Computational background

Today, most peptide and protein chemists are using one of the varieties of empirical force fields to study peptide and protein conformations. The results of such methods are essentially educated guesses of the potential energy hyper-surface (PEHS) that describes the energetics of folding. The validity of such empirical methods, of course, is dependent on the accuracy of parameterization. The advantages of such computational methods are that they are fast while their disadvantage is that they are not completely reliable. Consequently, there is a calculated risk in their usage, as they could produce misleading results in certain cases but one would not know when this would occur. The method is about a quarter of a century old and no one has had neither the commitment nor the money to polish up the parameterization to the level that is desirable for scientific precision.

In contrast to the above, serious efforts have been made by thousands of university researchers during the past half a century to develop non-empirical or ab initio ("from the beginning") methods to compute PEHS from first principles (i.e. from quantum mechanics). The advantage of these methods is that they are reliable while their disadvantage is that they are computationally intensive. In the 1960s up to four or six atoms could be handled at the ab initio level of theory (for example, molecules such as HCOF [9–11] or even the isoelectronic $HCONH_2$ [12], the latter structure containing one peptide bond). However, all calculations were performed on a fixed structure, without geometry optimization. Nevertheless, by the turn of the millennium, the largest peptide on which ab initio calculations have been carried contains 12 amino acids [13].

The present paper aims to propose a *mega-project* which will rely on the use of this second method, carrying out computations on relatively small peptides at the reliable ab initio level of theory.

Both methods yield gas phase results, which implies that the intrinsic properties of the molecules are computed in the absence of environmental interactions. Thus, in order to gain an understanding of the role that solvents may play in peptide and protein folding, the effect of water needs to be subsequently simulated.

The reason ab initio methods are being suggested, as the method of investigation, is twofold:

1. the result would be reliable for a considerable length of time;
2. the results would provide a primary standard for future force-field parameterization.

In spite of these good intentions, one may wonder if it is prudent to set our standards so high. Software and hardware development during the past half-century suggest that our computational development in the *next half-century will supercede all imagination*. Consequently, it seems not only desirable but also practical as well to aim for the more reliable method, even if it is computationally intensive.

Fig. 1 shows the century of development, from 1950 to 2050. In 1950, Boys suggested the use of Gaussian-type orbitals and Roothaan published his SCF procedure in 1957. However, in the 1950s and early 1960s, all digital computers were vacuum-tube based. The first GAUSSIAN calculation on an organic molecule (HCOF), which in fact is isoelectronic to formamide ($HCONH_2$), was carried out in 1963 on an IBM 709 computer that was still using vacuum tubes. The leading software was POLYATOM at that time. The first transistorized mainframe computer (such as the IBM 7090 and 7094) arrived a bit later on the scene.

The vertical bars in Fig. 1 show the hardware development, after transistorization, in units of FLOPS (Floating point Operations Per Second). Clearly there was a dramatic change from mega (10^6) FLOPS through giga (10^9) FLOPS, all the way to the current tera (10^{12}) FLOPS. However, it has been suggested that future computers can solve problems in 30 s—what today's 10^{12}-FLOPS supercomputers would take 10 billion years to solve. The ratio (10^{10} years: 30 s) is of the order of 10^{16}. Thus, we

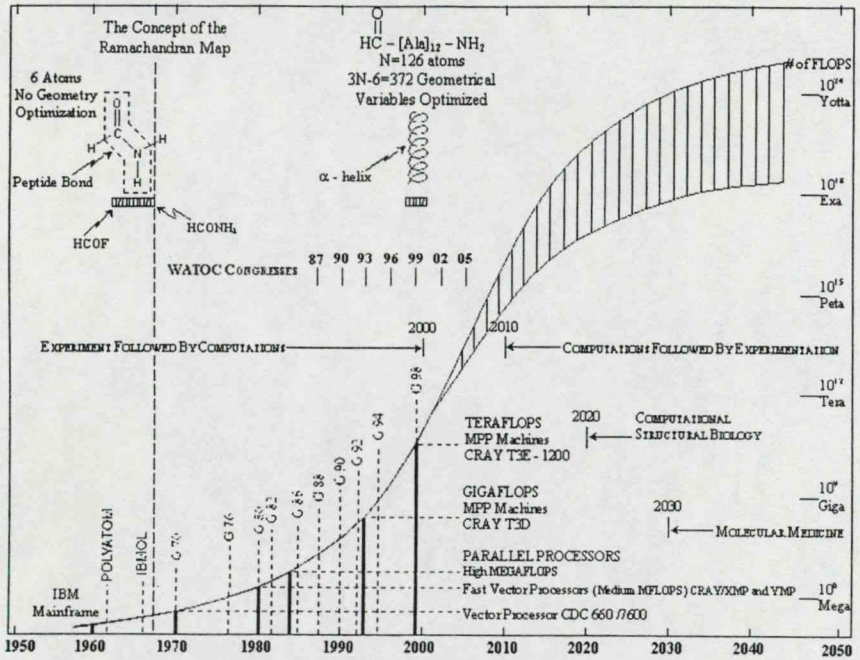


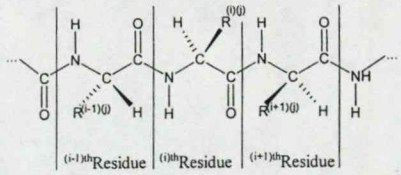
Fig. 1. Development of computer hardware and software as the basis of biomolecular computations. Note that after this figure has been completed (July 1999), IBM announced the creation of the "Blue Gene" ultra supercomputer with one petaflops capability, to be ready by 2005. This makes the optimistic (upper) growth curve a conservative estimate of the future development.

may consider the limit of the growth-curve to be $10^{12} \times 10^{16} = 10^{28}$ FLOPS. Fig. 1 shows two curves, the lower of which is the pessimistic prediction and the upper of which is the optimistic one. Note that even the optimistic curve levels off at about 10^{25} FLOPS, which is about 1000 times more conservative than the prediction currently lingering around (10^{28} FLOPS).

This dramatic hardware development is paralleled by a heroic effort in software development. Such development is not quantified by values of benchmark computations; they are simply presented at their time of appearance. The most durable package is GAUSSIAN [14], which has had several editions from 1970 to 1998. However, other softwares (e.g. SPARTAN, JAGUAR, etc.), not shown in Fig. 1, are now challenging the GAUSSIAN's dominance. The dates of the WATOC congresses (from 1987 to 2005) where molecular computational chemists and biologists are reporting their progress are also indicated. Finally, the dates of the first and therefore the smallest [$HCONH_2$] [12] and last and therefore largest [$HCO–(Ala)_{12}–NH_2$] [13] ab initio peptide calculations of the 20th century are also marked.

3. Peptide conformational background

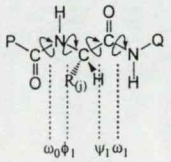
A polypeptide chain on its own, or as part of a protein molecule, consists of a string of amino acid residues as shown below in **I**:



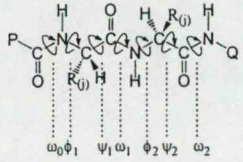
I

where $R^{(i)}$ specifies the side chain of the ith amino acid residue.

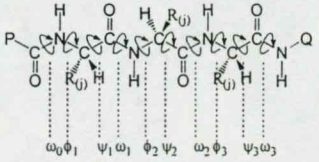
The second index, j, specifies that one of the 20 different side-chains of the 20 naturally occurring amino acid residues[2] [15] may be involved ($1 \le j \le 20$). Mono-peptides (**II**), dipeptides (**III**) and tripeptides (**IV**) contain 1, 2 and 3 amino acid residues, respectively. In such models, the chain may be terminated by methyl groups or by hydrogens, as shown below for the above three cases (P = Q = CH_3 or H).



II



III



IV

Again, $R^{(j)}$ represents the side-chain of one of the 20 naturally occurring amino acids ($1 \le j \le 20$) given in Table 1.

The torsional angles (ω_0, ϕ_1, ψ_1, ω_1,...) are responsible for folding. The energy of folding is a multi-variable function where the torsional angles are the independent variables. The PEHS for the mono(1°)-, di(2°)- and tri(3°)-peptides, given below, are functions containing 4, 7 and 10 independent variables, respectively.

$$E(1°) = E[\omega_0, \omega_1, \phi_1, \psi_1] \quad (1)$$

$$E(2°) = E[\omega_0, \omega_1, \omega_2, \phi_1, \psi_1, \phi_2, \psi_2] \quad (2)$$

$$E(3°) = E[\omega_0, \omega_1, \omega_2, \omega_3, \phi_1, \psi_1, \phi_2, \psi_2, \phi_3, \psi_3] \quad (3)$$

Clearly, for a degree of polymerization of n amino acids, there are n torsional angle pairs of ϕ_i and ψ_i ($1 \le i \le n$), two terminal peptide functionalities (ω_0 and ω_n) and ($n - 1$) mid-chain peptide bonds [ω_i for $1 \le i \le (n - 1)$]. Thus, the total number of folding variables for a polypeptide containing n amino acids is

$$N = [(n - 1) + 2]\omega + n\phi + n\psi = 3n + 1 \quad (4)$$

The *trans*-peptide bond is more stable than the *cis*-peptide bond. Consequently, it has been traditional to set $\omega_i = 180°$ for all i. This limitation reduces the dimensionality of the problem substantially

$$E(1°) = E_{trans}[\phi_1, \psi_1] \quad (5)$$

$$E(2°) = E_{trans}[\phi_1, \psi_1, \phi_2, \psi_2] \quad (6)$$

$$E(3°) = E_{trans}[\phi_1, \psi_1, \phi_2, \psi_2, \phi_3, \psi_3] \quad (7)$$

The total number of folding variables, after the reduction of the dimensionality, becomes

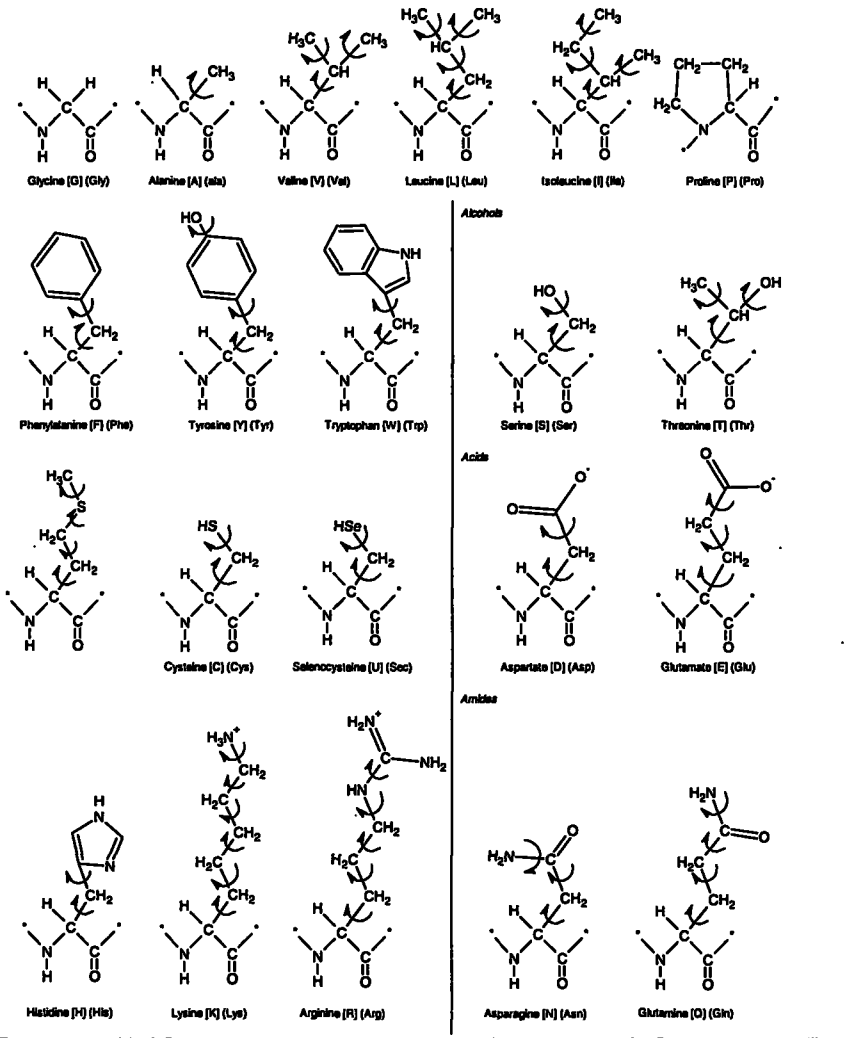
$$N = n\phi + n\psi = 2n \quad (8)$$

This reduction in dimensionality does not represent a denial of importance of the *cis*-configuration of any given peptide bond; it only means that we are partitioning the problem. First, we study the backbone conformations for *trans*-peptide bonds as this represents the primary problem and subsequently we may study the same problem for any peptide bond being in the *cis*-conformation.

Most of the study carried out so far, has been centered on the generation and analysis of the $E(1°)$ potential energy surface (PES). The contour diagram of this type of PES is frequently referred to as the "Ramachandran Map", in honour of the Indian

[2] The 20 naturally occurring amino acids have both DNA and RNA codons. Selenocysteine nowadays are called the 21st amino acid because it has an RNA codon (UCA stop codon) [15].

Table 1
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



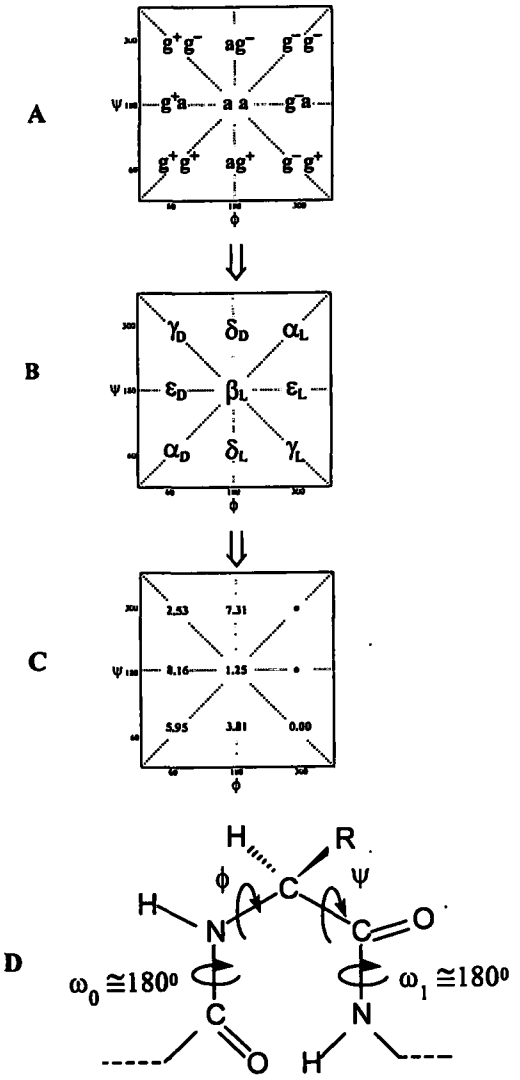


Fig. 2. Topology of a 2D-Ramachandran map. (A) conformational assignments (B) names of conformers (C) relative energies (kcal/mol) of *N*-formyl-L-alanineamide computed at HF/3-21G level of theory (D) structure of a general amino acid residue with the relevant torsional angles.

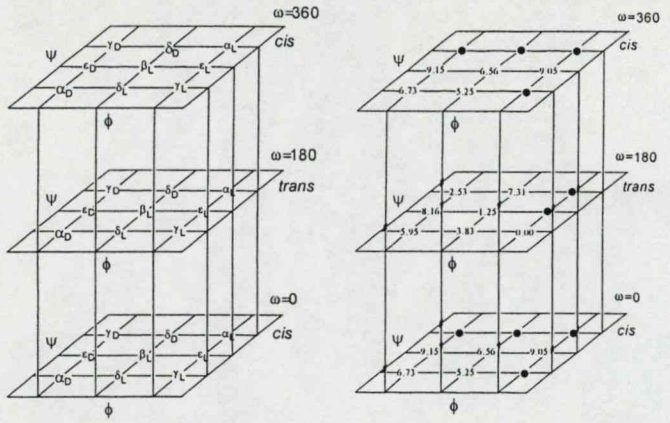


Fig. 3. Topology of a 3D-Ramachandran PEHS. (A) name of conformers (B) relative energies of HCO–(ω)–NH– (ϕ)–CHMe–(ψ)–$CONH_2$.

chemist, Professor Ramachandran, who first called attention to the importance of such a PES.

A topological representation of the Ramachandran map for $R^j = CH_3$ (i.e. for the alanine residue) is shown in Fig. 2. The various minima of the PES, representing the stable conformers, are marked by subscripted Greek letters. Although not all 20 amino acids were subjected to ab initio computational conformational analysis, several amino acids were investigated, and it was recognized that side-chain and backbone conformation can be studied independently, at least in the first approximation.

The relative energies of *N*-formyl alaninamide with *trans*-peptide bond are given in Fig. 2C in the form of a PES topology. The corresponding conformational energy hyper-surface (PEHS) topology, involving both the *trans*- as well as the *cis*-isomers is depicted in Fig. 3.

As yet, these residues have been considered in the absence of conformationally variable side-chains. However, side-chains make contributions to the total energy and they are also involved in backbone/side-chain as well as side-chain/side-chain interactions, thus they ultimately help to determine protein folding. To include these further degrees of freedom in the analytic evaluations requires the extension of the above equations. Labeling of the torsion angle variables in the side-chain is accomplished using χ_1, χ_2, χ_3,... and so on, beginning at the C_α. As the side-chain is the only differentiating structural element between amino acids, each analytic equation also becomes unique. The equations then become:

$$E(1°) = E_{\text{trans}}[\phi_1, \psi_1, (\chi_1^1, \chi_2^1, \ldots, \chi_k^1)] \quad (9)$$

$$E(2°) = E_{\text{trans}}[\phi_1, \psi_1, (\chi_1^1, \chi_2^1, \ldots, \chi_k^1), \phi_2, \psi_2, (\chi_1^2, \chi_2^2, \ldots, \chi_k^2)] \quad (10)$$

$$E(3°) = E_{\text{trans}}[\phi_1, \psi_1, (\chi_1^1, \chi_2^1, \ldots, \chi_k^1), \phi_2, \psi_2, (\chi_1^2, \chi_2^2, \ldots, \chi_k^2), \phi_3, \psi_3(\chi_1^3, \chi_2^3, \ldots, \chi_k^3)] \quad (11)$$

where $(\chi_1, \chi_2, \ldots, \chi_k)$ is specific to each amino acid and its side-chain, with the superscript denoting which residue that side-chain belongs to in the polypeptide chain.

Extending this concept to a generalized form, we obtain the following multi-variable function:

$$E(n°) = E_{\text{trans}}[\phi_1, \psi_1, (\chi_1^1, \chi_2^1, \ldots, \chi_k^1), \ldots, \phi_n, \psi_n, (\chi_1^n, \chi_2^n, \ldots, \chi_k^n)] \quad (12)$$

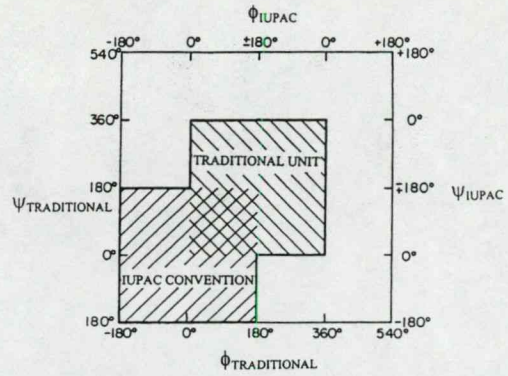


Fig. 4. Two kinds of partition of the PES. The central square corresponds to the traditional cut (0° → 180° → 360°), while the lower left hand square represents the IUPAC conventional cut (−180° → 0° → +180°).

4. Conformational study of single amino acid peptide models

There are two ways to represent the rotation about a single bond. Traditionally, the 0° → 180° → 360° range is used but recently, IUPAC recommended the convention of −180° → 0 → +180°. This convention has the advantage of designating the 0° → +180° segment as a clockwise rotation and the 0° → −180° segment as a counter clockwise rotation. However, it has the disadvantage that certain minima fall on the edges or the corners of the Ramachandran map. The two representations are presented in Fig. 4.

Peptide models, such as $CH_3CONH–CHR–CONHCH_3$ or simply $HCONH–CHR–CONH_2$ can mimic the *i*th amino acid residue in a protein chain. The ϕ and ψ torsional angles are defined in Fig. 2D. The conformational assignments (g^+g^+, g^+a, ...etc.) are shown in Fig. 2A.

The names of the minima (Fig. 2B) are subscripted Greek letters. The Greek letters originate from earlier nomenclature (involving α, β and γ) while the L and D subscripts originate from the observation that L-amino acids favor L conformations while D-amino acids favor D conformations (c.f. lower part of Fig. 5). The names also suggest the combination of the chirality of a constitutional structure (*R* or *S* configuration) and the chirality of the conformational twist or folding. This is summarized in Fig. 6.

The top of Fig. 7 shows the symbolic representation of a conformational PES for two full cycles of rotation (−360° → 0° → +360°) of both ϕ and ψ. The PES can be partitioned into four quadrants, in the traditional way, or it can be partitioned according to IUPAC convention, as shown by the broken lines.

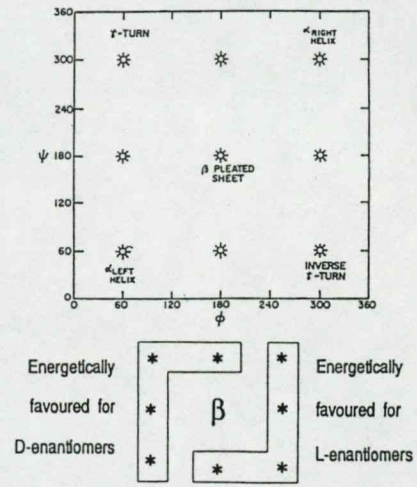


Fig. 5. Underlying principles for choosing subscripted Greek letter (e.g. α_L, α_D, β_L...etc.) as names for the peptide conformations.

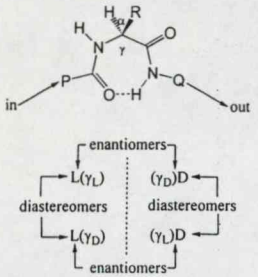
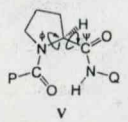


Fig. 6. Stereochemical relationships of γ-turns. Note, that not only the α carbon has chirality, but there is also chirality in the twisting of the backbone. The combination of these two types of chiralities leads to enantiomeric and diastereoisomeric structures. D and L denote the chirality of the C^{α} configuration, while γ_L and γ_D denote the chirality of the conformation.

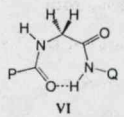
An energy contour diagram of the conformational PES of a peptide (PCONH–CHR–CONHQ), presenting two full cycles of rotation ($-360° \rightarrow 0° \rightarrow 360°$) of both ϕ and ψ, is shown at the bottom of Fig. 7. The central square is the IUPAC conventional cut, while the four quadrants are the traditional cuts. One of these traditional cuts (e.g. the upper right hand quadrant) is shown in pseudo-3D-representation in Fig. 8.

There are 20 naturally occurring amino acids. A total of 18 of them have the same type of backbone folding as shown in Fig. 2 (i.e. nine discrete conformations). The two other amino acids are exceptions.

One exception is proline, which is built into proteins like any other amino acid, but its N is locked in a five-membered ring. For proline residue V, ϕ can only be in the vicinity of $-60°$ (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 **VI**, which is achiral.



V



VI

In the case of the glycine residue, double degeneracy occurs in its conformational PES as shown in Fig. 9.

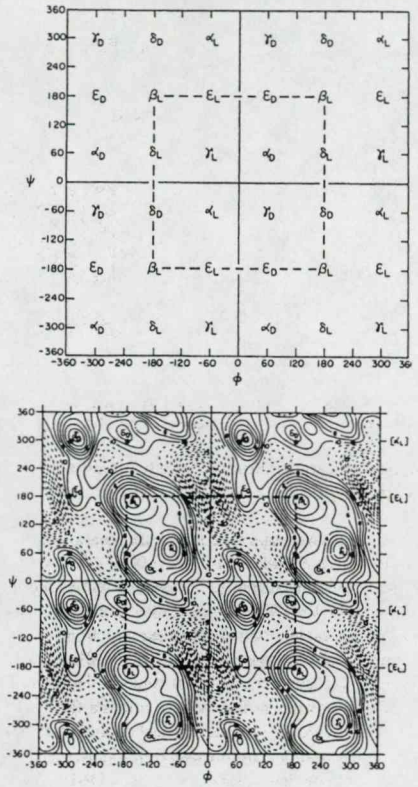


Fig. 7. (Top) A schematic representation of the conformational PES of a peptide (PCONH–CHR–CONHQ). Subscripted Greek letters symbolize the approximate locations of the conformations. (Bottom) Contour diagram of the 2D Ramachandran potential energy surface of $HCONH–CHCH_3–CONH_2$, presented in the $-360° \leq \phi \leq 360°$ and $-360° \leq \psi \leq 360°$ range of independent variables. The central square (broken lines) is the IUPAC conventional cut, while the four quadrants are the traditional cuts.

Finally, it should be mentioned that for certain molecular residues molecular computations have established the actual location of the nine minima, as shown above. The values of ϕ and ψ deviate somewhat from the ideal values, although not too significantly. Table 2 lists these numerical values for

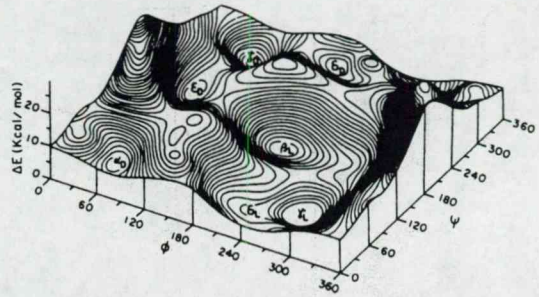


Fig. 8. Pseudo-three-dimensional Ramachandran potential energy surface of $HCONH–CHCH_3–CONH_2$, presented in the $0° \leq \phi \leq 360°$ and $0° \leq \psi \leq 360°$ range of independent variables. This represents one of the four equivalent quadrants in Fig. 7.

alaninediamide. This information is also presented graphically in Fig. 10.

4.1. Previously published single amino acid peptide models

The topological representation of the conformational PES (Ramachandran map) of For-Ala-NH_2 is shown in Fig. 2A, while that of the conformational PEHS is shown in Figs. 7 and 8. Work has been completed for the following N- and C-protected amino acids containing a *trans*-peptide bond: Gly [16,17], Ala [16,17], Val [18], Phe [19,20] and Ser [21–23]. Preliminary studies have been completed on Pro [24], Asp [25], Asn [26], Cys [27], and Sec (selenocysteine) [15].

The following protected amino acid residues, again with *trans*-peptide bonds, are currently under investigation: Arg, Lys, His, Tyr, Leu, Thr and Trp.

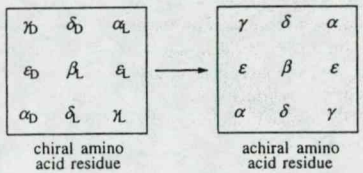


Fig. 9. The development of double degeneracy of the PES when a chiral amino acid residue is changed to an achiral amino acid residue.

Proline is fundamentally different from all the other 18 chiral amino acids in more than one respect:

(i) the R group forms a five-member ring with the backbone;
(ii) there is no peptidic N–H group in the residue to be involved in hydrogen bonding, and
(iii) since there are two carbon atoms connected to the nitrogen, there is a greater chance of *cis*/*trans* isomerization in the peptide bond.

The potential energy cross-sections of the type $E = E(\psi)$, for the Ramachandran map of HCO–Pro–NH_2 containing *cis*- and *trans*- peptide bonds [28] are

Table 2
Optimized ϕ, ψ torsional angle pairs for alanine diamide ($HCONH–CHMe–CONH_2$). 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	−60	−60	ag^-
ϵ_L	−74.7	+167.8	+60	+180	g^-a
ϵ_D	+64.7	−178.6	−180	−180	g^+a

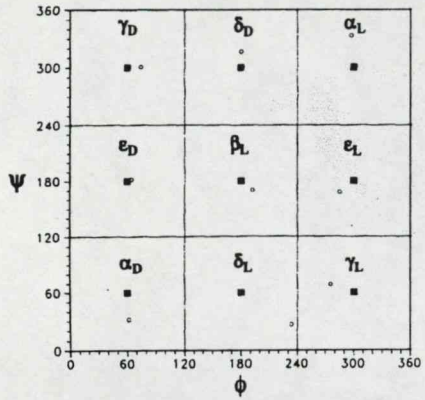


Fig. 10. A schematic illustration of the PES of an average amino acid residue, obtained from the calculations carried out so far on mono-, di- and tri-peptides. The idealized positions are marked by shaded squares and the computationally determined positions are shown as open circles. The names of the conformers are given as subscripted Greek letters. Note that a single amino acid residue might not be able to take on all of the shown conformations.

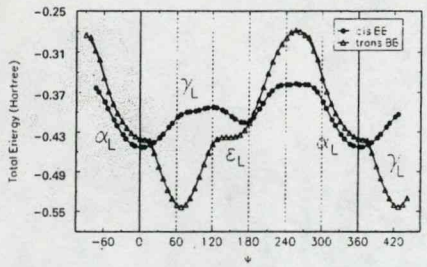


Fig. 11. Conformational Potential Energy Curves $E = E(\psi)$ for *cis* and *trans*-N formyl-L-prolinamide.

shown in Fig. 11. Preliminary investigation on the *cis*-peptide bond has been completed and the *cis*-Ramachandran map is currently under construction [29].

4.2. Single amino acid peptide models in progress

4.2.1. Aspartic acid and asparagine

Aspartic acid (Asp) and its side-chain deprotonated form, i.e. the aspartate [25] ion as well as its side-chain amide asparagine (Asn) [26] have been studied in a preliminary fashion. In both cases, the *N*-formyl-L-aspartic acidamide, *N*-formyl-L-aspartatamide and *N*-formyl-L-asparaginamide were investigated in their γ_L backbone conformation. The side-chain conformational PES for the three compounds are shown in the following illustrations: Figs. 12–14.

4.2.2. Arginine (Arg)

Arginine is an amino acid with a positive charge at neutral pH due to its polar side-chain, the guanidium group. The typical pK_a value of arginine is 12.0, the highest pK_a of all amino acids. It is, therefore, important in salt bridge coupling, which is part of certain docking processes.

Arginine is the source of nitric oxide (NO) in biological systems. NO is a free radical which serves as an intracellular second messenger and an intercellular messenger that regulates neighbouring, and possibly distant cells. NO takes part in many biological processes, such as vasodilation for example. NO is also involved in the central nervous system (CNS), such as the modification of pain perception. One of the two equivalent nitrogens of the terminal guanidine in L-arginine undergoes five-electron oxidation in the presence of the enzyme NO synthase (NOS), yielding NO and L-citrulline. *N*-methyl guanidine has been used as a model compound for arginine to study the mechanism of NO release using ab initio molecular computations [30].

Ethylguanidine or ethylguanidium ion is the terminal portion of the side chain (R-group) of the arginine residue, as exemplified by *N*-formyl-argininamide and its side chain terminal *N*-protonated form respectively. Arginine can be divided into two portions: the backbone, which is analogous to glycine (or alanine) and the side chain, which is an alkyl guanidine.

When a planar moiety is twisted about a single bond with respect to a tetrahedral moiety, in principle, up to six minima may appear. In the case of ethylguanidine **[VII]** and ethylguanidium ion **[VIII]**, there are two torsional angles χ_3 and χ_4. These are two types of bonds, C–C and C–N, with respect to

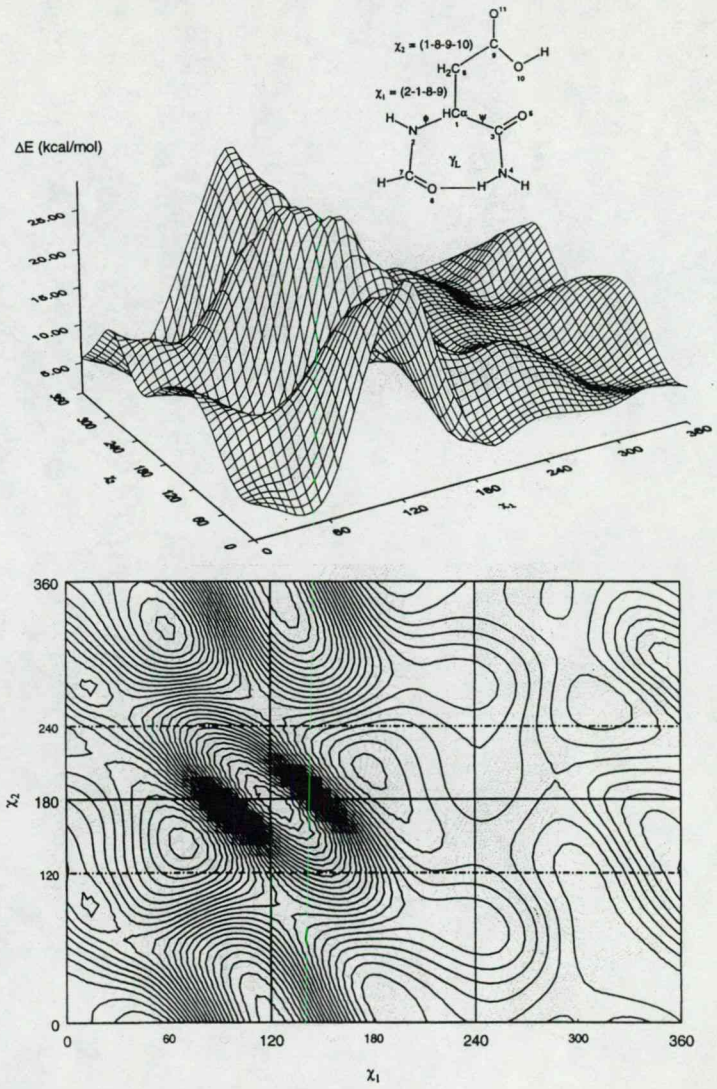


Fig. 12. Side-chain conformational PEHS of *N*-formyl-L-asparticacidemide. (Top) Landscape representation. (Bottom) Contour diagram representation.

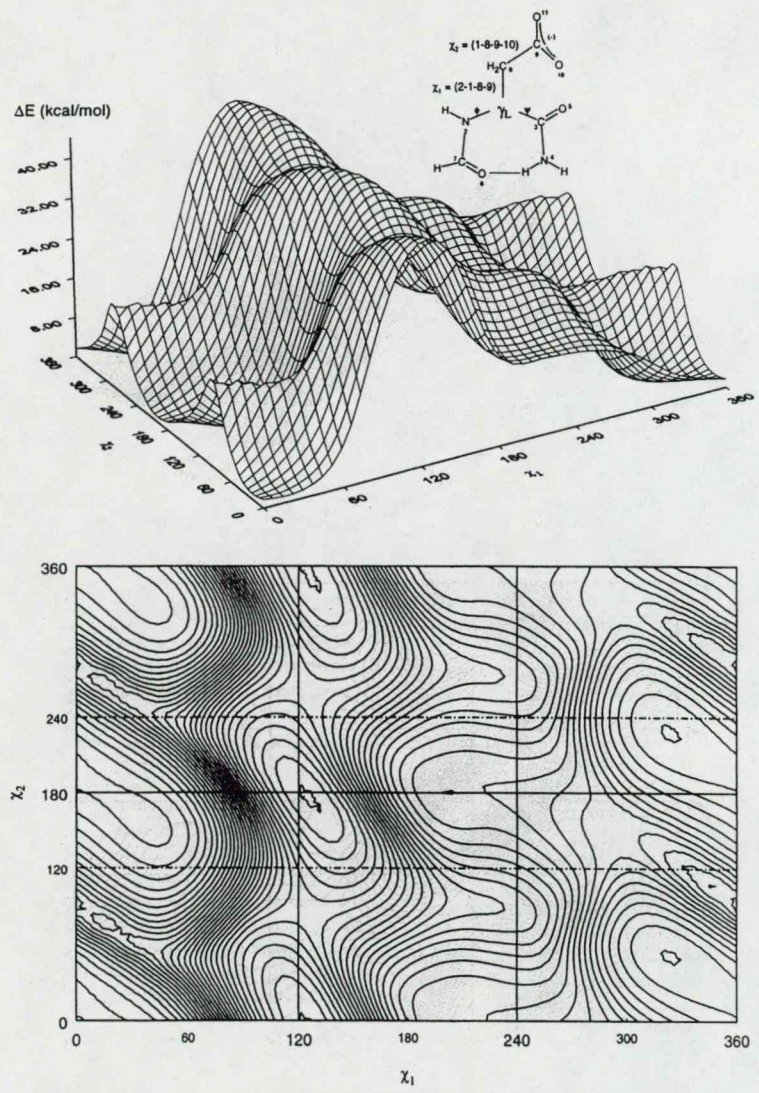


Fig. 13. Side chain conformational PES of *N*-formyl-L-aspartatamide. (Top) Landscape representation. (Bottom) Contour diagram representation.

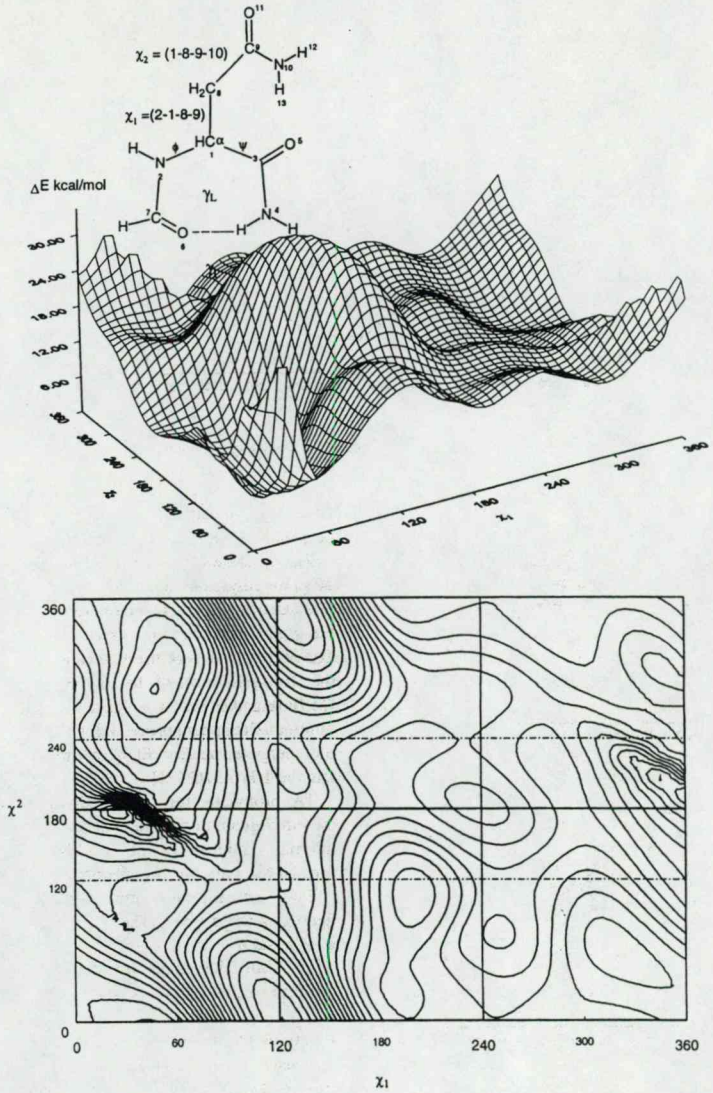


Fig. 14. Side-chain conformational PES of *N*-Formyl-L-asparaginamide. (Top) Landscape representation. (Bottom) Contour diagram representation.

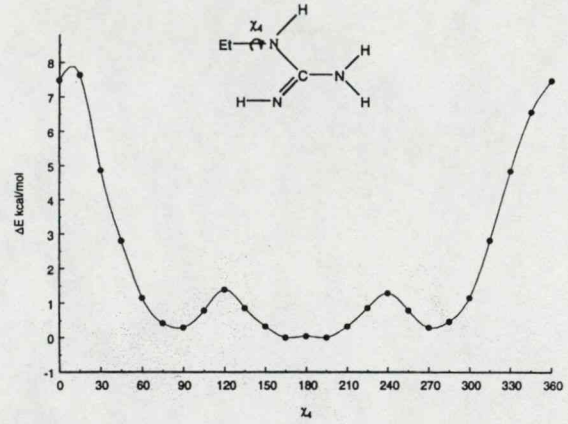


Fig. 15. Conformational PEC of ethyl guanidine, which may model the rotation (χ_4) about the C–N bond in neutral arginine.

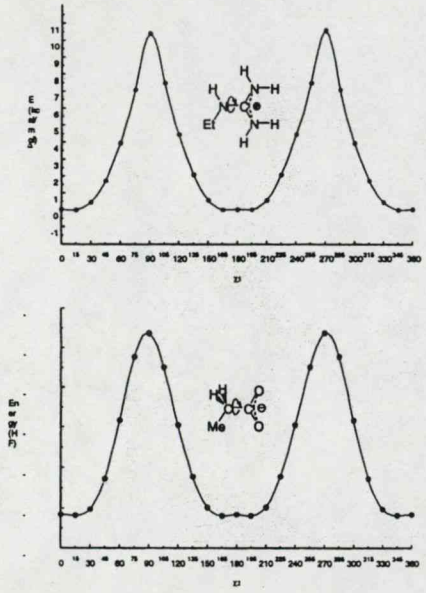


Fig. 16. Rotational potential energy curves of ethylguadinium ion (top) and propionate ion (bottom). Note that each molecule has only one discrete conformation.

which the guanidine group could be eclipsed or be perpendicular to it.

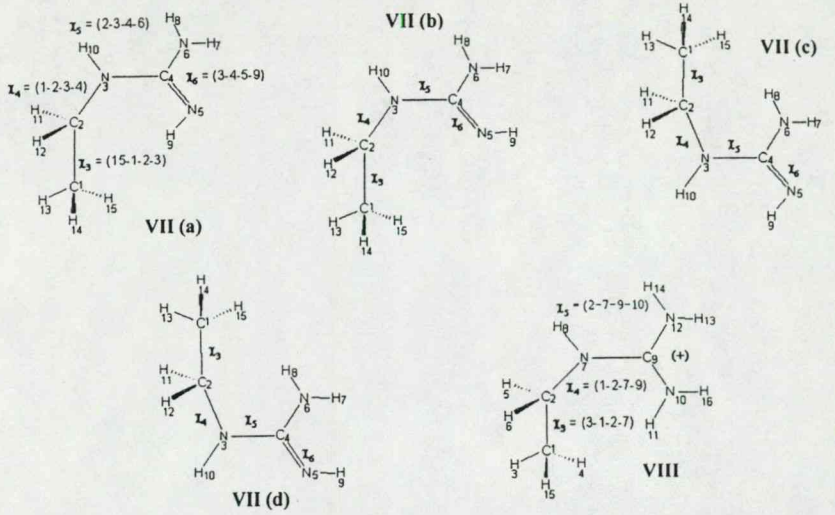
Fig. 15 shows the torsional potential, as a function of χ_4 for ethylguanidine in its *s-cis/exo* [**VII(a)**] structure calculated using two basis sets. For each basis set, the global minimum occurs at 180° (the anti form). The energy values of the 1D-scans were computed at the HF/6-31G level of theory. For all four structures [**Ia**,**Ib**,**Ic** and **Id**] g^+, a, g^- were shown to have energy minima. The upper portion of Fig. 16 shows the potential energy curve $E = E(\chi_5)$ of structure **VIII** calculated at HF/6-31G [31].

The protonated form of the ethyl guanidine [**VII**] is the ethyl guanidium structural ion [**VIII**]. Due to the internal symmetry of the $N–C(NH_2)_2^+$ moiety, this species has only one arrangement.

Two vertical proton affinity values may be identified using these PECs. The two vertical ones involve $g \rightarrow g$ and $a \rightarrow a$.

The full side-chain orientation of N- and C-protected arginine and its conjugate acid depends on χ_1, χ_2, χ_3, and χ_4 as illustrated by molecular structures **IX** and **XI**.

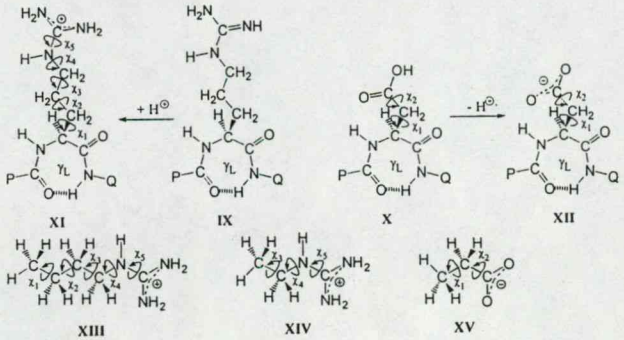
Consider the diamides of the two amino acids arginine **IX** and aspartic acid **X**. Under physiological conditions the amino acid side chains occur in the protonated **XI** and deprotonated **XII** forms



respectively. The side chain conformations could be mimicked for **XI** by **XIII** and to some extent **XIV** and for **XII** by **XV**. The potential energy curve, $E = E(\chi_5)$, for **XIV** and that for **XV**, $E = E(\chi_2)$, is shown in Fig. 16. The side chain PES of **XII** has also been reported and is shown in Fig. 13. It is clear that torsional angles within the amino acid side chains do increase the dimensionality of the conformational space. Nevertheless, they are very important, because the side chain makes each amino acid unique. Such distinction can be made because, to a good degree of approximation, the side chain conformational pattern is independent of the backbone conformational behavior. For all practical purposes, the backbone of all amino acids shows the same conformational pattern (c.f. Figs. 2, 7, 8 and 10).

4.2.3. Cysteine (Cys) and selenocysteine (Sec)

Each of serine [21–23,32], cysteine [27] and



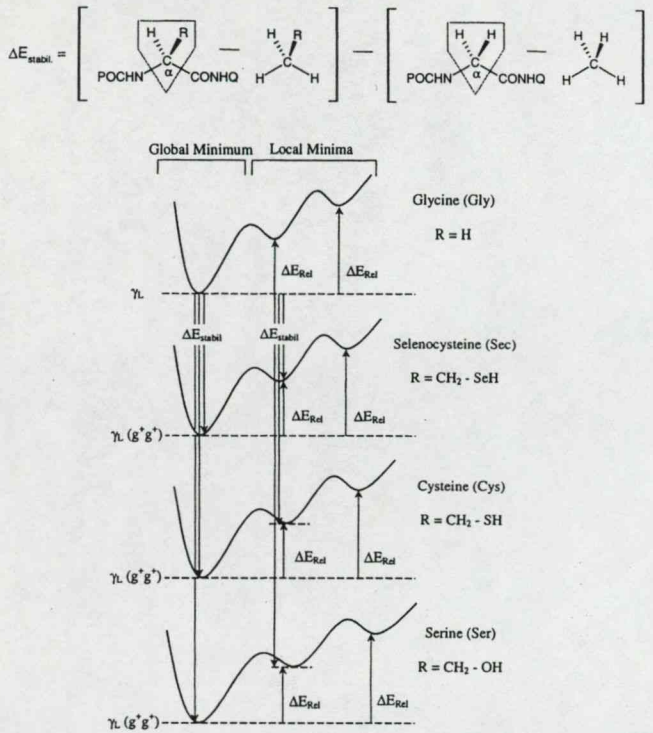


Fig. 17. Calculation of stabilization energies ($\Delta E_{stabil.}$) with various substituents (R may be CH_2-OH, CH_2-SH and CH_2-SeH) relative to the γ_L conformation of Glycine ($R = H$).

selenocysteine [15] is expected to have $9 \times 9 = 81$ conformations [$3 \times 3 = 9$ backbone: $\psi(g^+, a, g^-) \times \phi(g^+, a, g^-)$ and $3 \times 3 = 9$ side-chain: $\chi_1(g^+, a, g^-) \times \chi_2(g^+, a, g^-)$]. To investigate the effects of side-chain/backbone conformational interactions, all torsional modes of the side-chain (χ_1: rotation about the C^α–C^β and χ_2: rotation about the C^β–X bonds) were studied in the relaxed γ_L backbone [$(\phi, \psi) = (g^-, g^+)$] conformation. Six out of the nine expected minima for serine and seven out of the nine expected minima for cysteine and selenocysteine were found at the RHF/3-21G level of theory.

The stabilization energy exerted by the $-CH_2-SeH$ side-chain has been compared to that of $-CH_2-SH$ and $-CH_2-OH$. The stabilization energies were calculated with respect to the γ_L backbone conformation of N- and C-protected glycine [33,34] using the following isodesmic (same number of bonds) reaction, where P and Q may be H or CH_3 and $R = CH_2-XH(X = O,S,Se)$:

$$\underset{\text{conformation } \gamma_L}{PCONH\text{–}CH_2\text{–}CONHQ + CH_3\text{–}R} \rightarrow \underset{\text{conformation X}}{PCONH\text{–}CHR\text{–}CONHQ + CH_3\text{–}H} \quad (13)$$

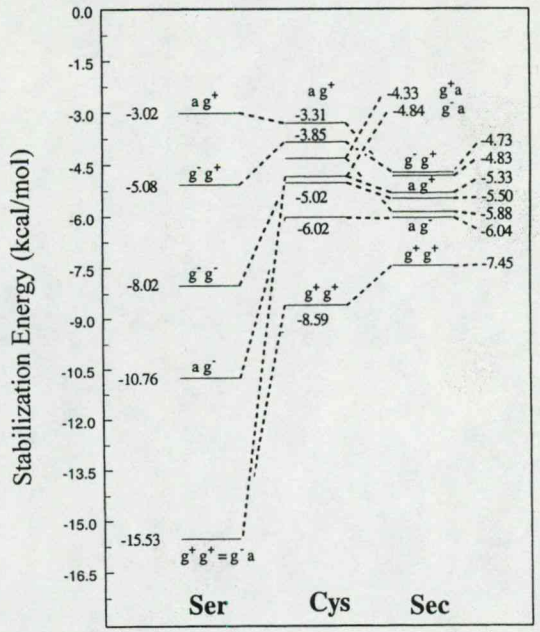


Fig. 18. Spectrum of conformational dependence of the side-chain stabilization energy on the γ_L backbone conformation of formyl Ser-, Cys- and Sec-amide.

The stabilization energy may be calculated as follows:

$$\Delta E_{stabilization} = \{E[PCONH\text{–}CHR\text{–}CONHQ]_x + E[CH_3\text{–}H]\} - \{E[PCONH\text{–}CH_2\text{–}CONHQ]_{\gamma L} + CH_3 - R]\} \quad (14)$$

These equations are also illustrated in Fig. 17.

From Fig. 18, it can be observed that the effect of side-chain orientation with respect to the peptide backbone, going from Serine to Selenocysteine, is gradually decreasing. This is due to the increase in the size of the atoms $O \rightarrow S \rightarrow Se$ and the resulting gradual lengthening of $C\text{–}O \rightarrow C\text{–}S \rightarrow C\text{–}Se$ bond lengths as well as the reduced hydrogen bonding ability in going from OH to SH and all the way to SeH.

Topological analysis of the electron density has been performed for cysteine and selenocysteine, using Bader's Atoms in Molecule (AIM) approach at the B3LYP/6-31G(d,p) level of theory. The hydrogen bonds were verified by the existence of bond critical points. Three conformations: (g^+ g^+), (g^+ a) and (a g^-) exhibited such interactions as illustrated graphically in Fig. 19 for For-Sec-NH_2.

4.2.4. Leucine (Leu)

Leucine is one of the naturally occurring essential amino acids. It is an important amino acid, having many implications in medicine, one of them being its role in Maple Syrup Urine disease (MSUD). Maple syrup disease occurs when plasma leucine, isoleucine, alloisoleucine and valine are elevated to very high levels. These amino acids will also be present in the urine and hence the urine releases

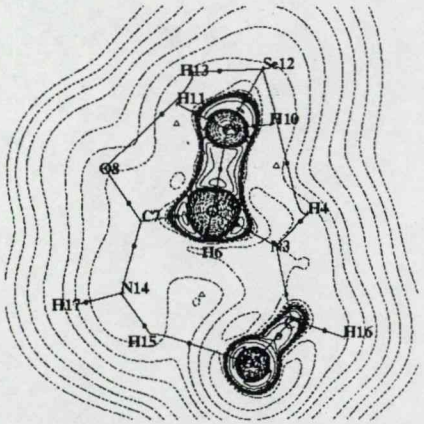


Fig. 19. The contour map of the Laplacian of the electron density for the (g^+g^+) conformation of For-L-Sec-NH_2 calculated at the B3LYP/6-31+G(d,p)//RHF/3-21G level of theory. Bond paths are denoted by lines, bond critical points (BCPs) are denoted by black dots, ring critical points (RCPs) are denoted by open triangles and the nuclei are denoted by crosses.

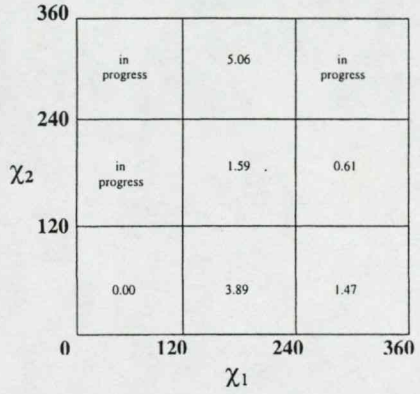


Fig. 20. Relative energies of the N- and C-protected leucine with varying side chain conformations for the β_L backbone conformation retaining *trans* peptide bonds calculated at HF/3-21G level of theory.

an odor similar to maple syrup. This disease has severe neurological and gastrointestinal effects. The current treatment for it involves intense control of the intake of leucine, isoleucine and valine. This careful restriction of leucine also includes monitoring of the branched chain amino acids in plasma.

Leucine is also found in some small proteoglycans. This group of small, interstitial proteoglycans has a high degree of homology in their protein core sequence. Each has between 10 and 12 highly conserved leucine rich tandem repeats which makes up the central portion of the core protein. The current members of the group are fibromodulin and lumican, which are keratan sulphate substituted, as well as decorin and biglycan, which are both chondroitin sulphate substituted. Other proteoglycans which have had functions ascribed to them are the small leucine rich interstitial proteoglycans decorin and fibromodulin which have been shown to modulate collagen fibrillogenesis.

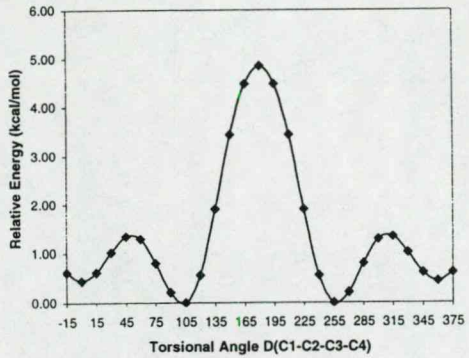
The knowledge of the conformational potential energy surface for leucine would allow a better understanding of diseases such as MSUD and the function of leucine in biological processes such as collagen fibrillogenesis. Each of the nine possible backbone conformations is associated with nine possible side chain conformations. In addition, there are four chain-end conformations associated with the *cis* and *trans* peptide bond that also must be explored. Thus the total number of conformations to be considered is $9 \times 9 \times 4 = 324$. The results of some calculations that have been carried out for the β_L backbone conformation are shown in Fig. 20.

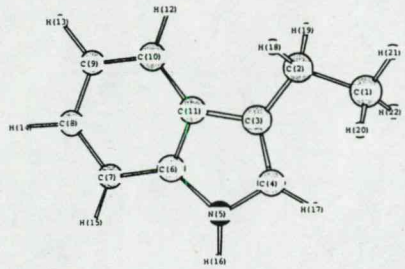
4.2.5. Tryptophan (Trp)

Tryptophan is one of the essential amino acids the body cannot manufacture itself. It is the least abundant in proteins and is destroyed easily by the liver. This amino acid is necessary for the production of the vitamin B niacin, which is essential for the brain to manufacture the key neurotransmitter serotonin. Low levels of serotonin have been linked with insomnia, anxiety and depression.

Tryptophan is believed to be effective for insomnia and jet lag. It is also believed to have anti-anxiety effects and control aggressive behavior in some individuals. Other studies have shown that its anti-depressant effect lasts longer than that of the popular anti-depressant drug, Imipramine. In addition, some preliminary studies of combined Vitamin B-6 and tryptophan show that they may be effective in reducing the severity of hyperventilation as well as the panic attacks produced by it.

Although it is safe to use, tryptophan is no longer available in supplement form, because of potential adverse reactions. Tryptophan supplements are not recommended in cases of pregnancy, asthmatics, or auto-immune disorders like Lupus or Scleroderma. As a preliminary investigation, a scan at HF/3-21G level of theory was performed on ethyl indole (Fig. 21).





	RHF Regular Optimization			RHF Tight Optimization
	STO-3G	3-21G	6-31G	3-21G
Planar (0 degrees)	-434.1966671 (hartree)	-437.0948503 (hartree)	-439.3773570 (hartree)	-437.0948513 (hartree)
Perpendicular	-	-437.0955447 [a] (hartree)	-439.3775512 [b] (hartree)	-437.0955634 [c] (hartree)

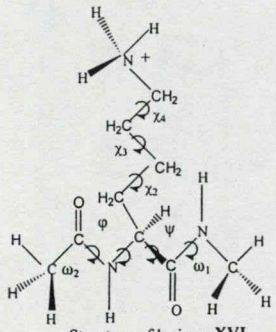
[a] Torsional angle was found to be 105.00 degrees.
[b] Torsional angle was found to be 104.46 degrees.
[c] Torsional angle was found to be 102.34 degrees.

Fig. 21. Potential energy curve of D(C1–C2–C3–C4) of ethyl indole.

Minima were found to exist at torsional angles (D) of 0 and in the vicinity of 105°.

4.2.6. Lysine (Lys)

Similar to arginine, lysine **XVI** is a basic hydrophilic amino acid. At neutral pH, it is positively charged, with a typical pK_a of 10.0. One can imagine that the two amino acids would be able to assume similar roles in biological systems. They may increase the solubility of proteins, furthermore, they can provide binding sites at the surface of proteins. However, there is a main stereochemical difference between the two molecules. Lysine has a hydrocarbon chain with a terminal tetrahedral $NH_3(+)$, whereas the terminal guanydilic ion of arginine is planar. This difference may be important in the fine-tuning of molecular interactions, for example, between substrate and the active site.



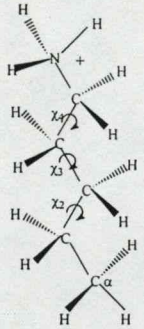
Structure of lysine **XVI**

Both, the side chain **XVII** and the lysine residue **XVI** with the protected backbone, are currently being studied.

5. Dipeptide, tripeptide, and tetrapeptide

Most of the oligopeptides studied so far at the ab initio level of theory involved alanine. All backbone conformers of dialanine diamides have been studied [35–38]. Analogously trialanine diamide (tripeptide) [39] and tetraalanine diamide (tetrapeptide) [40] have also been investigated at the ab initio level but only a selected few conformers were studied. In this paper, we wish to illustrate in the case of tripeptide Arg-Gly-Asp (or RGD for short) and in the case of tetrapeptide Pro-Pro-Thr-Pro (or PPTP for short) the medical importance of such small oligopeptide fragments.

Single amino acid diamides or "monopeptides" for short, have nine unique conformers. For diamino-acid diamides, or "dipeptides" for short, each amino acid residue has nine conformers. This leads to $9 \times 9 = 81$ conformers for dipeptides. In the case of tripeptides, for each of the 81 dipeptides conformations there are 9 conformers for the third amino acid. This results in $9 \times 9 \times 9 = 729$ conformers. The name of each of the 729 conformers, given in subscripted Greek letters, is summarized in Fig. 22. These 729 conformers are derived within the law of MDCA, and in that respect



Side chain of lysine **XVII**

they may be regarded as "legitimate" conformers. Not all 729 conformers are expected to exist. In certain folding patterns, the terminal groups may come too close and such minima may be annihilated due to steric repulsion. The pattern shown in Table 2 has been previously recognized for $HCON-(Ala)_n-NH_2$ via ab initio computations (Table 3).

Mono-peptides, dipeptides and tripeptides contain 1, 2 and 3 amino acid residues, respectively. In such models, the chain may be terminated by methyl groups or by hydrogens, ($P = Q = CH_3$ or H) as shown for the above three cases by

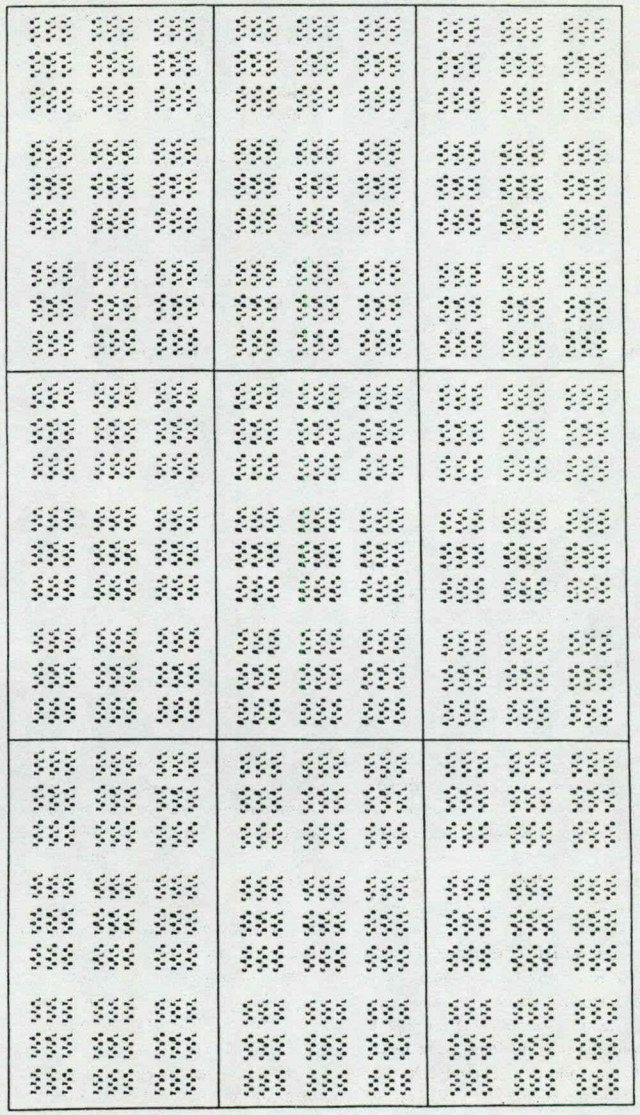


Fig. 22. Names, in terms of subscripted Greek letters, of the 729 legitimate backbone conformers of tripeptides.

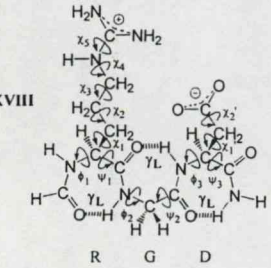
Table 3
Legitimate and existing conformers of small peptides

n	Legitimate conformers	Conformers found
1	9	7
2	81	49[a]
3	729	Not investigated

[a] These 49 conformers were found by optimizing the 81 "legitimate" conformers. Subsequently, Prof. Lothar Schafer and coworkers [Can. J. Chem. 76 (1998) 566] have scanned the conformational space computing 11 664 grid points at the HF/4-31G level of theory and located one more conformer. Now the number of conformers stands at 50.

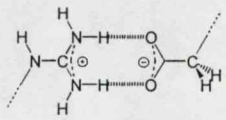
II, **III** and **IV**. The symbol $R^{(j)}$ represents the side-chain of one of the 20 naturally occurring amino acids ($1 \leq j \leq 20$).

XVIII



R G D

XIX



5.1. Tripeptide sequence, RGD

The tripeptide Arg-Gly-Asp (RGD) is an adhesive molecule which has a positively and a negatively charged side chain as shown in **XVIII** in accordance with **XI** and **XII**. The positively and negatively charged side chains may be engaged in salt bridge **XIX** formation either via an intra- or an intermolecular connection.

The structure shown in **XVIII** is the amide of the *N*-formyl derivative of the RGD-motif and it is given in its $\gamma_L\gamma_L\gamma_L$ backbone and extended (fully *anti*) side chain conformations. Once again, assuming that all peptide bonds are of *trans* configuration, the PES can be represented analytically by the following function **XX**.

The backbone may be represented by a 6D-conformation subspace (ϕ_1,ψ_1,ϕ_2,ψ_2,ϕ_3,ψ_3). Since we may expect only one distinct conformer along χ_5 and χ_2', the side chain may be represented by a 5D-conformational chain subspace (χ_1, χ_2, χ_3, χ_4, χ_1'). Consequently, we may expect $3^5 = 243$ side-chain conformations. The total number of distinct RGD conformers may, therefore, be $729 \times 243 = 177147$. In view of the enormity of the problem only a selected few conformations were subjected to molecular computations.

Table 4 summarizes the geometrical parameters (e.g. torsional angles), conformational assignments and relative stabilities of the RGD structures optimized at HF/3-21G level of theory.

The last two structures of the table are depicted in Fig. 23. Note that one of them is without the involvement of Ca^{2+}, and therefore, it exhibits intramolecular salt bridge. The other structure is with Ca^{2+}, implying that the previously negatively charged side chain ($-COO^-$) now contains a mono-positive ($-COO^-\cdots Ca^{++}$) ending. Consequently, there is no chance for intramolecular salt-bridge formation. This charge difference in side-chain has implication for the docking mechanism. Fig. 24 shows a possible docking mechanism without the involvement of Ca^{2+}.

Fig. 25 indicates that in the presence of Ca^{2+}, the receptor would need two negatively charged side chains (i.e. two carboxylate ions).

It is interesting to recall the result of the physiological study [41], involving GRGDS(β-OH) (c.f. Fig. 26) and GRGDS(β-O-SO_3^-). The final results are summarized in Table 5.

It seems that when Ca^{2+} ion is not present the mechanism given in Fig. 24 is operative and thus an extra negative charge at the C-terminal makes no

Table 4
Geometrical parameters, conformational assignment, total energies and relative stabilities of collected RGD conformers without (A–E) and with (F) the inclusion of Ca^{2+}, optimized at the HF/3-21G level of theory

Residue		A		B		C		D		E		F[a]	
		Geometry	Conformation	Geometry	Conformation	Geometry	Conformation	Geometry	Conformation	Geometry	Conformation	Geometry	Conformation
Arginine (R)	ω^0	177.5804	–	172.5255	–	−175.5717	–	−177.5620	–	−171.2285	–	−176.4709	–
	ϕ^1	−153.0601	δ_L	−155.4259	δ_L	46.2751	α_D	−81.9119	γ_L	−82.2705	γ_L	−153.6854	β_L
	ψ^1	47.7288		53.3474		41.4783		66.0202		66.0650		165.4315	
	ω^1	−179.4318	–	173.4531	–	−178.5048	–	−179.8780	–	−179.5433	–	175.9534	–
	χ^1	44.2825	g^+	58.1487	g^+	57.9267	g^-	57.1272	g^-	−61.4889	g^-	−60.7727	g^-
	χ^2	75.7941	g^-	123.3340	a	96.1472	g^+	142.7702	a	−64.8448	g^-	−66.1861	g^-
	χ^3	−92.5497	g^-	−166.2056	a	−144.5144	a	−64.6406	g^-	−57.8581	g^-	−175.4661	a
	χ^4	155.5904	a	79.6676	g^-	96.1651	g^+	169.7310	a	142.8879	a	−96.2058	g^-
Glycine (G)	ϕ^2	120.6121	β_L	96.0964	δ_L	123.2492	δ_D	−87.4209	γ_L	−65.5310	α_L	173.2842	β_L
	ψ^2	−131.5270		−103.2997		−34.2922		58.5616		−26.2768		179.6252	
	ω^2	−175.4871	–	171.9142	–	171.0358	–	−159.1781	–	−169.9989	–	−174.1726	–
Aspartate (D)	ϕ^3	−66.4906	ϵ_L	−84.2133	ϵ_L	−161.4273	δ_D	−93.1384	γ_L	−81.2064	γ_L	−83.6510	γ_L
	ψ^3	171.5693		−141.3300		−119.3301		53.6705		69.7315		52.2155	
	ω^3	−177.6658	–	−155.8824	–	−158.9678	–	−171.2286	–	173.9757	–	−172.5103	–
	χ^1	59.2826	g^+	64.5079	g^+	48.5048	g^+	37.1034	g^+	53.9209	g^-	168.2582	a
	χ^2	89.7330	g^+	106.2422	g^+	89.8874	g^+	29.3742	g^+	155.2624	a	37.4610	g^-
Energy (hartree)		−1328.6774261		−1328.7079568		−1328.6899608		−1328.6819226		−1328.69329353		−2001.8202808	
ΔE (kcal/mol)		19.16		0.00		11.29		16.34		9.20		N/A	
Backbone conformation		$\delta_L\delta_L\epsilon_L$		$\delta_L\delta_L\epsilon_L$		$\alpha_D\delta_D\delta_D$		$\gamma_L\gamma_L\gamma_L$		$\gamma_L\alpha_L\gamma_L$		$\beta_L\beta_L\gamma_L$	

[a] Includes a Ca^{2+} in the vicinity of the asparate side chaincarboxylate moiety.

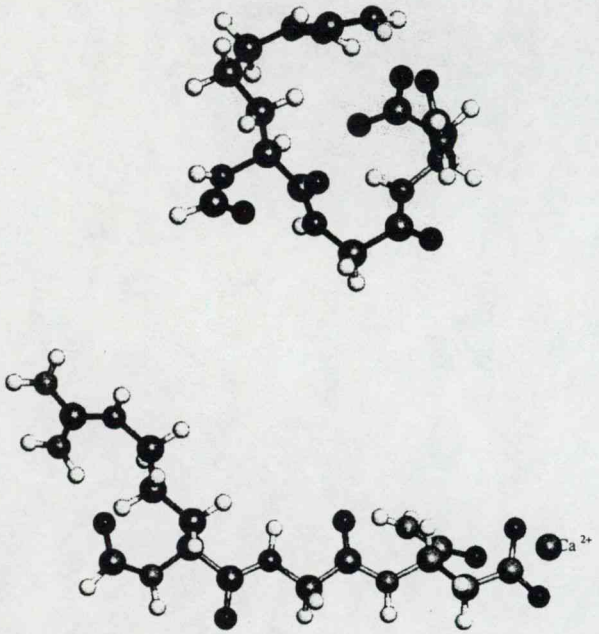
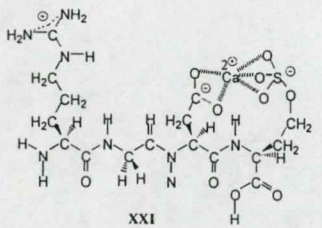


Fig. 23. Two conformers of RGD without (top) and with Ca^{2+} (bottom) optimized at the HF/3-21G level of theory corresponding to structures E and F respectively in Table 3.

significant change. However, when Ca^{2+} ion is present, GRGDS(β-OH) becomes very active, favoring the complex formation according to Fig. 25. However, the Ca-salt of GRGDS(β-OSO_3^-) has only one charge, a positive charge at the Arg side chain, but no charge at the DS(β-OSO_3^-) moiety since the two (+) and the two (−) will lead to a neutralization as illustrated in **XXI**. It is not surprising therefore that under these conditions the activity is reduced greatly. One structure using Mg ion rather than Ca ion is optimized at the HF/3-21G level of theory and is shown in Fig. 26.

$$E = E(\underbrace{\phi_1,\psi_1,\phi_2,\psi_2,\phi_3,\psi_3}_{\text{backbone}},\ \underbrace{\chi_1,\chi_2,\chi_3,\chi_4,\chi_5,\chi_1',\chi_2'}_{\text{side-chain}})$$

XX



XXI

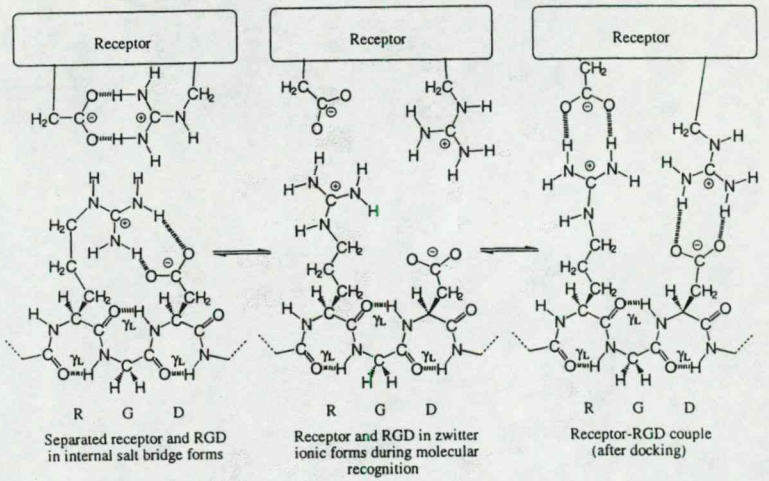


Fig. 24. A possible docking mechanism of RGD operative in the absence of Ca^{2+} ion. Note that the backbone is shown arbitrarily in its $\gamma_L\gamma_L\gamma_L$ conformation.

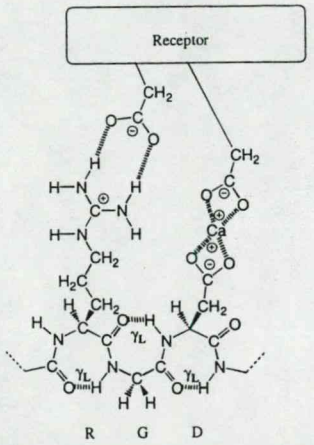


Fig. 25. Receptor-RGD-complex after docking, in the presence of Ca^{2+} ion.

5.2. Tetrapeptide sequence, PPTP

For a discussion of a tetrapeptide example, let us briefly review immunoglobulin A (IgA). Antibodies or immunoglobulins (Ig) represent the great defense force in the army of the immune system. Ig molecules are produced during a humoral immune response to bind and neutralize invading antigens. Subsequently, the antigen is labeled for removal by phagocytosis. The fight, however, is not one-sided. To help neutralize the role of Ig in antigen recognition, bacteria have developed effective deactivation mechanisms. One of them involves a site-specific cleavage of an Ig molecule at the hinge region with the aid of an extracellular protease produced by the bacteria.

Although there are differences in their catalytic mechanisms, all reported IgAses fall into one of three types: serine-, cysteine-, or metallo-proteases [42]. The IgAse from *N. gonorrhea* has been classified as a serine protease [43]. A comparison between the gene and amino acid sequences of *N. gonorrhea* and *H. influenzae* [44,45] showed that they are 50% identical and contain a fingerprint region for the chymotrypsin/trypsin family of serine proteases [42]. The most conserved sequence,

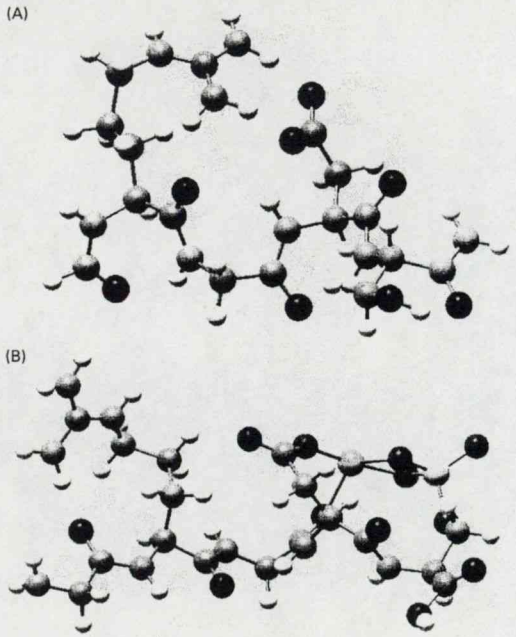


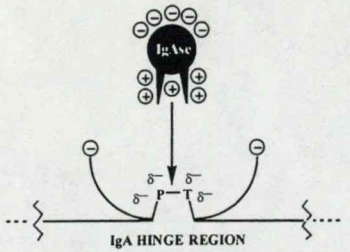
Fig. 26. (A) An internal salt-bridge forming conformer of RGDS(β-OH). Conformer optimized at the HF/3-21G level of theory. (B) An Mg^{2+} complex of RGD (β-OSO3-). The conformer shown was optimized at the HF/3-21G level of theory.

GDSGGPL (S = active serine site) is found in the vicinity of Ser 250 in *N. gonorrhea* IgAse.

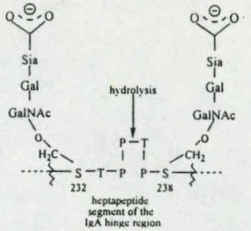
There is an equal amount of anionic forming (Glu, Asp, Tyr) and cationic forming (Arg, Lys, His) amino acid side chains present [44]. Based on the nature of the inhibitors, there would be reason to believe that these oppositely charged moieties are not uniformly distributed in the tertiary structure. Instead, it is likely that this enzyme contains concentrated areas of positive charge that would enable the IgAse to seek out substrates that are negatively charged **XXII**. The fact that serine proteases contain histidine at the active site suggests that the active site of the enzyme will also be positively charged.

Table 5
Summary of physiological effects of GRGDS(β-OH) and GRGDS(β-OSO_3^-) in the absence and in the presence of Ca ion

Biological phenomena	$[Ca^{2+}]$	Peptide sequence	
		GRGDS(β-OH)	GRGDS(β-OSO_3^-)
Platelet aggregation	Not present	Active	Active
Vasodilator effect	Present	Very active	Not active



XXII



XXIII

There have also been inhibition studies done on short peptide segments of the hinge region [46], such as Pro-Pro-Thr-Pro, as well as longer ones and these have been found to act as mild inhibitors. They exhibit IC_{50} values in the micro-molar (10^{-6} M) range or higher. Thus, they are weak inhibitors when compared with the ideal nano-molar (10^{-9} M) range values expected for a drug candidate.

Successful drug candidates, therefore, may need to mimic the heptapeptide segment of the hinge region containing the negative charges of the two sialylic acid moieties **XXIII**. The structure must be appropriately folded and should not exceed a certain molecular weight. Because bacteria became resistant to most antibiotics such an inhibitor may become a successful lead compound for the development of several possible antibacterial agents. While these target oriented research projects may be laudable nevertheless it is lamentable whether we have enough information to search for these highly desirable compound. What is needed is an extensive database, which includes peptide, cyclo-peptide, and peptidomimetic conformations.

To date, neither X-ray structure for the IgA hinge region nor for the IgA protease produced by *N. gonorrhea* are available. Until such structures become available, we can carry out some necessary conformational structure analyses. With the aid of the enzyme structure, the drug design process may proceed at an accelerated pace. The most obvious place to start is with the tetrapeptide Pro-Pro-Thr-Pro. The scission occurs at the Pro-Thr peptide bond and we have included at least the nearest neighbours at each end (XXII).

Clearly, this Pro-Pro-Thr-Pro tetrapeptide is more complicated than Ala-Ala-Ala-Ala [40] because Pro may exhibit *cis*- and *trans*-isomerism in its peptide bonds, as well as *syn*- and *anti*-ring puckering. Thus, $2 \times 4 = 4$ isomeric forms may exist for each backbone conformer. Up to three backbone conformers (γ_L, ϵ_L, α_L) may be expected; thus each proline moiety may have $4 \times 3 = 12$ conformations (Table 6). Consequently, for the dipeptide Me-CO-Pro-Pro-CO-NHMe, we may anticipate $12 \times 12 = 144$ conformations. However, even if one disregards ring puckering and considers only *trans*-peptide bonds, Pro-Pro-Thr-Pro may exhibit $3 \times 3 \times 9 \times 3 = 243$ backbone conformations. Including the $3 \times 3 = 9$ side chain conformations of the Thr side chain moiety, this will lead to a total of 243×9 or 2187 conformers. Thus, the 243 backbone conformations may be best studied on Pro-Pro-Ala-Pro and the Ala side chain may be extended to a Thr side chain subsequently. One of the 2187 conformers has been optimized in our preliminary study and the structure is shown in Fig. 27. The central Pro-Thr structure turned out to be γ_L γ_D, which has been previously identified as a new type of β-turn [36]. It is interesting to note that several of the oxygen atoms occupy a region of the space creating an electron-rich domain. Such a display of high electron density may play a role in the docking of the PPTP moiety within the hinge into the cavity of the IgAse.

Table 6
Side chain dimensionality of selected representative compounds and their corresponding amino acids—excluding methyl-group rotation (proline and glycine have no dimensionality; alanine has methyl rotation dimensionality only)

Model compound		Amino acid residue	
Dimensionality	Structure	Compound	Dimensionality
0	$CH_3-CH-(CH_3)_2$	Valine (Val)	1
1	$CH_3-CH_2-CH-(CH_3)_2$	Leucine (Leu)	2
1	$CH_3-CH(CH_3)-CH_2-CH_3$	Isoleucine (Ile)	2
1	CH_3-CH_2-Ph	Phenylalanine (Phe)	2
1	CH_3-CH_2-COOH	Aspartic Acid (Asp)	2
1	$CH_3-CH_2-CONH_2$	Asparagine (Asn)	2
1	$CH_3-CH_2-C=CH-NH-CH_2=N:$	Histidine (His)	2
1	CH_3-CH_2-she	Selnocysteine (Sec)	2
1	CH_3-CH_2-SH	Cysteine (Cys)	2
1	CH_3-CH_2-OH	Serine (Ser)	2
1	$CH_3-HCOH-CH_3$	Threonine (Thr)	2
1	$CH_3-CH_2-Indole$	Tryptophan (Trp)	2
2	$CH_3-(CH_2)_2-S-CH_3$	Methionine (Met)	3
2	$CH_3-CH_2-Ph-OH$	Tyrosine (Tyr)	3
2	$CH_3-(CH_2)_2-COOH$	Glutamic Acid (Glu)	3
2	$CH_3-(CH_2)_2-CONH_2$	Glutamine (Gln)	3
3	$CH_3-(CH_2)_4-NH_3^{(+)}$	Lysine (Lys)	4
3	$CH_3-(CH_2)_3-NH-C-(NH_2)_2(+)$	Arginine (Arg)	4

6. Cyclopeptide conformations for $C_6H_9N_3O_3$ and $C_8H_{12}N_4O_4$

Cyclopeptide structures **XXIV** have not been studied as extensively as their open chain counterparts. Once the ring size increases, it is more difficult to optimize all stable conformations in order to explore the conformational space of these types of cyclic molecules. Ab initio MO (HF/STO-3G, HF/3-21G, HF/6-31G**) and DFT/6-31G** (using B3LYP functional) computations are to be carried out on various cyclopeptides $(C_2H_3NO)_n$ for $3 \leq n \leq 4$. All-*cis*-CONH and all-*trans*-CONH structures were considered. Typical optimized structures are shown in Figs. 28 and 29.

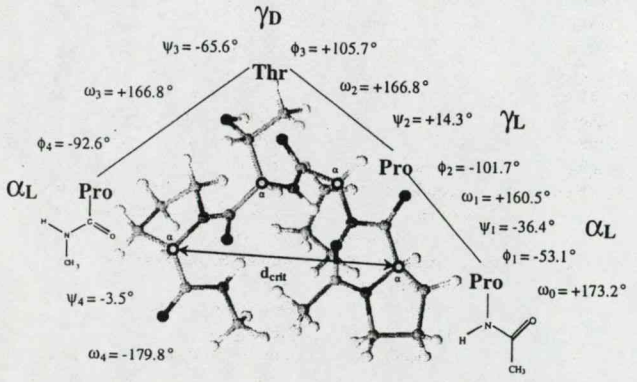
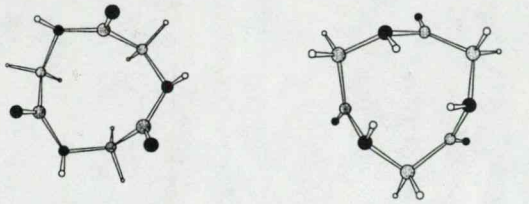
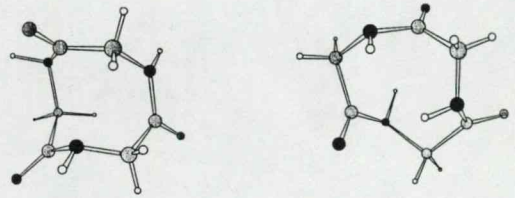


Fig. 27. An optimized structure of the tetrapeptide Pro-Pro-Thr-Pro (PPTP).

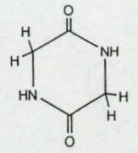


Conformer no. 1



Conformer no. 2

Fig. 28. Conformers of all-*cis*-cyclotriglycine and all-*trans*-cyclotriglycine.



XXIV

For small rings, such as diketropyperazine $(C_2H_3NO)_2$, **XXIV**, only *cis*-peptide bonds are allowed. As the ring size increases, the ring will tolerate *trans*-peptide bonds although for cyclic triglycine $(C_2H_3NO)_3$, the ring is very strained. The situation is somewhat better of cyclotetraglycine $(C_2H_3NO)_4$.

7. Conformational analysis of simple peptidomimetics

7.1. Dehydroalanine (Dha or ΔAla)

The α,β-dehydroamino acid residues occur widely in various classes of natural and synthetic drugs, although they are most commonly found in classes of antibiotics such as "lantibiotics" (lanthionine containing gene-encoded peptides with antibiotic properties such as nisin [47,48], subtilin [47,49] and ancovenin [47]) and the thiopeptide antibiotics [50,51] berninamycin A, B, C and D. They are also increasingly used in the synthesis of enkephalin analogues having different potencies and selectivities from the parent compound [52–54].

The simplest α,β-dehydroamino acid is ΔAla Fig. 30, having no substituents on its beta carbon. Because it is the only unsubstituted α,β-dehydroamino acid, its experimentally observed conformations differ significantly from those of the other dehydro residues [55,56]. In this sense, ΔAla is reminiscent of glycine, which also has very different properties from the rest of the naturally occurring amino acids, due to its lack of a side chain [16,17].

Through ab initio calculations, it has been found that *trans*-For-ΔAla-NH_2 can take on seven (β_{DL}, δ_D, δ_L, ϵ_D, ϵ_L, γ_D and γ_L) of the nine theoretical conformations of a

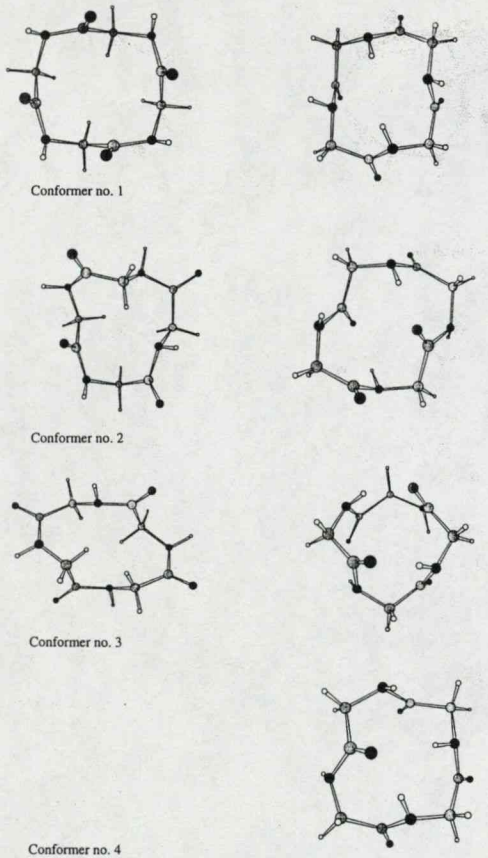


Fig. 29. Conformers of all-*cis*-cyclotetraglycine and all-*trans*-cyclotetraglycine.

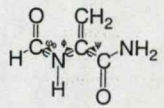


Fig. 30. Torsional angles ω_0, ϕ and ψ of N- and C-protected 2,3-didehydroalanine. ω_0 is approximately 180° for *trans*-2,3-didehydroalanine.

saturated amino acid residue Fig. 31, whereas, even N and C protected glycine, can actually only take on five.

Furthermore, it has also been found, by means of relative and stabilization energies, that although ΔAla is more rigid than glycine Fig. 32, it has a much greater stabilizing effect on the peptide chain as shown in Fig. 33. In fact, ΔAla has a greater stabilizing

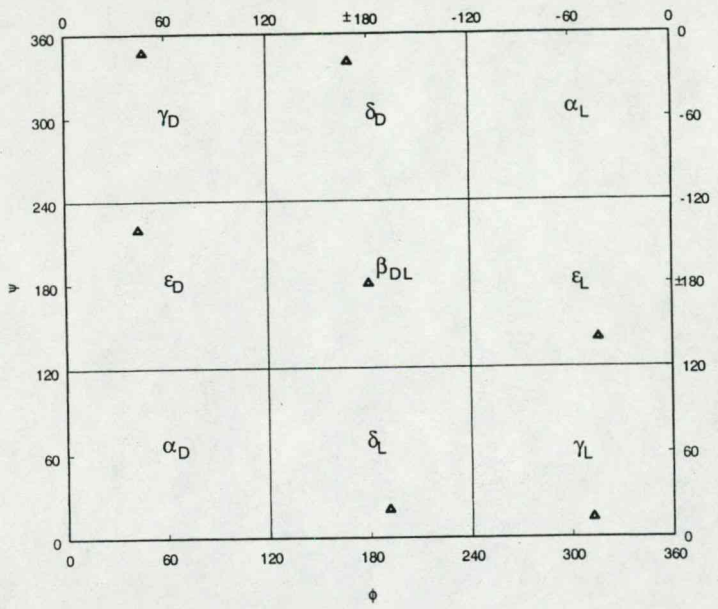


Fig. 31. The seven stable conformations of N- and C-protected *trans*-2,3-didehydroalanine as computed at B3LYP/6-31G*.

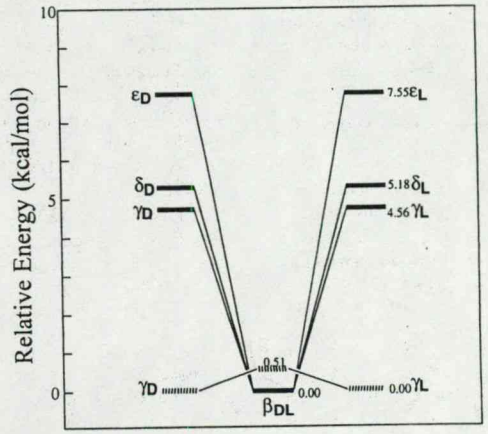


Fig. 32. Relative energies of the various conformations of N- and C-protected *trans*-2,3-didehydroalanine (solid line) and *trans*-glycine (dashed line) computed at B3LYP/6-31G* level of theory.

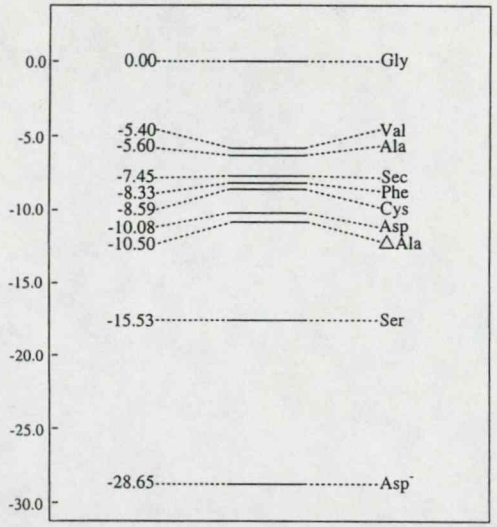


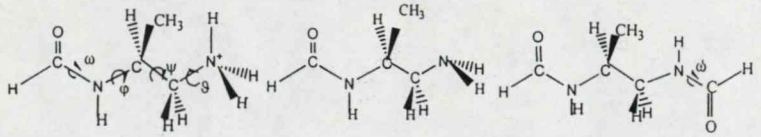
Fig. 33. Relative stabilization energies of various side chains for *trans*-For-CHR-NH_2, calculated according to the equation given at the top of Fig. 17. The most stabilizing conformer of *trans*-For-ΔAla-NH_2 is the β_{DL} one and hence its stabilization energy, instead of that of the γ_L conformer, is given. This value was calculated according to the following isodesmic equation: HCONH–CH_2–$CONH_2$ + $H_2C{=}CH_2$ → HCONH–C(=CH_2)–$CONH_2$ + CH_4.

effect than most of the saturated residues examined so far, except for serine and deprotonated aspartic acid [57]. These results can have great implications for the future of peptide based drug design.

7.2. Chiral pseudo-peptide and selected derivatives

Due to their newly discovered pharmacological properties, pseudo-peptides are able not only to substitute peptides, but also essentially improve their pharmacological action [58–63]. In order to assist the search for new compounds with desired pharmacological properties, conformational properties of three model pseudo-peptides were investigated by semi-empirical and ab initio MO methods. The three models include the protonated form of *N*-formyl pseudo-alaninamide HCONH–CHMe–$CH_2NH_3^+$, *N*-formyl pseudo-alanine itself HCONH–CHMe–CH_2NH_2 as well as the formylated *N*-formyl pseudo-alanine (i.e. propylene-di(*N*-formylamine)) HCONH–CHMe–CH_2NHCOH [64]: **XXV**, **XXVI**, **XXVII**.

For compound **XXV**, which is an *N*-protonated pseudo-peptide derived from *N*-formylalaninamide, both *trans*- and *cis*-configurations were considered (i.e. $\omega = 180°$ and 0°).



XXV **XXVI** **XXVII**

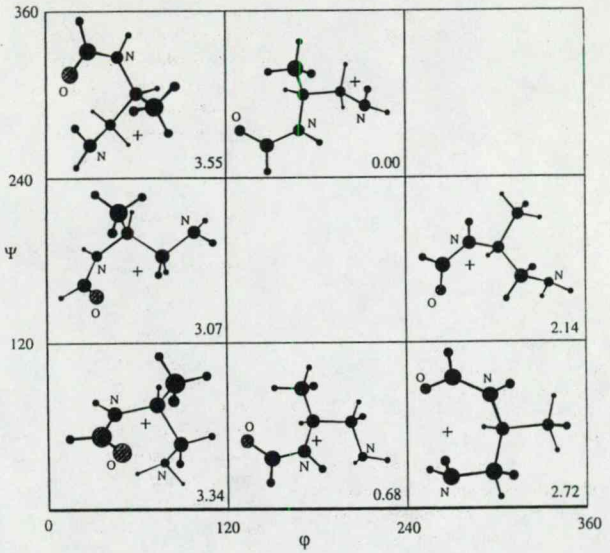


Fig. 34. Topology of a Ramachandran-type PES of the simplest chiral pseudo-peptide derived from *N*-formylalaninamide.

Taking into the account the dominant role of the peptide backbone, it was important to assess how the backbone changes under the transition from peptides to pseudo-peptides. To characterize the backbones for compounds **XXV** and **XXVI**, four dihedral angles were chosen, namely ω, φ, ψ, θ. In addition, compound **XVII** was characterized by a fifth dihedral angle ω. The conformations of the simplest pseudo-peptides were conveniently described by a 2D-Ramachandran map, i.e. a 2D-conformational potential energy surface (PES). This surface was defined in terms of two dihedral angles, φ and ψ, varying between 0° and 360° [16,17,57,64–66]. To describe pseudo-peptide conformations, the 9 expected legitimate minima for model compounds **XXV**, **XXVI** and **XXVII** were carefully optimized.

Both semi-empirical [67] and ab initio [14] RHF SCF calculations were performed on typical backbone conformations. For ab initio calculations, the RHF/3-21G level of theory was used for all compounds under consideration. To check the validity of the results in some cases, calculations were performed on RHF/6-31G(d) level of theory. It was found that the geometries of pseudo-peptides in their global minima are determined by H-bonds.

According to the computational results, pseudo-peptides, in general, can mimic the backbone of the parent peptides, as shown by the experimental data [62,63]. Pseudo-alanine was shown to be more flexible, than the parent alanine. Fig. 34 shows the topology of the E (ϕ, ψ) PES which is analogous to the Ramachandran map.

8. Non-peptidic bioactive molecules:

8.1. Antioxidants

During the metabolism of foodstuffs, carbon is oxidized to CO_2 with the help of the inhaled oxygen. Parallel to that, O_2 is reduced through a multistep process to H_2O.

$$O_2 \rightarrow O_2^- \longrightarrow\longrightarrow H_2O_2 \longrightarrow 2H_2O \qquad (15)$$

Somewhere along this line, hydroxy radical ($HO^{\cdot}$) is released in small amounts. Due to the non-discriminative reactivity of ($HO^{\cdot}$), it damages a number of key organic compounds in the human body. Among these are DNA, the damage of which leads to premature aging as well as numerous including cancer. Collectively, these processes are labeled as "oxidative stress" and can be relieved by free radical scavengers, called antioxidants, such as flavones, selenocysteine and lycopene. It has also been recognized recently that certain drugs can act as antioxidants.

8.1.1. Flavone conformations

Flavone is the parent molecule of a broad range of compounds belonging to the family of flavonoids. Flavonoids are polyphenolic compounds, which occur as yellow pigments in plants. They are believed to be effective against HIV, cancer, coronary artery disease and aging [68].

Conformational analyses were performed on flavone and three other related compounds, namely 2-phenyl pyranone, β-phenyl naphthalene and biphenyl. In particular, the phenyl rotation with respect to the rest of the molecule was studied using Hartree–Fock (HF) calculations performed on GAUSSIAN 94. The basis sets employed were STO-3G and 3-21G. It was concluded that the hydrocarbons (biphenyl and β-phenyl naphthalene) are noticeably different from oxygen containing heterocyclic analogues such as 2-phenyl pyranone and flavone, as far as phenyl rotation is concerned. One would also expect a difference in their reactivity towards free radicals. It seems that only the oxygen heterocyclic analogues can be reactive in nucleophilic and radical attack because they contain positively charged atoms, as shown by computed Mulliken charges. Nevertheless, the topology at the potential energy curves (number of minima and transitions structures) is the same for all four compounds.

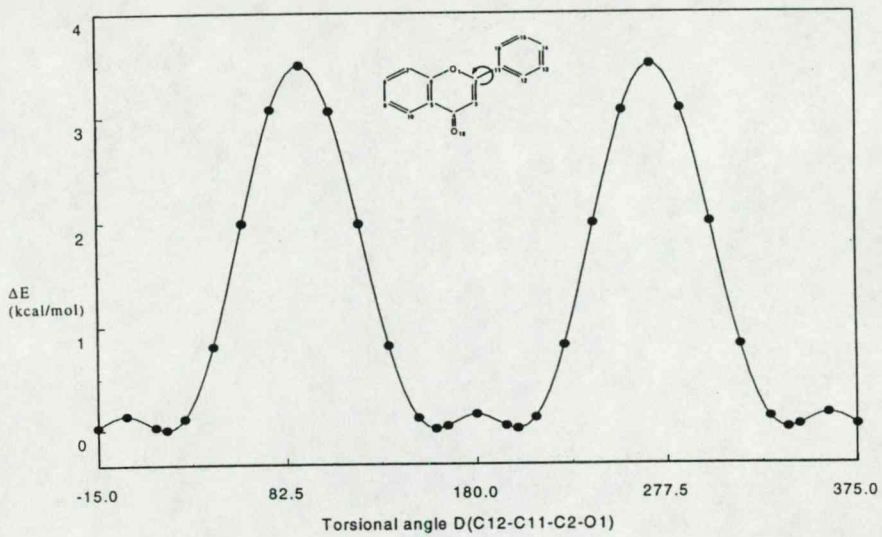
The potential energy curve of flavone is shown in Fig. 35. The antioxidant mechanism of flavone skeleton is now under investigation.

8.1.2. Selenocysteine as antioxidant

Selenocysteine is involved in a variety of biochemical interactions and has recently been considered as the 21st amino acid used by nature for RNA directed protein synthesis [69,70]. Several selenoproteins containing selenocysteine at their active sites have been identified using neutron activation analysis. At the end of the 20th century, possibly 28 of these proteins are known to exist. At present, however, we know very little about the structure and function of most of these proteins. Those that have been characterized and for which functions are known include the following:

(a) Four species of glutathione peroxidase [71–73] (cellular or classic, extracellular or plasma, gastrointestinal and phospholipid hydroperoxide glutathione peroxidase) have been identified. Glutathione (GSH) peroxidase catalyses (Fig. 36) the reduction of a variety of hydroperoxides including lipid hydroperoxides (which, in the presence of trace metals, can form toxic lipid radicals). This mechanism is believed to protect biomembranes and other essential cellular components against oxidative challenge [74–76], cancer [77] and gastric ulceration [78].
(b) Selenoprotein P is an extra-cellular protein presumably containing 10 Selenocysteine residues that are encoded by the UGA stop codon in the open reading frame of the mRNA. Some indirect evidence suggests that selenoprotein P acts as a free radical scavenger [79].
(c) Selenoprotein W is found in muscles [80].
(d) Iodothyronine deiodinase catalyzes the deiodination of thyroxine to the biologically more active thyroid hormone triidothyronine [81].
(e) Thioredoxin reductase catalyzes the NADPH-dependent reduction of the redox protein thioredoxin. Thioredoxin stimulates cell growth and is over-expressed in a number of human cancerous cells [82,83].
(f) Mitochondrial selenoprotein [84] was localized in the mitochondrial membranes of endocrine organs.
(g) Prostatic selenoprotein and testicular selenoprotein were found in spermatid nuclei. Recent findings suggest that it takes part in the process of replacing and condensation of the DNA, which occur in the spermatid nuclei, and thus may have an important function in sperm development [85,86].

In addition to the above, new evidence, based on sequence analysis, suggests that HIV-1 encodes a



	Regular Optimization	
	RHF	
	STO-3G	3-21G
Energy (hartree)	-714.5491401	-719.5533998
Deviation from coplanarity (degrees)	20.82	8.84

Fig. 35. Phenyl torsional potential of the flavone skeleton.

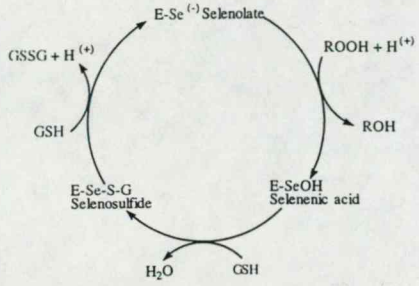


Fig. 36. Putative catalytic mechanism of glutathione peroxidase.

protein containing selenocysteine (Sec) [87–89]. It has been observed previously that plasma selenium and glutathione levels are subnormal in HIV-infected individuals [90] and that specifically, four ^{75}Se-containing proteins are lower in HIV-infected cell populations than in uninfected cell populations.

8.1.3. Lycopene and free radicals

Lycopene is a very important free radical scavenger [91,92]. It is a polyene with 15 conjugated double bonds. There are numerous *cis*- and *trans*-isomers of lycopene (c.f. Fig. 37). The all *trans* isomer is the predominant species in tomatoes and tomato products (≈95%). The 5 *cis*-, 13 *cis*- and 9 *cis*- lycopene are

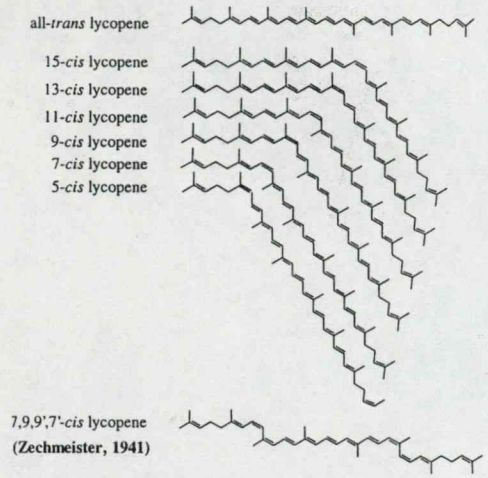


Fig. 37. Geometrical isomers of lycopene.

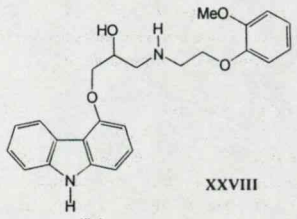
found in human serum, making up approximately half the total lycopene content. Isomers with multiple *cis*-double bonds are also known. For example 7,9,9′,7′ *cis*-lycopene is the naturally occurring form of lycopene in fresh Tangerine-type tomatoes is also shown in Fig. 37.

Two isomers of 1,3,5 hexatriene were used to model *cis*- and *trans*-lycopene. The central double bond (Δ [4–6]) of hexatriene was used in the *cis*- and *trans*-forms. For the computational convenience, free radical fluorine atom (˙F) was used instead of the biologically more important hydroxy radical (˙OH).

Through ab initio calculations, using ROHF/6-31++G(d,p) and UHF/6-31G(d,p) methods, it has been found that if C2 or C3 of the 1,3,5 hexatriene is fluorinated, both the *trans* and *cis* isomers have the same relative energies. On the other hand, fluorinating C1 makes a difference of 3.7 kcal/mol between the *trans* and the *cis* isomer, the *trans* isomer being more stable Fig. 38.

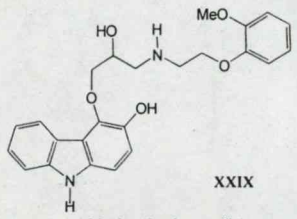
8.1.4. Carvedilol as a new antioxidant

The Merck Index recognizes carvedilol (**XXVIII**) as a nonselective β-blocker with vasodilating activity,



XXVIII

carvedilol



XXIX

3-hydroxylated carvedilol
4-[2-hydroxyl-3-[[2-(2-methoxy-phenoxy)ethyl]amino]propoxyl]
-9H-carbazol-3-ol

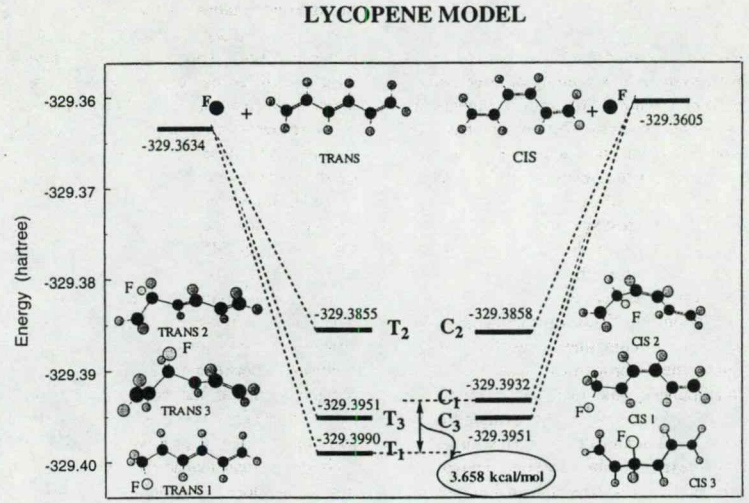


Fig. 38. Relative energies of radical addition products to lycopene models computed at RHF/3-21G level of theory.

which is mediated by its antagonizing effect on α-adrenoceptors. It was firstly patented in Germany in 1979 and subsequently in the US in 1985.

Carvedilol has been used in the treatment of hypertension [93], angina pectoris and congestive heart failure. In these fields, carvedilol is at least as effective as reference drugs (metoprolol [94], enalapril, verapamil [95]) but its side effect profile is better, and it has more beneficial effects.

According to clinical trials, carvedilol effectively reduced the morbidity and mortality rates in-patients suffering from chronic heart failure to such an extent that the US Data and Safety Monitoring Board stopped, for ethical reasons, the investigation before its completion [96]. This was the first time in the history of drug investigations that a clinical trial was stopped because the drug was too effective. It seemed unethical to deprive the medication from placebo group. Carvedilol also reduces the frequency of reversible myocardial ischaemic events after thrombolysis [97] and prevents cardiac remodeling in-patients suffering from left ventricular dysfunction after acute myocardial infraction [98]. This latter effect has been attributed to its scavenging action of oxygen free radicals and preventing the lipid peroxidation [97]. These reactive radicals are implicated in the process of programmed cardiac cells death (apoptosis) [99] and the concomitant loss of myocardial cells leading to progressive decrease of left ventricular mass and function [100].

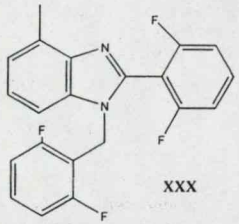
Originated from this antioxidant property, carvedilol may also prevent the development of nitrate tolerance in, patients receiving continuous nitrate therapy [101,102].

Apart from carvedilol, at least one of its 3-hydroxylated metabolites (**XXIX**) has marked antioxidant properties. The metabolite is approximately a thousandfold more effective than vitamin E in certain experimental setups and more effective than carvedilol itself. Although the exact antioxidant mechanism of these compounds is not known, recent structure activity relationships indicate that its activity resides in the carbazol ring. A computational mechanistic study on this exceptional antioxidant capability is now in progress.

8.2. Medicinal chemistry and drugs

Most of today's marketed medications were found by chance or by systematic screening of large collections of compounds (called "libraries") either from natural or man-made sources. Current emphasis on high-throughput screening and combinatorial chemistry suggests that an educated guess is pivotal for the discovery of tomorrow's drugs. Computer assisted molecular modeling (CAMM) is a relatively new and rapidly developing tool in drug design. Computer graphics techniques allow the transformation of complex data sets, obtained e.g. from theoretical chemical calculations or X-ray diffraction patterns, into an image on a computer screen. Chemical structures and their properties may thus be visualized, manipulated, and matched or combined with other relevant molecules. The availability of the three-dimensional architecture of HIV protease, thymidylate synthase, thrombin, carbonic anhydrase, and many more has led to at least an equal number of new chemical entities being evaluated and used now in clinics. In such cases, CAMM has certainly led to a more rational approach toward drug discovery, in particular when the binding domains for inhibitors and substrates are known from cocrystallization experiments. Knowing the "lock" definitely helps in designing the "key", although the interaction between newly synthesized ligands and the target macromolecule may appear sometimes unexpectedly.

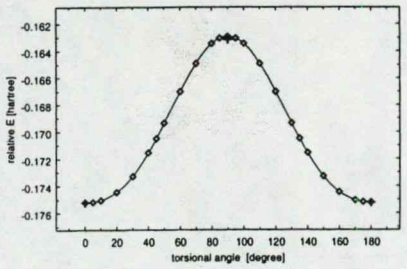
Locking into certain stereochemically important conformations is sometimes achieved by heavy substitution, such as o,o' difluorophenyl groups. This principle has been applied in one of the anti-HIV drugs where o,o' difluoride substitutions are used [103].

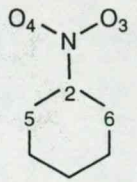


XXX

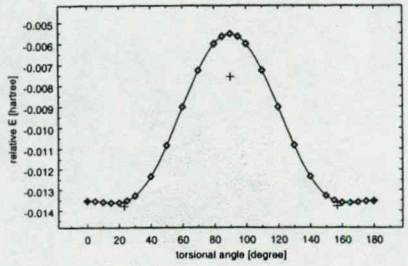
Calculations of the barriers and associated one-dimensional torsional potential are performed for the internal rotation of the nitro-group in nitrobenzene as well as for *o*-fluoro- and o,o'-difluoro-nitrobenzenes [104]. This structure is related to certain drugs, such as the one shown in **XXX**, that are under investigation. It is shown that the presence of substituents in *ortho*-positions forces the benzene ring to rotate about the C–N bond, out of the plane of the nitro group. In these conjugated molecules, the inclusion of electron correlation is shown to be necessary for reliable barrier prediction. The potential curves at HF level are shown on Fig. 39. Corresponding MP2 values are given by "+" symbol. It is found that the π-stabilization (resonance) is of relatively small contribution to the structure of the nitroaromatics, whereas the steric repulsion between the *ortho*-substituents and oxygen atoms of the NO_2 group is rather significant. The interplay between them accounts for the fact that for o,o'-$C_6H_3F_2NO_2$, the barrier height at 0° is larger than that at 90°, when electron correlation is included. The values for the barrier at 0° were 2.67 and 3.81 kcal/mol computed without and with electron correlation and the corresponding values at 90° were 1.42 and 0.91 kcal/mol, respectively.

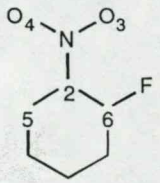
A fairly large area of drug design is related to G-protein coupled receptors. A preliminary structural analysis of the β_2 adrenergic G-protein coupled receptor was considered. This receptor is a transmembrane protein, and it usually consists of seven bundles of α-helices. Each of these seven helices consists of approximately 21–24 amino acids, as required to span the cell membrane. In order to determine the type of interaction between the helices, an α-helix made up of seven alanine residues was built using HYPERCHEM. Two other helices were built in the same manner, except one of the amino acids was changed to serine and tryptophan respectively. It was found that amino acids arrange so that many of the polar amino acids can face the cavity and non-polar amino acids can face away from the cavity. This is expected because a hydrophobic lipid membrane surrounds the outer circumference of the bundle of seven helices. Interactions between residues of a helix such as that between tryptophan and serine are practically thermo-neutral. Interaction between the Ala-Ala helices has a stabilization energy of −3.0 kcal/mol, while that between the



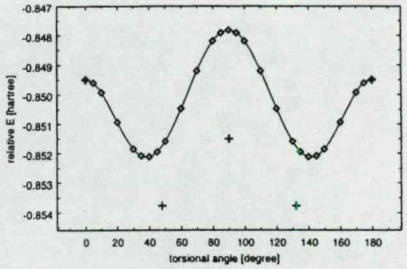


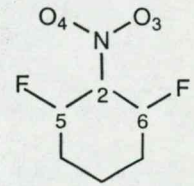
$C_6H_5NO_2$





$C_6H_4FNO_2$





$C_6H_3F_2NO_2$

Fig. 39. Phenyl rotation in nitrobenzene, o-fluronitrobenzene and o,o′-difluronitrobenzene.

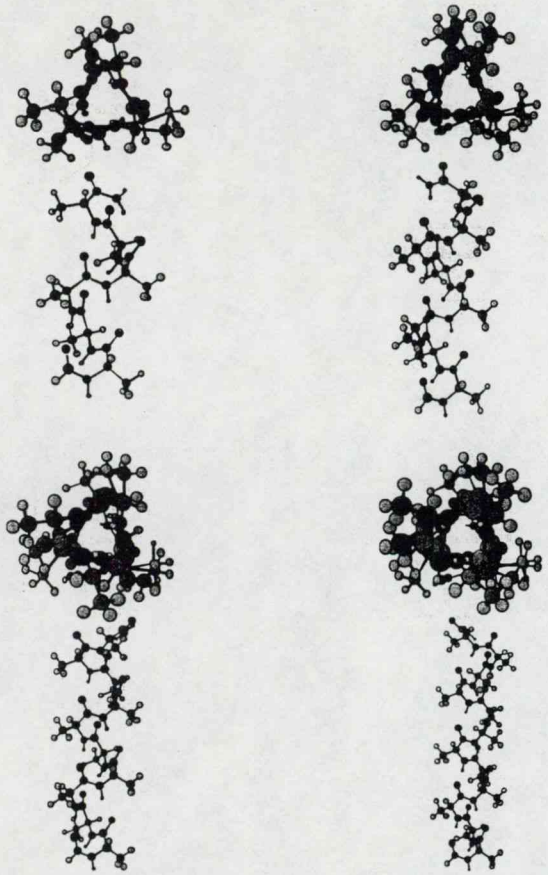


Fig. 40. Helical structure of N- and C-protected polyalanines HCO-$(Ala)_n$-NH_2 ($n = 6, 8, 10, 12$).

Ser-Ser helices has a stabilization energy of −8.2 kcal/mol [105]. The interaction between two serines result in a greater stabilization energy than that of the two alanine residues interacting. Thus, replacing the alanine with serine resulted in larger stabilization energy but replacing serine with tryptophan did not significantly change the stabilization energy. An ab initio study on the helical structure of For-$(Ala)_n$–NH_2 has been carried out for $n =$ 6, 8, 10 and 12 (c.f. Fig. 40).

Cell cycle inhibitors or modulators that halt uncontrollable tumor growth are regarded as highly promising new therapeutic agents against human cancer. Certain sulfonamides [106] inhibit microtubule assembly owing to its reversible binding to the colchicine-binding site on tubulin. They also

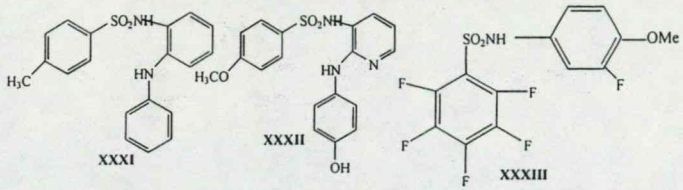
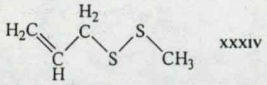


exhibit good in vivo antitumor activity against various rodent tumors and human tumor xenografts. The focus on making a sulfonamide compound library is based upon template **XXXI** ever since compound **XXXII** was found to inhibit cellular growth and mitosis in vitro, but not quite potent in vivo. In the design of this template, the sulfonamide moiety located between two aromatic rings was fixed as a basic motif, and the NH group at the *ortho* position of the sulfonamide was considered a key functionality for substantial antiproliferative activity in cell-based assays [106].

A substantially different aromatic sulfonamide (**XXXIII**) has also been reported [107] to interact with microtubules during cell division. Concerning the recently reported polyfluorodiarylsulfonamide (**XXXIII**), we are now involved in assessing its mechanism of action.

9. Multidimensional conformational analysis of allyl methyl disulfide: a key component of garlic

Garlic has been one of the most popular medically researched plants, with over 1300 research articles in only the last 100 years [108]. Organosulfur compounds in garlic, like allyl methyl disulfide, have been found to be involved in antimutagenic, anticarcinogenic, antithrombotic, and lipid-lowering activities, and it has also been found to act as an antioxidant. Ab initio molecular computations were performed on allyl methyl disulfide (**XXXIV**)

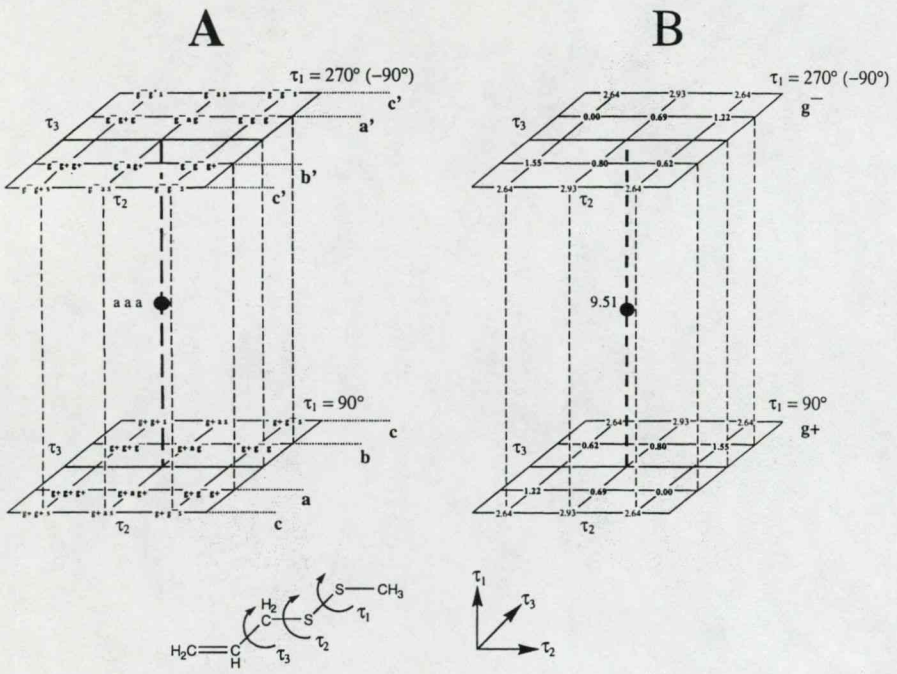


XXXIV

with respect to torsional angles $\tau_1 = \tau(H_3C_2{\cdot}CH_2{\cdot}S{-}S{\cdot}CH_3)$, $\tau_2 = \tau(H_3C_2{\cdot}CH_2{-}S{-}S{\cdot}CH_3)$, and $\tau_3 = \tau(H_3C_2{-}CH_2{\cdot}S{\cdot}S{\cdot}CH_3)$. Potential energy scans, resulting in potential energy curves (PEC), were performed along **a**, **b** and **c** as shown in Fig. 40, at the HF/3-21G level of theory. All conformations were optimized at the HF/6-31G* level of theory and their energy values are provided in Fig. 41 as well. The potential energy hypersurface surface (PEHS) of **XXXIV**, i.e. $E = E(\tau_1, \tau_2, \tau_3)$, revealed six lower energy pairs of enantiomeric minima (i.e., $[g^+g^+g^+|g^-g^-g^-]$; $[g^+ag^-|g^-ag^+]$; $[g^+g^-g^+|g^-g^+g^-]$; $[g^+g^+g^-|g^-g^-g^+]$; $[g^+ag^-|g^-ag^+]$; and $[g^+g^-g^-|g^-g^+g^+]$) as well as three higher energy minima (i.e. $g^+g^+s|g^-g^-s]$; $g^+as|g^-as]$; and $[g^+g^-s|g^-g^+s]$) were optimized at $\tau_1 = \pm 90°$using the two HF/6-31G* and B3LYP/6-31G* methods at this level of theory (Fig. 40). In our previous study on the multidimensional conformational analysis of ethyl benzene [109], we found that very often two minima result when a planar moiety is rotated against a tetrahedral moiety, one of them being higher (0° rotation) and the other being lower (90° rotation) in energy.

As illustrated in Fig. 41, it is apparently clear that the center of symmetry of the PEHS of **XXXIV** is at $\tau_1 = \tau_2 = \tau_3 = 180°$, i.e. at the fully *anti–anti–anti* orientation. Therefore, if an imaginary line can be drawn through this fully *anti* center from one energy platform to the other on the PEHS of **XXXIV**, then the structures along such a line are enantiomers. Conversely, if an imaginary line joining two conformers cannot be drawn through fully *anti* center, then the structures are diastereomers. Therefore, although there are no stereocentres in **XXXIII**, there is chirality in the conformational twist with respect to the fully symmetrical *anti-anti-anti* structure [*aaa*] through $\tau_1 = \tau_2 = \tau_3 =$ 180°.

In order to denote the enantiomeric relationship of each pair of conformers, a convenient notation has been adopted, with the symbol "|" representing a



Allyl Methyl disulfide (**XXXIV**)

Fig. 41. Topology of the conformational the PEHS of **XXXIV** (allyl methyl disulfide), where $E = E(\tau_1, \tau_2, \tau_3)$. (A) The notation of the 6 pairs of lower energy conformers (along a, a′, b and b′), the 3 pairs of higher energy conformers (along c and c′) and the fully symmetrical *anti* structure. (B) The relative energies in kcal/mol of the 6 pairs of lower energy conformers, the 3 pairs of higher energy conformers, and the fully symmetrical *anti* structure. (All energy values obtained at the HF/6-31G* level of theory). Note that the centre of symmetry is denoted by a solid dot and corresponds to the $\tau_1 = \tau_2 = \tau_3 = 180°$, fully *anti* conformation (*a a a*) as shown in Fig. 42. [Cross-sections were studied at a,a′,b,b′,c and c′]. All conformations were optimized at the HF/6-31G* level of theory.

mirror between two enantiomers. The global minimum was determined to be the $[g^+g^-g^+|g^-g^+g^-]$ conformer, and the fully symmetrical *anti* [*aaa*] conformer was determined to be a second order saddle point (Fig. 42) [126].

Based on the energies and MO diagrams of **XXXIV**, the HOMO and LUMO + 1 orbitals were determined to be involved in electron donating and accepting activity of **XXXIV**. The putative anti-carcinogenic and cholesterol lowering mechanism of activity of **XXXIV** is presented in Fig. 43 [126].

10. Macromolecular interactions

10.1. Protein–protein interactions

Molecular recognition forms the basis of protein–protein interactions. Proteins that contain covalently bound oligosaccharides, called glycoproteins, are often tailored to be specific recognition sites. They form a diverse group that includes enzymes, hormones structural proteins, and transport proteins.

The presence of one or more oligosaccharide chains

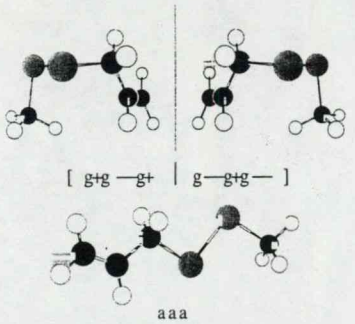


Fig. 42. Structures of the degenerate $g^+g^-g^+$ and $g^-g^+g^-$ conformers ($\lambda = 0$), and the fully symmetric *a a a* conformer ($\lambda = 2$). Note that the global minima conformers, $g^+g^-g^+$ and $g^-g^+g^-$, are enantiomeric and that this relationship is denoted with the symbol "|", indicating a mirror plane.

on a protein can alter its physical properties, such as its size, solubility and stability. These changes can affect biological activities such as immunogenicity. A number of mammalian hormones are dimers of glycoproteins whose oligosaccharide chains assist in the assembly of the dimer. As a result, they confer resistance to proteolysis. Another role of glycoproteins is the recognition of cell by another cell in processes such as cell migration or oocyte fertilization. These events are dependent partly on the surface of one cell binding to the sugar portions of specific glycoproteins on another cell's surface. Studies of a wide range of lectins have shown that oligosaccharide chains of proteins are always located on the outer surface of a cell membrane, and not to the cystolic surface.

10.2. Protein–lipid interactions

The backbone conformation of a polypeptide chain may follow some well-defined pattern, which could be drastically modified by various side chains. These will lead to side-chain side-chain or side-chain backbone interactions in a polypeptide chain, which very often involve hydrogen bonding. Side chains of arginine and lysine could play important roles on many occasions.

The overwhelming presence of arginine and lysine (approximately 30 out of the 170 residues) in myelin basic protein (MBP) is crucial in the protein's role in stabilizing myelin. The positively charged side chains can interact with the hydrophilic head groups of the lipid bilayer. Consequently, MBP may be involved in bridging distant points on surfaces of myelin. MBP exhibits charge microheterogeneity as a result of various processes such as post-translational deamination, phosphorylation and the deamination of arginine to citrulline. Such changes lead to a decrease in the stability of the lipid-MBP complex [109]. Eighteen out of the nineteen arginine residues are citrullinated (MBP Citl8) in the fulminant form of multiple sclerosis (MS), known as Maburg's Disease [110]. The interaction between MBP and the lipid bilayer could be studied experimentally using either NMR or atomic force microscopy (AFM), or a combination of both. It may be expected that the spectroscopic pattern associated with the complex will be different from the sum of the isolated components. The altered observation of the complex could be regarded as diagnostic of the interactions between the two components.

The interaction of the two components could also be modeled at the molecular level. The simplest interaction that one can imagine between the guanidinium ion functionality of the arginine side chain and the carboxylate ion functionality is shown in **XXXV**. The actual interaction is, of course, far more complicated because phospholipids (see structures **XXXVI–XXIX**, where X = head group) have a number of electron acceptor sites. In particular, binding could occur between the two NH moieties of the guanidinium ion and one of the following three pairs of proton acceptor, as illustrated in **XXXV**. Binding occurs in the fashion shown in **XXXV**; the pairs of arrows in **XXXVII–XXXIX** indicate the locations of the binding sites.

10.3. Protein–nucleic acids interactions and protein biosynthesis

Protein–nucleic acid interaction is of special importance. It occurs even at the birth of a protein molecule at its biosynthesis. Translation of the genetic message into proteins implies the precise correspondence between the 64 base triplets and the 20 canonical amino acids. In this process, tRNA plays a central

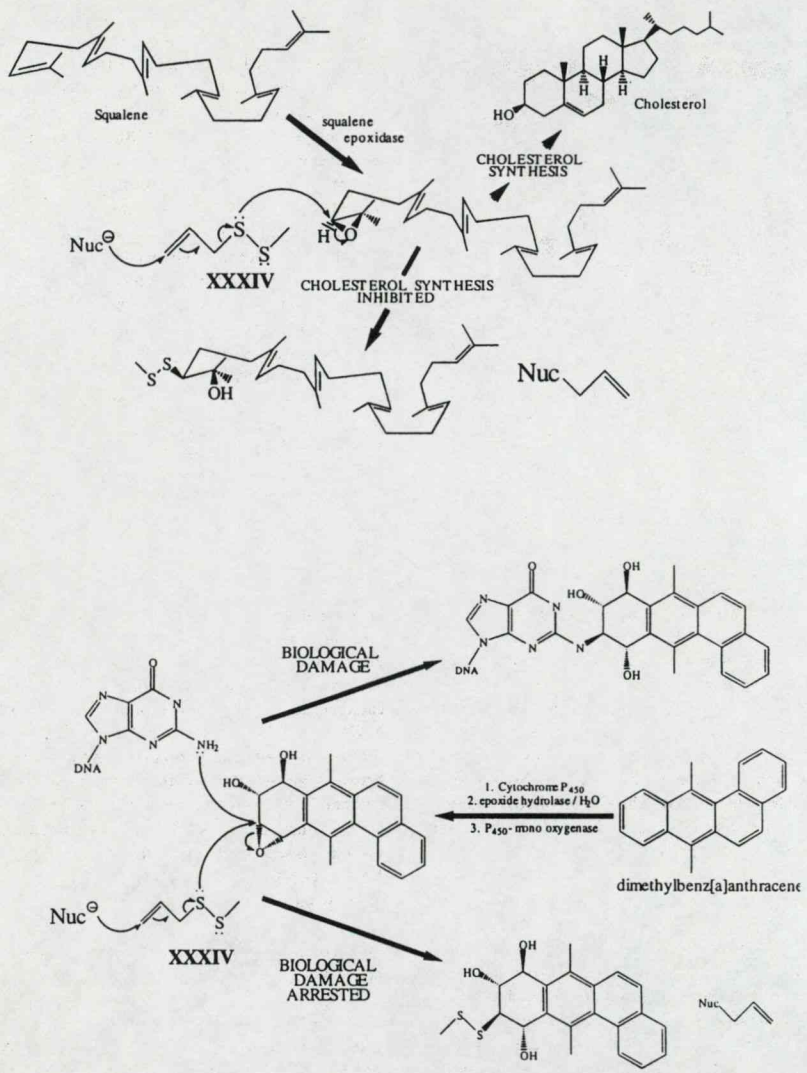
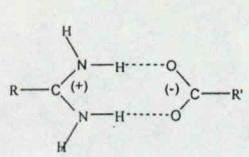
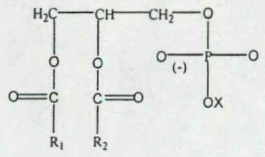


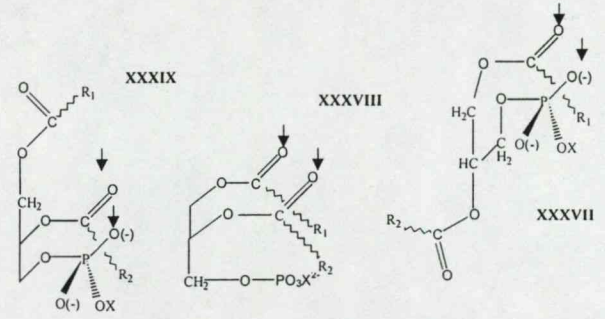
Fig. 43. Putative mechanism of the (A) anticholesterol and the (B) anticarcinogenic activity of **XXXIV** (allyl methyl disulfide).



XXXV



XXXVI



XXXIX XXXVIII XXXVII

role by providing the nascent polypeptide with the amino acids to which they are esterified in response to codons on RNA. In this step, the amino acylated tRNAs decipher the base triplet (codon) using Watson–Crick base pairings between mRNA and a complimentary region on the tRNA. The anticodon is independent of the nature of the amino acid esterified to the tRNA. Therefore, the accuracy of the tRNA aminoacylation reaction, ensured by the aminoacylation tRNA synthetases aaRS (a given aminoacyl-tRNA synthetase is abbreviated by its cognate amino acid three-letter code followed by RS; e.g., GlnRS stands for glutaminyl-tRNA synthetase), is of primary importance in all living cells, since it will govern to a large extent the fidelity of the translation process [111].

Aminoacylation of tRNA proceeds through a two step mechanism. The first step involves the formation of a stable aaRS-amino acyladenylate complex resulting from the specific binding and reaction with ATP:Mg^{2+} of the amino acid; pyrophosphate is subsequently released. In the following step the 3′ terminal adenosine of enzyme-bound tRNA reacts with the aminoacyladenylate, leading to both the esterification of the tRNA and the release of AMP. Notably, the binding of tRNA occurs in the presence of a polyamine or a divalent cation, usually magnesium [112].

$$\text{aaRS} + \text{ATP} : \text{Mg}^{2+} + \text{aa} \rightarrow \text{aaRS} : \text{aa} \sim \text{AMP} + \text{PPi}; \text{Mg} \quad (16)$$

$$\text{aaRS} : \text{aa} \sim \text{AMP} + \text{tRNA} \rightarrow \text{aaRS} + \text{aa} - \text{tRNA} + \text{AMP} \quad (17)$$

where aa stands for amino acid; aa ~ AMP for aminoacyladenylate.

Because the two steps involve the specific recognition of both the amino acid and the tRNA they impose specificity throughout the overall reaction. As a general rule, there is one synthetase for each of the 20 canonical amino acids in a given organism or organelle. Two main functions are carried out by an aaRS: the activation of the amino acid and the recognition of the tRNA molecule. In the activation of the amino acid, the carboxylate of the amino acid becomes linked to the AMP moiety of an ATP molecule to form the aminoacyladenylate, which remains firmly bound to the synthetase [113].

The mixed anhydride bond between the carboxylate and the phosphoryl group of AMP is highly susceptible to the attack by a hydroxyl group, thereby facilitating further transfer of the amino acid moiety onto the tRNA. Consequently, the reaction of the activated amino acid with tRNA can be viewed as the result of a productive positioning of the 3′-terminal adenosine of the nucleic acid within the active center of the synthetase. This allows the attachment of the amino acyl-AMP by either the 2′- or 3′ hydroxyl group of the terminal ribose. In this context, the accuracy of the activation reaction is of importance in the further specificity of the tRNA amino acetylation reaction. In all cases the rapid formation of a ternary complex of the enzyme with both ATP:Mg^{2+} and amino acid was evident. The amino acid acyladenylate is produced through a rate-limiting isomerization of this ternary complex PP:-Mg, one of the products, is then released.

The other function of the aaRS is the recognition of the tRNA molecule. Many recent studies have pointed out that the occurrence of discrete nucleotidic recognition elements on a tRNA molecule dictates the specificity of aminoacylation by the corresponding aaRS. Briefly, three major recognition sites have been identified:

(i) the acceptor stem, including the "discriminator base";
(ii) the three bases of the anticodon;
(iii) the D-loop.

In some systems, the three sites are important for correct aminoacylation by the corresponding amino acid, whereas in others, one site is enough. However, at least two sites appear to be required systematically for achievement of maximal aminoacylation rates. Because a limited number of recognition sites are present on the tRNA molecule, it appears that there may be a limited number of recognition sites (peptides regions) on the cognate aaRS, too. The aaRS–amino acid interaction is based on protein structure rules, but other kinds of rules govern the tRNA–aaRS interaction. An unsolved question in the molecular biology of gene expression is the determination of these rules. Discrimination between tRNAs by the aaRS enzymes is complex because all tRNAs are about 76 nucleotide residues in length and form the typical tRNA cloverleaf secondary structure [114–117].

In the process of investigating protein synthesis in general, an exploratory set of ab initio molecular orbital computations were carried out on the various conformations of ribose and deoxy-ribose model compounds to establish the nature of possible interactions between the 5′-hydroxyl and the oxygen of the ribose ring. Moreover, the investigation was extended to the hydrogen bonding between the two adjacent hydroxyls, 2′- and 3′-, to explore their effect on the stabilization of the whole system. The consistent presence of a rather unusual weak hydrogen bond between the proton of $C_{2'}$–H and the $O_{5'}$ of the hydroxyl methyl group for the g^-g^- conformation of the $-CH_2$-OH functional group has been studied using Bader's atoms in molecules (AIM) method.

Density functional calculations coupled with Bader's Atoms in Molecule (AIM) method are becoming a very important tool in examining intramolecular interactions that are responsible for the system's stabilization. As has been demonstrated [118], an unusual hydrogen bond between $C_{5'}$–$O_{5'}\cdots H_{2'}$ is present for the specific g^-g^- orientation of the $-CH_2$-OH moiety in the ribose system.

10.4. Carbohydrates and molecular recognition

Carbohydrates are sometime referred to as *information-molecules*. This characterization of carbohydrates is due to the fact that carbohydrates have a very large number of structural variants. First of all there are several hexoses and they can connect to each other to form oligosaccharides in different sequences. The chain can be branched or unbranched and the connection can be via α- or β-glycosidic linkage. Furthermore, the numerous OH groups can assume g^+, a and g^- rotamers. Such tremendous structural variation can code a great deal of information. Oligosaccharides are frequently attached to proteins via O- or N-glycosidic linkage. The former one involves either serine or threonine while the latter utilizes almost exclusively asparagine. These carbohydrate antennas are involved in molecular recognition processes. During the molecular recognition process, hydrogen bonds change from intramolecular to intermolecular type. These processes could be studied in terms of glycopeptides. However, before embarking to the field of glycopeptide research one must understood the conformational intricacies of carbohydrate conformations.

We have modeled hydrogen bonding between adjacent OH groups in ethylene glycole [119,120]. We have also investigate internal hydrogen bonding in a 6-dehydro hexose: L-fucose [121,122]. D-glucose [123,124] as well as D-mannose [124] have also been studied in some details. An explanatory study of an oligosaccharide: the Lewis X trisaccharide [125] has just been completed. We are now engaged to study glycopeptides.

11. The mystery of protein folding

Protein synthesis occurs sequentially, one amino acid at a time. Initially, it was tacitly assumed that once the synthesis of the primary sequence is completed, the nascent macromolecule is folded to its bioactive conformation. Originally, it was further assumed that the folded structure is the global minimum, and thermodynamics will drive the structure to this uniquely folded structure. Relatively soon, it became obvious that it takes an enormous amount of time "to search the entire conformational space available for a polypeptide chain" [8]. This principle became known as the Levinthal's paradox [7]. Latter, it was envisaged that the bioactive form may not be the global minimum, but in fact it could be one of the local minima. Thus, thermodynamics could not be treated as "Maxwell's demons" to guarantee such a specifically folded structure, even if the necessary time was available. At that time, chaperones were invoked as an explanation of post biosynthesis folding. This view effectively allows the interaction between the chaperone, and the nascent polypeptide while it is being synthesized. Thus, the folding could in fact occur while the synthesis is in progress.

Nowadays, it is customary to look upon folding in terms of statistical ensembles of states. However, such ensembles of states are nothing more than energetically and conformationally closely spaced minima on a rather complex multivariable potential energy hypersurface (PEHS). Professor Honig greets the "light at the end of the tunnel" with the following acclamation [8]:

> It is likely that a complete solution to all aspects of the protein folding problem will require that the information available from the analysis of large structural and sequence databases be combined with the principles and methods of physical chemistry and computational chemistry.

We concur with his view. However, we wish to add that computational chemistry should go beyond currently used empirical force fields. It must start with ab initio molecular computations, which will eventually lead to an accurate analytic function that encompasses all possible protein conformers as the roots of the gradient of such an analytic function. The search for the secret of folding patterns can really start in the possession of such roots. The mystery associated with protein folding will gradually fade in the light of new understanding.

12. Proposal for a millennial mega-project and conclusion

Once a complete set of data exists for all 21 amino acids and all of their 'found' backbone conformations, a database may be constructed. This large data set may then be used in the construction of di-, tri- and even large arrays of peptides, with the ability of accepting changes at any time. However, as the number of units in a polypeptide chain grows linearly, the number of possible combinations and conformations grows exponentially. Therefore, without question, an automated procedure is necessary for both the construction of input files as well for the extraction and tabulation of relevant output parameters.

Even with a successful and efficient digital script or program to perform these tasks, the current problem

of insufficient computational power and other hardware resources would still exist. This is the largest and omnipresent limitation in computational chemistry, amicably referred to as the 'limiting reagent' in the synthesis of ab initio solutions to chemical problems. The present solution of continually upgrading to larger and faster hardware will not bring about the elimination of this low yield.

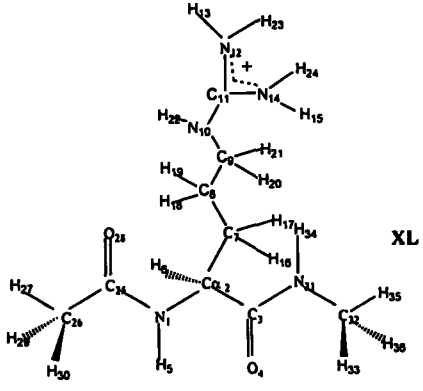
The implication of truly parallelized software and more efficient optimization algorithms is one part of the solution. The other lies in the creation of a distributed network, through which calculations may be sent out to various individuals, institutions and organizations willing to donate free CPU down-time.

As PC's have come to the forefront of the cost to performance ratio, many nodes are now available to run large arrays of computations, both quickly and efficiently. All that is required is a means to manage the incoming and outgoing data, in an organized and error-free manner. The fragmentation of jobs, for parallelization, may be accomplished by sending a copy of the electron-density and Fock matrices to each node in the array. Portions of the Hessian are then constructed and assembled into a complete set afterwards.

This may be accomplished using a standardized numbering method for the molecules to be studied. With this common input parameter definition, polypeptides may be constructed without error and optimized quickly.

A novel and sound proposal has been made to optimize all possible tri-peptide (IV) combinations and conformations. The reason for choosing tri-peptides lies in that fact that all *N*-acetyl and *N*-methyl interactions and resultant conformations possible with a central peptide would be evaluated. Hence, once completed, two tri-peptide blocks could then be bound together and optimized more quickly than a single calculation of a hexapeptide. This is due to the fact that the internal structure of each tri-peptide unit would not change dramatically, upon biding to a second one. These 'blocks' could subsequently be used to give insight into the most stable geometry of much larger polypeptide chains, both quickly and efficiently.

For the automation of this input generation and subsequent data extraction from the log files, the numbering of the atomic sequence in the peptides must be standardized. In this way, scripts and programs may be used to reorganize immediate patterns necessary for one's chemical analysis. As an example, we use Arginine **XL** for this:



XL

As all amino acids contain the same backbone sequence, the numbering is started from the nitrogen attached to the alpha C, following through to the carbonyl carbon all the way to the carbonyl oxygen. The groups on the nitrogen and subsequently the alpha C are numbered next, with the side chain the hydrogen attached to the alpha C. The *N*-Acetyl or *N*-Methyl groups are defined separately and then added at any time, as their coordinates do not disturb the definitions of the internal parameters specific to the core (i.e. amino acid residue).

In the case of di-peptides or larger polypeptides lengths, the numbering is left 'open', in the form of variables, so that the files may be modified at any time at any of these groups without disturbing the structure of the remaining coordinates, or the molecular environment it defines. As the internal coordinate system relies on fixing subsequent atoms to previously defined ones, the next amino acid would join to the carbonyl carbon of the first. Hence, any of the standardized amino acids in this database, numbered beginning at the N attached to the C_α, would then 'mesh' with the internal coordinates of the first without problem. In fact, any of the groups or atoms attached to the backbone may be denoted by a variable of choice and modified or truncated accordingly.

Although an extensive explanation of the above numbering sequence would be required to show all the inherent benefits of this system, the reader is encouraged to attempt trials on various input files.

In this way, it is possible to make an additive or integrateable database for the 20 amino acids, with the point of attachment in a polypeptide chain being defined as the #1 atom, in the coordinate system used. Other previously optimized structures can then be added to the database, including the *N*-acetyl and *N*-methyl portions, or even long chain fatty acids, provided they use the standardized numbering system. Simply adding the coordinate definitions onto the internal coordinate system of that original structure may then extend an optimized structure. Any polypeptide combination of any possible conformation can be easily created and submitted for optimization efficiently and can be easily automated. With the new millennium, a massive mega-project is proposed to use the above process to explore the large number of potential energy surfaces created by this gigantic number of peptide structures. For example, a tripeptide sequence would contain many millions of possible combinations and conformations to be optimized; this number grows exponentially with the extension of the polypeptide chain.

To create a surface large enough to help in the unraveling of the protein folding problem, one must continue to explore such surfaces with progressively larger sections of the amino acid polymer. This creates a massive demand for both CPU time, hard storage space and data management and extraction; the latter being a much less significant obstacle. The all too common problem of CPU power being the 'bottle neck' in the creation of meaningful ab initio results may be solved via a different avenue. Rather than throwing massive amounts of hardware in tightly coupled networks at the challenge, instead it is proposed to farm-out the CPU time to the millions of stand alone PC's and workstations residing on the loosely coupled network of the internet.

Some computational problems and data analysis ventures have indeed used this approach with much success. SETI@home is one example, whereby radio-telescope data from Arescebo in Puerto Rico are analyzed by millions of individuals in their homes, using their PC's. At last count, over 1.3 million people were contributing computer time toward this effort. The process is automated to use 'down CPU time', whereby the individual's computer usage is not affected by the running program, nor are they burdened by any direct responsibility of managing their chunk of the data. Once the user needs their computer, the calculating program is stopped and the CPU is entirely free for the user. Another such venture is the DES/RC5 project, whose goal is to show how US export limitations on the strength of cryptographic algorithms is naive. They are undertaking this by having contributors' computers try every possible decryption key in sequence against a cipher. With over 100,000 CPUs working on the job, cracking such encryption is relatively easy. The megaproject could indeed make use of a similar such automated and 'loosely coupled' network, whereby surfaces of the many independent variables may be analyzed. Those who wish to participate may register at the following e-mail address: **protein@velocet.ca.**

It is part of human nature to ask what good will come out of a project that aims to decode the secret of protein folding. Such a discovery may well be regarded as the greatest human achievement in history. Its social impact would be greater than the combined social impact of:

the discovery of fire;
the discovery of writing;
the discovery of the wheel.

Even though this topic is very important for the whole human race, it cannot be done due to lack of funds unless individual contributions come from millions of people.

The deciphering of the law of protein folding would open new dimensions to all life sciences, including medicine. For example, the drug discovery process would be shortened dramatically, by a more cost-effective method for the production of pharmaceutical products. Not only would the cure be found reasonably soon (say by the year of 2025) for cancer, HIV, as well as other diseases, preventive medicine would also reach new heights whereby pathogen infections (bacteria, viruses and prions) could be eliminated.

Anyone, who has lost a loved one to an incurable disease, appreciates the prospects of eliminating all human sufferings. This heartache includes the pain of the ill, as well as the anguish of those who would

mourn after the patient has passed away. Needless to say, a person lost prematurely by a family is also lost by society as a whole, with all of his/her talents and productivity. Global hunger could also be eliminated by improved agricultural and food-processing technologies, in which the knowledge of protein folding also plays a significant role.

The peptide folding project, which one can participate in, at the above e-mail address, does not offer a quick fix, but gradually it will lead to a new understanding of protein folding, which in turn will produce a healthier society enjoyed by generations to come.

Acknowledgements

The authors wish to acknowledge the continuous support of the National Research Council (NRC), and its funding-successor, the National Sciences and Engineering Research Council (NSERC) of Canada for the period of 1964 to 1999. We would like to thank Dr. Carlos Sosa of Silicon Graphics Inc. for providing computational resources at their facility in Eagan, MN. We are also grateful to the National Cancer Institute (NCI) for the use of services at Frederick Biomedical Supercomputing Center. Finally, we also wish to thank Velocet Communications Inc. for database management, network support, software and distributive processing development.

References

[1] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, in: Highlights of Modern Biochemistry, Proceedings of the 14th International Congress of Biochemistry, vol. 5, VSP, Utrecht, The Netherlands, 1988, p. 47.
[2] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, in: Proceedings of the 10th International Symposium of Medicinal Chemistry, Budapest, Hungary, 1988, p. 288 (Abstract P-168).
[3] A. Furka, F. Sebestyen, M. Asgedom, G. Dibo, Int. J. Pept. Protein Res. 37 (1991) 487.
[4] K.S. Lam, S.E. Salmon, E.M. Hersh, V.J. Hruby, W.M. Kazmierski, R.J. Knapp, Nature 354 (1991) 82.
[5] R.A. Houghten, C. Pinilla, S.E. Bloudelle, J.R. Appel, C.T. Dooley, J.H. Cuervo, Nature 354 (1991) 84.
[6] B.A. Bunin, J.A. Ellman, J. Am. Chem. Soc. 114 (1992) 10 997.
[7] C. Levinthal, J. Chim. Phys. 65 (1968) 44.
[8] B. Honig, J. Mol. Biol. 293 (1999) 283.
[9] I.G. Csizmadia, M.C. Harrison, B.T. Sutcliffe, Q Prog. Rep. (MIT-SSMTG) 50 (1963) 1.
[10] I.G. Csizmadia, M.C. Harrison, B.T. Sutcliffe, Q Prog. Rep. (MIT-SSMTG) 59 (1966) 43.
[11] I.G. Csizmadia, M.C. Harrison, B.T. Sutcliffe, Theor. Chim. Acta 6 (1966) 217.
[12] M.A. Robb, I.G. Csizmadia, Theor. Chim. Acta 10 (1968) 269.
[13] I.A. Topol, S.K. Burt, E. Deretey, T.-H. Tang, A. Perczel, I.G. Csizmadia, in preparation.
[14] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez, J.A. Pople, Gaussian 94, Revision D.2, Gaussian, Inc., Pittsburgh, PA, 1995.
[15] J.C. Vank, C.P. Sosa, A. Perczel, I.G. Csizmadia, Can. J. Chem. 178 (2000) 395–408.
[16] A. Perczel, J.G. Angyan, M. Viviani, J.-L. Rivail, J.-F. Marcoccia, I.G. Csizmadia, J. Am. Chem. Soc. 113 (1991) 6265.
[17] M.A. McAllister, A. Perczel, P. Császár, W. Viviani, J.L. Rivail, I.G. Csizmadia, J. Mol. Struct. (Theochem) 288 (1993) 161.
[18] W. Viviani, J.-L. Rivail, A. Perczel, I.G. Csizmadia, J. Am. Chem. Soc. 115 (1993) 8321.
[19] Ö. Farkas, M.A. McAllister, J.H. Ma, A. Perczel, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 369 (1996) 105.
[20] A. Perczel, Ö. Farkas, I.G. Csizmadia, Can. J. Chem. 75 (1997) 1120.
[21] Ö. Farkas, A. Perczel, J.F. Marcoccia, M. Hollósi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 331 (1995) 27.
[22] A. Perczel, Ö. Farkas, I.G. Csizmadia, J. Comp. Chem. 17 (1996) 821.
[23] A. Perczel, Ö. Farkas, I.G. Csizmadia, J. Am. Chem. Soc. 118 (1996) 7809.
[24] H.A. Baldoni, A.M. Rodriguez, G. Zamarbide, R.D. Enriz, Ö. Farkas, P. Csaszar, L.L. Torday, C.P. Sosa, I. Jakli, A. Perczel, M. Hollosi, I.G. Csizmadia, J. Mol. Struct. (Theochem) 465 (1999) 79.
[25] S.J. Salpietro, A. Perczel, Ö. Farkas, R.D. Enriz, I.G. Csizmadia, J. Mol. Struct. (Theochem) 497 (2000) 39–63.
[26] M. Berg, S.J. Salpietro, I.G., Csizmadia, J. Mol. Struct. (Theochem) (2000) (in press) paper 6593.
[27] M.A. Zamora, H.A. Baldoni, A.M. Rodindez, R.D. Enriz, C.P. Sosa, J.C. Vank, A. Perczel, A. Kucsman, Ö. Farkas, E. Deretey, I.G. Csizmadia, in preparation.
[28] Ö. Farkas, G.N. Zamarbide, H.A. Baldoni, L.L. Torday, A.M. Rodriguez, R.D. Enriz, C.P. Sosa, I. Jakli, A. Perczel, I.G. Csizmadia, in: Proline, the Maverick Amino Acid WATOC 99, Fifth World Congress of Theoretically Oriented Chemists, Imperial College, London, UK, 1–6 August, 1999, p. 212.

[29] H.A. Baldoni, A. Perczel, R.D. Enriz, I.G. Csizmadia, J. Mol. Struct. 315 (2000) 97–111.
[30] M.B. Santillan, G.M. Ciuffo, E.A. Jauregui, I.G. Csizmadia, J. Mol. Struct. (Theochem) 463 (1999) 237–250.
[31] M.L. Mak, S.J. Salpietro, R.D. Enriz, I.G. Csizmadia, Can. J. Chem. 78 (2000) in press.
[32] A. Perczel, O. Farkas, I. Jakli, I.G. Csizmadia, Theochem 315 (1998) 455.
[33] W. Viviani, J.-L. Rivail, A. Perczel, I.G. Csizmadia, J. Am. Chem. Soc. 115 (1993) 8321.
[34] M.A. McAllister, G. Endredi, W. Viviani, A. Perczel, P. Csaszar, J. Ladik, J.L. Rivail, I.G. Csizmadia, Can. J. Chem. 73 (1995) 1563.
[35] M.A. McAllister, A. Perczel, P. Csaszar, I.G. Csizmadia, J. Mol. Struct. (Theochem) 288 (1993) 181.
[36] A. Perzcel, M.A. McAllister, P. Csaszar, I.G. Csizmadia, J. Am. Chem. Soc. 115 (1993) 4849.
[37] A. Perzcel, M.A. McAllister, P. Csaszar, I.G. Csizmadia, Can. J. Chem. 72 (1994) 2050.
[38] M. Ramek, C.H. Yu, L. Schafer, Can. J. Chem. 76 (1998) 566.
[39] M. Cheung, M.E. McGovern, T. Jiu, D.C. Zhao, M.A. McAllister, P. Csaszar, O. Farkas, I.G. Csizmadia, J. Mol. Struct. (Theochem) 309 (1994) 151.
[40] G. Endredi, M.A. McAllister, O. Farkas, A. Perzcel, J. Ladik, I.G. Csizmadia, J. Mol. Struct. (Theochem) 331 (1995) 11.
[41] L. Torday, J. Patrica, G.E. Balogh, B. Penke, J.Gy. Papp, J. Pharm. Pharmacol. 50 (1998) 667.
[42] A.G. Plaut, W.W. Bachovchin, Methods Enzymol. 244 (1994) 137.
[43] N.D. Rawlings, A.J. Barett, Methods Enzymol. 244 (1994) 19.
[44] K. Poulsen, J. Brandt, J.P. Hjorth, H.C. Thogersen, M. Kilian, Infect. Immun. 57 (1989) 3097.
[45] J. Pohlner, R. Halter, K. Bayreuther, T.F. Meyer, Nature 325 (1987) 458.
[46] S.G. Wood, M. Lynch, A.G. Plaut, J. Burton, J. Med. Chem. 32 (1989) 2407.
[47] H. Sahl, R.W. Jack, G. Bierbaum, Eur. J. Biochem. 230 (1995) 827.
[48] E. Gross, J. Morell, J. Am. Chem. Soc. 89 (1967) 2791.
[49] E. Gross, J. Morell, L.C. Craig, Proc. Natl. Acad. Sci. (USA) 62 (1969) 952.
[50] R.C.M. Lau, K.L. Rinehart, J. Am. Chem. Soc. 117 (1995) 7606.
[51] C.J. Pearce, K.L. Rinehart, J. Am. Chem. Soc. 101 (1979) 5069.
[52] A.F. Spatola, in: B. Weinstein (Ed.), Chemistry and Biochemistry of Amino Acids, 7, Marcel Dekker, New York, 1983, pp. 295–345.
[53] Y. Shimohigashi, C.H. Stammer, Int. J. Pept. Protein Res. 20 (1982) 199.
[54] Y. Shimohigashi, M.L. English, C.H. Stammer, Biochem. Biophys. Res. Commun. 104 (1982) 583.
[55] G. Pietrzynski, B. Rzeszotarska, E. Ciszak, M. Lisowski, Z. Kubica, G. Boussard, Int. J. Pept. Protein Res. 48 (1996) 347.
[56] G. Pietrzynski, B. Rzeszotarska, Z. Kubica, Int. J. Pept. Protein Res. 40 (1992) 524.
[57] A.K. Fuzery, I.G. Csizmadia, J. Mol. Struct. (Theochem) (2000) (in press).
[58] Y. Ohnishi, H. Fujii, K. Murakami, T. Sakamoto, K. Tsukada, M. Fujimaki, M. Kojima, I. Saiki, Cancer Lett. 124 (1998) 157–163.
[59] E. Okuda-Ashitaka, T. Minami, S. Tashibana, Y. Yoshihara, T. Nishiuchi, T. Kimura, S. Ito, Nature 392 (1998) 286–289.
[60] G. Calo, R. Guerrini, R. Bigoni, A. Rizzi, C. Bianchi, D. Regoli, S. Salvadori, J. Med. Chem. 41 (1998) 3360–3366.
[61] J.-L. Butour, C. Moisand, C. Mollereau, J.-C. Meunier, Eur. J. Pharmacol. 349 (1998) R5–R6.
[62] G. Guichard, S. Calbo, S. Muller, Ph. Kourilsky, J.-P. Briand, J.-P. Abastsdo, J. Biol. Chem 270 (N3) (1995) 26 057–26 059.
[63] W.M. Kazmierski, R.D. Ferguson, R.J. Knapp, G.K. Lui, H.I. Yamamura, V.J. Hruby, Int. J. Pept. Protein Res. 39 (1992) 401–404.
[64] I.V. Repyakh, E. Deretey, I.G. Csizmadia, J. Mol Struct. (Theochem) 503 (2000) 81–96.
[65] G. Endredi, A. Perczel, Ö. Farkas, M.A. McAllister, G.I. Csonka, J. Ladik, I.G. Csizmadia, J. Mol. Struct. (Theochem) 391 (1997) 15.
[66] A.M. Rodriguez, H.A. Baldoni, F. Suvire, R.N. Vazquez, G. Zamarbide, R.D. Enriz, Ö. Farkas, A. Perczel, M.A. McAllister, L.L. Torday, J.G. Papp, I.G. Csizmadia, J. Mol. Struct. (Theochem) 455 (1998) 275.
[67] J.J.P. Stewart, MOPAC 6.00: A Semiempirical Molecular Orbital Program. 6th ed., F.J. Seiler Research Lab., US Air Force Academy, Boulder, CO, USA, 1990.
[68] J. Anderson, B. Deskins. The Nutrition Bible, William Morrow & Company, Inc., 1995.
[69] A. Bock, K. Forchhammer, J. Heider, W. Leinfelder, G. Sawers, B. Veprek, F. Zinoni, Mol. Microbiol. 5 (1991) 515.
[70] T.C. Stadtman, Annu. Rev. Biochem. 59 (1990) 111.
[71] K. Wingler, M. Bocher, L. Flohe, H. Kollms, R. Brigelius-Flohe, Eur. J. Biochem. 259 (1/2) (1999) 149.
[72] V. Gladyshev, T.C. Stadtman, D.L. Hatfield, K.T. Jeang, Proc. Natl. Acad. Sci. (USA) 96 (3) (1999) 835.
[73] B.A. Zachara, J. Trace. Elem. Electrolytes Health Dis. 6 (1992) 137.
[74] L. Flohe, Glutathione: Chemical, Biochemical and Medical Aspects, Wiley, New York, 1989.
[75] C. Little, J.P. O'Brien, Biochem. Biophys. Res. Commun. (1968) 145.
[76] B.O. Christophersen, Biochim. Biophys. Acta 176 (1969) 463.
[77] V.N. Gladyshev, V.M. Factor, F. Housseau, D.L. Hatfield, Biochem. Biophys. Res. Commun. 251 (2) (1998) 488.
[78] S. Maity, J.R. Vedasiromoni, D.K. Ganguly, Jpn J. Pharmacol. 78 (3) (1998) 285.
[79] Y. Saito, T. Hayashi, A. Tanako, Y. Watanabe, M. Suzuki, E. Saito, K. Takahashi, J. Biol. Chem. 274 (5) (1999) 2866.
[80] S.C. Vendeland, M.A. Beilstein, C.L. Chen, O.N. Jensen, E. Barofsky, P.D. Whanger, J. Biol. Chem. 268 (23) (1993) 17 103.

[81] C. Buettner, J.W. Harney, P.R. Larsen, J. Biol. Chem. 273 (50) (1998) 33374.
[82] M.M. Berggren, J.F. Mangin, J.R. Gasdaska, G. Powis, Biochem. Pharmacol. 57 (2) (1999) 187.
[83] V.N. Gladyshev, K.T. Jeang, T.C. Stadtman, Proc. Natl. Acad. Sci. (USA) 93 (1996) 12.
[84] A. Kyriakopoulos, C. Hammel, H. Gessner, D. Behne, Am. Biotech Lab 114 (1966) 22.
[85] M. Kalklosch, A. Kyriakopoulos, C. Hammel, D. Behne, Biochem. Biophys. Res. Commun. 217 (1995) 162.
[86] D. Behne, A. Kyriakopoulos, M. Kalcklosch, C. Weiss-Nowak, H. Pfeifer, H. Gessner, C. Hammel, Biomed. Enviom. Sci. 10 (1997) 340.
[87] E.W. Taylor, C.S. Ramanathan, R.K. Jalluri, R.G. Nadimpalli, J. Med. Chem. 37 (1994) 2637.
[88] E.W. Taylor, R.G. Nadimpalli, C.S. Ramanathan, Biol. Trace Element Res. 56 (1997) 63.
[89] E.W. Taylor, A. Bhat, R.G. Nadimpalli, W. Zhang, J. Kececioglu, J. AIDS Hum. Retrovirol. 17 (1997) 1977.
[90] M.K. Baum, G. Shor-Posner, S. Lai, G. Zhang, H. Lai, M.A. Fletcher, H. Sauberlich, J.B. Page, J. AIDS Hum. Retrovirol. 15 (1997) 370.
[91] P.H. Gann, J. Ma, E. Giovannucci, W. Willet, F.M. Sacks, C.H. Hennekens, M.J. Stampfer, Cancer Res. 59 (1999) 1225–1230.
[92] M.L. Nguyen, S.J. Schwartz, Food Technol. 53 (2) (1999) 38–45.
[93] S. Basu, R. Senior, E.B. Raftery, A. Lahiri, Eur. Heart J. 17 (1996) 43.
[94] R. van der Does, U. Hauf-Zachariou, E. Pfarr, W. Holtbrugge, S. Konig, M. Griffiths, A. Lahiri, Am. J. Cardiol. 83 (1999) 643.
[95] U. Hauf-Zachariou, R.A. Blackwood, K.A. Gunawardena, J.G. O'Donnell, S. Garnham, E. Pfarr, Eur. J. Clin. Pharmacol. 52 (1997) 95.
[96] M. Packer, M.R. Bristow, J.N. Cohn, W.S. Colucci, M.B. Fowler, E.M. Gilbert, N.H. Shusterman, N. Engl. J. Med. 334 (1996) 1349.
[97] B. Tadolini, F. Franconi, Free Radic. Res. 29 (1998) 337.
[98] R. Senior, S. Basu, C. Kinsey, S. Schaeffer, A. Lahiri, Am. Heart J. 137 (1999) 646.
[99] R.R. Ruffollo Jr. G.Z. Feuerstein, J. Cardiovasc. Pharmacol. 32 (Suppl. 1) (1998) S22.
[100] G. Feuerstein, T.L. Yue, X. Ma, R.R. Ruffolo, Prog. Cardiovasc. Dis. 41 (Suppl. 1) (1998) 17.
[101] H. Watanabe, M. Kakihana, S. Ohtsuka, Y. Sugishita, J. Am. Coll. Cardiol. 32 (1998) 1201.
[102] H. Watanabe, M. Kakihana, S. Ohtsuka, Y. Sugishita, J. Am. Coll. Cardiol. 32 (1998) 1194.
[103] G. Milligan, G-Protein signal transduction and diseases, Academic Press. Harcourt Brace Jovanovich, UK. 1992 pp. 323–343.
[104] M. Staikova, I.G. Csizmadia, J. Mol. Struct. (Theochem) 466 (1999) 181–186.
[105] M. Patel, E. Deretey, I.G. Csizmadia, J. Mol. Struct. (Theochem) 492 (1999) 1–18.
[106] T. Owa, H. Yoshino, T. Okauchi, K. Yoshimatsu, Y. Ozawa, N.H. Sugi, T. Nagasu, N. Koyanagi, K. Kitoh, J. Med. Chem. 42 (1999) 3789.
[107] B. Shan, J.C. Medina, E. Santha, W.P. Frankmoelle, T.-C. Chou, R.M. Learned, M.R. Narbut, D. Stott, P. Wu, J.C. Jaen, T. Rosen, P.B.M.W. Timmermans, H. Beckmann, Proc. Natl. Acad. Sci. (USA) 96 (1999) 5686–5691.
[108] K.C. Agarwal, Med. Res. Rev. 16 (1996) 111.
[109] Ö. Farkas, S.J. Salpietro, P. Császár, I.G. Csizmadia, J. Mol. Struct. (Theochem) 367 (1996) 25.
[110] J.M. Boggs, et al., Biochemistry 36 (16) (1997) 5065–5071.
[111] L. Cao, R. Goodin, D. Wood, M.A. Moscarello, J.N. Whitaker, Biochemistry 38 (1999) 6157.
[112] C.W. Carter, Annu. Rev. Biochem. 62 (1993) 715.
[113] J. Cavarelli, B. Rees, M. Ruff, J.C. Thierry, D. Moras, Nature 362 (1993) 181.
[114] P. Schimmel, Annu. Rev. Biochem. 56 (1987) 125.
[115] D. Söll, L. RajBhandary, L. Uttam (Eds.), tRNA: Structure, Biosynthesis, and Function ASM Press, Washington, DC, 1995 chap. 14,16,18–20.
[116] V.M. Mazza, Divalent Metal Ion Catalysis in the Hydrolysis of Aminoacyl Alkyl Phosphates, Master of Science Thesis, University of Toronto, 1996.
[117] S.M. Hecht (Ed.), Bioorganic Chemistry of Nucleic Acids *Oxford University Press, New York*, 1996.
[118] H. Henry-Riyad, T. Tang, I.G. Csizmadia, J. Mol. Struct. (Theochem) 492 (1999) 67.
[119] G.I. Csonka, I.G. Csizmadia, Chem. Phys. Lett. 243 (1995) 419.
[120] G.I. Csonka, N. Anh, J. Angyan, I.G. Csizmadia, Chem. Phys. Lett. 245 (1995) 129.
[121] G.I. Csonka, K. Elias, I.G. Csizmadia, J. Comp. Chem. 18 (1997) 330.
[122] G.I. Csonka, K. Elias, I. Kolossvary, C.P. Sosa, I.G. Csizmadia, J. Phys. Chem. 102 (1998) 1219.
[123] G.I. Csonka, K. Elias, I.G. Csizmadia, Chem. Phys. Lett. 257 (1996) 49.
[124] G.I. Csonka, I. Kolossvary, P. Csaszar, K. Elias, I.G. Csizmadia, J. Mol. Struct. (Theochem) 395/396 (1997) 29.
[125] G.I. Csonka, C.P. Sosa, I.G. Csizmadia, J. Phys. Chem. 104 (2000) 3381–3390.
[126] A.C. Lin, S.J. Salpietro, E. Deretey, I.G. Csizmadia, Can. J. Chem. 78 (2000) 362–382.

THEOCH7129

ELSEVIER

Journal of Molecular Structure (Theochem) 000 (2001) 000–000

THEO CHEM

www.elsevier.nl/locate/theochem

Conformational potential energy surfaces of a Lycopene model

Gregory A. Chasse[a,b,*], Kenneth P. Chasse[a], Arpad Kucsman[c], Ladislaus L. Torday[d], Julius G. Papp[d]

[a]*Velocet Communications Inc., 210 Dundas St. W., Suite 800 Toronto, Ont., Canada M5G 2E8*
[b]*Department of Chemistry, University of Toronto, Toronto, Ont., Canada M5S 3H6*
[c]*Department of Organic Chemistry, Eötvös Loránd University of Budapest, P.O. Box 32, H-1518 Budapest 112, Hungary*
[d]*Department of Pharmacology and Pharmacotherapy, Albert Szent-Gyorgy Medical University, H-6701 Szeged, Hungary*

Received 30 October 2000; accepted 5 February 2001

Abstract

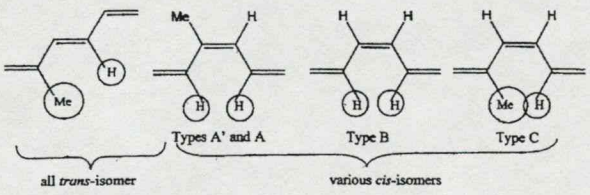
Ab initio conformational analysis has been carried out at the RHF/3-21G level of theory. Computations were performed on a tail-end lycopene (Model B). Both the all-*trans* and the 5-*cis*-isomers were studied. The fully planar structure turned out to be a second-order saddle point, which indicated that lycopene itself is not planar. Most of the conformers of the 5-*cis*-isomer are more stable than the corresponding conformers of the all-*trans*-isomer. This stability is in agreement with the observation that even though lycopene is biosynthesized in plants as the all-*trans* form, in the human body over 65% exists in one of the *cis*-forms and less than 35% remains in its all-*trans* form. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Conformational analysis; All *trans*-lycopene tail-end model; Selected *cis*-isomers of lycopene tail-end model; Conformational potential energy surfaces; 2D scans; Ab initio MO theory

1. Introduction

1.1. Biological background

Several chronic diseases, amongst which are cancer and cardiovascular diseases, are related to oxidative stress which is now recognized as an important etiological factor. Antioxidants are effective in reducing the damaging effect of oxygen containing radicals and radicaloids such as $^{\cdot}OH$, O_2^{-} and H_2O_2 as well as that of singlet oxygen $^1O_2$. Lycopene is a carotenoid type antioxidant, present in tomatoes and other fruits and vegetables. Studies have shown that lycopene acts as an antioxidant in vivo, providing protection against the oxidation of lipids, proteins and DNA [1,2]. A recently published review [3] indicated a significant inverse correlation between the intake of lycopene and therefore the serum concentrations of lycopene and the risk of several diseases including cancer. In the absorption and bio-availability the isomeric forms of lycopene appear to play an important role.

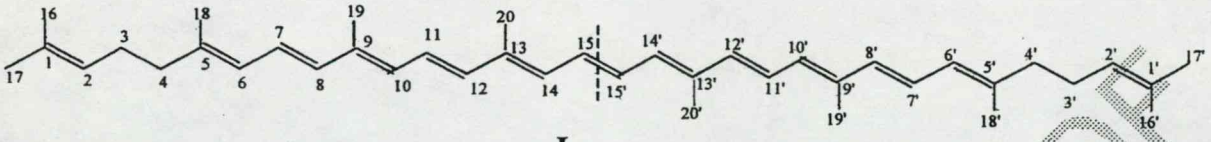


Scheme 1.

* Corresponding author. Address: Velocet Communications Inc., 210 Dundas St. W., Suite 800 Toronto, Ont., Canada M5G 2E8. Tel.: +1-416-978-3598; fax: +1-416-978-3598.
E-mail address: gchasse@fixy.org (G.A. Chasse).

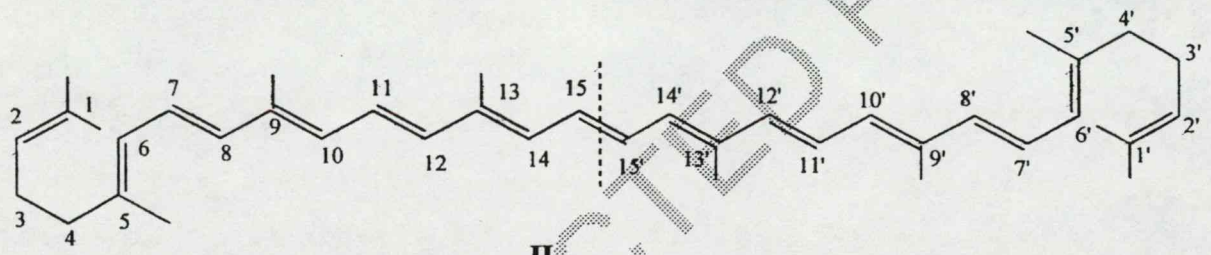
1.2. Structural background

The skeleton of lycopene [$C_{40}H_{56}$] consists of eight isoprenic units, thus, it is related to tetraterpenes [$C_{40}H_{64}$] even though it contains fewer hydrogens and therefore more double bonds. Its composition is shown in **I** in its fully extended form which is the all *trans*-lycopene. As such, lycopene is closely related to β-carotenes.



I

Sometimes, the structure of lycopene is presented in a pre-folded form (**II**) to show its structural similarity to β-carotene. Due to the internal molecular symmetry, it has been traditional to number the chain from the two ends as shown in **I** and **II**.



II

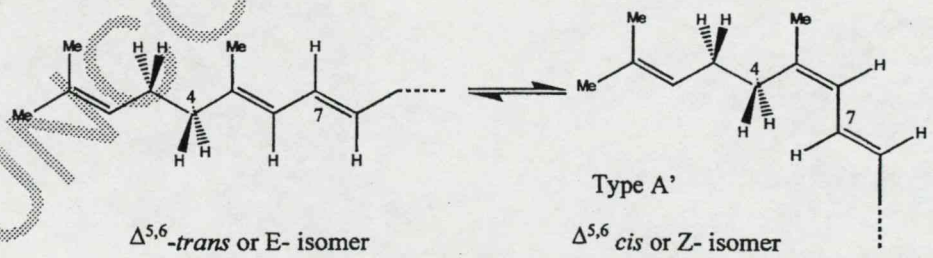
The history of lycopene [1–14] reveals an interesting story from the initial curiosity of a colourful substance in the 1910s to the medical application in the 1990s as an antioxidant.

In 1943, Pauling pointed out [14] not all *cis*-isomers may be of equal stability due to a number of possible 1,4 interactions as shown by the following structures. Clearly, on the basis of relative group sizes, the $-CH_3-H-$ interaction (type C) appears to be the most destabilizing (Scheme 1).

No X-ray structure of lycopene has been determined as yet. There are, however, two X-ray structures of β-carotene in the literature [15,16].

1.3. Computational background

In fruits and vegetables Lycopene is present mostly in its all-*trans* isomeric form. However, *cis*-isomers constitute the predominant form present in the serum and tissues [17,18]. This observation suggests that at least some of the *cis*-isomers may be energetically comparable to the stability of the all-*trans* isomers. Fig. 1 shows the all-*trans*- and selected *cis*-isomers of lycopene. The all-*trans*- and the selected six *cis*-isomers (5-, 7-, 9-, 11-, 13- and 15-) are yet to be explored computationally in their molecular



Type A'

$\Delta^{5,6}$-*trans* or E- isomer

$\Delta^{5,6}$ *cis* or Z- isomer

Scheme 2.

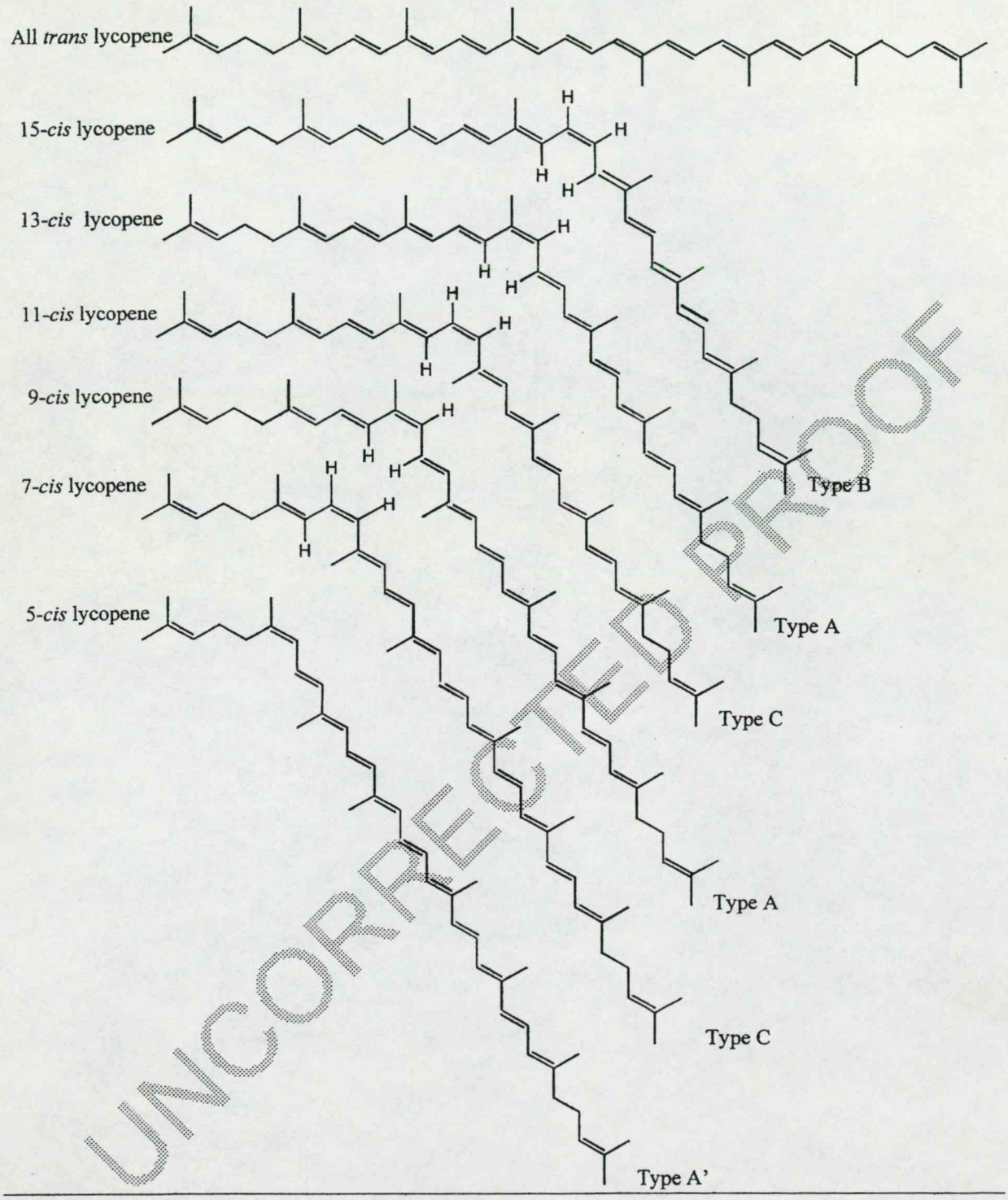
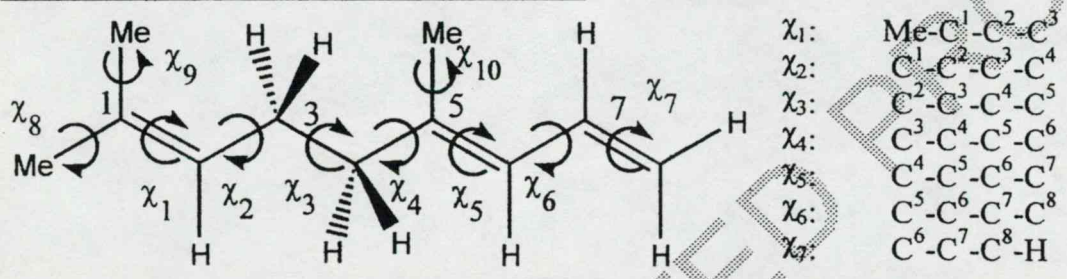


Fig. 1. Geometrical isomers of lycopene.

entirety (containing 296 electrons and 96 atoms, which corresponds to 282 geometrical parameters to be optimized). Thus, lycopene may well be among the largest organic molecules to be investigated, using ab initio molecular computation, with the current computational technology. The relative energies for the four types of structures, are depicted in Fig. 1, are expected to exhibit four categories of stability. Consequently, we may anticipate four energy ranges.

Type A′ involves the 5-*cis*-isomer.
Type A involves the 9-*cis* and 13-*cis*-isomers.
Type B involves the 15-*cis*-isomer.
Type C involves the 7- and 11-*cis*-isomers.

Type A′ is extremely similar to Type A (see Fig. 1) except that the hydrogens, involved in 1,4 interactions (attached to carbons 4 and 7), are not eclipsed but staggered. This is illustrated in greater detail by Scheme 2.



III

2. Scope

The presence of three consecutive carbon–carbon single bonds (C^2–C^3–C^4–C^5) suggests conformational flexibility of the tail-ends of lycopene. It was deemed desirable therefore to study the conformational intricacy of the tail-end of lycopene by choosing some model compound. If one looks at the lycopene structure, one can cut out from the full molecule shorter segments and terminate them with hydrogen atoms. These model compounds are labelled in Scheme 3 as Model A, Model B, Model C, Model D, etc.

For the present study, we have chosen Model B in order to be able to mimic conformationally, the tail-end of the all-*trans* isomer, as well as the 5-*cis* isomer, as depicted in Scheme 4.

3. Method

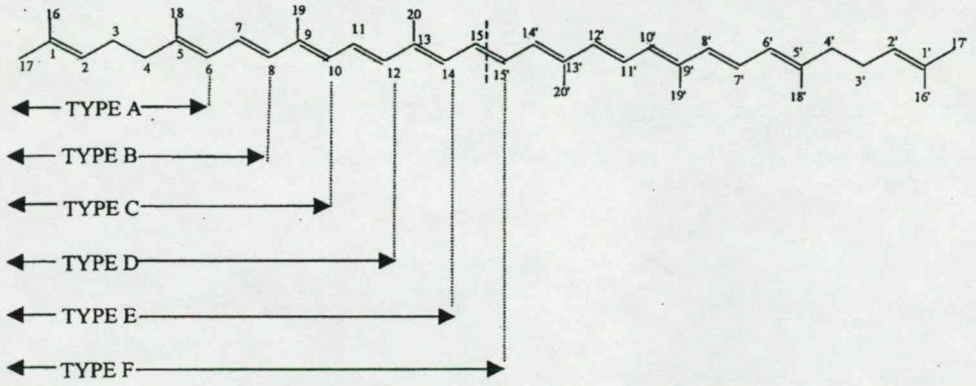
In order to learn about the conformational behaviours of the two tail-end conformations of lycopene, Model B (**III**) was studied in more detail. This truncated lycopene model (C^1–C^8 segment) contains of the first three double bonds of lycopene. Structure **III** shows the fully extended form of Model B.

When a planar moiety is rotated about a tetrahedral carbon, it may be either eclipsed with a tetrahedral bond or perpendicular to that bond. This has been revealed by the study on ethyl benzene [19]. These idealized conformers are illustrated by Scheme 5.

Single scans can be carried out to study the rotation about the C^2–C^3(χ_2), C^3–C^4(χ_3) and C^4–C^5(χ_4) single bonds of **III**. However, the conformational problem is better studied in the form of a potential energy hyper-surface (PEHS), which involves all three torsional angles.

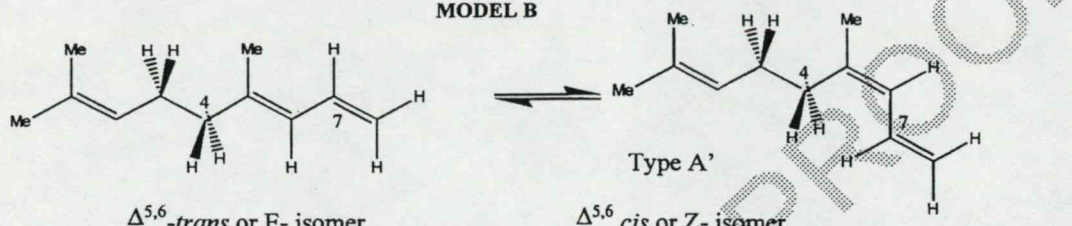
$$E = f(\chi_2 \text{å} \chi_3 \text{å} \chi_4) \text{æ} \quad (1)$$

Since the rotation about the single C–C bond (χ_3) is expected to be in g^+, a, g^- configurations, therefore three potential energy surfaces (PES)



Scheme 3.

MODEL B



$\Delta^{5,6}$-*trans* or E- isomer $\Delta^{5,6}$ *cis* or Z- isomer

Scheme 4.

may be generated at these three orientations of χ_3.

$$E = f(\chi_2 \text{å} \chi_4) \quad \text{at} \quad \chi_3 = +60°\text{å} \quad (2a)$$

$$E = f(\chi_2 \text{å} \chi_4) \quad \text{at} \quad \chi_3 = +180°\text{å} \quad (2b)$$

$$E = f(\chi_2 \text{å} \chi_4) \quad \text{at} \quad \chi_3 = +300°(-60°)\text{æ} \quad (2c)$$

The levels of these 2D cross-sections are shown in Fig. 2.

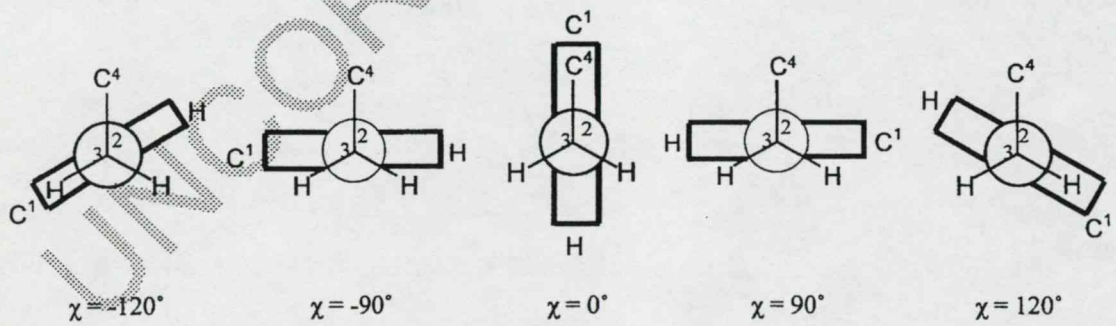
One may also generate 1D cross-sections such as

$$E = f(\chi_2) \quad \text{at} \quad \chi_3 = 180°\text{å} \quad \chi_4 = \pm 90°\text{å} \quad (3a)$$

$$E = f(\chi_3) \quad \text{at} \quad \chi_2 = \pm 90°\text{å} \quad \chi_4 = \pm 90°\text{å} \quad (3b)$$

$$E = f(\chi_4) \quad \text{at} \quad \chi_2 = \pm 90°\text{å} \quad \chi_3 = 180°\text{æ} \quad (3c)$$

The location of these conformational potential



Scheme 5.

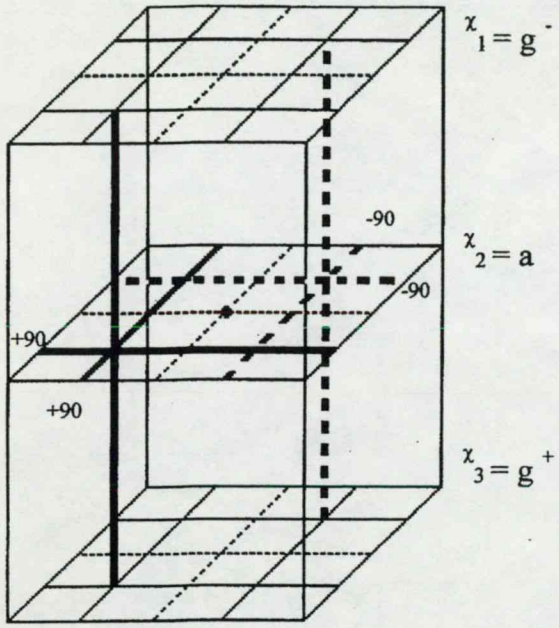


Fig. 2. A schematic representation of the conformational PEHS, $E = f(\chi_2 \hat{a} \chi_3 \hat{a} \chi_4)$ of lycopene Model B. The heavy dot, at the centre, illustrates the location of the fully symmetric conformation as drawn in **III**. The three levels indicate the three 2D cross-sections (PES) and the three perpendicular heavy broken and solid lines specify the locations of the three 1D cross-sections (PEC) investigated.

energy curves (PEC) are indicated by three heavy solid lines and three heavy broken lines in Fig. 2.

The computations were carried out using the GAUSSIAN 98 program system [20]. Standard geometry optimizations were performed on the all-*trans* as well as six selected *cis*-isomers.

4. Results and discussion

4.1. Molecular conformations

The conformational PEHS, as specified by Eq. (1), illustrated schematically in Fig. 2, is applicable for both the all-*trans*- and for the 5-*cis*-isomer of lycopene Model B (**III**).

4.1.1. The all-trans-structure

The all-*trans*-lycopene Model B (**III**) was subjected first to conformational analysis. The conformational PEHS (1) of three independent variables, was investigated in terms of three PESs (2) of two independent variables. These three PESs are depicted in landscape and contour representations in Figs. 3–5. For Fig. 3, χ_3 was kept in its *anti*-position (**2b**), for Fig. 4, χ_3 was kept in g^+ and for Fig. 5, χ_3 was kept in g^- position. Clearly the central level of Fig. 2, for which the PES (**2b**) is depicted in Fig. 3, is the most symmetric. By inspection, we may anticipate the presence of nine minima on the PES. These are listed in Scheme 6.

Those minima located at the edges are repeated twice and the structure at the corner is repeated four times. These repeated structures are shown in square brackets, leaving nine unique conformers (which are not in square brackets). Since there are three levels, as shown in Fig. 2, the three PESs (Figs. 3–5) may contain up to $3 \times 9 = 27$ stable conformations. The 1D scans leading to the three PECs (shown at the left-hand side of Fig. 6) indeed suggest the existence of three unique minima, on each of the three PECs. This reconfirms that $3^3 = 27$ minima may be anticipated. However, sometimes expected minima are annihilated from the PES. Geometry optimizations have been initiated on these 27 conformations.

The optimized torsional angles are summarized in Table 1. Indeed a few minima, involving some *syn*-orientations, were annihilated. The torsions about the single bonds (χ_2, χ_3 and χ_4) are shown in bold. Al other torsional angles are associated with double bonds. Since this is the all-*trans* form, these double bonds should have dihedral angles in the vicinity of $\pm 180°$. The last double bond, however, is ending in CH_2, so one of the two hydrogens are *trans* (χ_7) and the other is *cis*. In Table 1, the last three dihedral

[s(a)s]	**[g^+(a)s]**	**[g^-(a)s]**	**[s(a)s]**
s(a)g^-	**g^+(a)g^-**	**g^-(a)g^-**	**[s(a)g^-]**
s(a)g^+	**g^+(a)g^+**	**g^-(a)g^+**	**[s(a)g^+]**
s(a)s	**g^+(a)s**	**g^-(a)s**	**[s(a)s]**

Scheme 6.

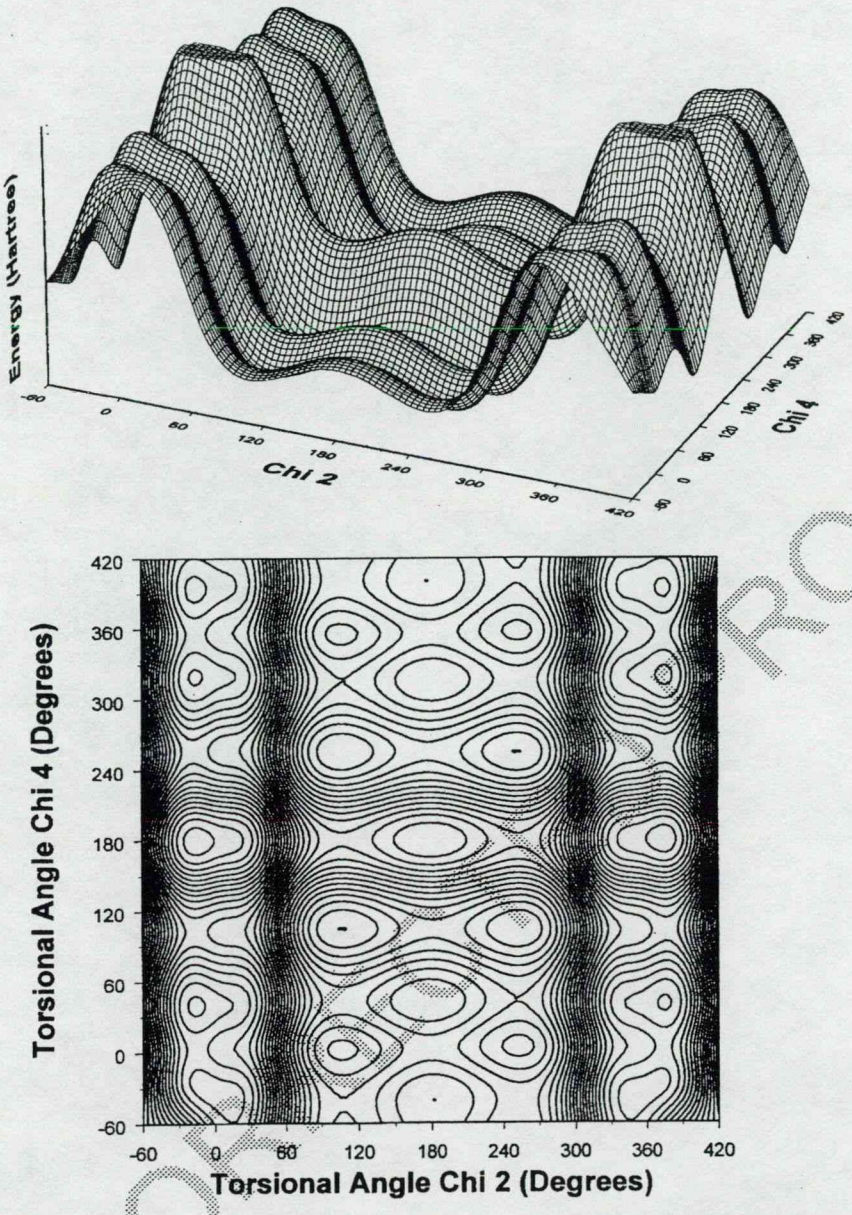


Fig. 3. All-*trans*-lycopene Model B PES, $E = f(\chi_2 \text{å} \chi_4)$ at χ_3 = *anti* position. Top: Landscape representation. Bottom: Contour representation.

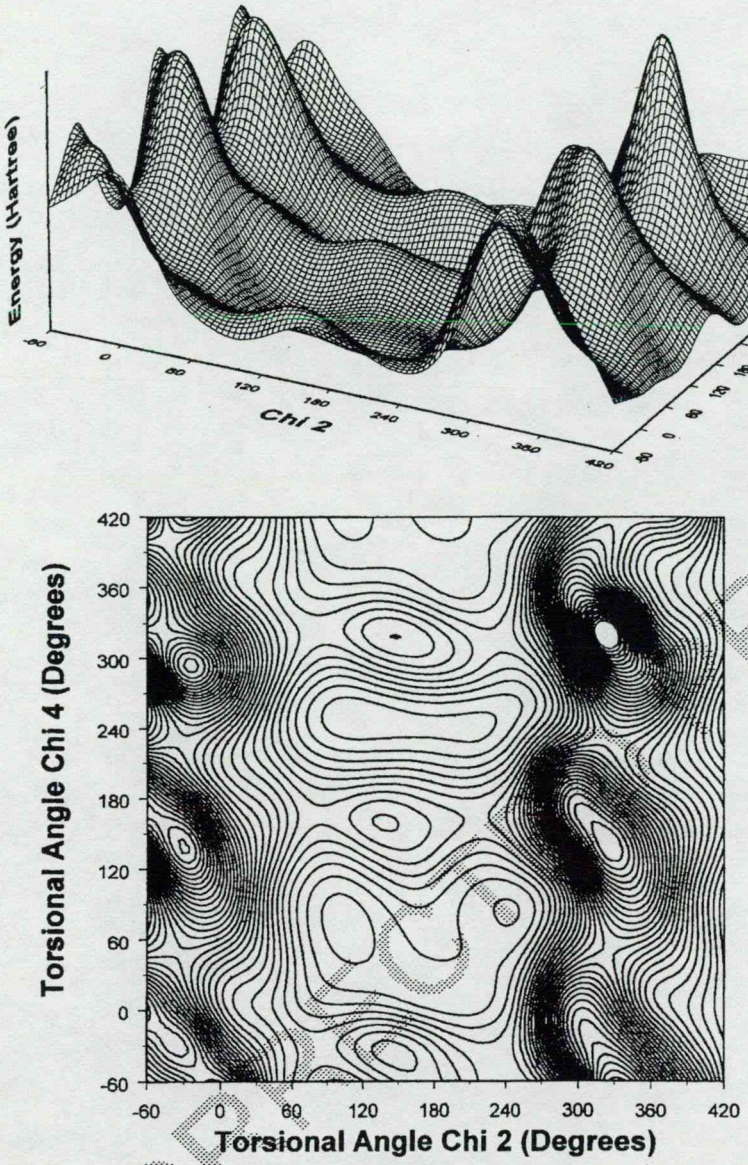


Fig. 4. All-*trans*-lycopene Model B PES, $E = f(\chi_2 \hat{a} \chi_4)$ at $\chi_3 = g^+$ position. Top: Landscape representation. Bottom: Contour representation.

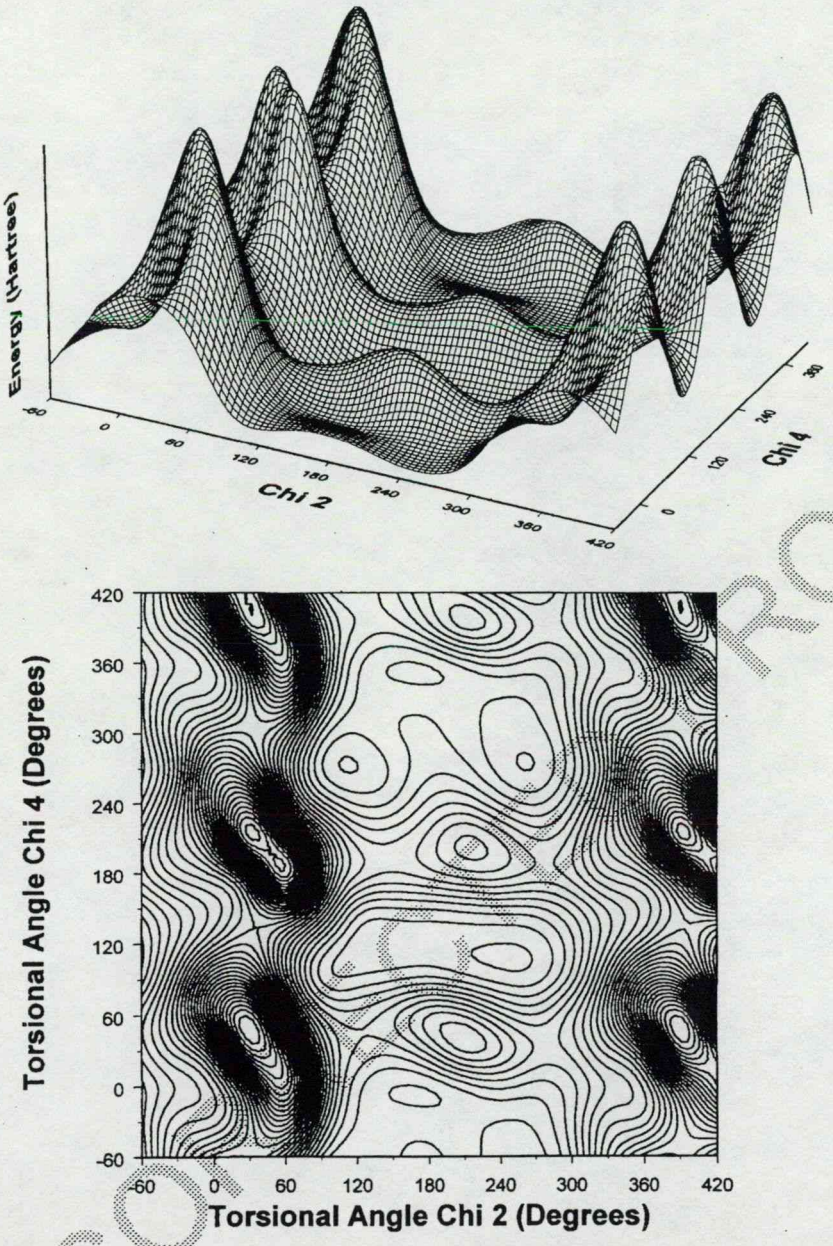


Fig. 5. All-*trans*-lycopene Model B PES, $E = f(\chi_2 \text{å} \chi_4)$ at $\chi_3 = g^-$ position. Top: Landscape representation. Bottom: Contour representation.

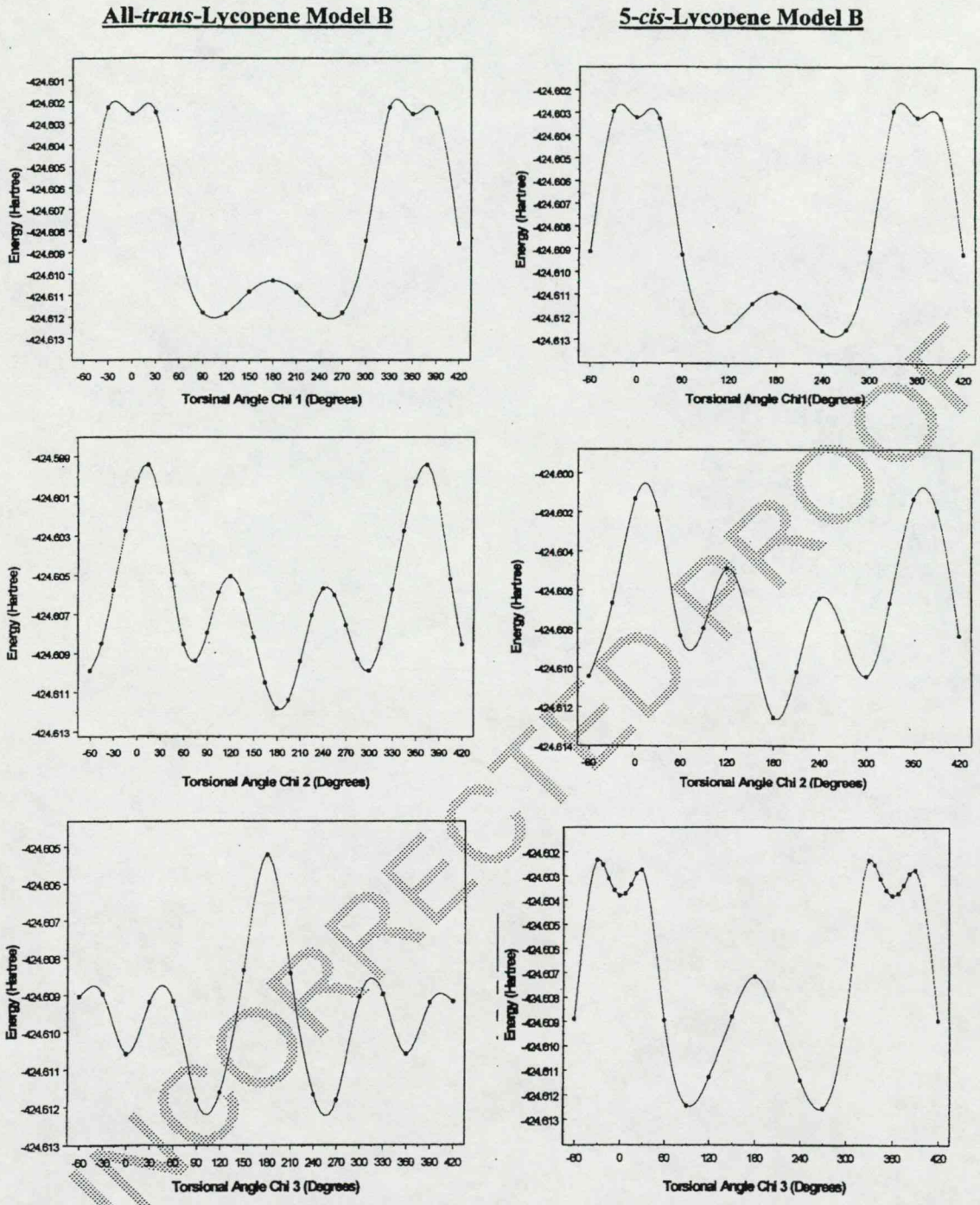


Fig. 6. Conformational PECs of lycopene tail-end Model B $E = g(\chi_2)$å$E = f(\chi_3)$å$E = f(\chi_4)$åaccording to Eq. (3a)–(3c), respectively. Left-hand side: all-*trans*-isomers. Right-hand side: 5-*cis*-isomers.

Table 1
Torsional angles of the optimized all-*trans* lycopene Model B conformers

Conformer			χ_1	χ_2	χ_3	χ_4	χ_5	χ_6	χ_7	χ_8	χ_9	χ_{10}
χ_2	χ_3	χ_4										
s	g^+	s	175.825	**8.219**	**73.456**	**−2.696**	178.530	179.109	−180.118	−1.567	155.285	179.822
s	g^+	g^+			**NOT**	**FOUND**		GOES	TO	$g^+g^+g^+$		
s	g^+	g^-			**NOT**	**FOUND**		GOES	TO	$g^+g^+g^-$		
g^+	g^+	s	180.620	**132.366**	**70.240**	**13.795**	178.533	180.080	−179.996	0.201	178.837	185.729
g^+	g^+	g^+	180.030	**99.689**	**59.014**	**79.958**	180.859	180.113	−180.005	0.247	179.439	172.338
g^+	g^+	g^-	180.584	**117.884**	**66.164**	**253.403**	179.837	180.202	−180.009	0.358	179.405	179.754
g^-	g^+	s	180.784	**−167.031**	**68.838**	**−0.291**	179.071	179.299	−180.081	−0.315	−179.499	
g^-	g^+	g^+	181.444	**−114.775**	**69.896**	**84.896**	181.140	180.181	−179.979	0.441	185.744	183.034
g^-	g^+	g^-	178.865	**−105.966**	**74.625**	**−98.969**	179.045	180.560	−179.990	0.277	176.476	−173.688
s	a	s	179.961	**−1.072**	**180.693**	**−0.550**	180.039	179.983	−180.001	0.016	179.745	179.679
s	a	g^+			**NOT**	**FOUND**		GOES	TO	$g^+a\,g^+$		
s	a	g^-			**NOT**	**FOUND**		GOES	TO	$g^-a\,g^-$		
g^+	a	s	180.835	**103.795**	**177.115**	**0.315**	179.902	179.969	−179.998	−0.037	179.214	180.142
g^+	a	g^+	181.065	**106.356**	**176.434**	**100.248**	181.363	180.202	−179.961	0.069	179.791	176.165
g^+	a	g^-	181.088	**104.305**	**177.254**	**−101.338**	178.588	179.917	−180.049	0.018	179.623	181.733
g^-	a	s	179.149	**−103.829**	**182.815**	**−0.257**	180.099	180.038	−180.001	0.083	180.774	179.863
g^-	a	g^+	178.912	**−104.659**	**182.837**	**101.431**	181.412	180.089	−179.959	0.006	180.344	178.335
g^-	a	g^-	178.956	**253.526**	**183.549**	**259.703**	178.637	179.818	−180.037	−0.067	180.284	183.768
s	g^-	s	184.130	**−8.143**	**−73.563**	**2.816**	181.474	180.903	180.116	1.777	155.308	−179.569
s	g^-	g^+			**NOT**	**FOUND**		GOES	TO	$g^-g^-g^+$		
s	g^-	g^-			**NOT**	**FOUND**		GOES	TO	$g^+g^-g^-$		
g^+	g^-	s	179.087	**167.376**	**−68.605**	**−0.093**	180.981	180.528	−179.921	0.683	179.219	180.034
g^+	g^-	g^+	181.140	**105.930**	**−74.655**	**98.949**	180.966	179.429	−180.010	−0.234	183.711	173.668
g^+	g^-	g^-	178.624	**114.557**	**−69.915**	**−84.922**	178.840	179.807	−180.014	−0.341	174.794	183.126
g^-	g^-	s	179.388	**−131.690**	**−70.369**	**−13.840**	181.446	179.922	−179.997	−0.460	180.771	174.221
g^-	g^-	g^+	179.402	**−117.654**	**−66.185**	**106.482**	180.151	179.865	−179.996	−0.269	180.616	179.818
g^-	g^-	g^-	−180.001	**−99.757**	**−58.865**	**−79.906**	179.135	179.874	−179.997	−0.381	180.530	187.383

angles (χ_8, χ_9 and χ_{10}) measure the spatial orientations of the three methyl groups of **III**.

The four central minima shown in Scheme 6 represent the two most stable conformations. They are pairwise equivalent, but they all practically have the same stability as shown in terms of ΔE (kcal mol^{-1}) in Scheme 7. The relative energies, together with the computed dipole moments, are summarized in Table 2.

In addition to the nine structures in Scheme 6,

$g^+(a)g^-$ $g^-(a)g^-$ 0.071 0.000

$g^+(a)g^+$ $g^-(a)g^+$ 0.000 0.071

Scheme 7.

where the central letter in parentheses represents *anti* orientation, (*a*), along χ_3, there are two additional sets of 9 structures; one set with (g^+) and the other with (g^-).

It is interesting to compare Figs. 4 and 5. They are centrosymmetric to each other through the fully symmetric (*a*, *a*, *a*) focal point, denoted as a heavy dot in Fig. 2. For this reason, the two sides of the 1D cross-section (**3b**), corresponding to the central left-hand side PEC, in Fig. 6, along with the heavy vertical line in Fig. 2, are not completely symmetrical. The g^+ and g^- minima differ slightly because the two surfaces at $\chi_2 = \chi_4 = 90°$å are not identical.

The centrosymmetric arrangement can also be seen from Scheme 8.

The discrepancies (1.063 versus 1.062 and 1.301 versus 1.302) are the result of regular optimization,

Table 2
Dipole moments, total energies and relative energies of the optimized all-*trans* lycopene Model B conformers

Conformer			Dipole	Energy (hartree)	ΔE (kcal mol^{-1})
χ_2	χ_3	χ_4			
s	g^+	s	0.7430	−424.6006056	7.381
s	g^+	g^+	Not found	GOES TO $g^+g^+g^+$	N/A
s	g^+	g^-	Not found	GOES TO $g^+g^+g^-$	N/A
g^+	g^+	s	0.8374	−424.6089342	2.155
g^+	g^+	g^+	1.0844	−424.6102939	1.301
g^+	g^+	g^-	0.6797	−424.6118287	0.338
g^-	g^+	s	0.6463	−424.6093099	1.919
g^-	g^+	g^+	0.8069	−424.6106742	1.063
g^-	g^+	g^-	0.5608	−424.6110997	0.796
s	a	s	0.8259	−424.604468	4.957
s	a	g^+	Not found	GOES TO g^+ag^+	N/A
s	a	g^-	Not found	GOES TO g^-ag^-	N/A
g^+	a	s	0.8872	−424.6108053	0.9805
g^+	**a**	g^+	**0.6568**	**−424.6123678**	**0.000**
g^+	*a*	g^-	*0.9860*	*−424.6122554*	*0.071*
g^-	a	s	0.8869	−424.6108055	0.980
g^-	*a*	g^+	*0.9861*	*−424.6122554*	*0.071*
g^-	**a**	g^-	**0.6571**	**−424.6123677**	**0.000**
s	g^-	s	0.7423	−424.6006059	7.3807
s	g^-	g^+	Not found	GOES TO $g^-g^-g^+$	N/A
s	g^-	g^-	Not found	GOES TO $g^+g^-g^-$	N/A
g^+	g^-	s	0.6492	−424.6093103	1.919
g^+	g^-	g^+	0.5602	−424.6110997	0.796
g^+	g^-	g^-	0.8070	−424.6106747	1.062
g^-	g^-	s	0.8412	−424.6089341	2.155
g^-	g^-	g^+	0.6794	−424.6118286	0.338
G^-	g^-	g^-	1.0843	−424.6102937	1.302

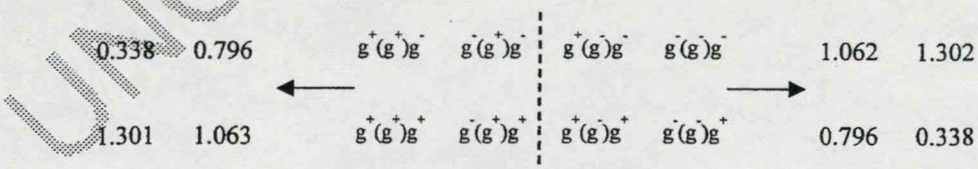
which are expected to disappear when the optimizations are performed to a tight convergence threshold.

4.1.2. The 5-cis-structure

The conformational PEHS (1) of three independent variables, shown schematically in Fig. 2 for the all-*trans*-isomer, is also valid for the 5-*cis*-isomer. In the case of the 5-*cis*-isomer, just as before, three PESs of two independent variables were generated for the $\chi_2 = aa$ g^+ and g^- orientations. These are shown in Figs. 7–9, respectively. The 1D-scans produced three PECs (shown at the right-hand side of Fig. 6, Table 3).

The centrosymmetric arrangement can be seen, analogously to the all-*trans*-form, in the case for $\chi_2 =$ *anti*à in Scheme 9.

Similarly, the centrosymmetric character of the PEHS can be further demonstrated by comparing energies for the $\chi_3 = g^+$ and $\chi_3 = g^-$ cases. This is illustrated by Scheme 10.



Scheme 8.

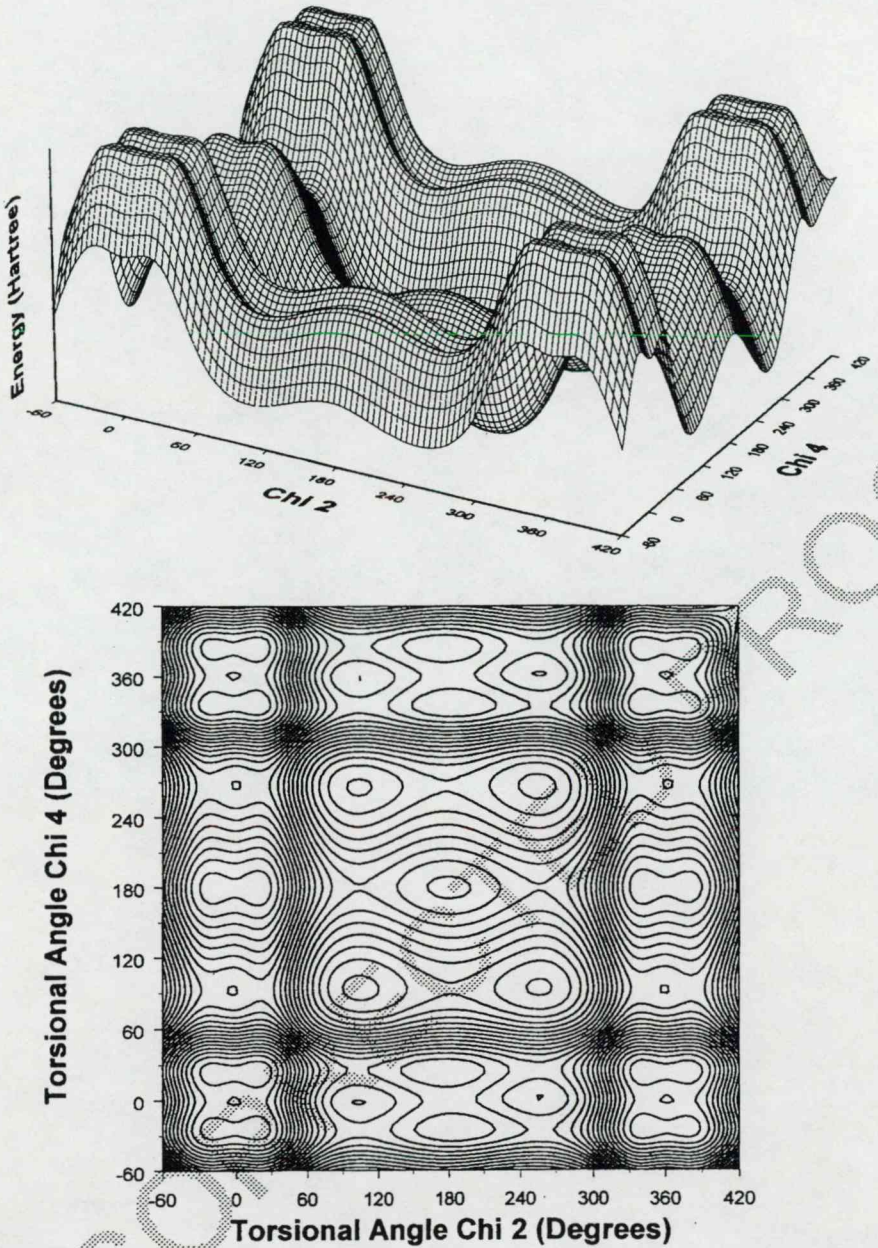


Fig. 7. The 5-*cis*-lycopene Model B PES, $E = f(\chi_2 \hat{a} \chi_4)$ at χ_3 = a conformation. Top: Landscape representation. Bottom: Contour representation.

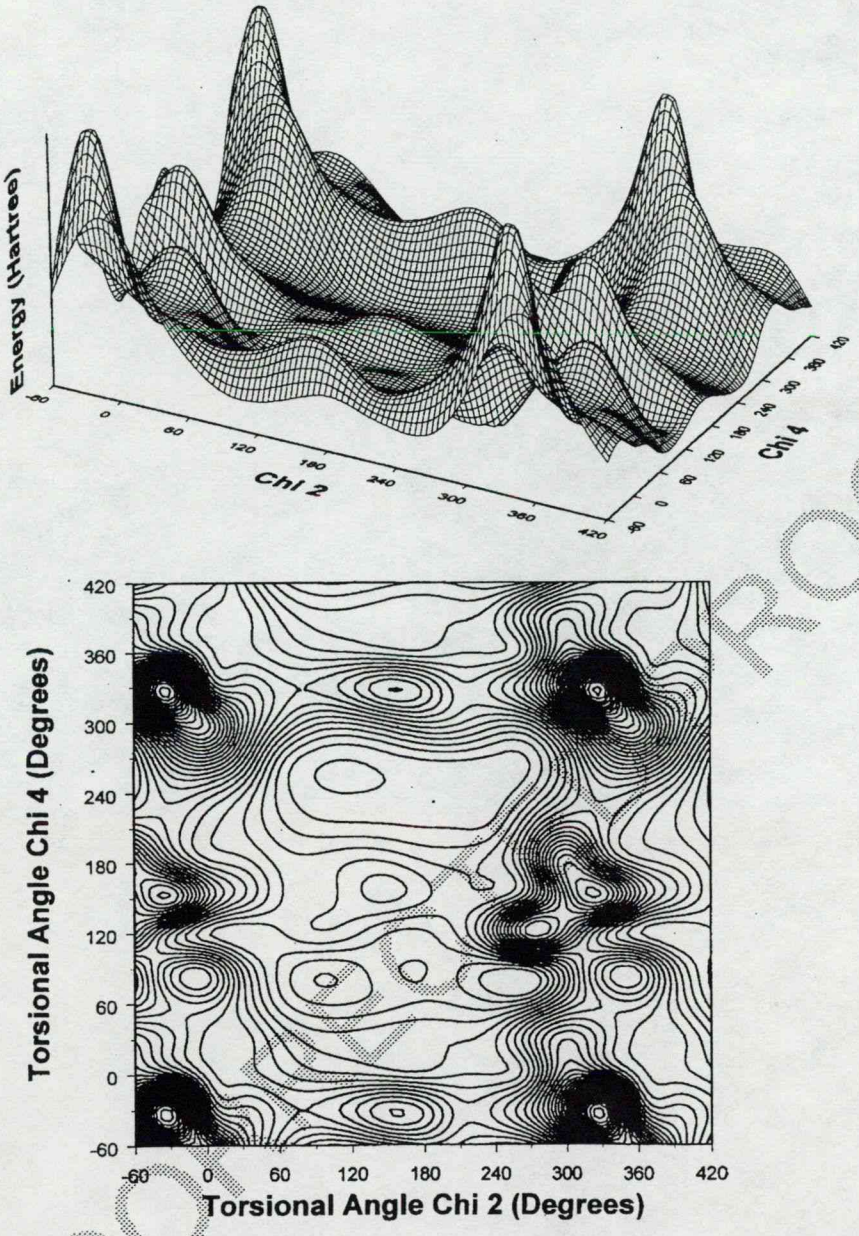


Fig. 8. The 5-*cis*-lycopene Model B PES, $E = f(\chi_2 \text{å} \chi_4)$ at $\chi_3 = g^+$ conformation. Top: Landscape representation. Bottom: Contour representation.

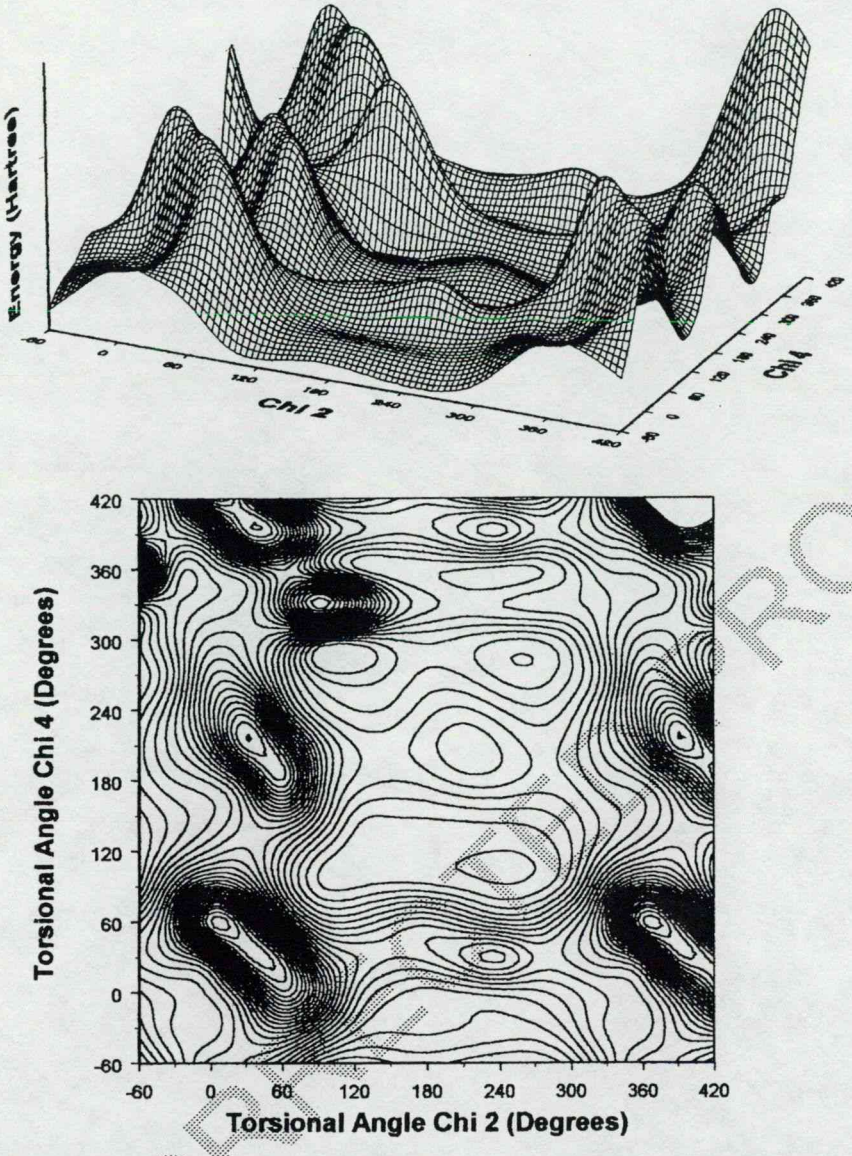


Fig. 9. The 5-*cis*-lycopene Model B PES, $E = f(\chi_2 \hat{a} \chi_4)$ at $\chi_3 = g^-$ conformation. Top: Landscape representation. Bottom: Contour representation.

Table 3
Torsional angles of the optimized 5-*cis* lycopene Model B conformers

Conformer			χ_1	χ_2	χ_3	χ_4	χ_5	χ_6	χ_7	χ_8	χ_9	χ_{10}
χ_2	χ_3	χ_4										
s	g^+	s	176.805	**−9.080**	**88.036**	**−8.820**	−4.471	170.478	−181.311	−1.685	163.845	178.980
s	g^+	g^+			**NOT**	**FOUND**		GOES	TO	s g^+s		
s	g^+	g^-			**NOT**	**FOUND**		GOES	TO	g^-a g^-		
g^+	g^+	s	180.452	**170.359**	**67.397**	**8.997**	−3.028	170.953	−180.864	−0.132	179.930	182.914
g^+	g^+	g^+	180.041	**99.783**	**58.797**	**79.371**	1.400	182.284	−179.580	0.255	179.460	175.630
g^+	g^+	g^-	180.717	**115.906**	**67.994**	**258.053**	0.596	181.810	−179.796	1.003	179.588	181.870
g^-	g^+	s			**NOT**	**FOUND**		GOES	TO	g^-a g^-		
g^-	g^+	g^+	181.120	**247.018**	**68.508**	**80.618**	0.963	182.754	−179.686	0.296	185.079	178.317
g^-	g^+	g^-	178.111	**−105.361**	**75.434**	**−100.127**	−0.038	181.286	−180.039	−2.349	−182.219	−177.214
s	a	s			**NOT**	**FOUND**		GOES	TO	g^+a g^+		
s	a	g^+			**NOT**	**FOUND**		GOES	TO	g^+a g^+		
s	a	g^-			**NOT**	**FOUND**		GOES	TO	g^+a g^+		
g^+	a	s	180.860	**102.780**	**174.547**	**4.853**	0.122	180.257	−179.872	−0.077	179.099	181.457
g^+	a	g^+	181.099	**104.375**	**175.566**	**91.422**	1.315	181.324	−179.966	−0.001	179.400	176.023
g^+	a	g^-	181.052	**103.802**	**179.675**	**−91.919**	−1.255	179.048	−180.005	0.009	179.441	183.829
g^-	a	s	179.155	**−102.865**	**185.316**	**−4.575**	−0.114	179.658	−180.125	0.129	181.131	178.668
g^-	a	g^+	178.939	**−103.962**	**180.368**	**91.951**	1.255	180.915	−179.994	0.014	180.608	176.131
g^-	a	g^-	178.916	**−104.231**	**184.418**	**−91.459**	−1.320	178.750	−180.030	0.021	180.516	184.011
s	g^-	s	183.218	**9.068**	**−88.104**	**8.891**	4.444	189.461	181.309	1.637	196.198	181.053
s	g^-	g^+			**NOT**	**FOUND**		GOES	TO	g^- g^- g^+		
s	g^-	g^-			**NOT**	**FOUND**		GOES	TO	g^- g^- g^-		
g^+	g^-	s			**NOT**	**FOUND**		GOES	TO	g^- g^- g^+		
g^+	g^-	g^+	181.885	**105.325**	**−75.373**	**100.166**	0.061	178.649	−179.960	−2.389	181.962	176.975
g^+	g^-	g^-	178.875	**113.110**	**−68.478**	**−80.610**	−0.968	177.325	−180.320	−0.360	174.822	181.661
g^-	g^-	s	179.535	**−170.224**	**−67.490**	**−8.822**	3.006	188.995	−179.135	0.295	180.655	177.220
g^-	g^-	g^+	179.289	**−115.815**	**−68.005**	**101.908**	−0.598	178.149	−180.201	−0.988	180.411	178.083
g^-	g^-	g^-	179.946	**−99.765**	**−58.836**	**−79.395**	−1.406	177.812	−180.418	−0.159	180.651	184.473

The optimized torsional angles are summarized in Table 4 and the computed properties are given in Table 4.

4.2. Molecular configurations

It is of considerable interest to optimize the fully symmetrical *a,a,a* ($\chi_2 = \chi_3 = \chi_4 = 180°$) conformation. This is done for both the all-*trans* and the 5-*cis* structures, in two steps. First, these three dihedral angles were kept frozen and then subsequently relaxed. In addition to the optimized torsional angles (Table 5), the energies and the imaginary frequencies (Table 6) were also tabulated. These fully symmetric points (marked as a solid dot at the centre of Fig. 2) are second order saddle points. All these indicate that neither the all-*trans*, nor the 5-*cis* isomers are ever fully symmetric. Consequently, neither the all-*trans* nor the 5-*cis* isomers are planar.

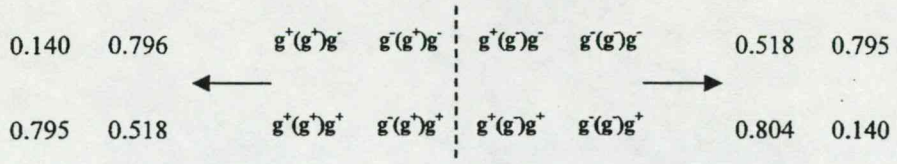
$g^+(a)g^-$ $g^-(a)g$ -0.157 -0.314

$g^+(a)g^+$ $g^-(a)g^+$ -0.314 -0.157

Scheme 9.

It is interesting to compare the relative stability of the all-*trans*- and the 5-*cis*-isomers. A graphical comparison of the relative energies given in Tables 2 and 4, is shown in Fig. 10. Intuitively, one would have guessed that all 5-*cis*-conformers have to be of higher energy than their corresponding all-*trans*-conformers. However, this is not the case.

Only four conformers of the 5-*cis*-isomer (sg^+s,



Scheme 10.

Table 4
Dipole moments, total energies and relative energies of the optimized 5-*cis* lycopene Model B conformers

Conformer			Dipole	Energy (hartree)	ΔE (kcal mol^{-1})
χ_2	χ_3	χ_4			
s	g^+	s	0.8089	− 424.5945699	11.168
s	g^+	g^+	NOT FOUND	GOES TO s g^+s	N/A
s	g^+	g^-	NOT FOUND	GOES TO g^-a g^-	N/A
g^+	g^+	s	0.8657	− 424.6023298	6.2989
g^+	g^+	g^+	1.0293	− 424.6111013	0.7947
g^+	g^+	g^-	0.7313	− 424.6121441	0.1404
g^-	g^+	s	NOT FOUND	GOES TO g^+a g^+	N/A
g^-	g^+	g^+	0.7078	− 424.6115428	0.5177
g^-	g^+	g^-	0.5608	− 424.6110997	0.7957
s	a	s	NOT FOUND	GOES TO g^+a g^+	N/A
s	a	g^+	NOT FOUND	GOES TO g^+a g^+	N/A
s	a	g^-	NOT FOUND	GOES TO g^- g^- g^+	N/A
g^+	a	s	0.9719	− 424.6040333	5.2300
g^+	**a**	**g^+**	**0.6513**	**− 424.6128689**	**− 0.3140**
g^+	**a**	**g^-**	**0.9975**	**− 424.6126178**	**− 0.1570**
g^-	a	s	0.9716	− 424.6040331	5.2301
g^-	**a**	**g^+**	**0.9976**	**− 424.6126178**	**− 0.1570**
g^-	**a**	**g^-**	**0.6515**	**− 424.6128689**	**− 0.3140**
s	g^-	s	0.8091	− 424.5945699	11.1680
s	g^-	g^+	NOT FOUND	GOES TO g^- g^- g^+	N/A
s	g^-	g^-	NOT FOUND	GOES TO g^- g^- g^-	N/A
g^+	g^-	s	NOT FOUND	GOES TO g^- g^- g^+	N/A
g^+	g^-	g^+	0.6145	− 424.6110874	0.8035
g^+	g^-	g^-	0.7079	− 424.6115427	0.5178
g^-	g^-	s	0.8654	− 424.6023291	6.2994
g^-	g^-	g^+	0.7314	− 424.6121441	0.1404
g^-	g^-	g^-	1.0296	− 424.6111013	0.7947

Table 5
Torsional angles of the symmetric second order TS lycopene Model B isomers

Conformer	χ_1	χ_2	χ_3	χ_4	χ_5	χ_6	χ_7	χ_8	χ_9	χ_{10}
Trans	179.973	**180.000**	**180.000**	**180.000**	180.000	179.996	180.000	0.181	180.124	180.121
5-Cis	180.001	**180.000**	**180.000**	**180.000**	−0.0008	180.001	−179.999	0.025	180.005	179.981

Table 6
Imaginary frequencies, dipole moments, total energies and relative energies of the symmetric second order TS of lycopene Model B isomers

Conformer	Imaginary frequencies (cm^{-1})		Dipole	Energy (hartree)	ΔE (kcal mol^{-1})
Trans	115.5i	51.0i	1.1343	−424.6036305	5.4827
5-Cis	77.1i	19.6i	1.0753	−424.6055073	4.305

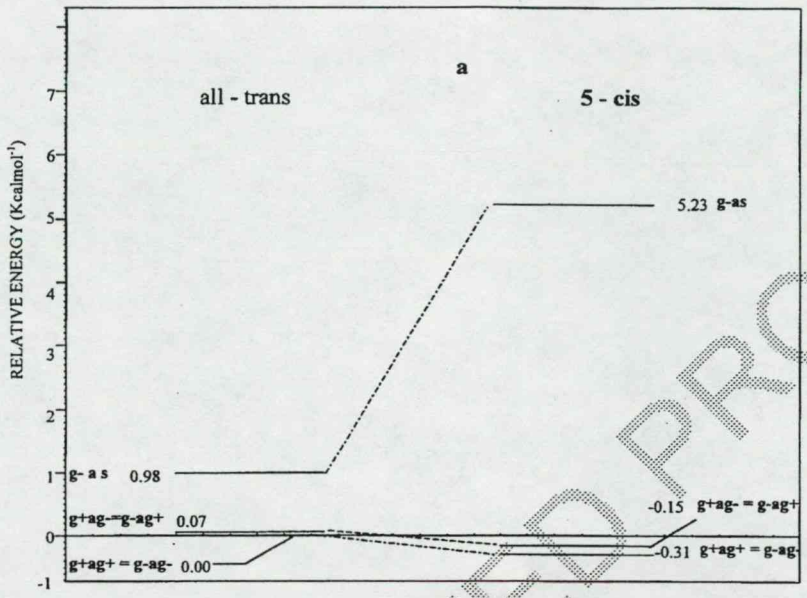


Fig. 10. Relative energy of the all-*trans* and 5-*cis*-isomers of lycopene Model B at the χ_3 = a conformation.

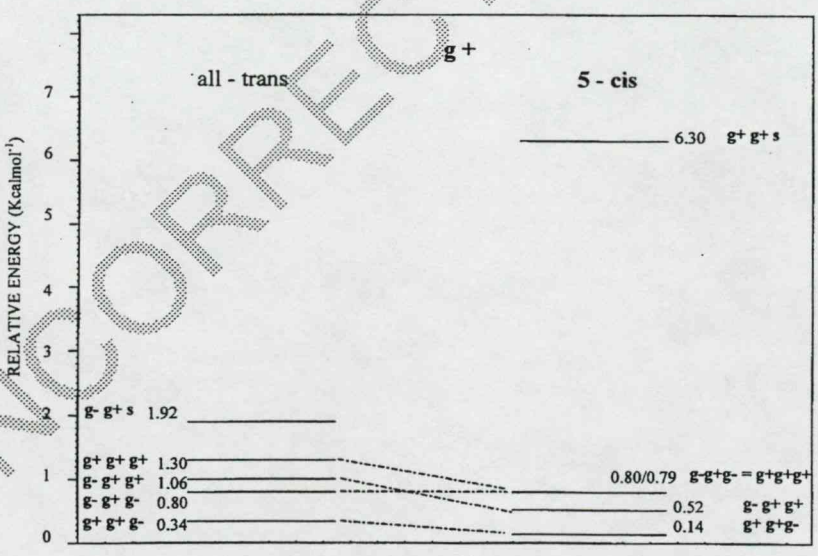


Fig. 11. Relative energy of the all-*trans* and 5-*cis*-isomers of lycopene Model B at the $\chi_3 = g^-$ conformation.

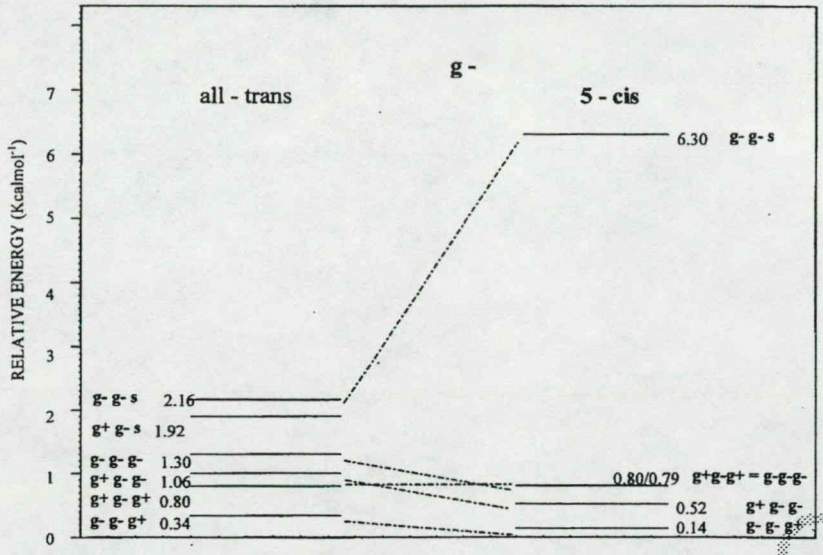


Fig. 12. Relative energy of the all-*trans* and 5-*cis*-isomers of lycopene Model B at the $\chi_3 = g^-$ conformation.

g^-g^+s, g^-as, g^-g^-s) had higher energy than the corresponding all-*trans* isomer. Two conformers ($g^-g^+g^-$ and $g^+g^-g^+$) had practically the same energies, while in the case of the other conformers, the 5-*cis*-isomers were slightly more stable, than their corresponding all-*trans*-isomers (Figs. 11 and 12).

5. Conclusions

Epidemiological studies have supported the hypothesis that consumption of heat processed tomatoes, such as in the Mediterranean diet, may reduce the risk of coronary heart disease by preventing the oxidation of the low-density lipoprotein [1,2]. Giovannucci et al. have also suggested that only the intake of processed tomato products was related to reduced risk of prostate cancer probably because of their high *cis* isomer content of lycopene [21]. The observation that high concentration of *cis* isomers are present in human serum and prostate tissue, also suggests that *cis* isomers might be biologically more active than the all-*trans* isomer.

In contrast to previous results [22], according to which the *cis*-isomer is less stable than the corresponding *trans*-isomer, the present study suggests at least the 5-*cis*-isomer is more stable than the all-*trans*-isomer. Such stability may be due to the favourable Type A′ 1,4-interaction (Scheme 4). The conformational study revealed that the fully planar structure of lycopene Model B, is a second-order saddle point. Such fully symmetric (i.e. planar) structure of lycopene would be expected to become a fourth-order critical point on its PEHS, because there are two negative eigenvalues at both ends (χ_2, χ_4 as well as $\chi_{2'}$, $\chi_{4'}$).

Acknowledgements

One of the authors (GAC) wishes to thank Graydon Hoare (graydon@pobox.com) for database management, network support, software and distributive processing development. A special thanks is also 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 computations.

References

[1] A.V. Rao, S. Agarwal, Nutr. Res. 19 (1999) 305–323.
[2] S.K. Clinton, Nutr. Res. 56 (1998) 35–51.
[3] E. Giovannucci, J. Natl Cancer Ins. 91 (1999) 317–331.

[4] R. Willstätter, H.H. Escher, Z. Physiol. Chem. 64 (1910) 47–61.
[5] P. Karrer, R.P. Widmer, Helv. Chim. Acta 11 (1928) 751–752.
[6] P. Karrer, A. Helfenstein, R.P. Widmer, Helv. Chim. Acta 11 (1928) 1201–1209.
[7] P. Karrer, W.E. Bachmann, Helv. Chim. Acta 12 (1929) 285–291.
[8] P. Karrer, A. Helfenstein, H. Wehrli, A.P. Wettstein, Helv. Chim. Acta. 13 (1930) 1084–1099.
[9] L. Zechmeister, A.L. LeRosen, W.A. Schroeder, A. Polgar, L. Pauling, J. Am. Chem. Soc. 65 (1943) 1940–1955.
[10] L. Zechmeister, P. Tuzson, Nature 141 (1938) 249–250.
[11] A. Polgar, L. Zechmeister, J. Am. Chem. Soc. 64 (1942) 1856–1861.
[12] L. Zechmeister, A. Polgar, J. Am. Chem. Soc. 65 (1943) 1522–1528.
[13] L. Zechmeister, R.B. Escue, J. Am. Chem. Soc. 66 (1944) 322–330.
[14] L. Pauling, Fortschr. Chem. Organ. Naturstoffe. 3 (1939) 203–235.
[15] C. Sterling, Acta Crystallogr. 17 (1964) 1224–1228.
[16] M.O. Senge, H. Hope, K.M. Smith, Z. Naturforsch. Teil C. 47 (1992) 474–480.
[17] M.L. Nguyen, S.J. Schwartz, Food Technol. 53 (1999) 38–45.
[18] S.K. Clinton, C. Emenhiser, S.J. Schwartz, D.G. Bostwick, A.W. Williams, B.J. Moore, J.W. Erdman Jr., Cancer Epidemiol Biomarkers Prev. 5 (1996) 823–833.
[19] Ö. Farkas, S.J. Salpietro, P. Császár, I.G. Csizmadia, J. Mol. Struct. (Theochem) 367 (1996) 25–31.
[20] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez, J.A. Pople, Gaussian, Inc., Pittsburgh PA, 1995.
[21] E. Giovannucci, A. Ascherio, E.B. Rimm, M.J. Stampfer, G.A. Colditz, W.C. Willett, J. Natl. Cancer Inst. 87 (1995) 1767–1776.
[22] M.A. Berg, G.A. Chasse, E. Deretey, A.K. Füzéry, B.M. Fung, D.Y.K. Fung, H. Henry-Riyad, A.C. Lin, M.L. Mak, A. Mantas, M. Patel, I.V. Repyakh, M. Staikova, S.J. Salpietro, Ting-Hua Tang, J.C. Vank, András Perczel, Ödön Farkas, Ladislaus L. Torday, Zoltán Székely, Imre G. Csizmadia, J. Mol. Struct. (Theochem) 500 (2000) 5–58.

THEOCH7138

ELSEVIER

Journal of Molecular Structure (Theochem) 000 (2001) 000–000

THEO CHEM

www.elsevier.nl/locate/theochem

An ab initio computational study on selected lycopene isomers

Gregory A. Chasse[a,b,*], Melody L. Mak[b], Eugen Deretey[b], Imre Farkas[b], Ladislaus L. Torday[c], Julius G. Papp[c], Dittakavi S.R. Sarma[d], Anita Agarwal[e], Sujatha Chakravarthi[e], Sanjiv Agarwal[e], A. Venket Rao[e]

[a]*Velocet Communications Inc., 210 Dundas St. W., Suite 800, Toronto, Ont., Canada M5G 2E8*
[b]*Department of Chemistry, University of Toronto, Toronto, Ont., Canada M5S 3H6*
[c]*Department of Pharmacology and Pharmacotherapy, Albert Szent-Gyorgy Medical and Pharmaceutical Center, University of Szeged, H 6701 Szeged, Hungary*
[d]*Laboratory of Medicine and Pathobiology, University of Toronto, Toronto, Ont., Canada M5S 1A8*
[e]*Department of Nutritional Science, University of Toronto, Toronto, Ont., Canada M5S 3E2*

Received 15 December 2000; accepted 12 February 2001

Abstract

Lycopene is an effective antioxidant in vivo. Although lycopene is present in its all-*trans* isomeric form in fruits and vegetables, serum and tissue samples show a predominance of various *cis*-isomers of lycopene. The present study was undertaken to investigate the molecular structure of several *cis*-isomers of lycopene using an ab initio molecular modeling procedure. The relative stability of selected *cis*-isomers of lycopene with respect to the all-*trans* isomer, was studied. The following sequence of stability was observed.

5-*cis* > all-*trans* > 9-*cis* > 13-*cis* > 15-*cis* > 7-*cis* > 11-*cis*

The first four of these isomers had relative energies within ±1 kcal mol^{-1}, but the fifth isomer (i.e. 15-*cis*) was within 3 kcal mol^{-1}. However, the last two isomers were less stable than the all-*trans*-isomer with more than 5 kcal mol^{-1}, difference in energy. The optimized molecular conformational study structures indicated that the central conjugated part of every lycopene isomer is planar. However, the two tail ends of the molecules, each containing three single C–C bonds, avoid coplanarity.

Keywords: All *trans*-lycopene; Selected *cis*-isomers of lycopene; Thermodynamic stability of lycopene isomers; Computations of lycopene isomers; Ab initio MO theory

1. Introduction

1.1. Biological background

Oxidative stress is now recognized as an important etiological factor in the causation of several chronic diseases, amongst which are cancer and cardiovascular diseases. Antioxidants are effective in mitigating the damaging effect of free radicals. Lycopene is a carotenoid antioxidant, present in tomatoes and other fruits and vegetables. Recent studies have shown that it acts as an antioxidant in vivo, providing protection against the oxidation of lipids, proteins and DNA [1,2]. A recently published paper [3] indicated a

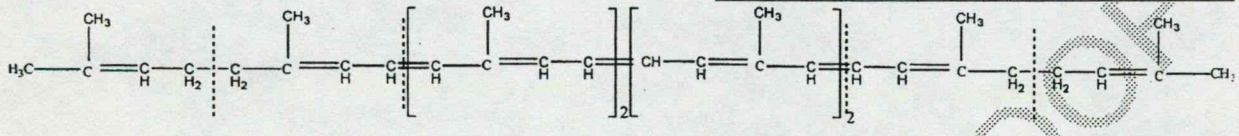
* Corresponding author. Tel.: +1-416-598-3229; fax: +1-416-598-7797.
E-mail address: gchasse@fixy.org (G.A. Chasse).

0166-1280/01/$ - see front matter
PII: S0166-1280(01)00424-9

significant inverse correlation between the intake of lycopene or serum concentrations of lycopene and the risk of several concerns. The isomeric forms of lycopene have been suggested to be important factors influencing its absorption and bioavailability.

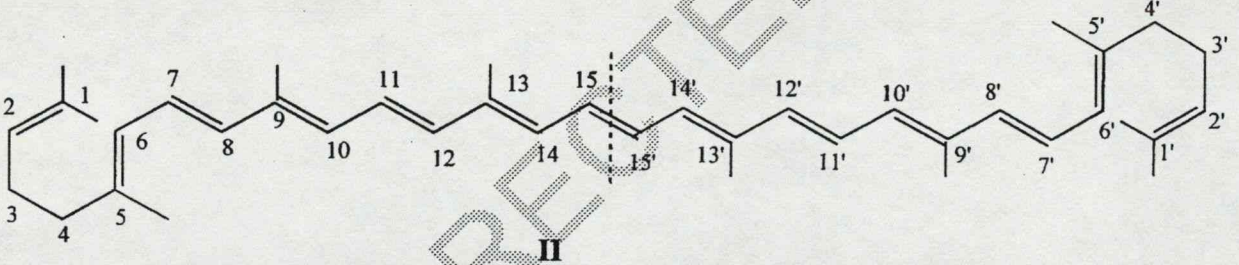
1.2. Structural background

Lycopene [$C_{40}H_{56}$] consists of eight isoprenic units, so its skeleton is related to tetraterpenes [$C_{40}H_{64}$], even though it contains fewer hydrogens. Its composition is shown in **I**.



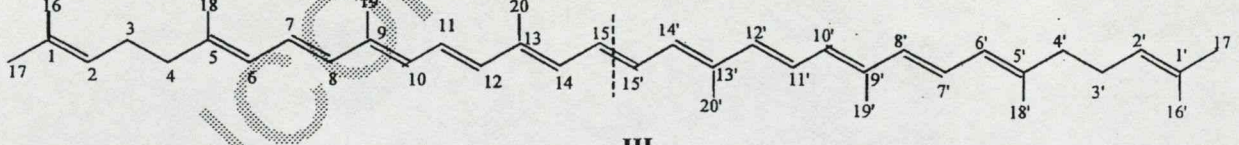
I

As such, lycopene is closely related to α-, β-, and γ-carotenes. Sometimes, the structure of lycopene is presented in a pre-folded form (**II**) to show its structural similarity to β-carotene. Due to the internal molecular symmetry, it has been traditional to number the chain from the two ends as shown in II.



II

The all-*trans* lycopene has also been presented, traditionally, in its fully extended form (**III**).



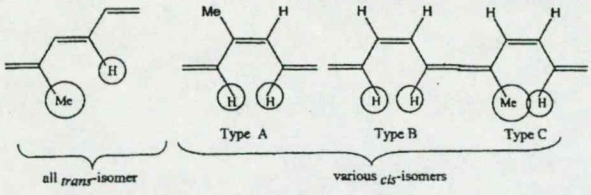
III

All of these representations imply a planar molecular form.

Isolation procedures for lycopene were first reported in 1910 [4] and its structure was determined during the period of 1928–1931 [5–8].

The spectral characteristics of stereoisomeric carotenoids, including lycopene, were studied [9] in 1943. The ease of isomerization of the $C_{40}H_{56}$ carotenes has been noticed by Zechmeister and Tuzson [10] as early as 1938. An isomerization equilibrium can only be reached very slowly but the process starts immediately after the crystalline carotene is dissolved. This indicates that the all-*trans*- and various *cis*-isomers are very close on the free-energy scale, but nevertheless require appreciable energy of activation for such a non-catalyzed inter-conversion. Subsequently, *trans* to *cis* isomerization of lycopene [11] and other compounds were studied using iodine as the catalyst. [11–13]. In 1943, Pauling pointed out [14] that *trans* to *cis* isomerization could be a result of the overlapping of the methyl group of a carbon atom adjacent to a double bond and the hydrogen. Not all *cis*-isomers may be of equal stability due to a number of possible 1,4 interactions as shown by the structures of Scheme 1. Clearly, on the basis of relative group sizes, the $-CH_3\cdots H-$ interaction (type C) appears to be the most destabilizing.

There are a number of possible geometrical



Scheme 1.

isomers. Zechmeister indicated [13] the possible presence of 20 β-carotene, 32 α-carotene, 64 γ-carotene and 72 lycopene geometrical isomers. No X-ray structure of lycopene has been determined as yet. There are, however, two X-ray structures of β-carotene in the literature [15,16].

2. Aim of study

Although lycopene is present in its all-*trans* isomeric form in fruits and vegetables, *cis*-isomers constitute the predominant form present in the serum and tissues [17,18]. The biological significance of lycopene isomerization is not well understood. The aim of our study was to investigate the structure and stability of different isomeric forms of lycopene using ab initio molecular modeling, and to gain a better understanding of lycopene isomerization requirements. In addition to geometrical *cis/trans* isomerism, the conformations of the two tails of the molecule were also analyzed.

3. Method

This computational study involved semi-empirical (AM1) geometry optimization, which was followed by an ab initio study carried out at the HF/3-21G level of theory.

The computations were carried out using the GAUSSIAN 94 program system [19]. Standard geometry optimizations were performed on the all-*trans*- as well as six selected *cis*-isomers. The optimized molecular conformations were also compared to the results obtained from a lycopene model compound [20].

Fig. 1 shows the all-*trans*- and selected *cis*-isomers of lycopene. The all-*trans*- and the selected six *cis*-isomers (5-, 7-, 9-, 11-, 13- and 15-) were explored in their molecular entirety (containing 296 electrons and 96 atoms, which corresponds to 282 geometrical parameters to be optimized). Thus, lycopene is among the largest organic molecules being investigated using ab initio molecular computation with the current computational technology.

4. Results

The relative energies, which measure thermodynamic stabilities of the lycopene isomers, computed at the HF/3-21G level of theory, are summarized in Table 1. The expected stability, on the basis of molecular structures, are given in Table 2. The values of the torsional angles ($\tau_{1'\,2'\,3}$ as well as $\tau_{1',2',3}$) of the three pairs of single bonds (C^2–C^3, C^3–C^4, C^4–C^5 as well as $C^{2'}$–$C^{3'}$ $C^{3'}$–$C^{4'}$ $C^{4'}$–$C^{5'}$) are summarized in Table 3,

5. Discussion

5.1. Molecular configurations

The relative energies, which measure thermodynamic stabilities of lycopene isomers, are summarized in Table 1 and are depicted in Fig. 2. We notice four categories for *cis*-isomers, which are summarized in Table 2. The computed relative stabilities of the four categories of the *cis*-isomers, are listed below.

Type A′ involves the 5-*cis*-isomer. This is the only *cis*-isomer which is slightly more stable (within -0.5 kcal mol^{-1}) than the all-*trans* form.
Type A involves the 9-*cis* and 13-*cis*-isomers. These *cis*-isomers are slightly less stable (within 1.0 kcal mol^{-1}) than the all-*trans*-form.
Type B involves the 15-*cis*-isomer, which is only about 3 kcal mol^{-1} less stable than the all-*trans* form.
Type C involves the 7- and 11-*cis*-isomers. These *cis*-isomers are about 6 kcal mol^{-1} higher on the energy scale than the all-*trans*-form.

Cis-isomers of lycopene in *Types A′* and *A* and the *trans* lycopene have similar energies, since we are dealing with triply substituted double bonds as shown in Table 2.

Type A structures (see Fig. 1) are only slightly

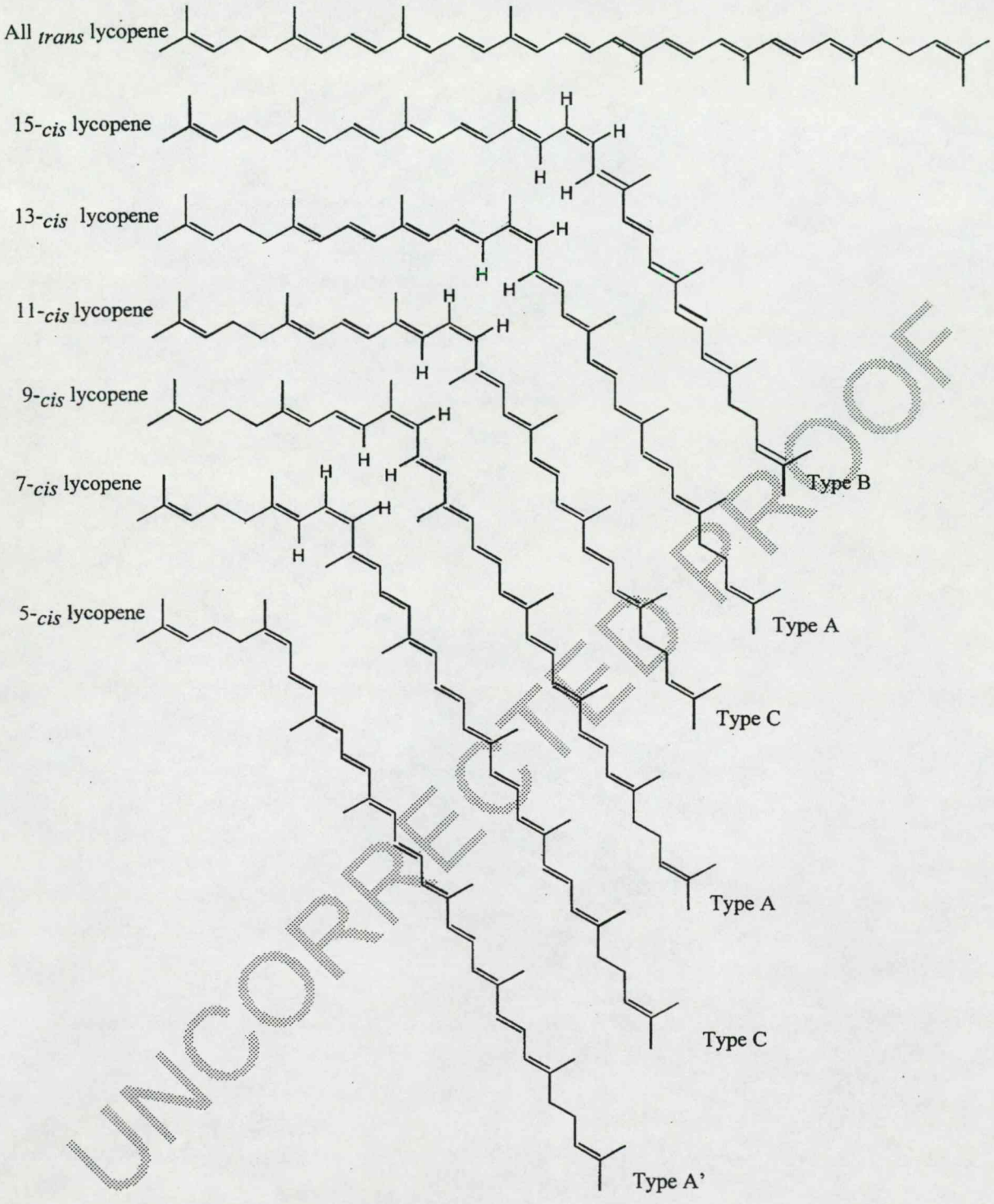


Fig. 1. Geometrical isomers of lycopene.

Table 1
Computed total energy and Relative Energy Values of Various Isomers of Lycopene, at HF/3-21G level of theory

Isomer	Energy (Hartree)	Relative energy (kcal mol^{-1})
All-*trans*	−1538.5731249	0.000
5-*Cis*	−1538.5737548	− (0.395)[a]
7-*Cis*	−1538.5636899	5.921
9-*Cis*	−1538.5722460	0.552
11-*Cis*	−1538.5634789	6.053
13-*Cis*	−1538.5721338	0.622
15-*Cis*	−1538.5693098	2.394

[a] Global minimum.

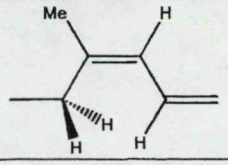
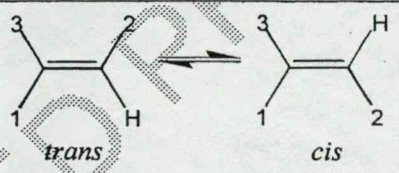
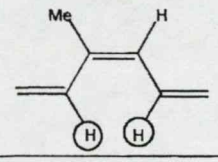
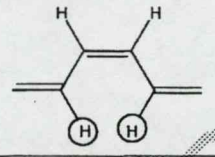
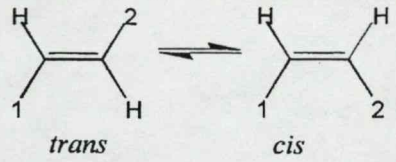
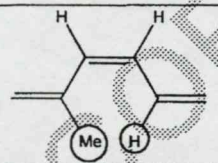
higher on the energy scale (see Fig. 2) than the all-*trans*-isomer. Type A′ is extremely similar to Type A, except that the hydrogens on the 1,4 positions (denoted as carbon 4 and 7) are not eclipsed but staggered. This is shown in Scheme 2.

Type B, there is only one example- the 15-*cis*-isomer. The 1,4 interaction involves the *cis* double bond which has only two alkyl substituents as shown in Table 2.

Type C has the two least stable structures, the 7-*cis* and 11-*cis*-isomers. In accordance with Fig. 2, the

Table 2
Classification of various *cis*-lycopene isomers according to 1,4-non-bonded interactions and extent of substitution about the carbon–carbon double bond involved. NB: Clearly, on the basis of relative group sizes, the $-CH_3\cdots H-$interaction (type C) appears to be the most destabilizing. The order of expected stability is the following: A′ > A > B > C

I

1,4-non bonded interaction		Isomers	Isomerisation with different degree of substitution
Type	*cis* structure		
A'		5-*cis*	
A		9-*cis* and 13-*cis*	
B		15-*cis*	
C		7-*cis* and 11-*cis*	

NB: Clearly, on the basis of relative group sizes , the $-CH_3$---H- interaction (type C) appears to be the most destabilizing. The order of expected stability is the following: A'>A>B>C.

Table 3
Optimized torsional angles and selected orbital energies for various isomers of lycopene, calculated at RHF/3-21G level of theory

Parameter[a]	Dihedral	all-*trans*	5-*cis*	7-*cis*	9-*cis*	11-*cis*	13-*cis*	15-*cis*
χ_1	D5	−106.596	103.040	−106.539	−106.605	−106.440	−106.409	−106.393
χ_2	D6	−176.424	176.517	−176.567	−176.479	−176.430	−176.524	−176.473
χ_3	D7	−100.364	90.063	−100.548	−100.368	−100.340	−100.393	−100.459
χ_1'	D31	106.368	106.440	106.515	106.460	106.394	106.474	106.475
χ_2'	D30	176.452	176.460	176.480	176.460	176.4637	176.504	176.456
χ_3'	D29	100.377	100.406	100.468	100.415	100.448	100.415	100.421
ϕ	D#	N/A	1.388	−5.854	−0.016	5.945	−0.0005	0.016
HOMO	–	−0.23712	−0.23320	−0.23387	−0.23347	−0.23508	−0.23422	−0.23459
LUMO	–	+0.06964	+0.05102	+0.05144	+0.05255	+0.06867	+0.05331	+0.05284
Δ	–	0.30676	0.28422	0.28531	0.28602	0.30375	0.28753	0.28743

[a] χ denotes the dihedral angle value for the single bonds, from the outsides inwards (∴χ_1 corresponds to the first single bond closest to the terminal end of the molecule). ϕ corresponds to the value of the dihedral angle about which the isomerisation has occured, for each particular *cis*-isomer. Thus, ϕ measures deviation from coplanarity of the *cis* double bond. HOMO is highest occupied MO energy, measuring the ionisation energy (IE) in terms of Koopman's theorem. LUMO is lowest unoccupied MO energy and Δ is the difference LUMO–HOMO in Hartree atomic units.

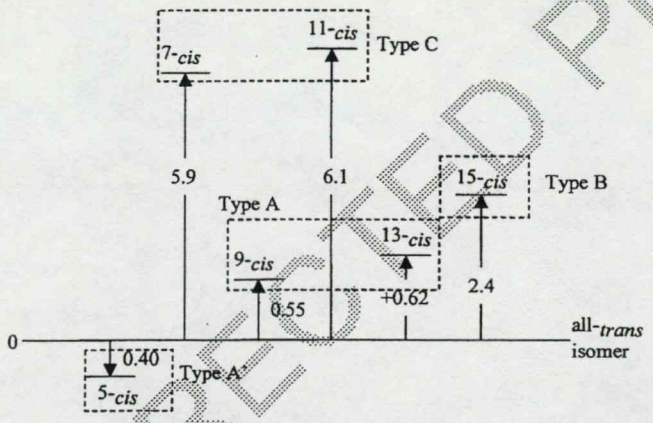
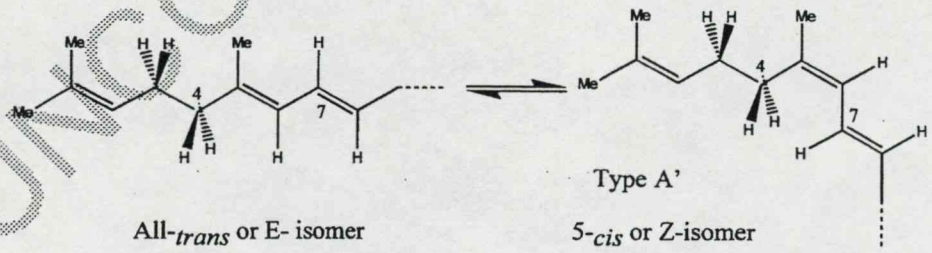
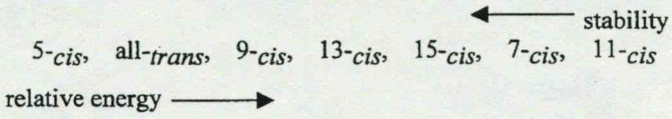


Fig. 2. A schematic illustration of the relative energies (kcal mol^{-1}) of the various *cis*-isomers of lycopene studied relative to the all-*trans* form.



Scheme 2.



Scheme 3.

relative energies measuring stabilities are therefore as follows in Scheme 3

In view of the foregoing, one only may wonder if the thermodynamic stability of the lycopene isomers, as well as the kinetic stability of the isomerization transition states, will predetermine the relative concentrations of these isomers in tissues and plasma.

5.2. Molecular conformations

When a planar moiety is rotated about a tetrahedral carbon, it may be either eclipsed with a tetrahedral bond or perpendicular to that bond. This has been revealed by the study on ethyl benzene [21].

In the case of lycopene isomers, in most cases, the olefinic moiety seems to favor a compromising position, having χ^1 and χ^3 in the range of 90–107°. Thus, it is either roughly perpendicular to the C^3–C^4 bond or near to a position of eclipsing with one of the two C–H bonds of the CH_2 moieties of C^3 and C^4. These conformers are shown in Scheme 4. The optimized values of the torsional angles are summarized in Table 3.

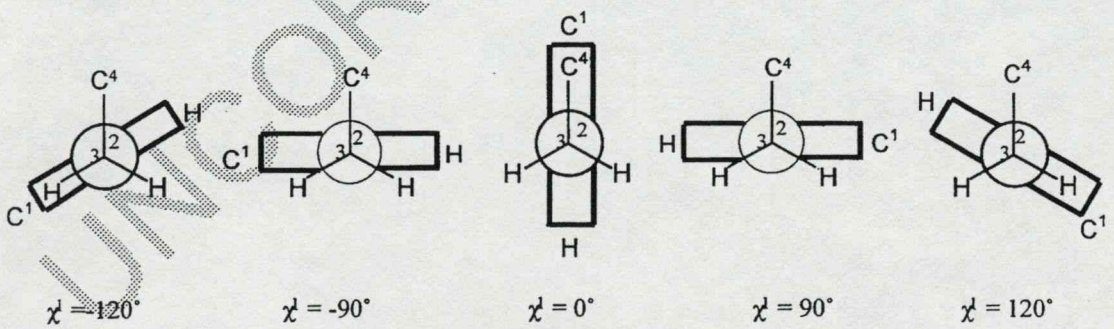
Fig. 3 shows the optimized structures of lycopene in the all-*trans* as well as in the selected *cis*-isomeric forms. The results (see Fig. 3 and Table 3) clearly indicate that while the conjugated double bonded segment of lycopene from C^5 to $C^{5'}$ is planar or nearly planar, the two terminal positions of the molecule (from C^1 to C^5 as well as from $C^{1'}$ to $C^{5'}$) are rotated out of plane.

5.3. Deviation from coplanarity

Sometimes the reactivity of an unsaturated compound is related to its not completely planar forms. This is particularly true for *cis*-double bonds. Deviation from coplanarity is within about 6° at the HF/3-21G level of theory as shown by the data in Table 3 by the entry ϕ. The deviation from coplanarity ($\Delta\phi$) at the corresponding double bonds in the all *trans*-form is considerably smaller as shown by the data given in Table 4. Such non-planarity implies a weakening of the π-bond, which in turn is related to reactivity. Such an enhanced reactivity is certainly related to *trans–cis* isomerization. However, it is not clear at this time if such structural differences may also imply enhanced radical scavenging ability.

6. Perspectives

Lycopene from natural plant sources exists predominantly in the all-*trans* configuration. For this reason, it may have been assumed to be the thermodynamically most stable form. However, all one can conclude is that the biosynthesis in the plants leads to



Scheme 4.

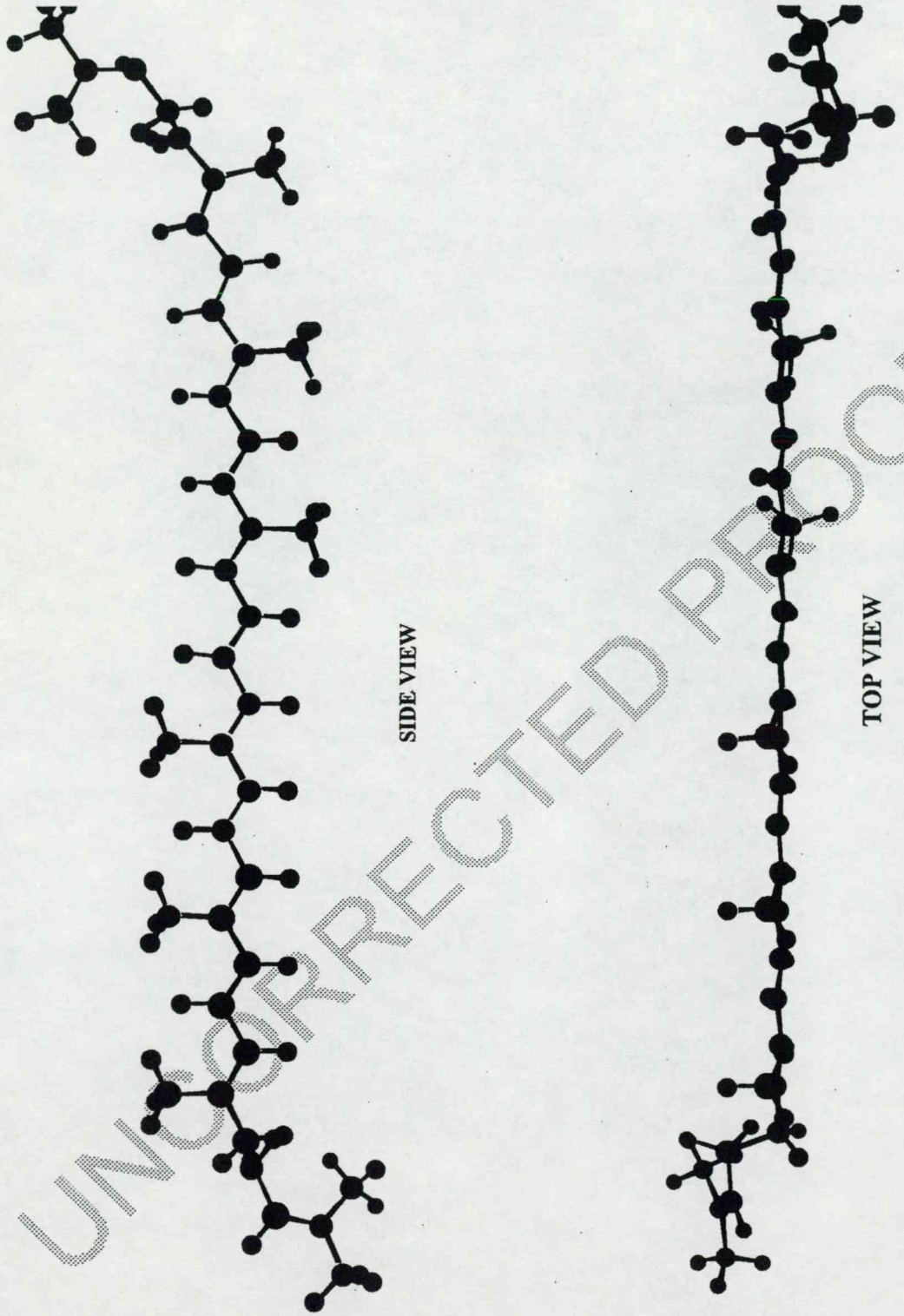


Fig. 3. (a) Side and top views of optimized all-*trans* lycopene Structure, (b) side and top views of optimized 5-*cis* lycopene structure.

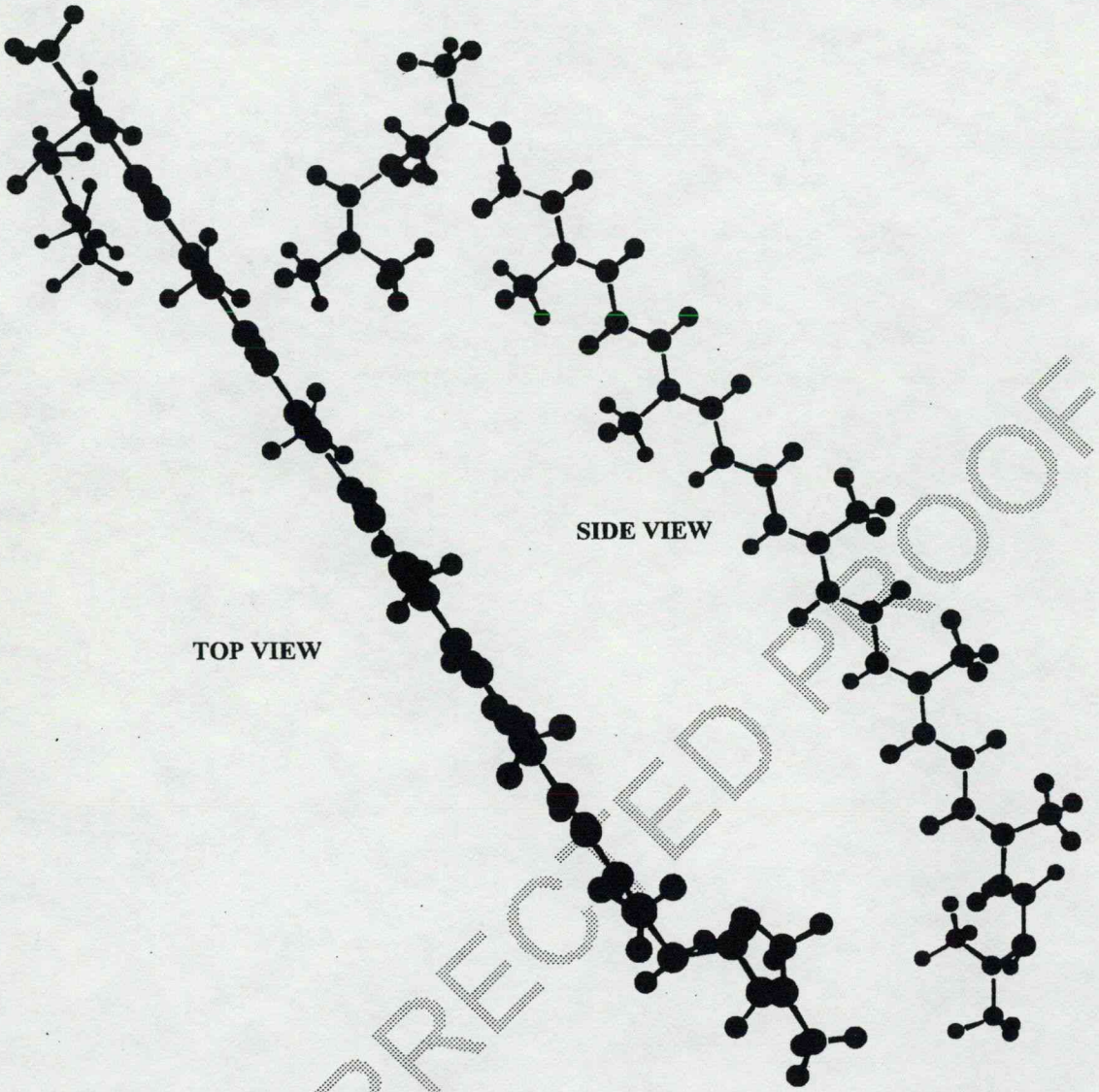


Fig. 3. (*continued*)

the all-*trans*-form and this is independent of its thermodynamic stability. In human plasma, lycopene is an isomeric mixture, containing at least 60% of the total lycopene as *cis*-isomers (Table 5). All-*trans*, 5-*cis*, 9-*cis*, 13-*cis* and 15-*cis* are the most commonly identified isomeric forms of lycopene [18]. The relative biological significance in terms of absorption and effectiveness of various geometric isomers of lycopene is unclear at present.

Epidemiological studies have supported the hypothesis that consumption of heat-processed tomatoes, such as in the Mediterranean diet, may reduce the risk of coronary heart disease by preventing the oxidation of the low-density lipoprotein [1,2]. Giovannucci et al. [22] have also suggested that only the intake of processed tomato products was related to reduced risk of prostrate cancer, probably because of their high content of lycopene

Table 4
Torsional angles for the *trans*-peptide bonds in the all-*trans* lycopene, computed at RHF/3-21G level of theory.

Bond	Dihedral (ϕ)	ϕ	$\Delta\phi$[a]
C^1–C^2	D4	178.955	1.045
C^5–C^6	D8	178.645	1.355
C^7–C^8	D10	179.943	1.057
C^9–C^{10}	D12	−179.999	−0.001
C^{11}–C^{12}	D14	179.999	0.001
C^{13}–C^{14}	D16	179.998	0.002
C^{15}–$C^{15'}$	D18	180.000	0.000
$C^{14'}$–$C^{13'}$	D20	179.997	0.003
$C^{2'}$–$C^{11'}$	D22	180.000	0.000
$C^{10'}$–$C^{9'}$	D24	180.000	0.000
$C^{8'}$–$C^{7'}$	D26	179.060	0.940
$C^{6'}$–$C^{5'}$	D28	−178.623	−0.377
$C^{2'}$–$C^{1'}$	D32	−178.925	−0.075

[a] Deviation from coplanarity measured as follows: $\Delta\phi = 180° - \phi$ (for $\phi > 0$), $\Delta\phi = -180° - \phi$ (for $\phi < 0$).

7. Conclusions

As far as configuration stability is concerned, the following sequence has been established at the HF/3-21G level of ab initio computations.

5-*cis* > *all-trans* > 9-*cis* > 13-*cis* > 15-*cis*7-*cis* > 11-*cis*

The conformational analyses indicated that the central skeleton of conjugated double bonds is practically planar for all seven isomers. However, the tail-end with C–C single bonds are flexible and their conformations avoid coplanarity with the central skeleton.

Acknowledgements

The authors are grateful to Professor A. Kucsman for helpful discussions. We also wish to thank Keneth P. Chasse (math@velocet.ca), Graydon Hoare (graydon@pobox.com) and Velocet Communications for database management, network support, software and distributive processing development. A special thanks is also extended to Andrew M. Chasse (fixy@fixy.org) for his continuing and ongoing development of novel scripting and coding techniques, helping to bring about a reduction in the necessary number of CPU cycles for each computations.

Table 5
Amounts of lycopene in various substances and the percentages of *trans*- and *cis*-isomers present

	Total lycopene[a]	% *trans* (of total)	% *cis* (of total)
Raw tomato	233.58±2.98 μM	89.70	10.30
Tomato juice	189.25±1.12 μM	90.62	9.38
Heated tomato juice	178.08 μM	71.94	28.06
Human serum	422.1±22.9 nM	32.37	67.63
Rat serum	36.41±5.62 nM	37.74	62.26
Rat prostate	316.66±55.88 nM	33.38	66.62

[a] Consider the density of animal and plant tissues as 1, MW of lycopene as 536.85.

[21] *cis*-isomer. The observation that high concentrations of *cis*-isomers are present in human serum and prostrate tissue, also suggests that *cis*-isomers might be biologically more active than the all-*trans*-isomer.

Of all the *cis*-isomers, the 5-*cis*-form appears to be the most stable. Furthermore, our calculations suggests that it is thermodynamically more stable than the all-*trans*-isomer. This suggests that the 5-*cis*-isomer has the highest concentration when equilibrium is reached.

All of these are in general agreement with the biological observations. Thus, one can project that in the future, quantum chemical computations will be able to enhance, considerably, our understanding of the antioxidant activity of lycopene.

References

[1] A.V. Rao, S. Agarwal, Role of lycopene as antioxidant carotenoid in the prevention of chronic diseases: a review, Nutr. Res. 19 (1999) 305–323.

[2] S.K. Clinton, Lycopene: chemisrty, biology, and implications for human health and disease: a review, Nutr. Res. 56 (1998) 35–51.
[3] E. Giovannucci, Tomatoes, tomato-based products, lycopene, and cancer: review of the epidemiologic literature, J. Natl. Cancer Ins. 91 (1999) 317–331.
[4] R. Willstätter, H.H. Escher, Über den Farbstoff der Tomate, Z. Physiol. Chem. 64 (1910) 47–61.
[5] P. Karrer, R. Widmer, Pflanzenfarbstoffe VII. Über Lycopin, Helv. Chim. Acta 11 (1928) 751–752.
[6] P. Karrer, A. Helfenstein, R. Widmer, Pflanzenfarbstoffe IX. Zur Kenntnis des Crocetins und Lycopins, Helv. Chim. Acta 11 (1928) 1201–1209.
[7] P. Karrer, W.E. Bachmann, Zur Kenntnis des Lycopins XI. Mitteilung über Pflanzenfarbstoffe, Helv. Chim. Acta 12 (1929) 285–291.
[8] P. Karrer, A. Helfenstein, H. Wehrli, A. Wettstein, Pflanzenfarbstoffe XXV. Uber die Konstitution des Lycopins und Carotins, Helv. Chim. Acta 13 (1930) 1084–1099.
[9] L. Zechmeister, A.L. LeRosen, W.A. Schroeder, A. Polgar, L. Pauling, Spectral charactersitics and configuration of some stereoisomeric carotenoids including prolycopene and procarotene, J. Am. Chem. Soc. 65 (1943) 1940–1955.
[10] L. Zechmeister, P. Tuzson, Spontaneous isomerization of lycopene, Nature 141 (1938) 249–250.
[11] A. Polgar, L. Zechmeister, Isomerization of -carotene. Isolation of a stereoisomer with increased adsorption affinity, J. Am. Chem. Soc. 64 (1942) 1856–1861.
[12] L. Zechmeister, A. Polgar, Cis–trans isomerization and spectral characteristics of carotenoids and some related compounds, J. Am. Chem. Soc. 65 (1943) 1522–1528.
[13] L. Zechmeister, R.B. Escue, A stereochemical study of methylbixin, J. Am. Chem. Soc. 66 (1944) 322–330.
[14] L. Pauling, Recent work on the configuration and electronic structure of molecules; with some applications to natural products, Fortschr. Chem. Organ. Naturstoffe. 3 (1939) 203–235.
[15] C. Sterling, Crystal structure analysis of β-carotene, Acta Crystallogr. 17 (1964) 1224–1228.
[16] M.O. Senge, H. Hope, K.M. Smith, Structure and conformation of photosynthetic pigments and related compounds 3, crystal structure of β-carotene, Z. Naturforsch. Teil C. 47 (1992) 474–480.
[17] M.L. Nguyen, S.J. Schwartz, Lycopene: chemical and biological properties, Food Technol. 53 (1999) 38–45.
[18] S.K. Clinton, C. Emenhiser, S.J. Schwartz, D.G. Bostwick, A.W. Williams, B.J. Moore, J.W. Erdman Jr., *Cis–trans* lycopene isomers, carotenoids and retinol in the human prostate, Cancer Epidemiol. Biomarkers Prev. 5 (1996) 823–833.
[19] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez, J.A Pople, Gaussian, Inc., Pittsburgh PA, 1995.
[20] G.A. Chasse, K.P. Chasse, A. Kucsman, L.L. Torday, J.G. Papp, Conformational potential energy surfaces of a lycopene model, J. Mol. Struct. (Theochem) (2001) (in press).
[21] Ö. Farkas, S.J. Salpietro, P. Császár, I.G. Csizmadia, Conformations of ethyl benzene (CH3–CH2–Ph), an ab initio study, J. Mol. Struct. (Theochem) 367 (1996) 25–31.
[22] E. Giovannucci, A. Ascherio, E.B. Rimm, M.J. Stampfer, G.A. Colditz, W.C. Willett, Intake of carotenoids and retinol in relation to risk of prostate cancer, J Natl. Cancer Inst. 87 (1995) 767–1776.

THEOCH7153

ELSEVIER

Journal of Molecular Structure (Theochem) 000 (2001) 000–000

THEO CHEM

www.elsevier.nl/locate/theochem

Cationic intermediates in *trans*- to *cis*- isomerization reactions of allylic systems. An exploratory ab initio study

Jenny C. Yeung[a], Gregory A. Chass[a,b,*], Edwin J. Frondozo[b], Ladislaus L. Torday[c,d], Julius G. Papp[c]

[a]*Department of Chemistry, University of Toronto, Toronto, Ont., Canada M5S 3H6*
[b]*Velocet Communications Inc., 210 Dundas St. W., Suite 800, Toronto, Ont., Canada M5G 2E8*
[c]*Department of Pharmacology and Pharmacotherapy, Albert Szent-Gyorgy Medical and Pharmaceutical Center, University of Szeged, H 6701 Szeged, Hungary*
[d]*Faculty of Medicine, University of Calgary, 3330 Hospital Drive N.W., Calgary, Alta, Canada T2N 4M1*

Received 20 December 2000; accepted 9 March 2001

Abstract

Computational analyzes were undertaken to investigate the geometrical isomerization mechanism of a truncated tail-end model (C_1–C_{10}) of the full lycopene molecule, the products of which are the 5-*cis* and 7-*cis*forms. The global conformational minima were identified for the neutral all-*trans* reactant, the cationic isomerization intermediates and that of these two *cis*-isomeric products. Energies and stabilities were compared during different stages of the isomerization mechanism of the segments. The C_4 allylic hydride affinity values of the all-*trans* to the 5-*cis*, and the 7-*cis* isomers of this lycopene model, are in the range of 255–260 kcal mol^{-1}, which are within the expected limits of such hydride affinity values (210 kcal mol^{-1} for weak and 360 kcal mol^{-1} for strong hydride affinity) at the RHF/3-21G level of theory. The bond lengths involving alternating single and double carbon–carbon bonds roughly coincide with the expected changes along each intermediate of the putative isomerization pathway. The following sequence of stability was observed in the tail-end model of neutral lycopene isomers:

5-*cis* > all-*trans* > 7-*cis*

For the cation, the order of stability was different:

all-*trans* > 5-*cis* > 7-*cis*

Keywords: Lycopene; Antioxidant; *trans*- to *cis*-isomerization; Carotenoids; Oxidative stress; Ab initio molecular orbital computations; 1,4-pentadiene; Tail-end model (C_1–C_{10}) of lycopene

* Corresponding author. Tel.: +1-416-598-3229; fax: +1-416-598-7797.
E-mail addresses: jenny.yeung@utoronto.ca (J.C. Yeung), gchasse@fixy.org (G.A. Chass), efrondoz@velocet.ca (E.J. Frondozo), ltordai@ucalgary.ca (L.L. Torday), papp@phol.szote.u-szeged.hu (J.G. Papp).

0166-1280/01/$ - see front matter
PII: S0166-1280(01)00441-9

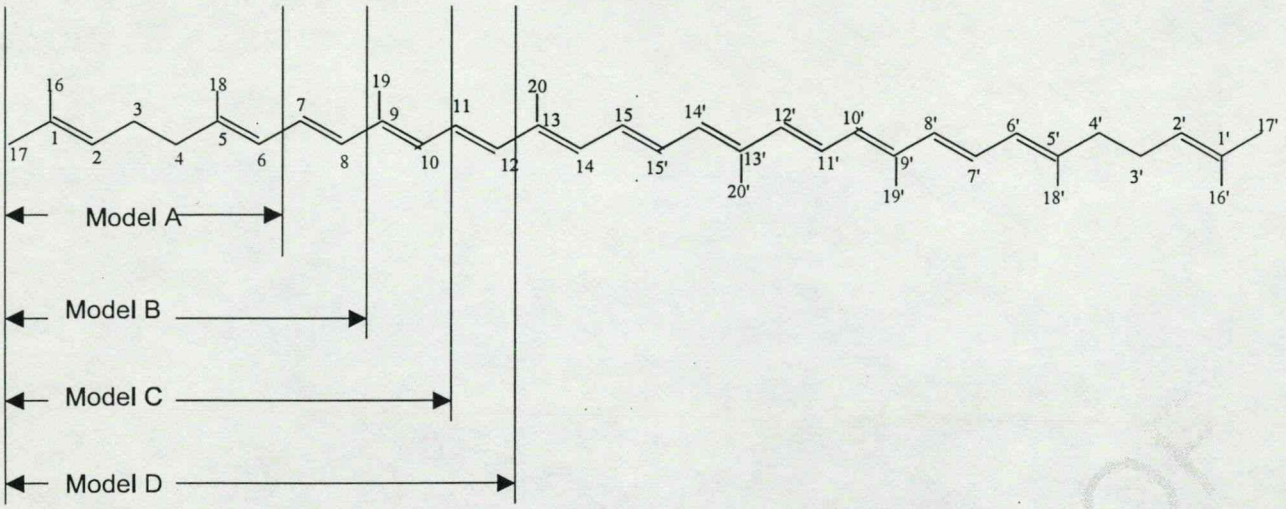


Fig. 1. Structure and numbering of lycopene and four experimental models.

1. Preamble

Lycopene (Fig. 1) belongs to a family of natural lipophillic pigments, known as carotenoids. Carotenoids, present in various fruits and vegetables, are fundamental constituents of plant photosystems. Perhaps the most extensively studied functions of carotenoids are their light-harvesting abilities during photosynthesis and protective effects against photosensitization, which could lead cancer. Today, approximately 700 carotenoids have been identified in nature [1], and their chemical structures have been unveiled. All carotenoids are characterized by an extended system of conjugated double bonds that account for their color and antioxidant activities. Carotenoids can be classified into hydrocarbon carotenoids containing solely carbons and hydrogens, and oxocarotenoids, which include oxygen atoms in their structure.

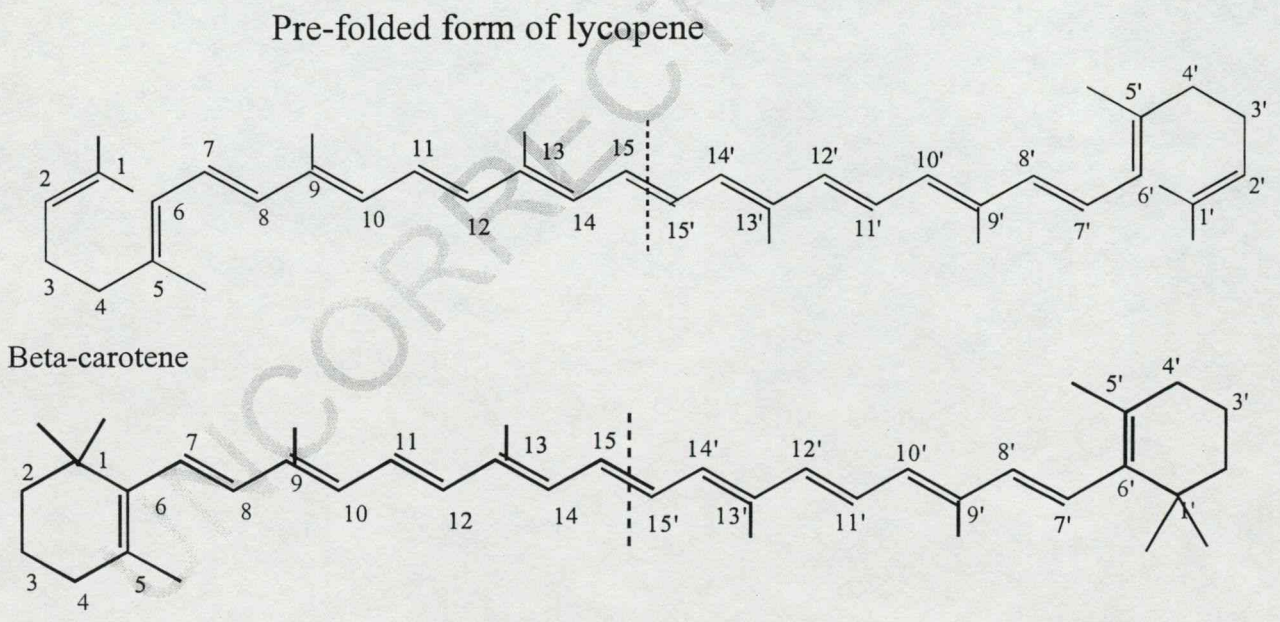


Fig. 2. Similarity in structure between lycopene (pre-folded) and beta-carotene.

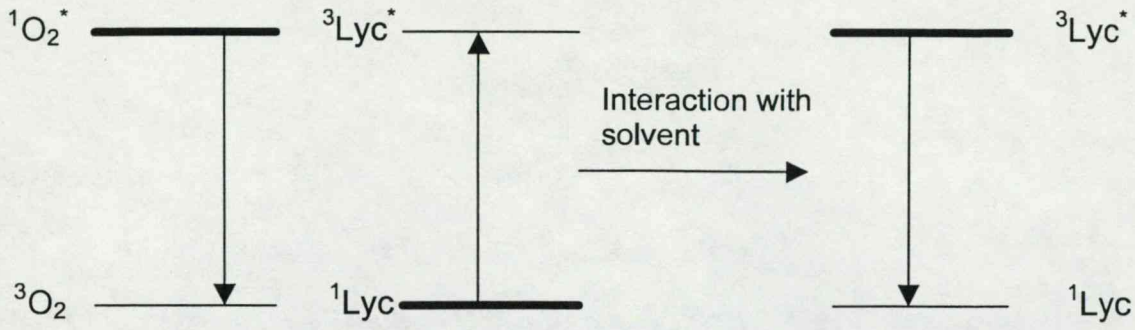


Fig. 3. Energy level diagram showing the energy transfer during $^1O_2$ scavenging (left side) the spin forbidden de-excitation of triplet lycopene (right side).

Another method of classification is based on the presence or absence of provitamin A activity. Originally, carotenoids were considered important only as precursors of vitamin A. More recently, there has been significant interest in evaluation of carotenoids for roles that are unrelated to their conversion to vitamin A. Beta-carotene has been the primary research focus of earlier studies because of its role in cancer prevention and its potential vitamin A activity. Although both lycopene and beta-carotene have very similar chemical structures (Fig. 2), and the former is even the precursor of the latter in the bio-synthetic pathway [2], the effects of lycopene and beta-carotene on health differ drastically. It is only recently discovered that instead of protecting humans against cancer, the supplement form of beta-carotene, when ingested, can increase the risk of lung cancer and heart disease [3]. Thus, the relationship between molecular structure and health is clearly demonstrated – even a diminutive difference in chemical structure can lead to unpredictable effects on health.

Although used as a natural food colorant for many years, it was only recently that lycopene became the subject of intense study with respect to its antioxidant activity and potential in alleviating chronic diseases such as prostate cancer [4]. A growing body of scientific evidence indicates that antioxidants are important in the human body's defences against harmful free radicals, which induce structural damages to cells and possibly contribute to aging and diseases that occur more frequently with advancing age. Lycopene, an antioxidant in vivo, provides protection against the oxidation of lipids, proteins and DNA by free radicals. The collective degenerative effects of these free radicals are termed 'oxidative stress'.

In fact, the most dangerous free radicals come from oxygen, known as the reactive oxygen species. They include superoxide (O_2^-), hydrogen peroxide (H_2O_2), hydroxyl radicals ($\cdot OH$) and singlet oxygen ($^1O_2$). Singlet oxygen is a very high energy form of oxygen with a shorter life than an oxygen radical, but damage cells and tissues much more quickly. Lycopene may well be the most potent of all singlet oxygen quenchers. Its efficiency for radical quenching resides in its extended system of conjugated double bonds. The greater the number of conjugated double bonds, the higher the ability of a carotenoid to quench free radicals [5]. Lycopene antioxidant activity begins with the transfer of energy of singlet oxygen to the carotenoid, yielding a triplet ground state oxygen and

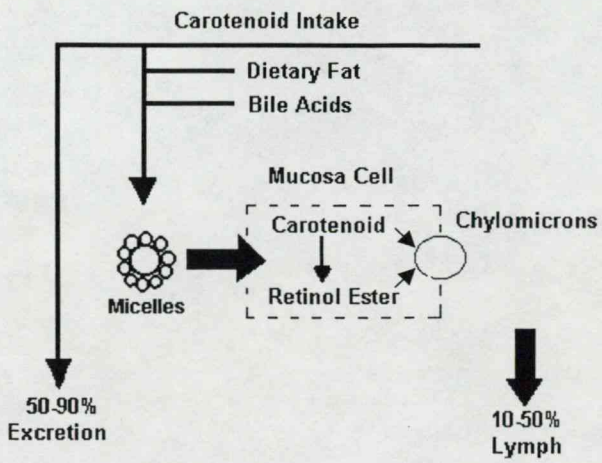


Fig. 4. Absorption of carotenoids (lymphatic pathway).

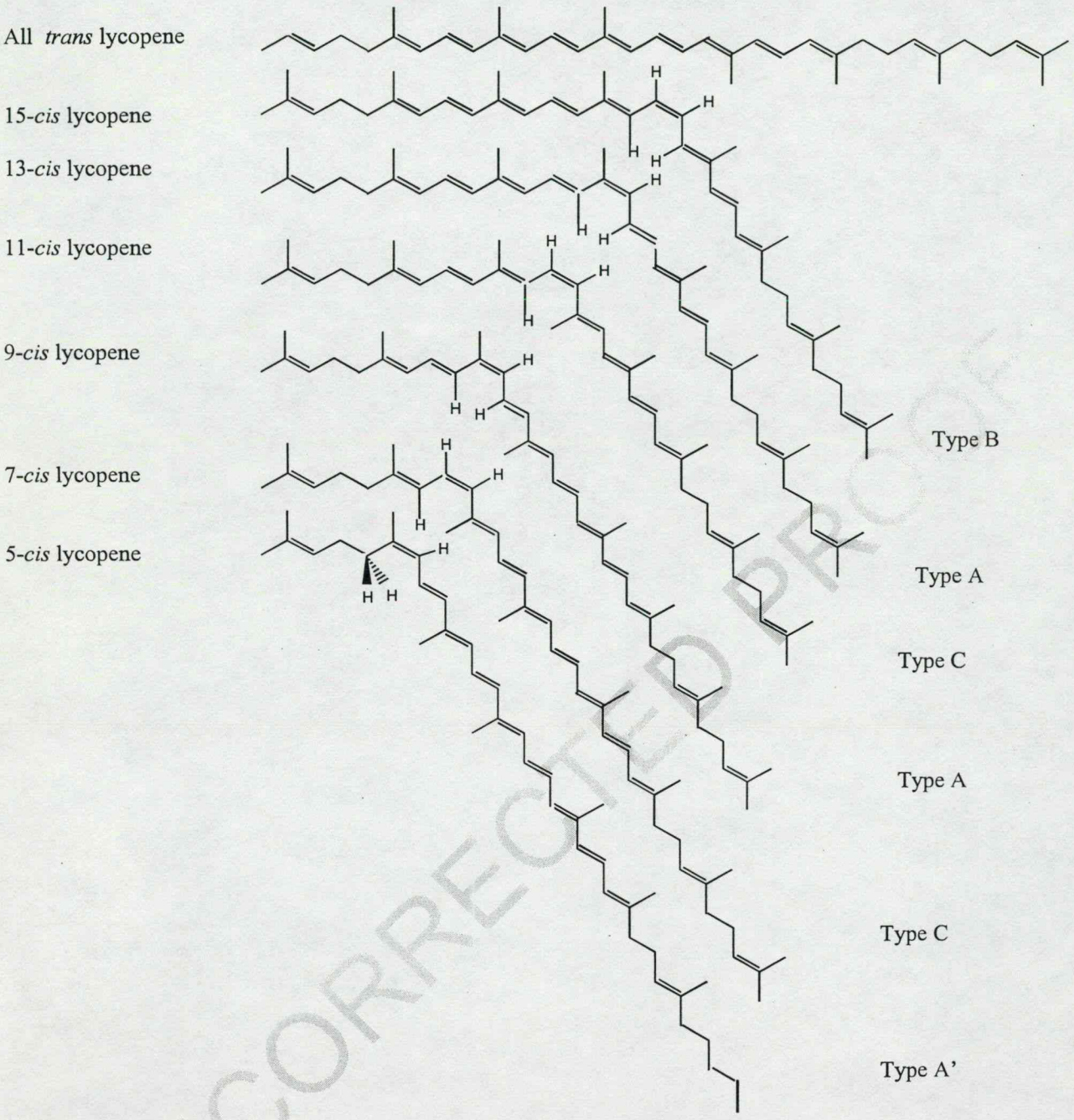


Fig. 5. Geometrical isomers of lycopene.

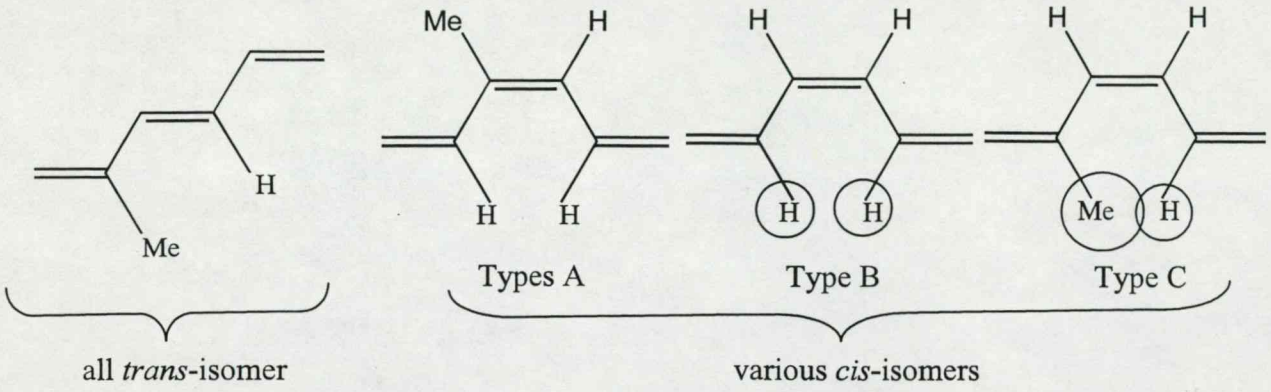


Fig. 6. Interaction between a methyl group and hydrogen group in lycopene.

a triplet-excited carotenoid (Fig. 3). Subsequently, the carotenoid returns to ground state by dissipating its energy through interaction with the surrounding solvent [6]. In fact, lycopene acts like a catalyst since it is left intact in this quenching process; also, once its job is finished, the physical quenching process begins again when lycopene further quench another singlet oxygen molecule. Furthermore, lycopene may also undergo isomerization during quenching, via the lowest triplet state of the molecule [5].

Together with dietary lipids and bile acids, lycopenes are incorporated into lipid micelles in the small intestine via the lymphatic pathway when it is taken into the body (Fig. 4). It is possible that configurations of different lycopene isomeric forms determine the ease of micelle formation, thus, aiding absorption of carotenoids into the intestinal mucosa cell. However, whether an in vivo isomerization mechanism exists is still unknown. The intact carotenoid is subsequently incorporated into chylomicrons, which are released into the lymphatic system for transport to the liver. The highest levels of lycopene have been found in the testes, adrenal glands and prostate [7]. The bioavailability of lycopene after its absorption and distribution in the body depends on various factors: food processing and presence of dietary lipids. Interestingly, processing such as heating has traditionally been considered as contributing towards the loss of nutritional quality of foods. In contrast, research confirmed that processing breaks down cell walls, which weakens the bonding forces between lycopene and tissue matrix, thus making lycopene more easily absorbed. Isomerization occurs during processing, converting a portion of the all-*trans* isomers to a mixture of *cis*-isomers. Lastly, since lycopene is a fat-soluble substance, it must be consumed with some fat in order to be absorbed through the intestine to achieve best results.

2. Introduction

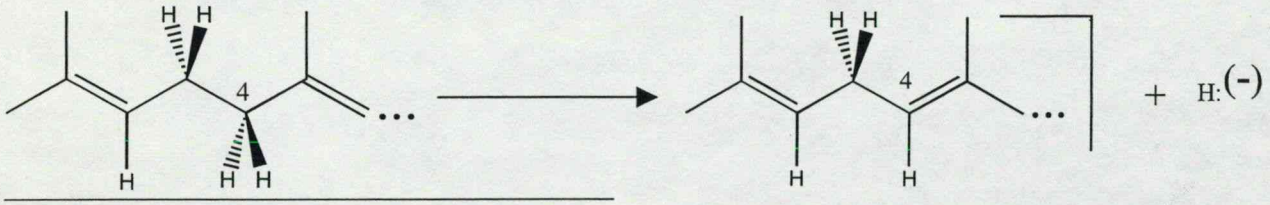
2.1. Structural background

Lycopene has a chemical formula of $C_{40}H_{56}$. It is an acyclic and internally symmetrical molecule with 11 conjugated double bonds. Zechmeister [8] pointed out the possible existence of 72 geometrical isomers of lycopene. However, the most predominant species in tomatoes and tomato products is the all-*trans* isomer. Lycopene is present naturally in human plasma as an isomeric mixture containing over 60% of the total lycopene as *cis* isomers [7].

The all-*trans*, 5-*cis*, 9-*cis*, 13-*cis*, and 15-*cis* lycopene (Fig. 5) are commonly identified in human serum. Not all *cis*-isomers are of equal stability due to a number of possible 1,4 interactions between neighbouring methyl and hydrogen groups (Fig. 6). Pauling called attention to the fact that *trans*- to *cis*-isomerization can be a result of overlapping of the methyl group of a carbon atom adjacent to a double bond and the hydrogen [9]. Apparently, the steric interaction between a methyl group and hydrogen atom (Fig. 6) is the most destabilizing which is based on their relative group sizes.

2.2. Computational background

Conformational analysis has been performed on a truncated (C_1–C_8) tail-end model (see Model B in Figs. 1 and 8) of lycopene [10]. The global



minimum on the conformational Potential Energy Hyper-Surface (PEHS): $E = E(\chi 2, \chi 3, \chi 4)$, corresponds to the ($g^+$, a, g^+) structure, or its degenerate pair (g^-, a, g^-), for both the all-*trans* and 5-*cis* isomers. For the full lycopene isomers, the sequence of stability:

5-*cis* > all-*trans* > 9-*cis* > 13-*cis* > 15-*cis* > 7-*cis* > 11-*cis*

has been observed from previous ab initio calculations [11]. Apparently, the 5-*cis* form is the energetically most stable of all isomers not only in the case of the full lycopene, but also its truncated tail-end model (Model B) [10].

2.3. Scope

The aim of the present research is to study the isomerization mechanism (Fig. 7) of a truncated (C_1–C_{10}) tail-end model (Model C) of the full lycopene (Figs. 1 and 8), the products of which are two *cis* forms (5-*cis* and 7-*cis*). The use of this truncated tail-end model, which is considerably smaller than the full lycopene, permits more convenient and manageable molecular computations as well as a detailed examination of the two terminal portions of lycopene (from C_1 to C_5 and from C'_1 to C'_5) that rotate out of plane. Previous studies concluded [10,11] that lycopene has a rigid central skeleton (from C_5 toC'_5) and two flexible tails (from C_1 to C_5 and from C'_1 to C'_5).

As it can be seen from the proposed mechanism (Fig. 7), the cation formation converts C_4 from its nearly tetrahedral (sp^3), to a trigonal planar (sp^2) geometry. Since the positive charge is dispersed throughout the π-network, C_3 now has two trigonal planar carbons attached.

For this reason, it became necessary to study the conformational pattern of divinyl methane (1,4-pentadiene), before the cationic intermediates were investigated.

3. Methods

Completely relaxed geometric optimizations were performed at RHF/3-21G level of theory for the neutral all-*trans* as well as the 5-*cis* and 7-*cis* isomers, of lycopene Model C (Fig. 8) using the conformational information of Model B, published earlier [10]. The three cationic intermediates (i.e. the all-*trans*, the 5-*cis* and 7-*cis* form) were also subjected to the same analysis.

An ab initio (RHF/3-21G) 2D-scan was carried out to generate the $E = E(\phi, \psi)$ potential energy surface of divinyl methane, where ϕ and ψ represent rotations about the C(sp^2)–C(sp^3) and C(sp^3)–C(sp^2) single bonds, respectively. The visually observed minima were subjected to geometry optimizations at the RHF/3-21G level of theory. The results were used in studying the cationic reaction intermediates. GAUSSIAN 94 has been used to carry out all the computations [12].

4. Results and discussion

4.1. Conformations of divinyl methane

Divinyl methane (1,4-pentadiene) has been studied first, as a conformational model of the cationic intermediates. The essential question here is whether the two unsaturated moieties, joined to the central CH_2 unit, are coplanar or not. In other words, do minima

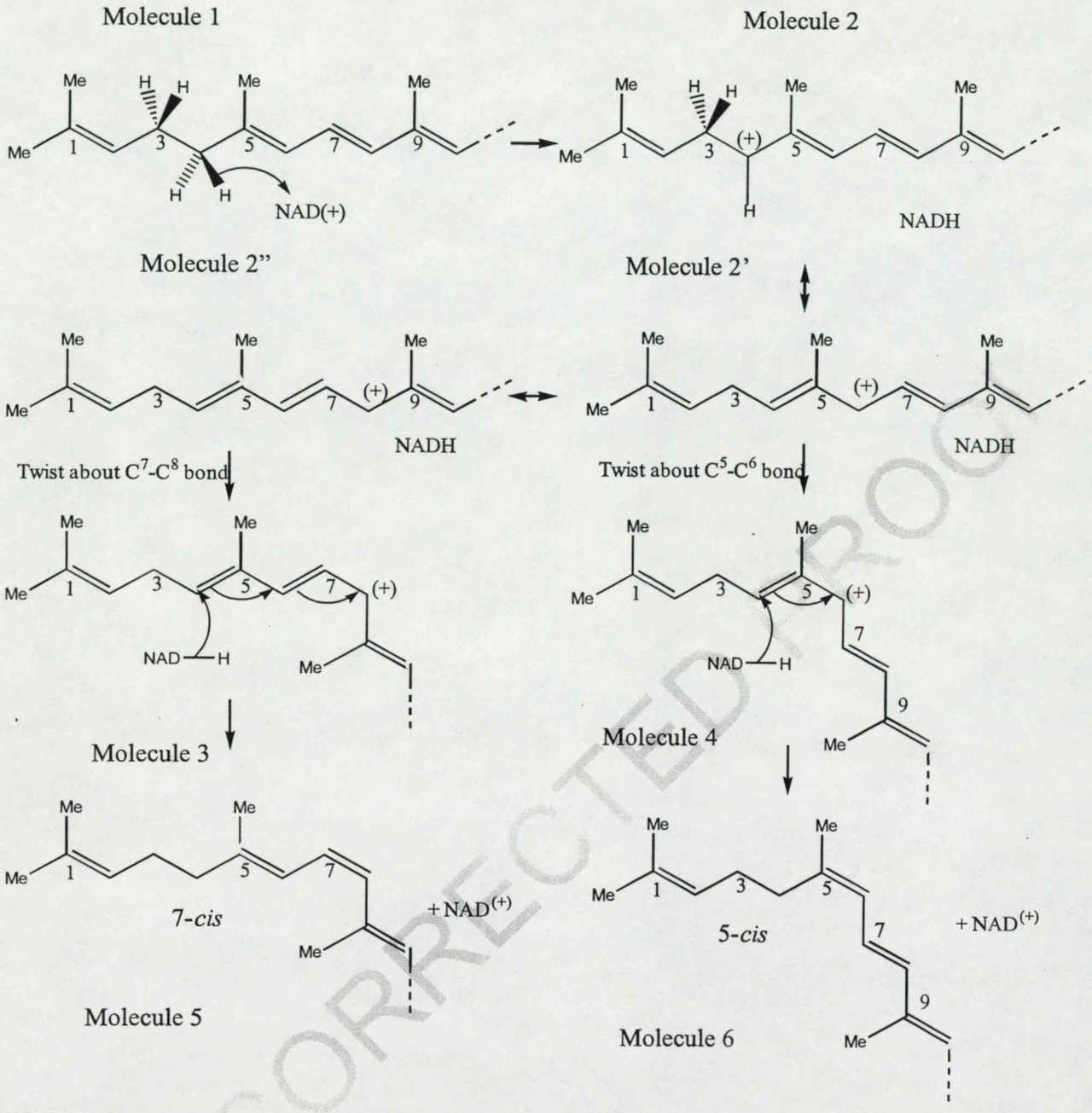
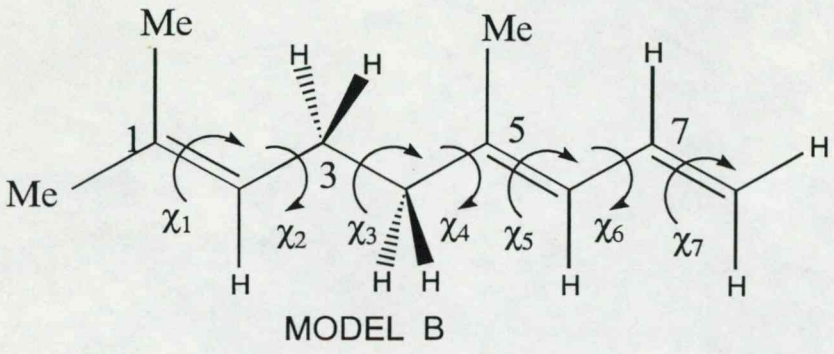
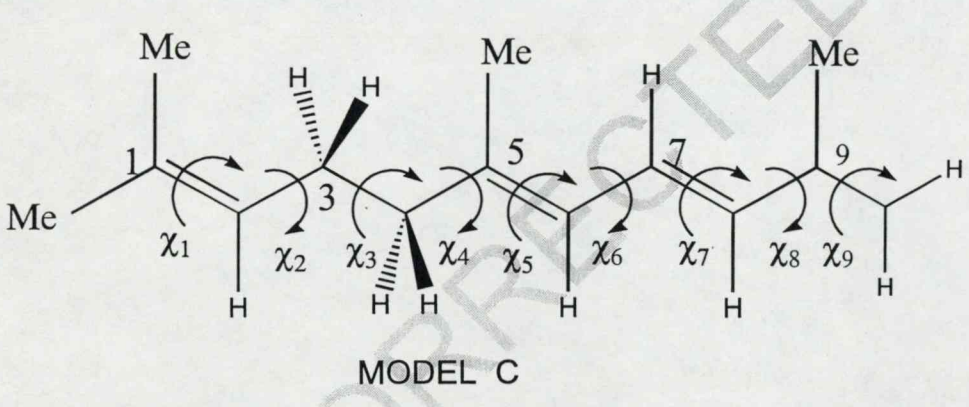


Fig. 7. Putative mechanism for lycopene isomerization.



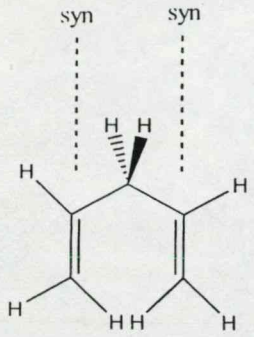
X_1: Me-C_1-C_2-C_3
X_2: C_1-C_2-C_3-C_4
X_3: C_2-C_3-C_4-C_5
X_4: C_3-C_4-C_5-C_6
X_5: C_4-C_5-C_6-C_7
X_6: C_5-C_6-C_7-C_8
X_7: C_6-C_7-C_8-H



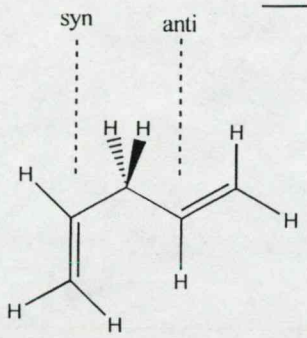
X_1: Me-C_1-C_2-C_3
X_2: C_1-C_2-C_3-C_4
X_3: C_2-C_3-C_4-C_5
X_4: C_3-C_4-C_5-C_6
X_5: C_4-C_5-C_6-C_7
X_6: C_5-C_6-C_7-C_8
X_7: C_6-C_7-C_8-C_9
X_8: C_7-C_8-C_9-C_{10}
X_9: C_8-C_9-C_{10}-H

Fig. 8. Structure and numbering of the tail-end models of lycopene. Model B includes the first eight carbons of lycopene's carbon backbone. Model C includes the first ten carbons of lycopene's carbon backbone.

occur either in the *syn/syn*, *syn/anti*, or in the *anti/anti* form.



$(0^0, 0^0)$



$(0^0, 180^0)$

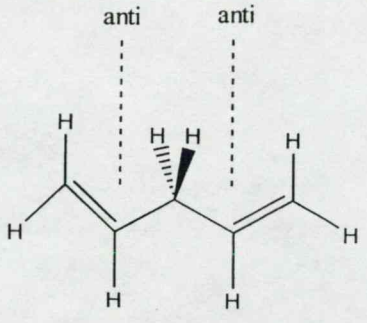
The conformational PEHS of divinyl methane is shown in Fig. 9. The eight minima located were optimized and the results are summarized in Table 1.

The minima occurred in the vicinity of ±16 and ±120°. That means that the unsaturated moieties are either nearly perpendicular (corresponding to ±30°) to one or the other of the C–H bonds, or in alignment, i.e. eclipsed, (±120°) with one or the other of the C–H bonds. However, no minimum was observed at 0 or ±90°, when the olifinic groups would be eclipsed with or perpendicular to the C–C bond.

Torsional potential energy curves (PECs) of the type

$$E = E(\psi),$$

were generated at $\phi = 0°$ as well as $\phi = 120°$ and $\phi = -120°$, to see the alternation of minima and maxima along the torsional mode of motion. The results are shown in Fig. 10A and B, respectively.



$(180^0, 180^0)$

4.2. Structures and stabilities, along the isomerization mechanism pathway of lycopene Model C

The *trans–cis* isomerization mechanism (Fig. 7) of Model C (Fig. 8B), as well as the conformational torsional angles and relative energies of its intermediates have been examined in detail. The geometries, energies and stabilities of all species involved are presented in Table 2.

Results obtained here for the neutral compounds are expected to follow the sequence of energy and stability of the full lycopene. The observed order of stability is clearly in agreement with earlier results obtained in the case of full lycopene [11]:

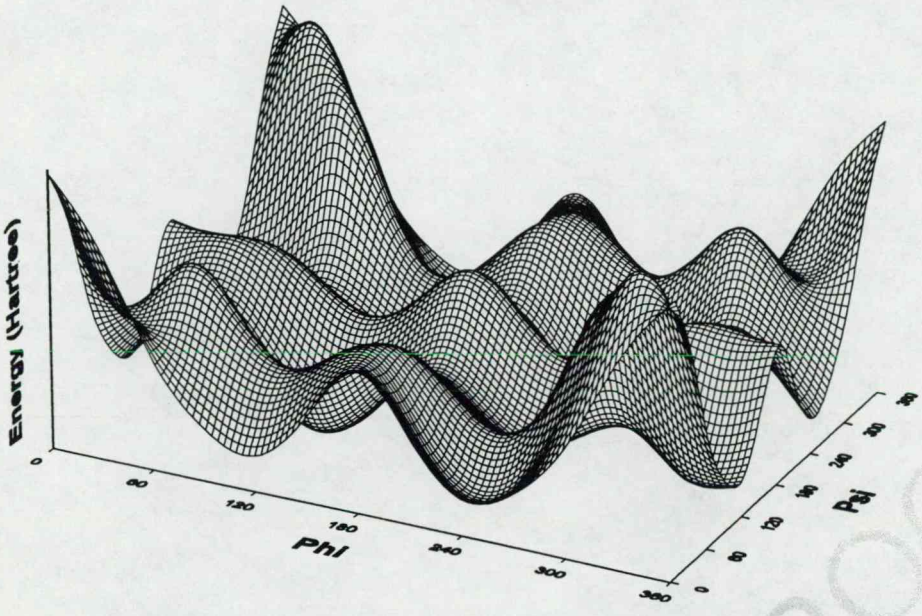
5-*cis* > all-*trans* > 7-*cis*

For the cationic intermediates, however, the order of stability is different:

all-*trans* > 5-*cis* > 7-*cis*

Table 1
Optimized dihedral angles, bond lengths, total energies and relative energies of various conformers of divinyl methane

Conformer	ϕ	ψ	C_3–H_9	C_3–H_{10}	Energy	ΔE (kcal mol^{-1})
a	116.85982686	116.8582153	1.08566470	1.08566516	−192.8732017	0.00000
a′	−116.86066755	−116.85996391	1.08566521	1.08566492	−192.8732017	0.00000
b	119.87843183	−119.87227647	1.08815294	1.08353286	−192.8718829	0.82756
b′	−119.87318394	119.87789676	1.08353187	1.08815307	−192.8718829	0.82756
c	16.10175702	−115.86964444	1.08760599	1.08673326	−192.8720363	0.73130
c′	−16.10317418	115.87135594	1.08673431	1.08760496	−192.8720363	0.73130
d	−115.86972268	16.1032023	1.08673701	1.0875990	−192.8720363	0.73130
d′	115.86753178	−16.10561373	1.08759825	1.08673763	−192.8720363	0.73130



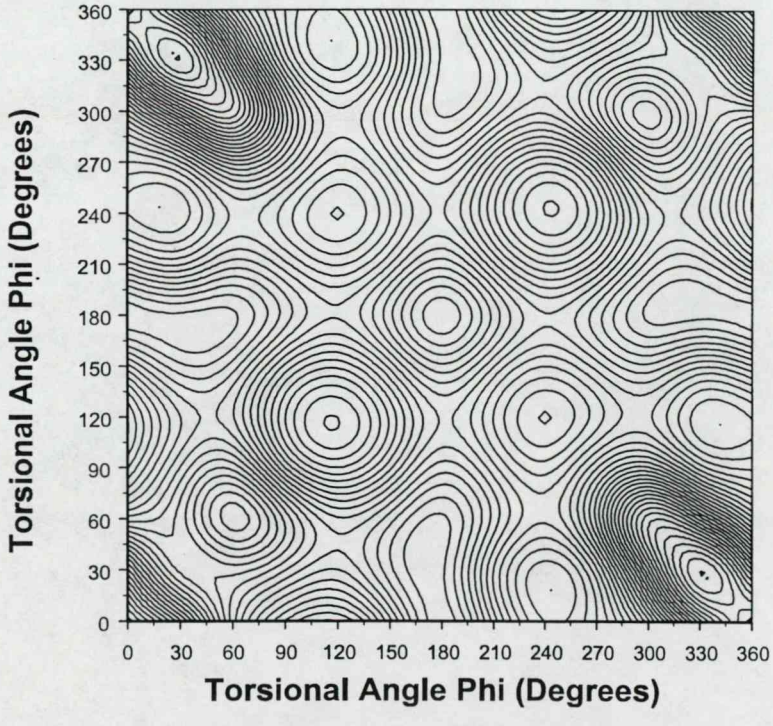


Fig. 9. Conformational potential energy surface of divinylmethane; $H_2C{=}CH{-}CH_2{-}CH{=}CH_2$; $E = E(\phi, \psi)$ where ϕ and ψ are the rotations about the two single bonds. Top: Landscape; Bottom: Contour representation.

J
A
ARAJ
PIA

The rotation PEC along ψ, of divinyl methane is structurally analogous to the rotational potential along χ_3 in the cationic intermediates of Model C. For the three cationic intermediates (i.e. for the all-*trans*, the 5-*cis* and 7-*cis* cations), the

$$E = E(\chi_3).$$

PECs are shown in Fig. 11. It is interesting to note that the cationic intermediates show a somewhat different torsional potential (Fig. 11) with respect to that obtained in the case of divinylmethane (Fig. 10). Although the increased size of Model C may be a factor, nevertheless it is more likely that the presence of the positive charge redistributes the electron density sufficiently, which in turn translates to a somewhat different conformational PECs.

4.3. Bond length interpretations

One of the structural features of the isomerization process is the changing carbon–carbon bond length. During the reaction, a double bond may change to a single bond and vice versa. Key bond lengths at various stages of isomerization are summarized in

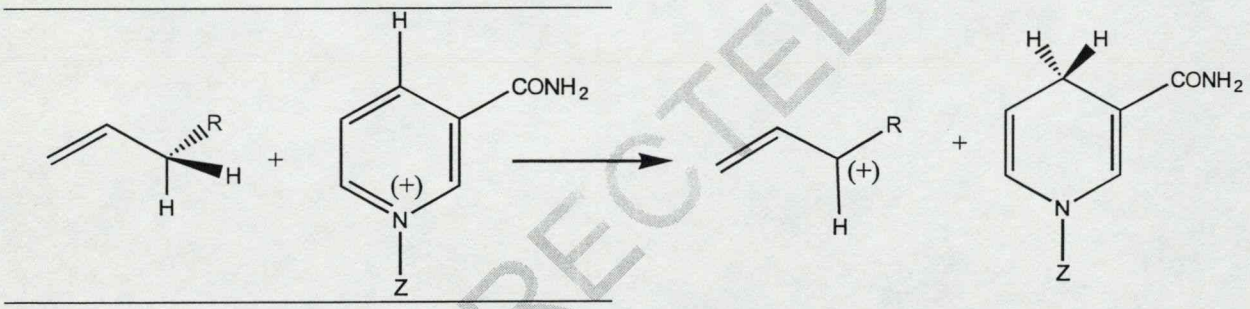


Table 3. The C–C single and C=C double bond lengths, obtained at this level of theory, are given, for the sake of comparison, in the footnote of Table 3.

Three types of CC bond lengths were identified:

double bond (approximately: 1.32 Å),
conjugated single bond (approximately: 1.47 Å),
genuine single bond (approximately: 1.54 Å).

On the basis that, a conjugated stabilization pattern has emerged, which is shown in Scheme 1. The broken line -- passing through three bonds from C_5 to C_8 are neither genuine single or genuine double bonds. Thus rotations may occur relatively easily along the C_5–C_6 and C_7–C_8 bonds. The results (Table 3 and Scheme 1) roughly coincide with the expected changes, but they are somewhat different from conventional resonance hybrid representation (c.f. Structures 2, 2′ and 2″ in Fig. 7).

4.4. Energetics of isomerization mechanism

The rotational potential energy curves associated with the isomerization processes

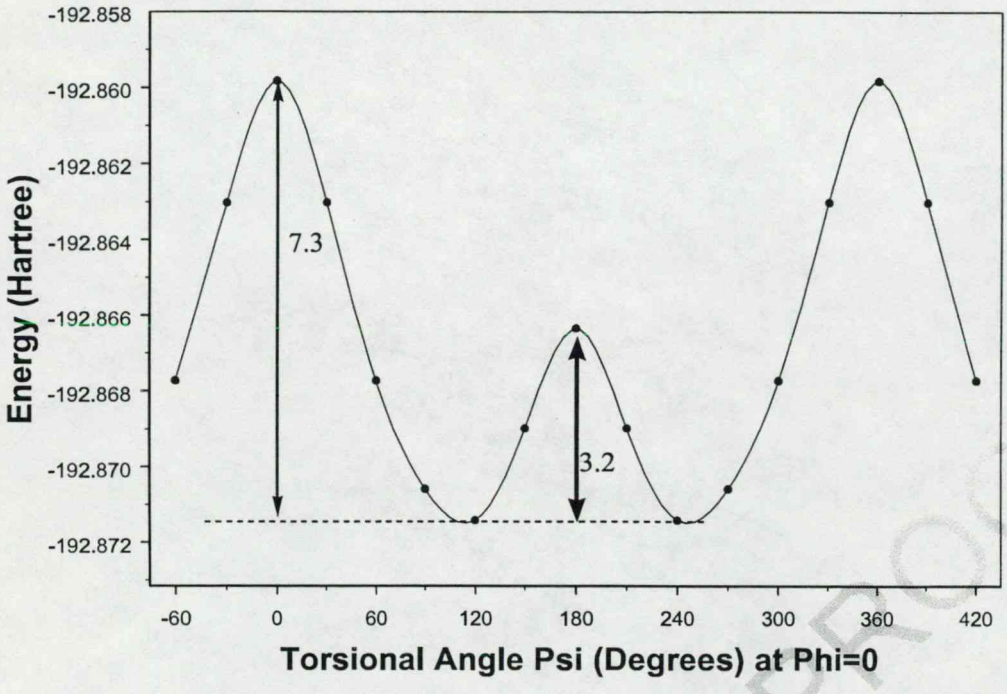
All-*trans cation* → 5-*cis cation*

All-*trans cation* → 7-*cis cation*

are shown in Fig. 12. The barriers to interconversions are different for the two processes. The two PECs clearly indicate that the isomerization to the 5-*cis* cation is kinetically more favourable than that to the 7-*cis* cation. However, the cations have to be formed in the first place. Hydride transfer to $NAD^{(+)}$ may lead to the formation of allylic cations.

The relative ease of hydride [$H:^{(-)}$] removal is measured by the hydride affinity of the molecular system. Hydride affinity values, as computed at the RHF/3-21G level of theory, are summarized in Table 4. These data suggest that the easiest removal of an allylic hydride is from the all-*trans* isomer and the most difficult to remove, is from the 5-*cis* isomer. These results are illustrated graphically by Fig. 13, in the form which preserves the mechanistic considerations given in Fig. 7.

As far as the thermodynamics of the reaction are concerned, a great deal depends on the hydride affinity of the hydride acceptor. This is clearly shown in



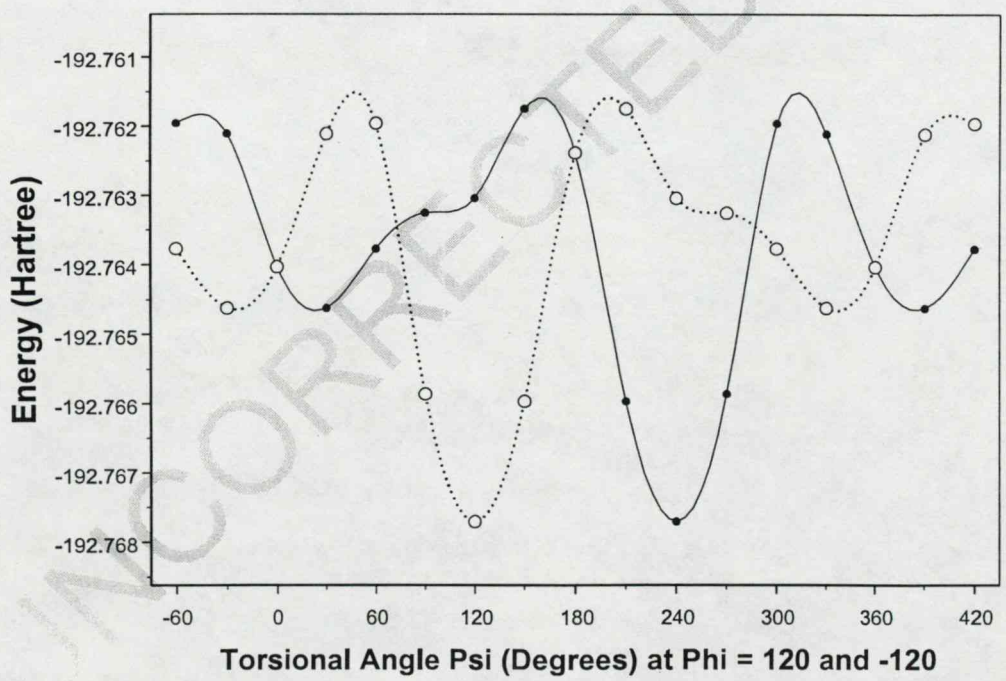


Fig. 10. Top(A): Torsional potential $E = E(\psi)$ at $\phi = 0°$, for divinylmethane. Bottom(B): Torsional potential $E = E(\psi)$ at $\phi = \pm 120°$, for divinylmethane.

Table 2
Torsional angles, total and relative energies for the reactant, intermediates and products of lycopene Model C isomerization

Isomer	Charge	χ_1	χ_2[a]	χ_3[a]	χ_4[a]	χ_5[b]	χ_6	χ_7[b]	χ_8	χ_9	χ_{10}[c]	χ_{11}[c]	χ_{12}[c]	Energy (Hartree)	ΔE (kcal mol^{-1})
All-trans	0	− 178.919	**106.090**	**176.414**	**100.211**	− 178.62	− 179.773	− 179.942	179.977	179.996	179.853	175.917	− 179.006	− 539.8927057	0.000000
5-*cis*	0	− 178.918	**103.058**	**176.448**	**90.186**	*1.362*	− 178.146	179.905	− 179.779	− 179.990	179.197	175.871	− 179.943	− 539.8933291	− 0.391190
7-*cis*	0	181.069	**106.427**	**176.351**	**99.600**	181.633	182.021	*0.772*	181.454	180.274	179.991	173.943	177.585	− 539.8833670	5.860127
>All-*trans*	+ 1	179.161	**97.958**	**99.284**	− 177.597	− 178.327	179.673	− 179.958	179.979	− 179.971	− 179.741	− 2.318	179.980	− 539.0853808	0.000000
5-*cis*	+ 1	179.030	**97.370**	**100.087**	− 177.362	*2.951*	− 179.406	180.000	− 0.003	− 0.001	− 179.933	− 178.288	179.894	− 539.0790704	3.959839
7-*cis*	+ 1	179.102	**98.443**	**98.994**	− 177.642	− 178.235	179.713	*0.098*	− 179.968	179.995	− 179.667	− 2.413	179.875	− 539.0750170	6.503388

[a] Conformational torsional angles, in bold.
[b] *cis*-isomeric torsional angles, in italics.
[c] Methyl rotations.

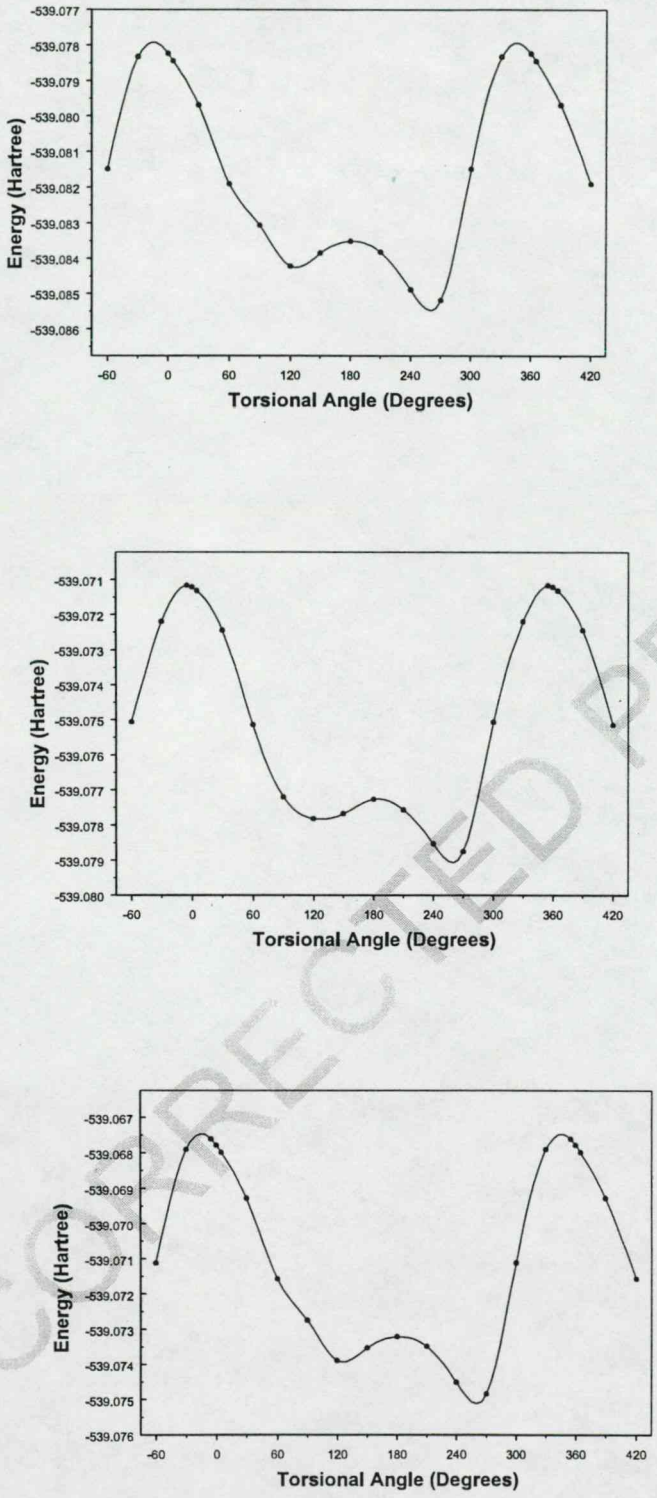
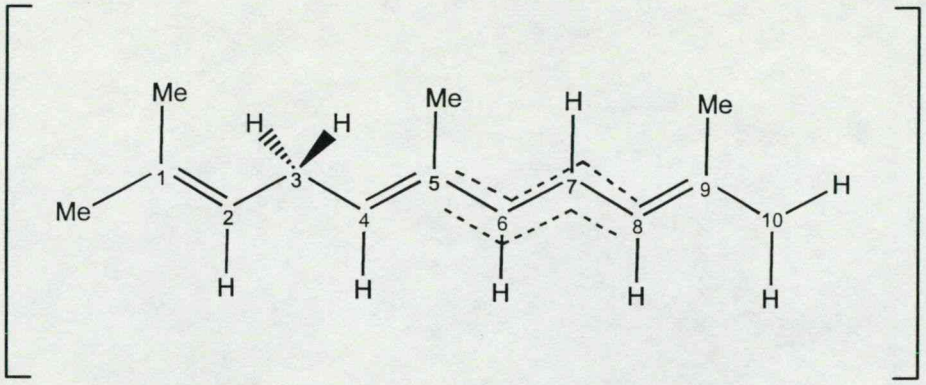


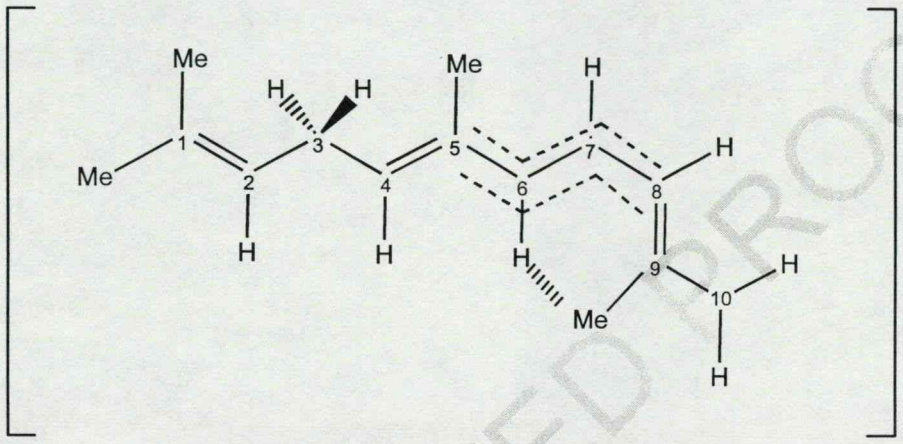
Fig. 11. Torsional potentials for the three cationic intermediates of lycopene Model C isomerization. Top: all-*trans*, middle: 5-*cis*, bottom: 7-*cis*.

all-trans



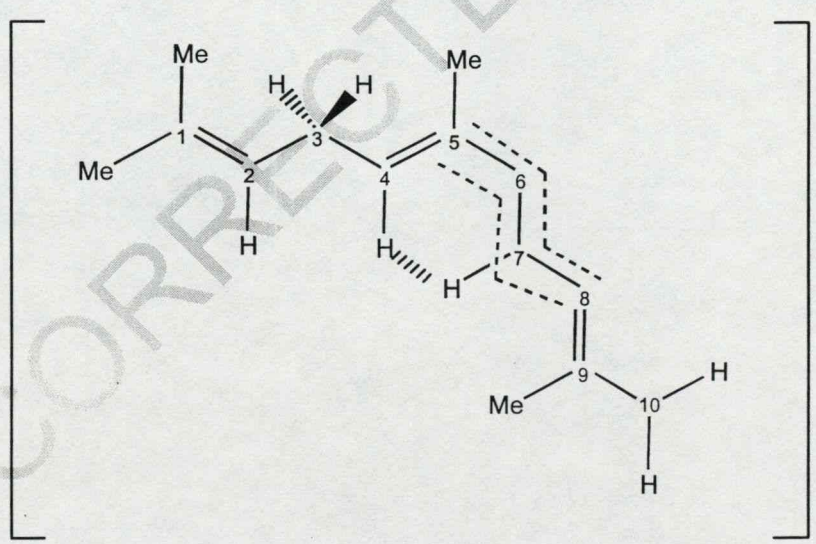
(+)

7-cis



(+)

5-cis



(+)

Scheme 1.

Table 3
Variation of carbon–carbon bond lengths along the isomerization pathway of lycopene Model C (conjugated single bonds in italic, double bonds in bold R[C–C] in ethane = 1.542, in butadiene = 1.467, R[C=C] in ethylene = 1.315, in butadiene = 1.320)

Molecule	Charge	R(Me-C_1)	R(C_1–C_2)	R(C_2–C_3)	R(C_3–C_4)	R(C_4–C_5)	R(C_5–C_6)	R(C_6–C_7)	R(C_7–C_8)	R(C_8–C_9)	R(C_9–C_{10})
All-*trans*	0	1.517	**1.321**	1.510	1.556	1.518	**1.328**	*1.462*	**1.328**	*1.472*	**1.325**
5-*cis*	0	1.517	**1.321**	1.510	1.556	1.517	**1.328**	*1.463*	**1.328**	*1.472*	**1.325**
7-*cis*	0	1.517	**1.321**	1.510	1.556	1.518	**1.329**	*1.463*	**1.332**	*1.478*	**1.326**
All-*trans*	1	1.516	**1.322**	1.529	1.495	**1.359**	*1.410*	*1.393*	*1.364*	**1.335**	1.516
5-*cis*	1	1.516	**1.323**	1.531	1.495	**1.358**	*1.412*	*1.399*	*1.362*	**1.334**	1.515
7-*cis*	1	1.516	**1.322**	1.529	1.495	**1.360**	*1.412*	*1.394*	*1.368*	**1.337**	1.516

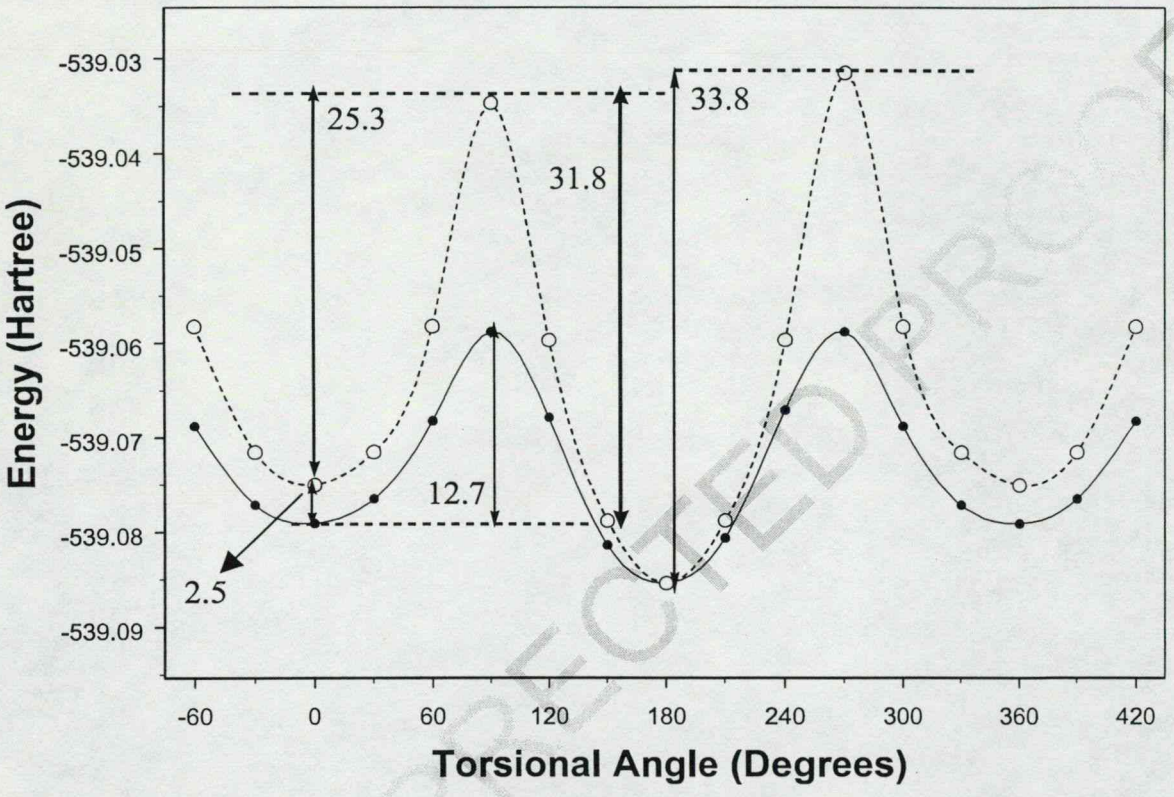


Fig. 12. Torsional potentials for the two isomerization processes Solid Line: all-*trans* → 5-*cis*, broken line: all-*trans* → 7-*cis*.

Table 4
Energy components ($E[H^{(-)}] = -0.4004207$ Hartree) for hydride affinities of selected compounds

Molecular system	Total energy (Hartree)		Hydride affinity (kcal mol^{-1})
	Cation	Neutral	
Weak hydride acceptor	−7.1870945 $Li^{(+)}$	−7.9298426 Li–H	214.813866774
Strong hydride acceptor	−39.0091291 $H_3C^{(+)}$	−39.9768768 H_3C–H	356.00336577
All-*trans* Model C	−539.0853808	−539.8927057	255.336454542
5-*cis* Model C	−539.0790704	−539.8933291	259.687483380
7-*cis* Model C	−539.0750170	−539.8833670	255.979715043

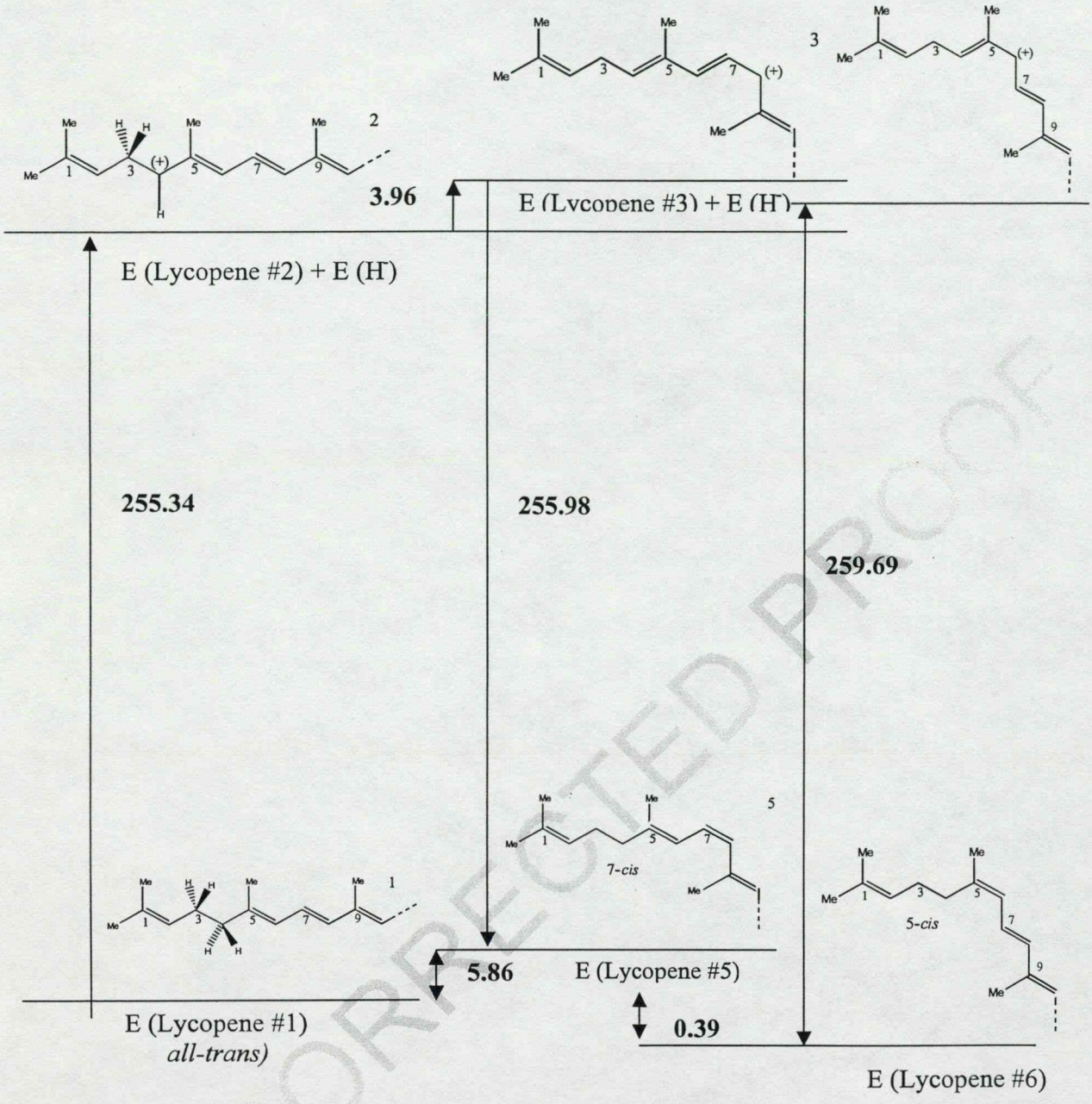


Fig. 13. Energy level diagram (kcal mol^{-1}) for *trans–cis* isomerization mechanism of the truncated lycopene model.

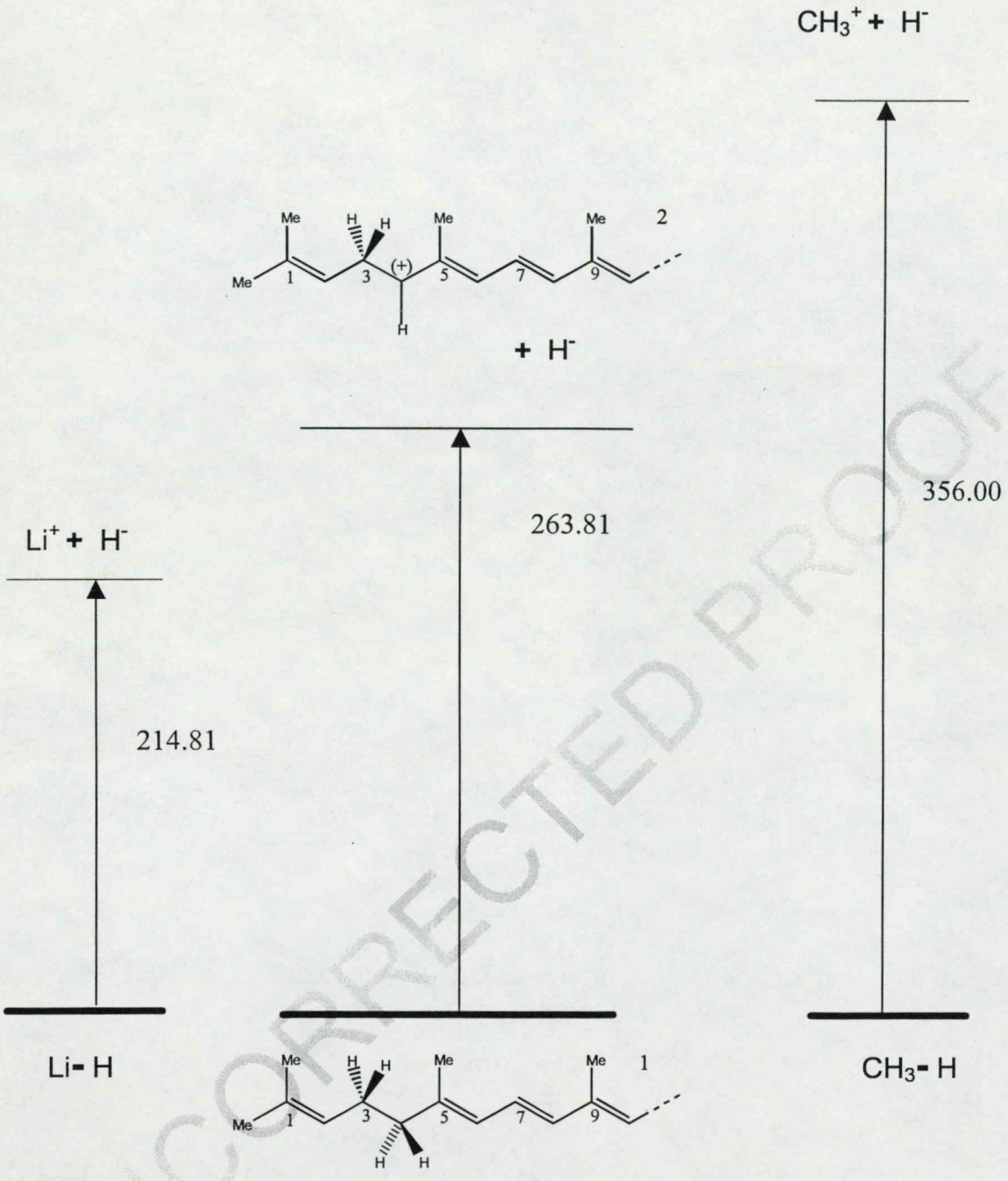


Fig. 14. Side-by-side comparison of hydride affinities (kcal mol^{-1}) for several small cationic systems.

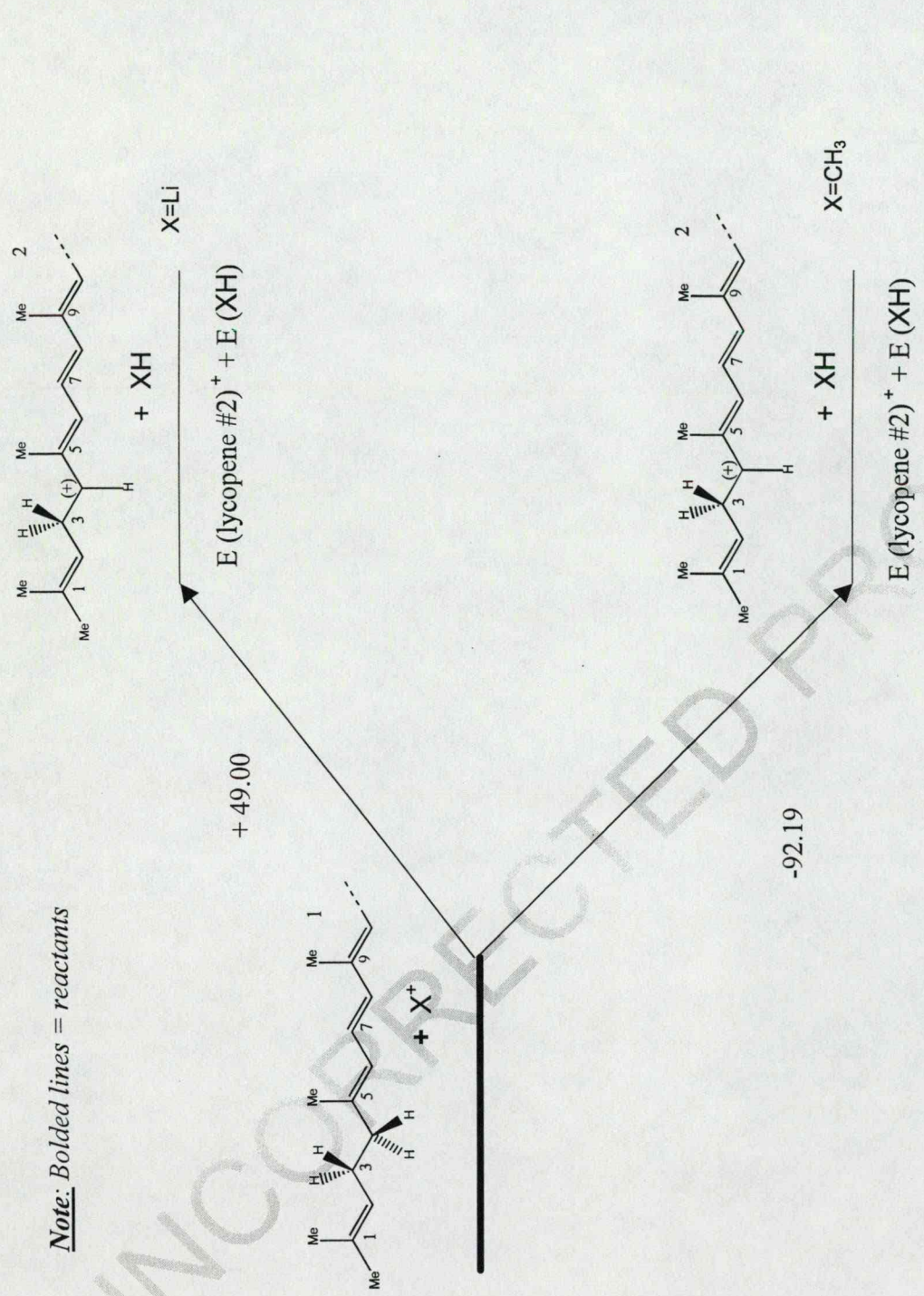


Fig. 15. Energy level diagram (kcal mol^{-1}) for hydride ion transfer from neutral all-*trans* lycopene Model C to selected cationic species.

Fig. 14 where a weak hydride acceptor [$Li^{(+)}$] is compared to a strong hydride acceptor [$H_3C^{(+)}$], as summarized in Table 4. The overall reaction may be endothermic for a weak hydride acceptor [$Li^{(+)}$], or exothermic for a strong hydride acceptor [$H_3C^{(+)}$], as shown in Fig. 15.

No attempt is made here to study $NAD^{(+)}$, because it is believed that the hydride affinity of $NAD^{(+)}$ is not just the hydride affinity of the pyridinium ring. It is quite likely modified by the carbohydrate attached, as well as by the remaider of the $NAD^{(+)}$ molecule.

5. Conclusions

All of the results obtained suggest that the isomerization mechanism proposed is indeed plausible.

The following sequence of stability is observed:

5-*cis* > all-*trans* > 7-*cis*

for the neutral tail-end (C_1–C_{10})model of lycopene (Model C).This is in line with the results from previous studies [11] on the various isomers of the entire lycopene. The relative stability of the intermediate cations generated from Model C, however exhibited a different stability:

all-*trans* > 5-*cis* > 7-*cis*.

6. Uncited reference

[13].

Acknowledgements

The authors are grateful to Professor A. Kucsman for helpful discussions. We also wish to thank Keneth P. Chasse (math@velocet.ca), Graydon Hoare (graydon@pobox.com)and Velocet Communications Inc. for database management, network support, software and distributive processing development. A special thanks is also extended to Andrew M. Chasse (fixy@fixy.org) for his continuing and ongoing development of novel scripting and coding techniques, helping to bring about a reduction in the necessary number of CPU cycles for each computation.

References

[1] F. Khachik, J.S. Bertram, M.T. Huang, J.W. Fahey, P. Talalay, in: L. Packer (Ed.), Antioxidant Food Supplements in Human Health, Academic Press, New York, 1999, pp. 203–228.
[2] D. DellaPenna, J. Pure Appl. Chem. 71 (1999) 2205–2212.
[3] M. Paolini, G. Cantelli-Forti, P. Perocco, G.F. Pedulli, S.Z. Abdel-Rahman, M.S. Legator, Nature 398 (1999) 760–761.
[4] P.M. Bramley, Phytochemistry 54 (2000) 233–236.
[5] W. Stahl, H. Sies, in: L. Packer (Ed.), Antioxidant Food Supplements in Human Health, 1999, pp. 183–202.
[6] H. Sies, W. Stahl, Proc. Soc. Exp. Biol. Med. 218 (1998) 121–124.
[7] H. Sies, W. Stahl, in: Ozben (Ed.), Free Radicals, Oxidative Stress, and Antioxidants, Plenum Press, New York, 1998, pp. 315–322.
[8] A.V. Rao, S. Agarwal, Nutr. Res. 19 (1999) 305–323.
[9] L. Zechmeister, R.B. Escue, J. Am. Chem. Soc. 66 (1944) 322–330.
[10] L. Pauling, Fortschr. Chem. Organ. Naturstoffe 3 (1939) 203–235.
[11] G.A. Chass, K.P. Chasse, A. Kucsman, L.L. Torday, J.G. Papp, Theochem (2001) (in press).
[12] G.A. Chass, M.L. Mak, E. Deretey, I. Farkas, L.L. Torday, J.G. Papp, D.S.R. Sarma, A. Agarwal, S. Chakravarthi, S. Agarwal, A.V. Rao, Theochem (2001) (in press).
[13] M.J. Frisch, G.W. Trucks, H.B. Schlegel, P.M.W. Gill, B.G. Johnson, M.A. Robb, J.R. Cheeseman, T. Keith, G.A. Petersson, J.A. Montgomery, K. Raghavachari, M.A. Al-Laham, V.G. Zakrzewski, J.V. Ortiz, J.B. Foresman, J. Cioslowski, B.B. Stefanov, A. Nanayakkara, M. Challacombe, C.Y. Peng, P.Y. Ayala, W. Chen, M.W. Wong, J.L. Andres, E.S. Replogle, R. Gomperts, R.L. Martin, D.J. Fox, J.S. Binkley, D.J. Defrees, J. Baker, J.P. Stewart, M. Head-Gordon, C. Gonzalez, J.A. Pople, GAUSSIAN 94, Revision E.2, Gaussian, Inc., Pittsburgh PA, 1995.