Theor Chem Acc (2003) 110:19–27
DOI 10.1007/s00214-003-0442-6

Theoretical
Chemistry Accounts

Regular article

Ab initio and density functional theory based studies on collagen triplets

R. Parthasarathi, B. Madhan, V. Subramanian, T. Ramasami

Chemical Laboratory, Central Leather Research Institute, Adyar, Chennai 600 020, India

Received: 13 December 2001 / Accepted: 19 January 2003 / Published online: 16 May 2003
© Springer-Verlag 2003

Abstract. An understanding of the amino acid sequence dependent stability of polypeptides is of renowned interest to biophysicists and biochemists, in order to identify the nature of forces that stabilize the three-dimensional structure of proteins. In this study, the role of various collagen triplets influencing the stability of collagen has been addressed. It is found from this study that proline can stabilize the collagen triplet only when other residues are also in the polyproline II conformation. Solvation studies of various triplets indicate that the presence of polar residues increases the free energy of solvation. Especially the triplets containing arginine residues displays a higher solvation free energy. The chemical hardness of all the triplets in collagen-like conformation has been found to be higher than that in the extended conformation. Studies on Gly–*X*–*Y*, Gly–*X*–Hyp, and Gly–Pro–*Y* triplets confirm that there will be local variations in the stability of collagen along the entire sequence.

Keywords: Collagen triplets – Collagen-like conformation – Extended conformation – Relative energy – Solvation free energy

Introduction

Collagen is an extremely important protein, which provides mechanical strength and structural integrity to various connective tissues of the body. Ramachandran [1] proposed the concept of a triple helical structure for collagen employing fiber diffraction theory with stereochemical consideration. The backbone torsional angles φ and ψ corresponding to collagen-like peptides fall in the region of –76°, 127° in the Ramachandran plot [2]. Independently Rich and Crick [3] also proposed a triple helical structure on the basis of model building studies using more stringent stereochemical criteria (single hydrogen bond between three polypeptide chains). It is now established that Ramachandran and Kartha [4] provided the first correct model for the structure of collagen. The role of water in the stabilization of collagen has also been clearly brought out by Ramachandran and Chandrasekaran [5]. The variety of chemical and biochemical properties of collagen strongly depend on its three-dimensional structure and unique amino acid composition [6]. Nineteen different types of collagen have been identified to date [7]. These types differ in the number and order of the amino acid residues present in their sequence, and the way in which they are associated with one another. Type I is the most abundant collagen, constituting almost 90% of the total collagen present in mammals. Type I collagen forms the principle component of skin, bone, and tendon [6]. The primary sequence analyses have revealed that collagen has a high content of glycine and imino acids, like proline and hydroxyproline. Gly–*X*–*Y* is the most common repeating triplet unit in collagen. In type I collagen approximately 340 triplets give rise to a twisted rodlike conformation of length 3,000 Å.

Collagen is a large protein and understanding of the various factors influencing the stability of collagen is derived on the basis of model studies from synthetic collagen-like peptide [8, 9, 10, 11]. In collagen, some of the triplets occur more frequently than others. The various triplet sequences occurring in type I collagen have been discussed by Heidemen and Roth [12]. Typically, Gly–*X*–Hyp and Gly–Pro–*Y* appear about 25% of the time each in the primary sequence and the other Gly–*X*–*Y* sequences occur about 50% of the time in the native collagen [12]. The stability and other properties of collagen have been related to the presence of these triplets; hence, several model studies have been initiated with a view to understanding the importance of various amino acids and triplets in the stabilization of collagen. Raines and coworkers [13, 14] have demonstrated that

Correspondence to: V. Subramanian
e-mail: subuchem@hotmail.com

the inductive effect due to the presence of an electron-withdrawing group on the pyrrolidine ring of proline residues has a significant influence on the structure of collagen. Recently Improta and coworkers [15] have addressed the role of stereoelectronic effects in the stability of collagen through quantum mechanical studies of dipeptide analogues in aqueous solution. It was found that the Hartree–Fock (HF)/6-31G* calculation reproduces the relative energy of proline, hydroxyproline, and fluoroproline in the down and up conformations.

In order to assess various properties leading to the stability of collagen triple helix and to the design of new synthetic collagen mimetics, understanding of the role of primary triplet sequences is necessary. Since the solvation phenomenon is intimately related to the various structure–function relationships in biological systems, an attempt to address the solvation properties of various triplets becomes important. Therefore in the present investigation efforts have been made to predict the stability and free energy of solvation of various collagen triplets by employing theoretical calculations based on ab initio quantum chemistry and density functional theory. Chemical hardness, a global quantum chemical descriptor, which provides information on the stability of chemical species, has been used to quantify the stability of various collagen triplets.

Computational details

Calculation of relative energy of triplets in the gas phase

The triplets with frequent occurrence in type I collagen were chosen for the study and are presented in Table 1. Standard three-letter codes for amino acids have been used throughout this article. The φ and ψ dihedral angles of triplets Gly–*X*–*Y*, Gly–Pro–*X*, and Gly–*X*–Hyp in extended and collagen-like conformation are given in Table 2. The energies of triplets in both collagen-like and extended conformations were computed in order to evaluate the stability of various collagen triplets. Two different geometries for various triplets have been generated. The geometry of the collagen triplets was generated with φ, ψ dihedral angles corresponding to the allowed regions of the collagen triple helix in the Ramachandran plot [2]. The other geometry of the triplets was generated in the extended conformation. In the construction of the extended conformation of various triplets, the dihedral angles of proline and hydroxyproline residues correspond to the polyproline conformation (Table 2). The total electronic energy of the triplets was calculated using both HF and Becke three-parameter Lee–Yang–Parr (B3LYP) methods employing a 6-31G* basis set. The relative energy of a triplet in the gas phase is calculated as

$$\Delta E_{\mathrm{g}} = E_{\mathrm{ext,g}} - E_{\mathrm{coll,g}}, \tag{1}$$

where $E_{\mathrm{ext,g}}$ and $E_{\mathrm{coll,g}}$ are the total energies of the triplets in the extended and collagen-like conformations in the gas phase, respectively.

Calculation of solvation free energy

The polarizable continuum model (PCM) developed by Tomasi and his group has been used to study the solvation of various collagen triplets in both collagen-like and extended conformations [16, 17, 18, 19, 20, 21] in water. Using the total energies obtained from the PCM calculation, the relative energy in the aqueous phase is calculated as

$$\Delta E_{\mathrm{sol}} = E_{\mathrm{ext,sol}} - E_{\mathrm{coll,sol}}, \tag{2}$$

where $E_{\mathrm{ext,sol}}$ and $E_{\mathrm{coll,sol}}$ are the total energies of the triplets in extended and collagen-like conformations in an aqueous medium, respectively.

In the PCM, the solute molecules are embedded in molecular-shaped cavities surrounded by a continuous dielectric medium whose polarization is reproduced by point charges distributed on the cavity surface. The cavity of molecular shape is defined by the united atom model for HF. The solvation free energy, ΔG_{S}, is expressed as

$$\Delta G_{\mathrm{S}} = \Delta G_{\mathrm{ES}} + \Delta G_{\mathrm{CAV}} + \Delta G_{\mathrm{DIS-REP}}, \tag{3}$$

where the subscripts stand for electrostatic (ΔG_{ES}) [17], cavitation (ΔG_{CAV}) [18], and dispersion–repulsion ($\Delta G_{\mathrm{DIS-REP}}$) [19, 20, 21] energies, respectively. It was observed from a recent study by Gontrani et al. [22] that the PCM in combination with the B3LYP/6-31G* level of calculation reproduced the stability of glycine and alanine in neutral and zwitterionic forms. Following the same method, the free energy of solvation of various triplets has also been computed using the PCM in the framework of HF and B3LYP approaches using the 6-31G* basis set. All calculations were made using the Gaussian 98 W suite of programs [23]. In the calculation of the solvation free energy, the standard default values of Gaussian 98 W were used for atomic radii that define the solute–solvent boundary, cavity size, and shape.

Assessment of stability of collagen-like peptides upon single-point mutation

The computation of the propensity of various amino acids to form the collagen-like conformation is an interesting area of research. Several thermodynamic quantities, such as the experimental free

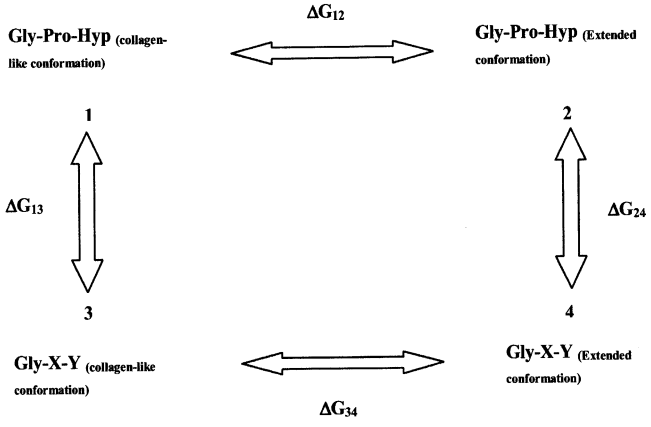
Table 1. Triplets chosen for the study and their frequency of occurrence in type I collagen [12]

Triplets Gly–*X*–*Y*	Number of occurrences	Triplets Gly–*X*–Hyp	Number of occurrences	Triplets Gly–Pro–*Y*	Number of occurrences
Gly–Ala–Arg	9	Gly–Ala–Hyp	20	Gly–Pro–Arg	8
Gly–Ala–Asp	5	Gly–Glu–Hyp	11	Gly–Pro–Gln	7
Gly–Ala–Lys	9	Gly–Gly–Hyp	1	Gly–Pro–Ile	4
Gly–Asp–Ala	5	Gly–Hyp–Hyp	1	Gly–Pro–Lys	7
Gly–Asp–Arg	4	Gly–Leu–Hyp	11	Gly–Pro–Met	3
Gly–Glu–Ala	7	Gly–Phe–Hyp	7	Gly–Pro–Thr	2
Gly–Glu–Arg	10	Gly–Ser–Hyp	10	Gly–Pro–Pro	1
Gly–Lys–Asp	3	Gly–Pro–Hyp	39	Gly–Pro–Ala	31
				Gly–Pro–Val	2
				Gly–Pro–Asp	1
				Gly–Pro–Ser	10

Table 2. ϕ and ψ angles of the triplets in extended and collagen-like conformations

Triplets	Extended conformation			Collagen-like conformation		
	$\phi_{(X/\mathrm{Pro})}$	$\psi_{(X/\mathrm{Pro})}$	$\phi_{(Y/\mathrm{Hyp})}$	$\phi_{(X/\mathrm{Pro})}$	$\psi_{(X/\mathrm{Pro})}$	$\phi_{(Y/\mathrm{Hyp})}$
Gly–*X*–*Y*	–180	–180	–180	–76.0	127	–76.0
Gly–Pro–*Y*	–71.14	–180	–71.14	–71.14	127	–71.14
Gly–*X*–Hyp	–71.14	–180	–71.14	–71.14	127	–71.14

energy, the experimental enthalpy, and the melting temperature, have been used as criteria to order the propensity of various amino acids to form collagen-like peptides. In this study the propensity is estimated by calculating the change in the free energy by considering triplets in the collagen-like conformation and the extended conformation. In the calculation of $\Delta\Delta G$, the sequence Gly–Pro–Hyp was used as the reference system. The change in the stability of the triplet upon replacement of the proline and hydroxyproline residues in the *X* and *Y* positions respectively by other residues were computed using Scheme 1.



Scheme 1

ΔG_{12} is the solvation free-energy change from the collagen-like conformation (1) to the extended conformation (2) of Gly–Pro–Hyp. ΔG_{13} is the free-energy change involved during mutation of the Gly–Pro–Hyp sequence to Gly–*X*–*Y* (3) in the collagen-like conformation. Similarly, processes 2 and 4 deal with the free-energy change (ΔG_{24}) during mutation of the extended conformation of Gly–Pro–Hyp (2) to extended Gly–*X*–*Y* (4). Processes 3 and 4 give the free-energy change (ΔG_{34}) for Gly–*X*–*Y* from the collagen-like conformation (3) to the extended conformation (4). The condition that the sum of the free-energy changes for the entire cycle must be zero can be written as

$$\Delta\Delta G = \Delta G_{34} - \Delta G_{34} = \Delta G_{13} - \Delta G_{24}. \quad (4)$$

From the free-energy values obtained from PCM calculations, the $\Delta\Delta G$ values for various mutated sequences were estimated to assess the propensity of various amino acids to form the collagen-like conformation.

Calculation of chemical hardness

It is well known that the stability of chemical species is related to the corresponding gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). The theoretical basis for these concepts has been provided by density functional theory [24]. Parr and Chattaraj [25] have provided formal proof for the maximum hardness principle (MHP). According to the MHP molecules arrange themselves in such a way so as to be as hard as possible. The inverse relationship between the MHP and the minimum polarization principle (MPP) has also been established [25]. It is evident from recent studies that maximum hardness and minimum polarizability are associated with greater stability [25]. There are numerous illustrations highlighting the usefulness of the MHP in complementing the minimum-energy criteria for stability [26, 27, 28, 29]. The Woodward–Hoffman rule in the light of the MHP has been analyzed by considering the electrocyclic transformation between *cis*-butadiene and cyclobutene as an example [27]. Chattaraj et al. [27] have analyzed the validity of the MHP and the MPP in molecular vibration and internal rotation. The chemical hardness has been employed as a parameter to understand the aromoaticity of many molecules [29]. Prompted by these studies, the stability of various triplets has been assessed on the basis of the calculation of chemical hardness. In addition an attempt has been made to quantify the difference in the hardness between the extended and collagen-like conformations of various triplets using computation of chemical hardness. Parr and Pearson [30] first provided the analytical definition of global hardness of a chemical species as

$$\eta = \left(\frac{\partial^2 E}{\partial N^2}\right)_{v(r)} = \left(\frac{\partial \mu}{\partial N}\right)_{v(r)}, \quad (5)$$

where E is the total energy, N is the number of electrons of the chemical species, and μ is the chemical potential, which is identified as the negative of the electronegativity (χ) as defined by Iczkowski and Margrave [31]. By applying the finite-difference approximation to Eq. (7), we get the operational definition for η as

$$\eta = (\mathrm{IP} - \mathrm{EA})/2, \quad (6)$$

where IP and EA are the ionization potential and electron affinity of the chemical species, respectively.

$$\mathrm{IP} = -E_{\mathrm{HOMO}} \text{ and } \mathrm{EA} = -E_{\mathrm{LUMO}} \quad (7)$$

IP and EA were calculated using Koopmans' theorem in the framework of both HF and B3LYP theory. The hardness of various triplets was calculated using the values of IP and EA.

Results and discussion

A designed primary sequence of a protein is expected to fold into a unique, well-defined structure only if it satisfies two important conditions. First, it must contain elements of positive design so as to thermodynamically stabilize the desired fold; in which folded conformations are generally stabilized by 4–10 kcal/mol [32]. Second, the sequence should contain elements of negative design to create a large energy gap between the native fold and any other folded conformation. Otherwise the protein would assume a molten globulelike ensemble of folds, rather than a unique native structure. Host–guest

peptides and proteins have been used successfully to understand the propensity of individual amino acid residues to form α-helix and β-sheet structures [33]. This approach has been used to probe the contribution of individual amino acids and pairs of amino acid residues to stabilize the triple helical structure of collagen [34, 35, 36]. The effective use of host–guest peptides in modeling the stability of the triple helix calls for knowledge of the frequently occurring triplets in collagen.

Relative energies of various triplets

The relative energies of various collagen triplets calculated in both the gas phase and in an aqueous medium using HF and B3LYP are given in Tables 3, 4, and 5. It is evident from the tables that collagen-like conformations of the triplets are more stable in both gaseous and aqueous media, when compared with the respective extended conformation. The molecular structure of the triplet Gly–Pro–Hyp in extended and collagen-like conformations is shown in Fig. 1a and b, respectively, and the superimposed structures of both conformations are shown in Fig. 1c. The energy difference between collagen-like and extended conformations of this peptide in the gas phase is 9.61 and 8.33 kcal/mol obtained from HF and B3LYP calculations, respectively. Though the Gly–Pro–Hyp sequence imparts greater stability to collagen, the possible reason for the small difference in the relative energy for Gly–Pro–Hyp may be due to the construction of the models. In the construction of the extended peptide model for imino acid containing sequences, the dihedral angles for these residues are fixed to the polyproline II like conformation. Since these

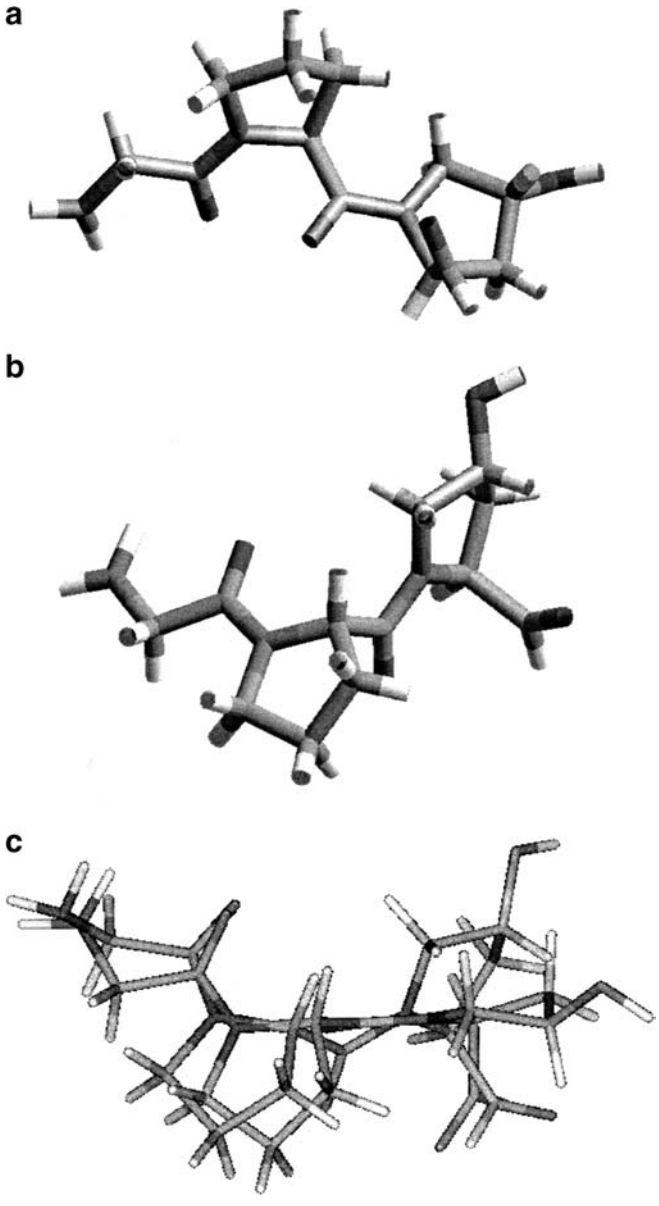


Fig. 1a–c. Geometry of the triplet Gly–Pro–Hyp. **a** Extended conformation, **b** collagen-like conformation, and **c** superimposed structure

Table 3. Relative energy of triplets containing proline residues in extended and collagen-like conformations

Triplets	Relative energy (kcal/mol)			
	HF/6-31G*		B3LYP/6-31G*	
	In vacuo[a]	In solution[b]	In vacuo[a]	In solution[b]
Gly–Pro–Arg	5.33	6.73	3.61	4.78
Gly–Pro–Gln	3.28	1.07	2.148	0.06
Gly–Pro–Ile	78.67	80.94	65.23	67.09
Gly–Pro–Lys	5.49	7.02	3.57	4.89
Gly–Pro–Met	37.43	40.16	30.09	32.18
Gly–Pro–Thr	33.61	35.11	27.52	29.06
Gly–Pro–Pro	9.19	3.34	8.12	3.183
Gly–Pro–Hyp	9.61	2.29	8.33	1.93
Gly–Pro–Ala	4.84	5.97	3.00	3.97
Gly–Pro–Val	62.31	62.47	51.55	51.51
Gly–Pro–Asp	1.95	4.87	0.54	3.11
Gly–Pro–Ser	2.48	1.33	2.56	0.65
Gly–Pro–Flp	9.91	3.11	8.54	2.55

[a]Calculated using Eq. (1)
[b]Calculated using Eq. (2)

Table 4. Relative energy of triplets containing hydroxyproline residues in extended and collagen-like conformations

Triplets	Relative energy (kcal/mol)			
	HF/6-31G*		B3LYP/6-31G*	
	In vacuo[a]	In solution[b]	In vacuo[a]	In solution[b]
Gly–Ala–Hyp	24.71	23.2	19.05	17.77
Gly–Glu–Hyp	25.74	23.49	19.89	17.77
Gly–Gly–Hyp	–4.32	–7.84	–4.08	–7.21
Gly–Hyp–Hyp	9.93	5.51	8.72	5.13
Gly–Leu–Hyp	91.61	88.37	76.63	73.29
Gly–Phe–Hyp	73.45	69.03	58.58	54.59
Gly–Ser–Hyp	27.69	27.96	21.88	22.07
Gly–Pro–Hyp	9.60	2.29	8.33	1.93

[a]Calculated using Eq. (1)
[b]Calculated using Eq. (2)

peptides are already twisted in the polyproline II conformation, the energy required for a further twist is only minimal.

The relative energies of collagen triplets containing proline in the *X* position (second position) and other amino acid residues in the *Y* (third position) are displayed in Table 3. It is seen from this table that these triplets prefer to form the collagen-like conformation than the extended conformation. The arginine residue in the *Y* position increases the stability of collagen-like triplets in the solvent owing to its polar nature. The glutamine residue in the *Y* position destabilizes the collagen-like conformation in solvent, when compared to the same in the gas phase. The relative energies of triplets containing Gly–*X*–Hyp sequences are given in Table 4. From the table it is clear that Gly–Gly–Hyp does not favor the collagen-like conformation. It is observed from the earlier work of Shah et. al. [34] that the presence of glycine in the second position leads to the destabilization of collagen. A similar destabilization effect of the glycine residue in the second position in the triplet Gly–Gly–Hyp was observed in the present study, where the extended conformation of the triplet was found to be more stable than the collagen-like conformation.

The relative energy for various collagen triplets containing other amino acids in the *X* and *Y* positions are given in Table 5. The HF calculation reveals that Gly–Glu–Arg has a small energy difference between the two conformations in both the gas and solvent phases. Considering the B3LYP values, it is possible to note that Gly–Glu–Arg and Gly–Ala–Asp have the smallest energy difference between the collagen-like and extended conformations. Comparison of the relative energies of all Gly–*X*–*Y* triplets reveals that Gly–Lys–Asp has a large energy difference between collagen-like and extended conformations in both the gas phase and the aqueous phase.

Table 5. Relative energy of other triplets containing other amino acids in *X* and *Y* positions, without proline and hydroxyproline residues in extended and collagen-like conformations

Triplets	Relative energy (kcal/mol)			
	HF/6-31G*		B3LYP/6-31G*	
	In vacuo[a]	In solution[b]	In vacuo[a]	In solution[b]
Gly–Ala–Arg	2.16	8.09	–0.55	4.33
Gly–Ala–Asp	–0.51	3.83	–3.01	0.24
Gly–Ala–Lys	2.45	8.83	–0.34	4.74
Gly–Asp–Ala	1.69	10.09	–2.33	4.48
Gly–Asp–Arg	1.49	6.75	–1.94	2.06
Gly–Glu–Ala	0.33	4.31	–1.76	1.34
Gly–Glu–Arg	0.85	2.01	–1.47	–0.85
Gly–Lys–Asp	15.87	24.39	9.69	16.89

[a]Calculated using Eq. (1)
[b]Calculated using Eq. (2)

The triplets containing proline or hydroxyproline are more stable in the collagen-like conformation than triplets consisting of other amino acid residues in the *X* and *Y* positions. It is well known that, proline being an imino acid with a five-membered ring, is sterically restricted in rotation around the N–C_α bond; thus it has a limited ϕ value of about $-63 \pm 15°$ [37]. Because of this, proline cannot be found in all the known main conformations, and proline disrupts regular secondary structural elements. Since proline can disrupt, induce, and stabilize the secondary structure, it is structurally an important amino acid. Proline can stabilize the secondary structure only when the allowed values of all other residues coincide with the allowed geometry of proline. The conformational energy minimum of proline has been calculated and found was to be $\phi = -75°$, $\psi = 145°$ for *trans*-proline [15]. In the case of the collagen triplet, if the other constituent of the amino acid residues has a geometry corresponding to proline, then it will stabilize the triplets. It is interesting to mention that all amino acid residues in collagen adopt the polyproline II like conformation. This clearly indicates the significance of proline and hydroxyproline residues in stabilizing the collagen triple helix. The relative energies of triplets containing proline and hydroxyproline in an aqueous environment are found to decrease when compared to gas-phase calculations. In the case of the

Table 6. Solvation free energy of various collagen triplets in extended and collagen-like conformations calculated using the polarizable continuum model

Triplets	B3LYP (kcal/mol)		HF (kcal/mol)	
	Extended	Collagen-like	Extended	Collagen-like
Gly–Ala–Arg	–28.04	–32.92	–34.56	–40.50
Gly–Ala–Asp	–20.63	–23.89	–26.33	–30.66
Gly–Ala–Lys	–17.82	–22.98	–22.49	–28.87
Gly–Asp–Ala	–17.36	–24.18	–22.66	–31.06
Gly–Asp–Arg	–31.42	–35.43	–39.09	–44.34
Gly–Glu–Ala	–21.09	–24.18	–26.98	–30.96
Gly–Glu–Arg	–34.71	–35.29	–42.84	–44.01
Gly–Lys–Asp	–20.72	–27.92	–26.49	–35.00
Gly–Ala–Hyp	–23.98	–22.69	–29.85	–28.34
Gly–Glu–Hyp	–27.92	–25.79	–34.91	–32.66
Gly–Gly–Hyp	–27.53	–24.41	–33.79	–30.27
Gly–Hyp–Hyp	–27.79	–24.22	–34.76	–30.33
Gly–Leu–Hyp	–19.22	–15.88	–24.67	–21.43
Gly–Phe–Hyp	–24.84	–20.85	–31.71	–27.29
Gly–Ser–Hyp	–27.03	–27.23	–33.97	–34.25
Gly–Pro–Arg	–28.52	–29.69	–35.29	–36.69
Gly–Pro–Glu	–22.21	–20.14	–28.39	–26.18
Gly–Pro–Ile	–10.39	–12.25	–14.62	–16.88
Gly–Pro–Lys	–19.85	–21.17	–24.96	–26.48
Gly–Pro–Met	–14.55	–16.64	–18.90	–21.63
Gly–Pro–Thr	–20.11	–25.76	–21.65	–27.27
Gly–Pro–Pro	–17.11	–12.16	–22.12	–16.27
Gly–Pro–Hyp	–24.37	–17.97	–30.34	–23.02
Gly–Pro–Ala	–16.12	–17.09	–20.77	–21.90
Gly–Pro–Val	–12.82	–12.78	–17.21	–17.39
Gly–Pro–Asp	–20.89	–23.45	–26.84	–29.77
Gly–Pro–Ser	–19.23	–22.44	–24.64	–28.45
Gly–Pro–Flp	–18.81	–12.83	–24.32	–17.52

other Gly–*X*–*Y* collagen triplet given in Table 5, the collagen-like conformation was found to be more stable in an aqueous environment when compared to the gas phase.

Free energy of solvation of various triplets

Solvation of biomolecules is an important phenomenon and various methods used to probe the solvation of biomolecules have been reviewed [38]. In this investigation the free energy of solvation for various triplets was calculated using HF/6-31G* and B3LYP/6-31G* methods by employing the PCM, developed by Tomasi and Persico [16].

The calculated free energy of solvation (ΔG) of various triplets is given in Table 6. From the table it is seen that the free energy of solvation is numerically less for triplets containing hydrophobic residues when compared to those containing polar residues in the *X* and *Y* positions. The presence of arginine in the third position significantly influences the solvation of triplets when compared to other triplets.

The total surface area of the triplets accessible to the solvent is observed to be larger in the case of the collagen-like conformation when compared to the extended conformation of the same triplet; hence, the solvation free energy of the triplet in the collagen-like conformation is higher than that of the corresponding extended conformation. It is interesting to monitor the effect of solvation on the solute, by looking at the values of the total dipole moment. It is evident from the calculation that solvent polarization leads to redistribution of the charge of the solute and hence the dipole moment changes. An increase in the dipole moment of all the triplets has been observed upon solvation. Gly–Asp–Arg is the triplet that has the highest dipole moment, both in extended and collagen-like conformations.

Table 7. Change in free energy of collagen triplets relative to Gly–Pro–Hyp. $-\Delta\Delta G$ was calculated from Eq. (4)

Triplets	$\Delta\Delta G$	
	HF (kcal/mol)	B3LYP (kcal/mol)
Gly–Pro–Hyp	0	0
Gly–Ala–Arg	13.26	11.28
Gly–Ala–Asp	11.65	9.66
Gly–Ala–Lys	13.7	11.56
Gly–Asp–Ala	15.72	13.22
Gly–Asp–Arg	12.57	10.41
Gly–Glu–Ala	11.3	9.49
Gly–Glu–Arg	8.49	6.98
Gly–Lys–Asp	15.83	13.6
Gly–Ala–Hyp	5.81	5.11
Gly–Glu–Hyp	5.07	4.27
Gly–Gly–Hyp	3.8	3.28
Gly–Hyp–Hyp	2.89	2.83
Gly–Leu–Hyp	4.08	3.06
Gly–Phe–Hyp	2.9	2.41
Gly–Ser–Hyp	7.6	6.6
Gly–Pro–Arg	8.72	7.57
Gly–Pro–Glu	5.11	4.33
Gly–Pro–Ile	9.58	8.26
Gly–Pro–Lys	8.84	7.72
Gly–Pro–Met	10.05	8.49
Gly–Pro–Thr	12.94	12.05
Gly–Pro–Pro	1.47	1.45
Gly–Pro–Ala	8.45	7.37
Gly–Pro–Val	7.5	6.36
Gly–Pro–Asp	10.25	8.96
Gly–Pro–Ser	11.13	9.61
Gly–Pro–Flp	0.52	0.42

Assessment of stability of collagen using $\Delta\Delta G$ calculations

The propensity of various triplets to form the collagen-like conformation was assessed using the calculated values of $\Delta\Delta G$. The $\Delta\Delta G$ values of various triplets calculated employing Scheme 1 using Gly–Pro–Hyp as reference are given in Table 7. It is evident from the table that Gly–Pro–Hyp has a high propensity to form the collagen-like conformation when compared to other triplets. The calculated $\Delta\Delta G$ values range from 0.0 to 15.8 kcal/mol at the HF/6-31G* level of calculation and from 0 to 13.6 kcal/mol at the B3LYP/6-31G* level of calculation. The $\Delta\Delta G$ values of Gly–Pro–Flp and Gly–Pro–Pro were found to be closer to the $\Delta\Delta G$ values of Gly–Pro–Hyp. The calculated values are in reasonable agreement with the experimental values derived from the melting temperature measurements of collagen-like peptides. A previous simulation study on the relative stabilities of glycine-to-alanine mutation in collagen-like peptides involved a relative free-energy change, $\Delta\Delta G$, of 10.76 kcal/mol [39]. It is evident from the experimental results that the two collagens differ in stability by 11.9 kcal/mol [40]. Though the $\Delta\Delta G$ values calculated in this study do not directly correspond to the glycine-to-alanine mutation, the free energies observed are similar to glycine-to-alanine mutation. In the present study, to obtain $\Delta\Delta G$, only the triplets were considered, whereas for realistic estimates of $\Delta\Delta G$, it is necessary to consider the triple helix model of collagen with appropriate triplets in the sequence. In addition, the translational, rotational, and vibrational contributions were not included in the present calculations of $\Delta\Delta G$. Future work is in progress to evaluate the importance of molecular motions in the prediction of propensity of various triplets to form the collagen-like conformation by considering the triple helical model of collagen.

Chemical hardness for various triplets

Density functional based reactive descriptors have been widely used to understand the stability and reactivity of various chemical systems. There are several reports on the use of chemical hardness to establish the preferred

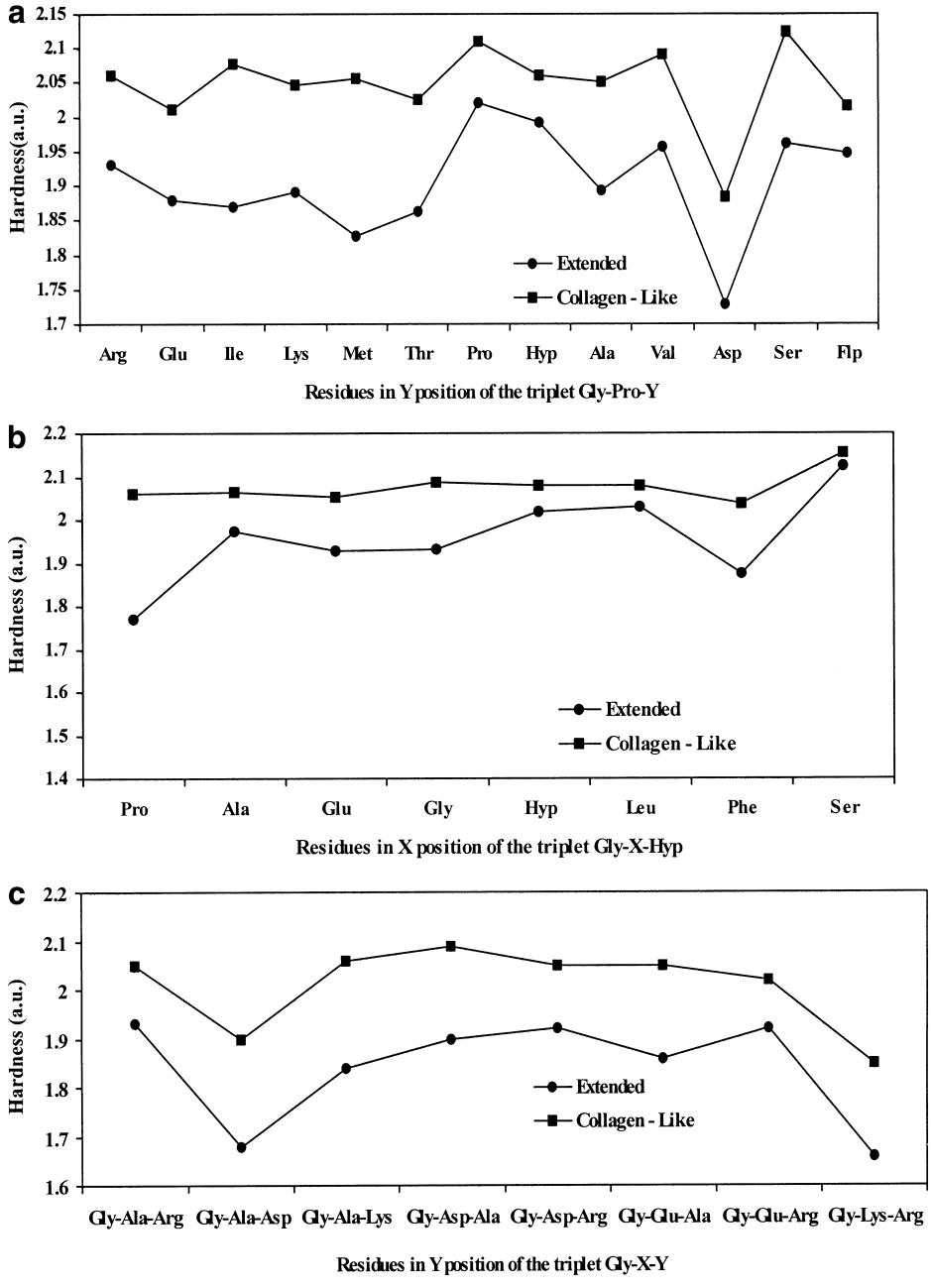


Fig. 2. Hardness values of the triplets in both extended and collagen-like conformations determined by the B3LYP method (6-31G*) **a** containing the sequence Gly–Pro–*Y*, **b** containing the sequence Gly–*X*–Hyp, and **c** containing the sequence Gly–*X*–*Y* (without Pro and Hyp)

conformations of molecules [26, 27, 29, 41]. Both the MHP and the MPP have been widely used to understand the stability and reactivity of several molecules. The variations of hardness and polarizability along the reaction coordinate (hardness and polarization profiles) have also been used to understand the stability and reactivity [29]. Kolandaivel and Senthilkumar [41] have used hardness as a parameter to identify the stablest conformation. In this investigation, the chemical hardness of various collagen triplets was calculated using the B3LYP method by employing the 6-31G* basis set. The calculated chemical hardness values for various triplets in both collagen-like and extended conformations are shown in Fig. 2a. From the figure it is evident that the triplets in the collagen-like conformation exhibit larger hardness values than those in the extended conformation. Hence, the triplets are more stable in the collagen-like conformation than the extended conformation.

The hardness of the triplets of sequences Gly–Pro–*Y* is shown in Fig. 2a. The triplets Gly–Pro–Hyp and Gly–Pro–Arg, which are experimentally proven to contribute towards the stability of collagen [11], have been found to have similar hardness values of around 2.05 eV in the collagen-like conformation. The B3LYP/6-31 G* level of theory predicted that the collagen-like conformation of Gly–Pro–Hyp is stabler than the extended conformation by 0.46 kJ/mol. The chemical hardness of the triplet sequences Gly–*X*–Hyp is shown in Fig. 2b. The hardness

values of the triplets containing hydroxyproline residues and proline residues were comparable, and were found to be around 2 eV in the collagen-like conformation. The hardness of triplets of sequences Gly–*X*–*Y* is shown in Fig. 2c and the hardness values were found to be lower than those of the triplets containing proline and hydroxyproline residues. In an earlier study by Ackerman et al. [36] it was shown that the presence of aspartic acid in the triplet sequence does not favor the folding of collagen-like peptides. It is also evident from the hardness calculation that the presence of aspartic acid in the *Y* position results in a lower hardness value, when compared to other triplets. This result further substantiates findings from folding experiments on collagen-like peptides.

Role of various triplets in the stability of collagen

Unraveling the importance of various amino acids in the *X* and *Y* positions will help to understand the forces that stabilize the collagen triple helix, to predict the regions of stability in the triple helix, and de novo design of collagen-like peptides. It is evident from the present study that both the relative energy, the free energy of solvation, and $\Delta\Delta G$ indicate that the presence of each amino acid in the triplet influences the propensities in forming the triple helix in collagen. The results presented in Tables 1, 2, 3, 4, 5, 6, and 7 clearly indicate variations in the stability and the solvation between triplets. Hence there will be local variations influencing the stability of the three-dimensional structure of the collagen triple helix [11]. The relative energy calculations indicate that the presence of proline or hydroxyproline in the triplets results in much higher relative energies for the formation of the collagen-like conformation than the triplets containing other residues. Previous comprehensive studies on various host–guest peptides in collagen-like peptides revealed *X* = Pro and *Y* = Hyp to be the stablest structure, with a melting temperature of 45.5 °C and the replacement of hydroxyproline by arginine resulting in a triple helical structure with the same melting temperature [11]. The relative energies and hardness values for both triplets were found to be similar. A triple helix containing 27 amino acid residues for every polypeptide chain was used for the experiments and apart from that there will be interchain and intrachain interactions involved in the experimental model system, but the theoretical calculations in the present study were carried out for single-chain triplets, which do not involve in any interactions as observed in the case of the experiments.

There has been a dispute regarding the role of hydroxyproline in providing stability to the collagen molecule; some of the literature cites the involvement of hydroxyproline in the water-mediated intramolecular hydrogen bond and certain other references state the inductive effect of the hydroxyproline. Collagen-like peptide containing triplets $(\text{Gly–Pro–Flp})_n$ also exhibit higher melting temperatures [42]. The relative energy value, $\Delta\Delta G$, and the hardness values of the triplet Gly–Pro–Flp have been found to be almost similar to the values of Gly–Pro–Hyp.

Conclusions

The relative energy calculations on the triplets have revealed that the collagen-like conformation is energetically stabler than the extended conformation. It is evident from the relative energy values that the triplets can stabilize collagen-like conformations only when the constituting residues adopt the polyproline conformation. This is mainly due to the role played by the presence of proline and hydroxyproline in the *X* and *Y* positions. On the basis of the changes in the relative stability of collagen-like triplets, it is possible to observe that the stability of collagen varies locally along the entire sequence. A recent experimental study on the establishment of the relative stabilizing effect of different Gly–*X*–Hyp and Gly–Pro–*Y* sequences confirms the previous finding regarding the local variation in the stability of collagen [11]. The $\Delta\Delta G$ values provided evidence for the stability offered by each triplet to form the collagen-like conformation. The calculated chemical hardness values of all the triplets are almost similar and they were found to be greater for the triplets in the collagen-like conformation than for those in the extended conformation. The study indicates that all the collagen triplets prefer to form the stable collagen-like conformation except Gly–Gly–Hyp. The results generated from this investigation have provided detailed knowledge on the stability and solvation of collagen triplets, and this will certainly help researchers working in the area of de novo protein design, and in particular synthesis of model collagen-like peptides.

Acknowledgement. The authors thank Professor J. Tomasi for his valuable scientific suggestions.

References

1. Ramachandran GN (1988) Int J Pept Protein Res 31:1
2. Ramachandran GN, Sasisekharan V (1968) In: Anfinsen CB, Anson ML, Edsall JT, Richards FM (eds) Advances in protein chemistry, vol 23. Academic, New York, p 283
3. Rich A, Crick FH (1961) J Mol Biol 3:483
4. Bella J, Eaton M, Brodsky B, Berman HM (1994) Science 266:75–81
5. (a) Ramachandran GN, Chandrasekaran R (1968) Biopolymers 6:1649; (b) Ramachandran GN, Kartha G (1955) Nature 176:593
6. (a) Nimni ME, Harkness RD (1988) In: Nimni ME (ed) Collagen. CRC, Boca Raton, p 1; (b) Berisio R, Vitagliano L, Mazzarella L, Zagari A (2002) Curr Pharm Des 9:107
7. Prockop DJ, Kivirikko KI (1995) Ann Rev Biochem 64:403
8. Goodman M, Melacini G, Feng Y (1996) J Am Chem Soc 118:10928
9. Fertala A, Sieron AL, Adachi E, Jimenez SA (2001) Biochemistry 40:14422
10. Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B (2002) J Mol Biol 316:385
11. Persikov AV, Ramshaw JAM, Kirkpatrick A, Brodsky B (2000) Biochemistry 39:14960

12. Heidemen E, Roth W (1982) Adv Polym Sci 43:143
13. (a) Jenkins CL, Raines RT (2002) Nat Prod Rep 19:49; (b) DeRider ML, Wilkens SJ, Waddell MJ, Bretscher LE, Weinhold F, Raines RT, Markley JL (2002) J Am Chem Soc 124:2497
14. Kersteen EA, Raines R (2001) Biopolymers 59:24
15. (a) Improta R, Mele F, Crescenzi O, Benzi C, Barone V (2002) J Am Chem Soc 124:7857; (b) Improta R, Benzi C, Barone V (2001) J Am Chem Soc 123:12568.
16. Tomasi J, Persico M (1994) Chem Rev 94:2027
17. (a) Miertus S, Scrocco E, Tomasi J (1981) Chem Phys 55:117; (b) Miertus S, Tomasi J (1982) Chem Phys 65:239
18. Cossi M, Tomasi J, Cammi R (1995) Int J Quantum Chem Quantum Chem Symp 29:695
19. Floris FM, Tomasi J (1980) J Comput Chem 10:616
20. Floris FM, Tomasi J, Pascual-Ahuir JL (1991) J Comput Chem 12:784
21. Cossi M, Mennucci B, Cammi R (1996) J Comput Chem 17:57
22. Gontrani L, Mennucci B, Tomasi J (2000) J Mol Struct (THEOCHEM) 500:113
23. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA Jr, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzalez C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Gonzalez C, Head-Gordon M, Replogle ES, Pople JA (1998) Gaussian 98, revision A.7. Gaussian, Pittsburgh, PA
24. Parr RG, Yang W (1989) Density functional theory of atoms and molecules. Oxford University Press, New York
25. (a) Parr RG, Chattaraj PK (1991) J Am Chem Soc 113:1854; (b) Ghanty TK, Ghosh SK (1996) J Phys Chem 100:17429; (c) Pearson RG (1993) Acc Chem Res 26:250
26. (a) Kolandaivel P, Jayakumar N (2000) J Mol Struct (THEOCHEM) 507:197; (b) Aliste MP (2000) J Mol Struct (THEOCHEM) 507:1
27. (a) Chattaraj PK, Fuentealba P, Gomez B, Contreras R (2000) J Am Chem Soc 122:348; (b) Chattaraj PK, Fuentealba P, Jaque P, Toro-Labbe A (1999) J Phys Chem A 103:9307
28. Subramanian V, Sivanesan D, Amutha R, Padmanabhan J, Ramasami T (1998) Chem Phys Lett 294:285
29. Proft FD, Geerlings P (2001) Chem Rev 101:1451; (b) Ghanty TK, Ghosh SK (2002) J Phys Chem 106:4200
30. Parr RG, Pearson RG (1983) J Am Chem Soc 105:7512
31. Iczkowski RP, Margrave JL (1961) J Am Chem Soc 83:3547
32. DeGardo WF, Summa CM, Povane V, Nastri F, Lombardi A (1999) Annu Rev Biochem 68:779
33. (a) Blaber M, Zhang X-J, Lindstrom JD, Pepiot SD, Baase WA, Matthews BW (1994) J Mol Biol 235:600; (b) Kim CA, Berg JM (1993) Nature 362:267
34. Shah, NK, Sharma M, Kirkpatrick A, Ramshaw JAM, Brodsky B (1997) Biochemistry 36:5878
35. Yang W, Chan VC, Kirkpatrick A, Ramshaw JAM, Brodsky B (1997) J Biol Chem 272:28837
36. Ackerman MS, Bhates M, Shenoy N, Beck K, Ramshaw JAM, Brodsky B (1999) J Biol Chem 274:7668
37. Cantor CR, Schimmel PR (1980) Biophysical chemistry – The conformation of biological macromolecules. Freeman, San Francisco
38. (a) Cramer CJ, Truhlar DG (1999) Chem Rev 99:2161; (b) Orozco M, Luque FJ (2000) Chem Rev 100:4187.
39. Mooney SD, Huang CC, Kollman PA, Klein TE (2001) Biopolymers 58:347
40. Bella J, Eaton M, Brodsky B, Berman H (1994) Science 266:75
41. Kolandaivel P, Senthilkumar K (2001) J Mol Struct (THEOCHEM) 535:61
42. (a) Eberhardt ES, Panasik N Jr, Raines RT (1996) J Am Chem Soc 118:12261