J Mol Model (2013) 19:1901–1911
DOI 10.1007/s00894-013-1754-7

ORIGINAL PAPER

Investigations of dipeptide structures containing pyrrolysine as N-terminal residues: a DFT study in gas and aqueous phase

Gunajyoti Das

Received: 21 November 2012 / Accepted: 2 January 2013 / Published online: 19 January 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract A set of six dipeptides containing pyrrolysine invariably at their N-terminal positions is studied in gas and aqueous phase using a polarizable continuum model (PCM). The molecular geometries of the dipeptides are fully optimized at B3LYP/6-31++G(d,p) level of theory and a second derivative (frequency) analysis confirms that all the optimized geometries are true minima. The effects of solvation and identity of the varying C-terminal residue on the energetics, structural features of the peptide planes, values of the ψ and ϕ dihedrals, geometry around the α-carbon atoms and theoretically predicted vibrational spectra of the dipeptides are thoroughly analyzed. Solvation effects are found to modify the gas phase conformation of the dipeptides around ψ dihedrals while the identity of the varying C-terminal residue affect the values of ϕ, planarity of the peptide planes and geometry around the α-carbon atoms. The presence or absence of three types of intramolecular H-bonds, namely O…H–N, N…H–N and O…H–C that leave noticeable signatures in the IR spectra, play crucial roles in influencing the geometry of the peptide planes and in determining the energetics of the dipeptides.

Keywords Dipeptides · Peptide plane · Pyrrolysine · Solvation effects · Vibrational frequencies

Introduction

Pyrrolysine (Pyl) is an important constituent in the active site of methylamine methyltransferases involved in methylamine metabolism in methanogenic archaea. In this group of archaea, pyrrolysine is co-translationally inserted into protein in response to a canonical stop codon (UAG) which in other organisms functions as a terminator signal during the translation process of protein biosynthesis [1]. Pyrrolysine, a lysine homologue with chemical identity N^{ϵ}-[(4*R*,5*R*)]-4-methyl-1-*pyrro*line-5-carbonyl]-L-*lysine*, is therefore considered as the 22nd genetically encoded natural amino acid [2–7]. Since its discovery in 2002, there have been prolific studies to understand the biosynthetic pathway which is still unclear [8], on metal-binding affinity/selectivity of pyrrolysine [9] and on the scope of synthesizing pyrrolysine analogues [10, 11]. However, literature survey reveals that the structural aspects of pyrrolysine containing dipeptides are not explored yet which to a large extent determine the dynamic properties and functional specificity of the proteins and polypeptides containing pyrrolysine.

A set of highly controlled reactions dictates the polymerization of solitary amino acid residues on ribosome of the cell forming the polypeptides or proteins which are at the center of action in almost all the biological processes. Since it is difficult to implement theoretical or computational approaches directly to the large systems such as polypeptides or proteins, model systems have to be studied first to understand the properties of the building units like the solitary amino acids or dipeptides. Again, it is of fundamental importance to determine the structural details of a biological molecule in aqueous solution since the vast majority of biochemical processes occur in an aqueous environment. It has now been realized that computational techniques are indispensable in elucidating atomic level structural information about biologically active molecules owing to certain limitations of experimental techniques as pointed out in the literature [12–14]. The low-energy structures and their related properties derived from such computations have a meaningful relationship to their presence and functional activities performed in the macromolecular context of real

Electronic supplementary material The online version of this article (doi:10.1007/s00894-013-1754-7) contains supplementary material, which is available to authorized users.

G. Das (✉)
Department of Chemistry, North Eastern Hill University, Shillong 793022, India
e-mail: guna_das78@yahoo.co.in

life systems. Gas phase computational studies on dipeptides arising from the genetically encoded amino acids [15–18] have been carried out with a view toward understanding the structural features of small amino acid sequences and their possible roles in imparting the three dimensional structure to proteins. Structural studies [15, 16] on a series of dipeptides have pointed out that in most of the dipeptides the amide planes are never perfectly planar; and these observations have been explained in terms of the cumulative effect of steric hindrance of –R group and H-bonding. It is therefore important to take into account the intramolecular H-bond interactions existing in the dipeptides while studying their structures. Besides serving as model systems, dipeptides themselves have been shown to play numerous key biological roles [19–23].

The effects of solvation on the conformations and energies of dipeptides have been well documented in literature [24–29]. In these studies the energetics and structural features of the dipeptides are analyzed in gas and in various solvent phases in order to understand the effects of the surrounding environment on the stabilities and conformational preferences of the dipeptides. It has been found that in a strong polar solvent like water the interactions among the nearest-neighbor residues of the dipeptides are dramatically modified as compared to those in gas phase, which consequently affects the Ramachandran dihedrals (ψ, ϕ) [30, 31] conferring markedly different conformations to the dipeptides in the aqueous phase. It has also been reported that solvation effects can enhance the planarity of the peptide planes [25].

The focus of this study is to obtain full knowledge about the effects of solvation and identity of the varying C-terminal residue on the energetics, structural features of the peptide planes, geometry about the α-carbon atoms, values of the ψ and ϕ dihedrals, theoretically predicted vibrational spectra, dipole moments, rotational constants and types of intramolecular H-bonding interactions that may play crucial roles in determining the structure and stability of a series of dipeptides containing pyrrolysine at their N-terminal positions. For the purpose of the current work, six dipeptides are constructed by keeping pyrrolysine as a fixed component at their N-terminal positions while the C-terminal positions are varied with six different combinations. The six different amino acids chosen for the C-terminal positions are alanine (Ala), leucine (Leu), aspartic acid (Asp), serine (Ser), asparagine (Asn) and histidine (His). All these amino acid residues are taken as neutral (non-ionic) species. The standard three letter abbreviations are used to represent the amino acids while a particular dipeptide is named by listing the N-terminal residue first. Thus, Pyl-Ala dipeptide corresponds to a structure in which pyrrolysine is in the N-terminal position and alanine in the C-terminal position. Figure 1 schematically represents the chemical structures of the six dipeptides studied here. The atom numbering of the pyrrolysine molecule is given in accordance with the schemes used earlier in various literatures [11, 32]. The C_{13}–N_{16} is the peptide bond of a given dipeptide structure while C_{12} and C_{17} are the α-carbon atoms of the N- and C-terminal residues respectively. To facilitate a clear representation of the intramolecular H-bonding interactions present in the dipeptides some of the hydrogen atoms are named as H_a or H_b. This DFT study is expected to provide the opportunity to know the structural features of pyrrolysine containing dipeptides at an atomic level which in turn may help us to understand the dynamics and functional specificity of proteins containing pyrrolysine, to synthesize a new generation of pyrrolysine analogues, in discovering the biosynthetic pathway of pyrrolysine and in understanding the nature of the genetic code or amino acid code which is still evolving [33].

Computational methodology

The molecular geometries of all the selected pyrrolysine dipeptides are subjected to full geometry optimization and vibrational frequency calculations using the B3LYP/6-31++G(d,p) level of theory [34, 35] of Gaussian 03 package [36]. These computations are conducted in gas as well as in aqueous phase using a polarizable continuum model (PCM) [37]. The accuracy of self-consistent reaction field (SCRF) model in predicting the structures and energetics of dipeptides has already been justified in literature [38]. Absence of imaginary frequency values in the vibrational frequency calculations proves that the optimized geometries are precise minima. Zero point energy (ZPE) corrections are applied to the total energies of all the dipeptides using a correction factor 0.97 [39]. The vibrational frequencies below 1800 cm^{-1} are scaled with 0.977 and for those above 1800 cm^{-1} a correction factor 0.955 is used [40, 41]. Use of diffuse functions is important to take into account the relative diffuseness of lone pair of electrons when a molecule under investigation contains the same [42] while the polarization functions are useful in studying the conformational aspects where stereoelectronic effects play an important role [43].

Results and discussion

Investigations of the numerous parameters involved in dipeptide structure prediction have now been regarded as a pivotal part of the computational studies concerning the structure of proteins and energetics of protein folding [44]. The geometrical parameters considered in this study are expected to give a clear account of the effects of solvation and identity of the varying C-terminal residue on the structural features of the peptide planes, geometry about the α-carbon atoms, values of the ψ and ϕ dihedrals and theoretically predicted vibrational spectra of the pyrrolysine

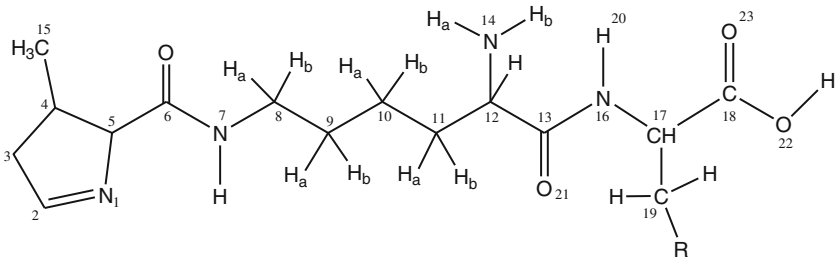

Fig. 1 Schematic representation of chemical structures of the six dipeptides studied

$\psi = N_{14}–C_{12}–C_{13}–N_{16}$

$\phi = C_{13}–N_{16}–C_{17}–C_{18}$

-R = -H (Pyl-Ala)

$-CH(CH_3)_2$ (Pyl-Leu)

$-CO_2H$ (Pyl-Asp)

-OH (Pyl-Ser)

$-CONH_2$ (Pyl-Asn)

$-C_3N_2H_3$ (Pyl-His)

containing dipeptides. This DFT study predicts that there are considerable changes in the structure and stability of the dipeptides in the aqueous phase as compared to those in the gas phase. Table 1 presents the gas and aqueous phase data on total energies, rotational constants and dipole moments of the dipeptides calculated at B3LYP/6-31++G(d,p) level of theory. Tables 2 and 3 list the values of the bond lengths and bond angles of the amide planes of the dipeptides respectively (the gas phase values are given in brackets). The four dihedral angles considered to monitor the planarity of the peptide planes of the dipeptides, viz. $C_{12}–C_{13}–N_{16}–C_{17}$, $O_{21}–C_{13}–N_{16}–H_{20}$, $C_{12}–C_{13}–N_{16}–H_{20}$ and $O_{21}–C_{13}–N_{16}–C_{17}$, are listed in Table 4. Table 4 also lists the two well known Ramachandran backbone dihedral angles ψ ($N_{14}–C_{12}–C_{13}–N_{16}$) and ϕ ($C_{13}–N_{16}–C_{17}–C_{18}$) which are useful in studying the effects of solvation on the dipeptide structures as well as in predicting the overall structure of proteins. Table 5 represents the gas and aqueous phase data on the geometrical parameters considered to examine the geometry around the α-carbon atoms. Table 6 lists some important intramolecular H-bonding interactions that play crucial roles in the energetics and in conferring the observed conformations to the dipeptides in both the phases. Table 7 lists some of the characteristic frequency and intensity values (given in brackets) of the dipeptides calculated at the B3LYP/6-31++G(d,p) level of theory. Figures 2, 3 and 4 depict the optimized geometries of the six dipeptides while Figs. 5, 6 and 7 represent their theoretical IR spectra (scaled with a correction factor 0.955).

Dipeptide structure

As listed in Table 1 all six dipeptide geometries exhibit large values of total dipole moments, ranging from 2.361 to 7.421

Table 1 Calculated total energies[a] (kcal mol^{-1}), rotational constants (GHZ) and dipole moments (Debye) of the pyrrolysine dipeptides in gas and solvent phases using B3LYP/6-31++G(d,p) level of theory

Dipeptides	Phases	Total energies	Rotational constants			Dipole moments
			A	B	C	
Pyl-Ala	Aqueous	−694596.48	0.53663	0.06229	0.05925	12.639
	Gas	−694582.44	0.92269	0.05179	0.05102	4.944
Pyl-Leu	Aqueous	−768560.15	0.50038	0.04313	0.04146	8.864
	Gas	−768546.80	0.39691	0.04534	0.04380	2.361
Pyl-Asp	Aqueous	−812925.34	0.48576	0.04369	0.04196	8.921
	Gas	−812908.35	0.41827	0.04483	0.04390	3.643
Pyl-Ser	Aqueous	−741793.41	0.52459	0.05313	0.04999	9.030
	Gas	−741777.38	0.60246	0.04988	0.04944	3.980
Pyl-Asn	Aqueous	−800454.12	0.50357	0.04451	0.04258	14.152
	Gas	−800434.40	0.42419	0.04490	0.04330	6.240
Pyl-His	Aqueous	−835780.47	0.44678	0.03749	0.03605	13.713
	Gas	−835762.73	0.32679	0.04014	0.03778	7.421

[a]ZPVE corrected; scaled with a correction factor 0.97

Table 2 Calculated bond lengths (in angstrom) for the peptide planes of the pyrrolysine dipeptides; the gas phase values are given in brackets

Dipeptides	C_{12}–C_{13}	C_{13}=O_{21}	C_{13}–N_{16}	N_{16}–H_{20}	N_{16}–C_{17}
Pyl-Ala	1.536 (1.544)	1.236 (1.231)	1.359 (1.360)	1.011 (1.015)	1.453 (1.446)
Pyl-Leu	1.537 (1.541)	1.237 (1.233)	1.357 (1.356)	1.012 (1.016)	1.457 (1.452)
Pyl-Asp	1.535 (1.540)	1.236 (1.232)	1.359 (1.357)	1.013 (1.017)	1.450 (1.446)
Pyl-Ser	1.536 (1.541)	1.234 (1.230)	1.362 (1.359)	1.012 (1.016)	1.453 (1.451)
Pyl-Asn	1.537 (1.541)	1.234 (1.230)	1.364 (1.360)	1.012 (1.016)	1.455 (1.458)
Pyl-His	1.536 (1.540)	1.236 (1.228)	1.359 (1.363)	1.013 (1.016)	1.452 (1.452)
Average	1.536 (1.541)	1.236 (1.231)	1.360 (1.359)	1.012 (1.016)	1.453 (1.451)
MD[a]	0.001 (0.003)	0.002 (0.003)	0.004 (0.004)	0.001 (0.001)	0.004 (0.007)

[a]Maximum deviation from average values

D in gas phase and 8.864 to 14.152 D in aqueous phase, indicating that they have greater polar character and consequently possess greater affinity to polar solvents. Thus, the data on the total energies of dipeptides correctly predicts that the dipeptide geometries are thermodynamically more stable in a strong polar solvent such as water than in gas phase by an energy difference that may range from 13.35 to 19.72 kcal mol^{-1}. The accuracy of DFT method in predicting the rotational constants of conformers of some aliphatic amino acids has been discussed in literature [45, 46]. In the absence of any experimental data on rotational constants and dipole moments these theoretically predicted values may assist experimentalists in determining the other conformers of the six dipeptides studied here.

It is evident from Table 2, which lists the gas and aqueous phase bond length values of the five bonds of the amide planes, i.e., C_{12}–C_{13}, C_{13}=O_{21}, C_{13}–N_{16}, N_{16}–H_{20} and N_{16}–C_{17}, that very little variance in the bond length values results as the identity of the C-terminal residue of a given dipeptide changes. Maximum deviations of 0.007 Å in the gas phase and 0.004 Å in aqueous phase from their respective average values indicate that the bond lengths are essentially fixed. However, due to solvation effects the aqueous phase bond length values of the above mentioned bonds deviate from their respective gas phase values. For example, in aqueous phase the C_{13}=O_{21} bonds are elongated up to 0.008 Å and the N_{16}–H_{20} bonds are shortened by a range of 0.003 to 0.004 Å. The C_{13}–N_{16} bonds are elongated in aqueous phase up to 0.004 Å for Pyl-Leu, Pyl-Asp, Pyl-Ser and Pyl-Asn systems while for Pyl-Ala and Pyl-His systems the same are shortened by 0.001 Å and 0.004 Å respectively. Table 3 lists the values of the six bond angles of the amide planes, i.e., C_{12}–C_{13}–O_{21}, C_{12}–C_{13}–N_{16}, O_{21}–C_{13}–N_{16}, C_{13}–N_{16}–C_{17}, C_{13}–N_{16}–H_{20} and H_{20}–N_{16}–C_{17}. These data in both the gas and aqueous phase indicate very little changes in the bond angle values as the individuality of the C-terminal residue of a given dipeptide changes. Maximum deviations of 1.2° in the gas phase and 1.0° in aqueous phase from their respective average values indicate that the bond angles are also essentially fixed. The solvent effects on these bond angles are quite apparent when their aqueous phase values are compared with the corresponding gas phase values; a maximum deviation up to 4.7° is observed for the angles C_{13}–N_{16}–H_{20} and H_{20}–N_{16}–C_{17} in Pyl-His system.

Investigating the four dihedral angles of the dipeptides viz. C_{12}–C_{13}–N_{16}–C_{17}, O_{21}–C_{13}–N_{16}–H_{20}, C_{12}–C_{13}–N_{16}–H_{20} and O_{21}–C_{13}–N_{16}–C_{17}, listed in Table 4, can provide valuable information regarding the planarity of the peptide planes. The values of the two dihedral angles C_{12}–C_{13}–N_{16}–C_{17} and O_{21}–C_{13}–N_{16}–H_{20} should be close to 180° and those for the other two, i.e., C_{12}–C_{13}–N_{16}–H_{20} and O_{21}–C_{13}–N_{16}–C_{17} should be close to 0° if indeed the amide plane is planar. The data presented in Table 4 shows that in aqueous phase the values of the four dihedral angles deviate

Table 3 Calculated bond angles (in degrees) for the peptide planes of the pyrrolysine dipeptides; the gas phase values are given in brackets

Dipeptides	C_{12}–C_{13}–O_{21}	C_{12}–C_{13}–N_{16}	O_{21}–C_{13}–N_{16}	C_{13}–N_{16}–C_{17}	C_{13}–N_{16}–H_{20}	H_{20}–N_{16}–C_{17}
Pyl-Ala	123.4 (121.4)	115.1 (115.0)	121.5 (123.6)	122.1 (122.3)	119.3 (115.5)	118.4 (121.1)
Pyl-Leu	122.7 (120.8)	114.7 (115.4)	122.5 (123.9)	123.1 (122.4)	119.7 (117.4)	117.1 (120.2)
Pyl-Asp	123.3 (121.2)	114.8 (115.5)	121.8 (123.3)	122.3 (122.1)	120.4 (117.8)	117.0 (120.0)
Pyl-Ser	123.0 (120.9)	114.5 (115.2)	122.4 (124.0)	123.1 (122.7)	119.5 (117.1)	117.4 (120.3)
Pyl-Asn	122.9 (120.9)	114.5 (115.1)	122.5 (124.0)	123.1 (122.6)	118.6 (116.2)	118.3 (121.0)
Pyl-His	123.2 (121.1)	114.6 (114.9)	122.1 (124.0)	122.6 (122.6)	120.2 (115.5)	117.1 (121.8)
Average	123.1 (121.1)	114.7 (115.2)	122.1 (123.8)	122.7 (122.5)	119.6 (116.6)	117.6 (120.7)
MD[a]	0.4 (0.3)	0.4 (0.3)	0.6 (0.5)	0.6 (0.4)	1.0 (1.2)	0.8 (1.1)

[a] Maximum deviation from average values

Table 4 Calculated dihedral angles (in degrees) for the peptide planes of the pyrrolysine dipeptides at B3LYP/6-31++G(d,p) level of theory; the gas phase values are given in brackets

Dipeptides	-SC Groups	C_{12}–C_{13}–N_{16}–C_{17}	O_{21}–C_{13}–N_{16}–H_{20}	C_{12}–C_{13}–N_{16}–H_{20}	O_{21}–C_{13}–N_{16}–C_{17}	ψ	ϕ
Pyl-Ala	$-CH_3$	176.0 (173.7)	−177.7 (−173.4)	0.8 (5.5)	−2.5 (−5.2)	145.9 (9.6)	−67.9 (−87.4)
Pyl-Leu	$-CH_2CH(CH_3)_2$	176.1 (−178.9)	178.5 (179.6)	−3.4 (0.6)	−2.0 (0.1)	134.9 (−14.6)	−142.4 (−143.7)
Pyl-Asp	$-CH_2CO_2H$	173.6 (179.3)	−178.3 (−178.4)	0.1 (2.9)	−4.8 (−2.1)	141.6 (−15.2)	−156.4 (−160.3)
Pyl-Ser	$-CH_2OH$	174.8 (−179.9)	178.4 (179.0)	−3.5 (−0.1)	−3.3 (−0.8)	138.8 (−13.6)	−136.8 (−142.7)
Pyl-Asn	$-CH_2CONH_2$	177.2 (−178.0)	177.9 (175.8)	−3.8 (−3.2)	−1.1 (0.9)	139.2 (−13.0)	−114.3 (−141.3)
Pyl-His	$-CH_2(C_3N_2H_3)$	174.1 (−177.4)	179.4 (178.7)	−2.7 (0.2)	−3.9 (1.0)	142.8 (−15.5)	−149.1 (−123.4)
MD[a]		6.4 (6.3)	2.3 (6.6)	3.8 (5.5)	4.8 (5.2)		

[a] Maximum deviation from expected values

up to a maximum value of 6.4° from the expected values whereas in the gas phase the maximum deviation observed is 6.6°. Thus, these dihedral angles do not deviate dramatically from their expected values in both the phases, however, the extent of deviations observed in the values of the four dihedral angles obviously suggest that the geometry of the amide planes are not perfectly planar, regardless of whether the systems are in gas phase or in strong polar solvents like water. A previous observation that solvation effects can enhance the planarity of the peptide planes [25] is found true only in the case of Pyl-Ala out of the six systems considered in this paper. For Pyl-Leu, Pyl-Ser, Pyl-Asp and Pyl-His systems the amide planes are predicted to be slightly more planar in gas than the aqueous phase. It is expected that the conformations of the six dipeptides predicted at B3LYP/6-31++G(d,p) level are reliable since it has been pointed out that full geometry optimization of gaseous tryptophan conformers at B3LYP/6-311G(d) and MP2/6-311++G(d,p) levels do not produce any noticeable structural changes, only the conformer energies change by small amounts [47]. Thus, it is reasonable to assume that solvation effects alone cannot drastically improve the planarity of the amide planes and the extent of the deviations from planarity primarily depends on two factors – (a) steric interactions of the side chain moieties of the C-terminal residues (–SC group) and (b) intramolecular H-bond formation by the H- and O-atoms of the amide planes with their adjacent moieties belonging to the C- and N-terminal residues. The intramolecular H-bond interactions that play crucial roles in deviating the amide planes from planarity and in imparting the observed conformations to the dipeptides in gas and aqueous phase are listed in Table 6 and a discussion on these interactions is also offered in a succeeding section of this paper.

Table 4 also lists the –SC groups of the C-terminal residues of the dipeptides as well as the values of the ψ and ϕ dihedrals in both the phases. A thorough analysis of the dipeptide structures reveals that both size as well as the type of functional groups present in a –SC group may influence the planarity of a given amide plane. A large sized –SC group may compete for its physical space requirements to accommodate itself in between the amide plane and carboxylic group of the C-terminal residue of a given dipeptide and consequently influence the planarity of the amide plane. The gas and aqueous phase values of the ϕ angles reveal that the value of ϕ

Table 5 Calculated bond angles (in degrees) for the α-carbon atoms of the pyrrolysine dipeptides; the gas phase values are given in brackets

Dipeptides	α-carbon atoms C_{12}			α-carbon atoms C_{17}		
	N_{14}–C_{12}–C_{11}	N_{14}–C_{12}–C_{13}	C_{11}–C_{12}–C_{13}	N_{16}–C_{17}–C_{18}	N_{16}–C_{17}–C_{19}	C_{19}–C_{17}–C_{18}
Pyl-Ala	110.7 (115.2)	109.5 (111.3)	109.1 (110.4)	113.4 (114.2)	110.3 (111.3)	109.7 (109.7)
Pyl-Leu	111.0 (111.1)	109.3 (111.2)	108.8 (108.8)	107.8 (108.0)	110.9 (111.1)	111.2 (111.1)
Pyl-Asp	110.8 (111.2)	109.4 (111.1)	108.8 (108.8)	108.1 (108.1)	111.3 (111.2)	111.5 (111.5)
Pyl-Ser	110.9 (111.3)	109.4 (111.3)	108.7 (108.9)	108.7 (108.5)	110.3 (110.4)	110.4 (110.0)
Pyl-Asn	110.9 (111.3)	109.3 (111.2)	108.7 (108.7)	110.5 (107.1)	111.8 (111.3)	111.8 (111.2)
Pyl-His	110.9 (111.3)	109.5 (111.1)	108.8 (108.8)	107.9 (109.0)	110.8 (111.7)	111.8 (112.8)
Average	110.9 (111.9)	109.4 (111.2)	108.8 (109.1)	109.4 (109.2)	110.7 (111.2)	111.1 (111.1)
MD[a]	0.2 (3.3)	0.1 (0.1)	0.3 (1.3)	4.0 (5.0)	1.1 (0.8)	1.4 (1.7)

[a] Maximum deviation from average values

Table 6 H-bond distances[a] (in angstrom) of the intramolecular H-bond interactions detected in the pyrrolysine dipeptides in gas and aqueous phases

Dipeptides	Phases	$N_{14}…H_{20}–N_{16}$	$O_{21}…H_b–C_{11}$	$O_{21}…H–C_{12}$	$O_{21}…H–C_{17}$	$O_{23}…H_{20}–N_{16}$	$O_{22}…H–C_{17}$	$O_{23}…H–C_{17}$
Pyl-Ala	Aqueous	*abs*	*abs*	*abs*	2.631	*abs*	*abs*	2.554
	Gas	2.141	2.470	2.762	2.439	*abs*	*abs*	2.666
Pyl-Leu	Aqueous	*abs*	*abs*	*abs*	2.455	2.426	2.485	*abs*
	Gas	2.184	2.664	2.579	2.454	2.487	2.494	*abs*
Pyl-Asp	Aqueous	*abs*	*abs*	*abs*	2.575	2.274	2.602	*abs*
	Gas	2.189	2.676	2.584	2.618	2.323	2.593	*abs*
Pyl-Ser	Aqueous	*abs*	*abs*	*abs*	2.425	2.539	2.451	*abs*
	Gas	2.173	2.653	2.584	2.468	2.578	2.453	*abs*
Pyl-Asn	Aqueous	*abs*	*abs*	*abs*	2.338	*abs*	2.602	*abs*
	Gas	2.160	2.647	2.591	2.452	*abs*	2.448	*abs*
Pyl-His	Aqueous	*abs*	*abs*	*abs*	2.506	2.338	2.512	*abs*
	Gas	2.145	2.672	2.584	2.350	2.414	2.589	*abs*

[a] Only the (B….H) distances are listed where B is H-bond acceptor; *abs*=absent

increases as the size of a given –SC group increases. This point has been well discussed in various literature [15, 16]. Among the six dipeptides studied in this paper the ϕ value of Pyl-Ala in aqueous phase is −67.9° (−87.4° in gas phase) while for the other five systems, which have much larger sized –SC groups compared to that of Pyl-Ala, the ϕ values are observed above −114.3° in aqueous phase (−123.4° in gas phase). On the other hand, the –SC groups, depending on the type of functional groups present in them, may exert electrostatic repulsive or electrostatic attractive forces on their neighboring atoms belonging to the peptide planes and the carboxylic group of the C-terminal residues of the dipeptides which may also influence the values of the ϕ as well as planarity of the amide planes. For example, the –SC groups of Asp and Asn are comparable in their size and the inconsistencies observed in the values of ϕ as well as in the values of the other four dihedral angles of Pyl-Asp and Pyl-Asn can be explained on the basis of the type of functional groups present in their –SC groups.

Regarding the solvent effects of the aqueous phase, the optimized structures depicted in Figs. 2, 3 and 4 and data on ψ listed in Table 4 suggest that the aqueous phase structures are appreciably different from those in gas phase around the dihedral angle ψ. It has already been recognized that polar solvents remarkably influence the conformational properties of dipeptides, by weakening the intraresidue hydrogen bonds and leading to the appearance of new energy minima [27–29]. The results of this DFT study predict that for the Pyl-Ala system the aqueous phase ψ value deviates from its corresponding gas phase value by 136.3° while in the other five systems the differences range from 149.5 to 158.3°. However, these changes in the values of ψ do not affect the planarity of the peptide planes but affects geometry around the C_{12} (discussed in the next section).

α-Carbon geometry

Since the protein structures usually contain thousands of amino acid residues, the geometries about the α-carbon atoms of the individual residues play important roles in deciding the overall structure of the proteins. The three bond angles considered to monitor the geometry around the C_{12} α-carbon atoms of the dipeptides are $N_{14}–C_{12}–C_{11}$, $N_{14}–C_{12}–C_{13}$ and $C_{11}–C_{12}–C_{13}$ while $N_{16}–C_{17}–C_{18}$, $N_{16}–C_{17}–C_{19}$ and $C_{19}–C_{17}–C_{18}$ are the same for the C_{17} α-carbon atoms. The α-carbon atoms of the amino acids are sp^3 hybridized and therefore the ideal bond angle should be 109.5°, however, this is not expected due to their stereogenic character. By monitoring the above mentioned bond angles around each α-carbon atom of the dipeptides one can get an idea about how the change in identity of the C-terminal residues can affect the geometries about these α-carbon atoms. This DFT study also provides the opportunity to probe the effects of solvation on the geometries of the α-carbon atoms. Table 5 lists the gas and aqueous phase data on the bond angles about the α-carbon atoms. Maximum deviations of 0.3° in aqueous and 3.3° in gas phase from their respective average values suggest that the geometry about a particular C_{12} atom does not change much with the change in the identity of the C-terminal residue. On the other hand, with maximum deviations up to 4.0° in aqueous and 5.0° in gas phase from their respective average values, the bond angles around the C_{17} change appreciably with the change in identity of the C-terminal residues of the dipeptides. These observations can be justified by invoking the

Table 7 Frequencies[a] (in cm^{-1}) and IR intensities (in km mol^{-1}) of various vibrational modes[b] obtained from the theoretical vibrational spectra of the pyrrolysine dipeptides in gas and aqueous phases. Intensities are given in brackets

Dipeptides	Phases	$\nu(C_{13}=O_{21})$	$\nu(N_{16}–H_{20})$	$\nu(C_{13}–N_{16})$	$\nu_s(N_{14}–H)$	$\nu_{as}(N_{14}–H)$	$Sis(N_{14}–H)$	$\nu(C_{17}–H)$	$\nu(C_{12}–H)$
Pyl-Ala	Aqueous	1665 (428.4)	3458 (68.9)	1502 (456.3)	3348 (0.7)	3425 (7.2)	1615 (77.9)	2958 (6.2)	2835 (92.7)
	Gas	1700 (241.2)	3399 (93.9)	1504 (453.3)	3349 (1.2)	3441 (5.6)	1625 (39.7)	2951 (9.0)	2870 (36.5)
Pyl-Leu	Aqueous	1656 (379.0)	3449 (95.6)	1510 (443.1)	3345 (0.2)	3421 (7.0)	1617 (66.4)	2992 (9.2)	2836 (101)
	Gas	1691 (291.2)	3395 (94.8)	1505 (402.9)	3363 (1.4)	3445 (7.4)	1626 (47.1)	2980 (15.9)	2867 (18.9)
Pyl-Asp	Aqueous	1662 (602.9)	3442 (124.5)	1504 (486.0)	3347 (0.7)	3424 (7.2)	1615 (70.1)	2951 (6.9)	2836 (93.1)
	Gas	1692 (288.3)	3385 (105.7)	1499 (437.7)	3366 (3.9)	3449 (8.4)	1626 (45.6)	2942 (5.7)	2868 (16.1)
Pyl-Ser	Aqueous	1666 (683)	3449 (87.5)	1508 (424.4)	3346 (0.7)	3422 (7.3)	1617 (68.6)	2998 (6.4)	2834 (97.7)
	Gas	1699 (318.3)	3389 (93.1)	1502 (370.9)	3366 (2.4)	3447 (7.9)	1628 (43.8)	2992 (4.8)	2869 (16.0)
Pyl-Asn	Aqueous	1671 (911.3)	3451 (66.9)	1511 (372.2)	3347 (0.7)	3423 (7.5)	1616 (70.9)	2967 (25.2)	2833 (99.2)
	Gas	1700 (320.3)	3387 (93.5)	1500 (374.1)	3363 (2.8)	3445 (7.9)	1626 (43.4)	2961 (3.6)	2869 (16.1)
Pyl-His	Aqueous	1660 (626.0)	3441 (112.7)	1501 (459.2)	3347 (0.7)	3424 (7.0)	1614 (75.2)	2972 (11.4)	2833 (97.7)
	Gas	1707 (284.3)	3383 (106.7)	1497 (407.8)	3362 (4.0)	3443 (8.2)	1626 (43.0)	2941 (26.7)	2879 (12.0)

[a] The frequencies below 1800 cm^{-1} are scaled with 0.977 and for those above 1800 cm^{-1} a correction factor 0.955 is used

[b] Vibrational modes: ν=stretching; Sis=scissoring; s=symmetric; as=asymmetric

two factors—size and the type of functional groups present in the –SC groups as previously mentioned while discussing the planarity of the peptide planes. The stereoelectronic effects of the varying –SC groups on the geometry of the C_{12} atoms are very little as they reside at a distance of four bonds away from these α-carbon atoms. On the contrary, since the varying –SC groups are situated adjacent to the C_{17} atoms the geometry around them are affected by the changing identity of the –SC groups. The solvation effects are more prominent on the geometry of the C_{12} atoms, a maximum deviation up to 4.5° is observed for the angle N_{14}–C_{12}–C_{11} in Pyl-Ala system, than that on the C_{17} atoms where the maximum deviation predicted is 3.4° for the N_{16}–C_{17}–C_{18} angle in Pyl-Asn. These alterations in the geometry around the C_{12} atoms of the dipeptides can be attributed to the changes in the values of ψ due to the solvation effects.

Intramolecular hydrogen bonds

Intramolecular hydrogen bonds (H-bonds), the strongest non-covalent interactions, play an important role in stabilizing the different conformations of a dipeptide molecule [18]. The strength of these H-bonds depends on two factors, (a) shorter is the distance A–H….B than the sum of their van der Waals radii and (b) closer the angle A–H….B to 180° [48], where A–H is H-bond donor and B is H-bond acceptor. Table 6 lists three types of intramolecular H-bonds, namely O…H–N, N…H–N and O…H–C, whose interplay is very crucial in imparting the observed deviations of the peptide planes from planarity as well as in determining the energetics of the pyrrolysine containing dipeptides. The gas phase intramolecular H-bond combinations of the dipeptides are noticeably different from those in the aqueous phase. The three H-bonds $N_{14}…H_{20}$-N_{16}, $O_{21}…H_b$-C_{11} and $O_{21}…H$-C_{12} observed in the gas phase structures of the dipeptides (with B….H distances ranging from 2.141 to 2.762 Å), disappear in the aqueous phase for all the dipeptide structures. This is because in aqueous phase the intra-residue interactions are modified appreciably so weak H-bond interactions like N…H–N and O…H–C are weakened further and as a result the ψ dihedral can afford to change from a range of 9.6 to −15.5° in the gas phase to a range of 134.9 to 145.9° in the aqueous phase (listed Table 4). The presence of the $O_{21}…H$-C_{17} H-bonds in both the phases indicates that solvent effects do not alter the conformation of the dipeptides about the ϕ dihedral angles (see Table 4). On the other hand, the gas and solvent phase data on the three H-bonds $O_{23}…H_{20}$-N_{16}, $O_{22}…H$-C_{17} and $O_{23}…H$-C_{17} clearly indicates the effects of size and the type of functional groups present in the –SC groups on the conformation of the dipeptides as well as on the number and type of H-bond interactions existing in the dipeptide molecules. For example, the

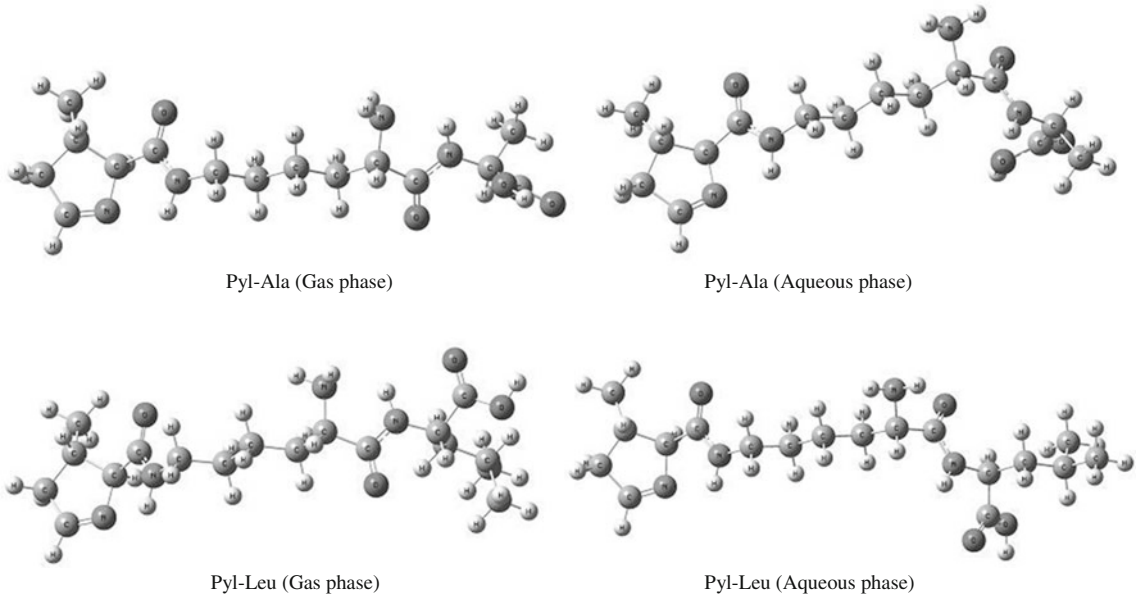

Fig. 2 The optimized structures of Pyl-Ala and Pyl-Leu in gas and aqueous phase

absence of $O_{23}…H_{20}$-N_{16} in Pyl-Ala and Pyl-Asn systems; absence of $O_{22}…H$-C_{17} and presence of $O_{23}…H$-C_{17} only in the case of Pyl-Ala can be explained on the basis of identify of the –SC groups of the C-terminal residues.

Vibrational spectra

The theoretically predicted vibrational spectra of the six pyrrolysine dipeptides in both the phases provide valuable information to understand the existence and nature of various types of intramolecular H-bonds in the dipeptides. Table 7 lists the characteristic frequency and intensity (given in brackets) values of only those vibrational modes which are sensitive to the structural changes caused by the varying C-terminal residues and solvent effects. It is evident from Table 7 that the vibrational frequencies shift invariably toward the lower side of frequency scale corresponding to the presence of intramolecular H-bond interactions. The presence of $N_{14}…H_{20}$-N_{16} bond in the gas phase structures is well reflected by the lowering in frequency values of the $\nu(N_{16}$-$H_{20})$ stretching by a range of 54 to 64 cm^{-1} than those in the aqueous phase. Solvent effects also lower the frequency values of the $\nu(C_{13}=O_{21})$, $\nu_s(N_{14}$-$H)$, $\nu_{as}(N_{14}$-$H)$ and Sis $(N_{14}$-$H)$ by a magnitude up to 47 cm^{-1} in the aqueous phase which can be due to elongation in the bond length values in solvent phase (the $C_{13}=O_{21}$ bonds are elongated up to 0.008 Å in the aqueous phase). The differences in the gas and aqueous phase values of $\nu(C_{12}$-$H)$ arise because of the

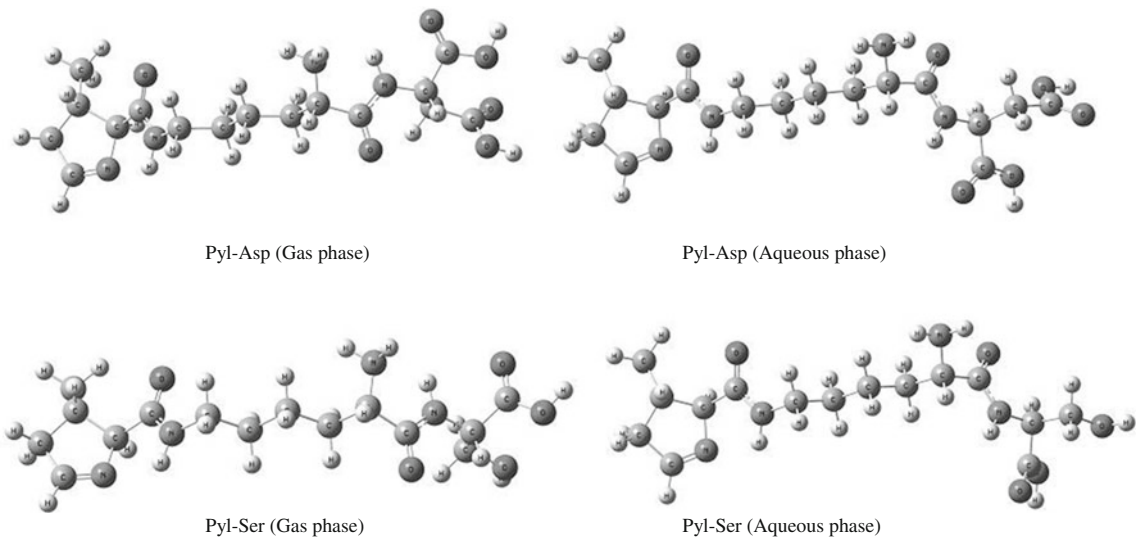

Fig. 3 The optimized structures of Pyl-Asp and Pyl-Ser in gas and aqueous phase

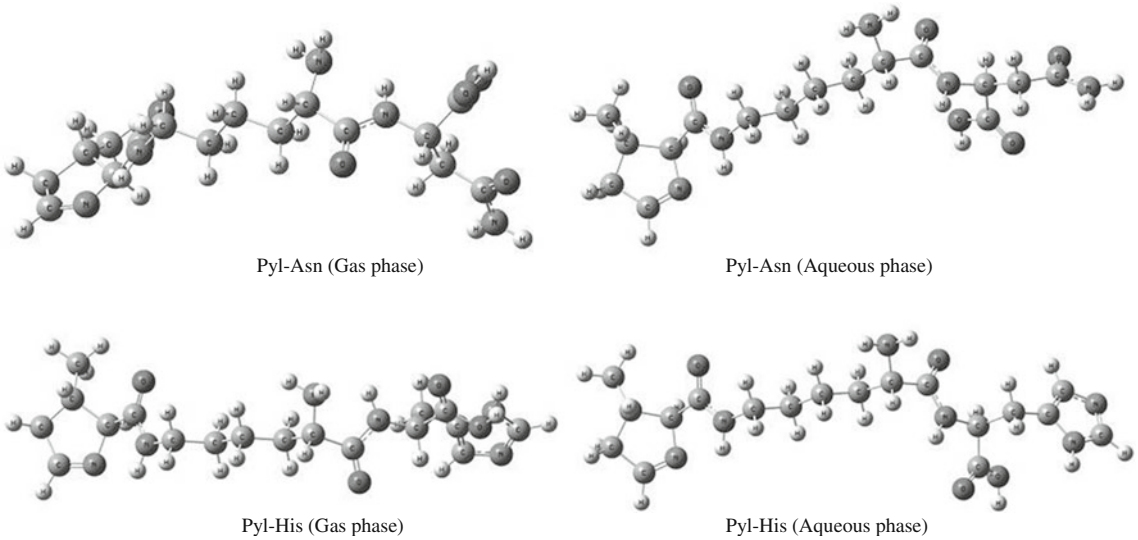

Fig. 4 The optimized structures of Pyl-Asn and Pyl-His in gas and aqueous phase

changes in conformation of the dipeptides around the ψ dihedrals due to solvent effects, while the variations in the $\nu(C_{17}\text{-H})$ values can be attributed to the effects of the changing–SC groups of the C-terminal residues.

Conclusions

This DFT study on dipeptides containing pyrrolysine as a fixed component at their N-terminal positions predicts large values of total dipole moments for the dipeptides, 2.361 to 7.421 D in gas phase and 8.864 to 14.152 D in aqueous phase, and as a consequence the aqueous phase structures show more thermodynamic stabilities by a range of 13.35 to 19.72 kcal mol^{-1} than those in the gas phase. The geometrical parameters like bond lengths and bond angles of the amide planes of the dipeptides show very little variance as the identity of the C-terminal residue of a given dipeptide changes. The geometry of the amide planes are not perfectly planar regardless of whether the systems are in gas or in strong polar solvents like

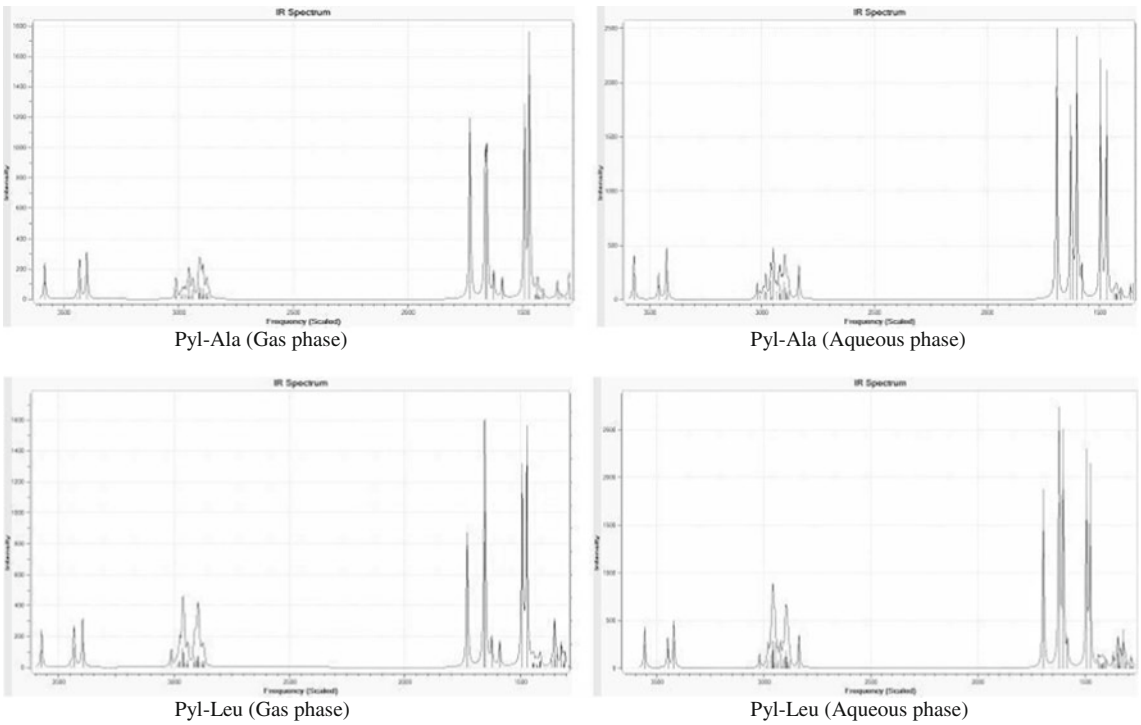

Fig. 5 Vibrational spectra of Pyl-Ala and Pyl-Leu in gas and aqueous phase

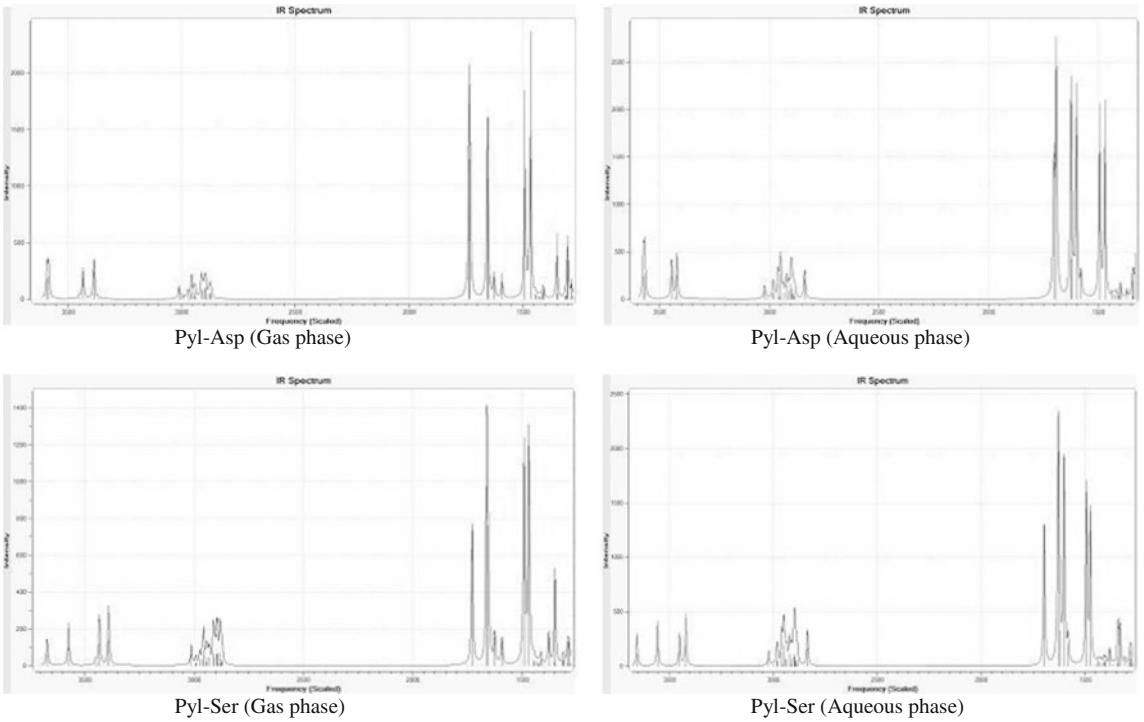

Fig. 6 Vibrational spectra of Pyl-Asp and Pyl-Ser in gas and aqueous phase

water and the deviations from planarity primarily depends on two factors – (a) steric interactions of the side chain moieties of the C-terminal residues and (b) intramolecular H-bond formation by the H- and O-atoms of the amide planes with their adjacent atoms belonging to the C- and N-terminal residues. The ϕ values depend on the size of a given –SC group which is evident from the fact that the ϕ value of Pyl-Ala in aqueous phase is −67.9° (−87.4° in gas phase) while for the other five systems, which have much larger sized –SC groups compared to that of Pyl-Ala, the ϕ values are observed

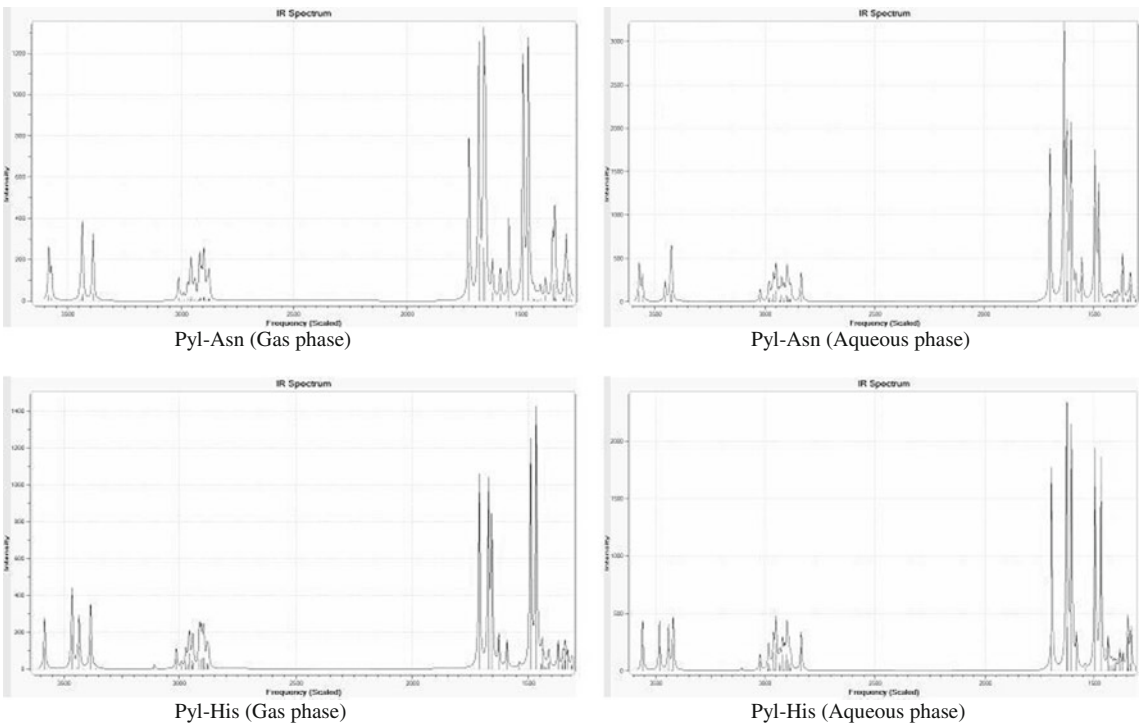

Fig. 7 Vibrational spectra of Pyl-Asn and Pyl-His in gas and aqueous phase

above −114.3° in aqueous phase (−123.4° in gas phase). The solvent effects appreciably modify the gas phase ψ values, for example in the Pyl-Ala system the difference of ψ values in gas and aqueous phase is 136.3° while in the other five systems the differences range from 149.5 to 158.3°. The geometry around the C_{17} atoms are affected by the changes in the identity of the –SC groups while the signs of solvent effects are evident on the geometry of C_{12} atoms. The presence or absence of three types of intramolecular H-bonds, namely O…H–N, N…H–N and O…H–C that leave noticeable signatures in the IR spectra, play crucial roles in influencing the geometry of the peptide planes and in determining the energetics of the pyrrolysine dipeptides.

Acknowledgments Financial assistance from the Special Assistance Program of the University Grants Commission to the Department of Chemistry, NEHU, is gratefully acknowledged.

References

1. Rother M, Krzycki JA (2010) Archaea doi: 10.1155/2010/453642
2. Atkins JF, Gesteland R (2002) Science 296(5572):1409–1410
3. James CM, Ferguson TK, Leykam JF, Krzycki JA (2001) J Biol Chem 276:34252–34258
4. Hao B, Gong W, Ferguson TK et al (2002) Science 296:1462–1466
5. Srinivasan G, James CM, Krzycki JA (2002) Science 296:1459–1462
6. Lukashenko NP (2010) Russ J Genet 46:899–916
7. Krzycki JA (2004) Curr Opin Chem Biol 8:484–491
8. Gaston MA, Zhang L, Green-Church KB, Krzycki JA (2011) Nature 471:647–650
9. Dudev T, Lim C (2009) J Phys Chem B 113:11754–11764
10. Fekner T, Li X, Lee MM, Chan MK (2009) Angew Chem 121:1661–1663
11. Fekner T, Li X, Chan MK (2010) Eur J Org Chem doi:10.1002/ejoc.201000204
12. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA (2002) Chem Rev 102:371–386
13. Foloppe N, Hartmann B, Nilsson L, MacKerell AD Jr (2002) Biophys J 82:1554–1569
14. Sponer J, Zgarbova M, Jurecka P, Riley KE, Sponer JE, Hobza P (2009) J Chem Theory Comput 5:1166–1179
15. Ghosh S, Mondal S, Misra A, Dalai S (2007) J Mol Struct (THEOCHEM) 805:133–141
16. Keefe CD, Pearson JK (2004) J Mol Struct (THEOCHEM) 679:65–72
17. Saada I, Pearson JK (2011) Computational and Theoretical Chemistry 969:76–82
18. Vargas R, Garza J, Hay BP, Dixon DA (2002) J Phys Chem A 106:3213–3218
19. Antohi O, Naider F, Sapse AM (1996) J Mol Struct (THEOCHEM) 360:99–108
20. Kapota C, Ohanessian G (2005) Phys Chem Chem Phys 7:3744–3755
21. Kolev T, Koleva BB, Spiteller M (2007) Amino Acids 33:719–725
22. Koleva BB, Kolev TM, Todorov S (2007) Chem Papers 61:490–496
23. Clavaguera C, Piuzzi F, Dognon JP (2009) J Phys Chem B 113:16443–16448
24. Tobiast DJ, Brooks CL III (1992) J Phys Chem 96:3864–3870
25. Wang ZX, Duan Y (2004) J Comput Chem 25:1699–1716
26. García-Prieto FF, Galvan IF, Aguilar MA, Martín ME (2011) J Chem Phys 135:194502–194509
27. Gould IR, Cornell WD, Hillier IH (1994) J Am Chem Soc 116:9250–9256
28. Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks CL III, Pople J (1991) J Am Chem Soc 113:5989–5997
29. Adamo C, Dillet V, Barone V (1996) Chem Phys Lett 263:113–118
30. Ramachandran GN (1963) Biopolymers 6:1494–1495
31. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) J Mol Biol 7:95–99
32. Hao B, Zhao G, Kang PT, Soares JA, Ferguson TM, Gallucci J, Krzycki JA, Chan MK (2004) Chem Biol 11:1317–1324
33. Osawa S, Jukes TH, Watanabe K, Muto A (1992) Microbio Rev 56:229–264
34. Becke AD (1993) J Chem Phys 98:5648–5652
35. Lee C, Yang W, Parr RG (1993) Phys Rev B 37:785–789
36. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, and Pople JA (2004) Gaussian 03, Revision D.01 Gaussian Inc, Wallingford, CT
37. Miertus S, Scrocco E, Tomasi J (1981) Chem Phys 55:117–129
38. Gould IR, Hillier IH (1993) J Chem Soc, Chem Commun doi:10.1039/C39930000951
39. Lambie B, Ramaekers R, Maes G (2004) J Phys Chem A 108:10426–10433
40. Kecel S, Ozel AE, Akyuz S, Celik S, Agaeva G (2011) J Mol Struct 993:349–356
41. Andersson MP, Uvdal P (2005) J Phys Chem A 109:2937–2941
42. Foresman JB, Frisch A (1996) Exploring chemistry with electronic structure methods, 2nd edn. Gaussian Inc, Pittsburgh, PA
43. Freeman F, Le KT (2003) J Phys Chem A 107:2908–2918
44. Chasse GA, Rodriguez AM, Mak ML, Deretey E, Perczel A, Sosa CP, Enriz RD, Csizmadia IG (2001) J Mol Struct (THEOCHEM) 537:319–361
45. Stepanian SG, Reva ID, Radchenko ED, Rosado MTS, Duarte MLTS, Fausto R, Adamowicz L (1998) J Phys Chem A 102:1041–1054
46. Stepanian SG, Reva ID, Radchenko ED, Adamowicz L (1998) J Phys Chem A 102:4623–4629
47. Huang Z, Lin Z (2005) J Phys Chem A 109:2656–2659
48. Kaur D, Sharma P, Bharatam PV, Kaur M (2008) Int J Quant Chem 108:983–991