J Mol Model (2013) 19:2613–2623
DOI 10.1007/s00894-013-1808-x

ORIGINAL PAPER

Structure of dipeptides having N-terminal selenocysteine residues: a DFT study in gas and aqueous phase

Shilpi Mandal · Gunajyoti Das

Received: 9 January 2013 / Accepted: 18 February 2013 / Published online: 15 March 2013

Abstract Over the last few decades, dipeptides as well as their analogues have served as important model systems for the computational studies concerning the structure of protein and energetics of protein folding. Here, we present a density functional structural study on a set of seven dipeptides having N-terminal selenocysteine residues (the component in the C-terminus is varied with seven different combinations viz. Ala, Phe, Glu, Thr, Asn, Arg and Sec) in gas and simulated aqueous phase using a polarizable continuum model (PCM). The molecular geometries of the dipeptides are fully optimized at B3LYP/6−311++G(d,p) level and subsequent frequency calculations confirm them as 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 investigated. Two types of intramolecular H-bonds, namely N…H-N and O…H-C, are found to play important roles in influencing the planarity of the peptide planes and geometry around the α-carbon atoms of the dipeptides. The identity of the varying C-terminal residue influences the values of ϕ, planarity of the peptide planes and geometry around the C_7 α-carbon atoms while the solvation effects are evident on the values of bond lengths and bond angles of the amide planes.

Keywords Dipeptides · Selenocysteine · Solvation effects · Vibrational frequencies

Electronic supplementary material The online version of this article (doi:10.1007/s00894-013-1808-x) contains supplementary material, which is available to authorized users.

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

Introduction

Most of the biological functions of proteins are specific to their three-dimensional (3D) structures [1] and these 3D structures are almost exclusively dependent on the primary amino acid sequences [2]. The primary amino acid sequences in proteins are directly determined by the genetic code which allows for the incorporation of 20 standard amino acids during protein biosynthesis. However, methanogenic archaea, characterized by their strict dependence on the process of methanogenesis for energy conservation, are known to synthesize proteins containing selenocysteine (Sec) or pyrrolysine (Pyl) [3]. Unlike pyrrolysine, whose occurrence is limited to methanogenic archea and certain bacteria [4], the distribution of selenocysteine in proteins is a widespread incident and has been observed in eubacteria, archaea and eukarya [5]. Selenocysteine is regarded as the 21st genetically encoded natural amino acid since it is co-translationally inserted into proteins corresponding to the opal codon UGA which generally ends the translation process of protein biosynthesis [6–8]. Structurally selenocysteine is similar to cysteine; the only difference is that in selenocysteine a selenium-containing selenol group replaces the sulfur-containing thiol group of cysteine. Selenocysteine is an important constituent in the active site of selenium-containing enzymes that catalyze many oxidation-reduction reactions [9] while in eukaryotic systems selenoproteins participate in anabolic processes and protect the cells from oxidative damage [10]. Successful attempts have been made to understand the synthesis and insertion of selenocysteine into protein in bacterial systems [8, 11], which are however, still less clear in eukaryotic organisms [5]. Considerable attention has also been paid to understand the metal-binding affinity/selectivity of selenocysteine [12, 13] and on the intrinsic conformational properties of non-ionic selenocysteine molecule in gas phase

[14]. However, the structural features of selenocysteine containing dipeptides are not explored yet which to a large extent determine the dynamic properties and functional specificity of the proteins and polypeptides containing selenocysteine.

Proteins are the principal ingredients for all living beings on earth. The functional diversity of proteins primarily depends on three factors—the primary amino acid sequences, post-translational modifications of these sequences, and folding of the resulting polypeptides. Therefore, understanding the conformational aspects of dipeptides seems to be the key to understand the structure of protein, mechanism of protein folding and ultimately their biological activities. Gas phase computational studies on dipeptides arising from the genetically encoded amino acids [15–18] have been performed with a view toward understanding the structural features of small amino acid sequences and their possible roles in imparting the 3D structure to proteins. The importance of gas-phase structural studies on dipeptides lies in the fact that such studies can provide us the opportunity to understand their intrinsic properties free from the solvent or crystal phase effects. Moreover, 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 [19–21]. Structural studies [15, 16] on a series of dipeptides have pointed out that in most of the dipeptides the amide plane is not completely planar; and this has been explained in terms of the cumulative effect of steric hindrance of –R group and H-bonding. Besides serving as model systems, dipeptides themselves have been shown to play numerous key biological roles [22–26].

It is of fundamental importance to determine the conformational details of a biological molecule in aqueous solution since the vast majority of biochemical processes occur in an aqueous environment. The effects of solvation on the conformations and energies of dipeptides have been well documented in the literature [27–32]. In these studies the energetics and structural features of the dipeptides in gas and solvent phases are analyzed to understand the effect of the surrounding environment on the stabilities and conformational preferences of the dipeptides. 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) [33, 34] conferring markedly different conformations to the dipeptides in aqueous phase. It has also been reported that solvation effects can enhance the planarity of the peptide planes [28].

The purpose of the present theoretical study is to obtain full knowledge about 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, 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 the selenocysteine containing dipeptides. The dipeptides are constructed by keeping selenocysteine as a fixed component in the N-terminus whereas the component in the C-terminus is varied with seven different combinations. The seven different amino acids chosen for the C-terminus position are Ala, Phe, Glu, Thr, Asn, Arg and Sec. All these amino acid residues are taken as neutral (non-ionic) species. The standard three letter abbreviations are used to represent an amino acid while a particular dipeptide is named by listing the N-terminal residue first. Thus, Sec-Ala dipeptide corresponds to a structure in which selenocysteine is in the N-terminal position and alanine is in the C-terminal position. Figures 1 and 2 schematically represents the chemical structures of the seven dipeptides studied here. The C_4–N_6 is the peptide bond of a given dipeptide structure while C_3 and C_7 are the α-carbon atoms of the N- and C-terminal residues respectively. To facilitate a clear representation of the intramolecular hydrogen bond interactions present in the selenocysteine dipeptides some of the hydrogen atoms are named H_a or H_b. This DFT study on dipeptides of selenocysteine in gas as well as in simulated aqueous phase is expected to provide the opportunity to know the structural features of the dipeptides at an atomic level which in turn may help us to understand the dynamics and functional specificity of proteins containing selenocysteine, in understanding the nature of the genetic code or amino acid code which is still evolving [35] and in enhancing this rapidly expanding area of research.

Computational methods

The molecular geometries of all the selected dipeptides are subjected to full geometry optimization and vibrational frequency calculations using the B3LYP/6−311++G(d,p) level of theory [36, 37] of Gaussian 03 package [38]. The efficacy of B3LYP/6−311++G(d,p) in studying conformational behavior and various other properties of amino acids has been explained in literature [39]. The computations are conducted in gas as well as in aqueous phase using a polarizable continuum model (PCM) [40]. The accuracy of self-consistent reaction field (SCRF) model in predicting the structure and energetics of dipeptides has already been justified in literature [41]. Absence of imaginary frequency value 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 conformers using a correction factor 0.9877 [42]. The vibrational frequencies below 1800 cm^{-1} are scaled with 1.01 and for those above 1800 cm^{-1} a correction factor

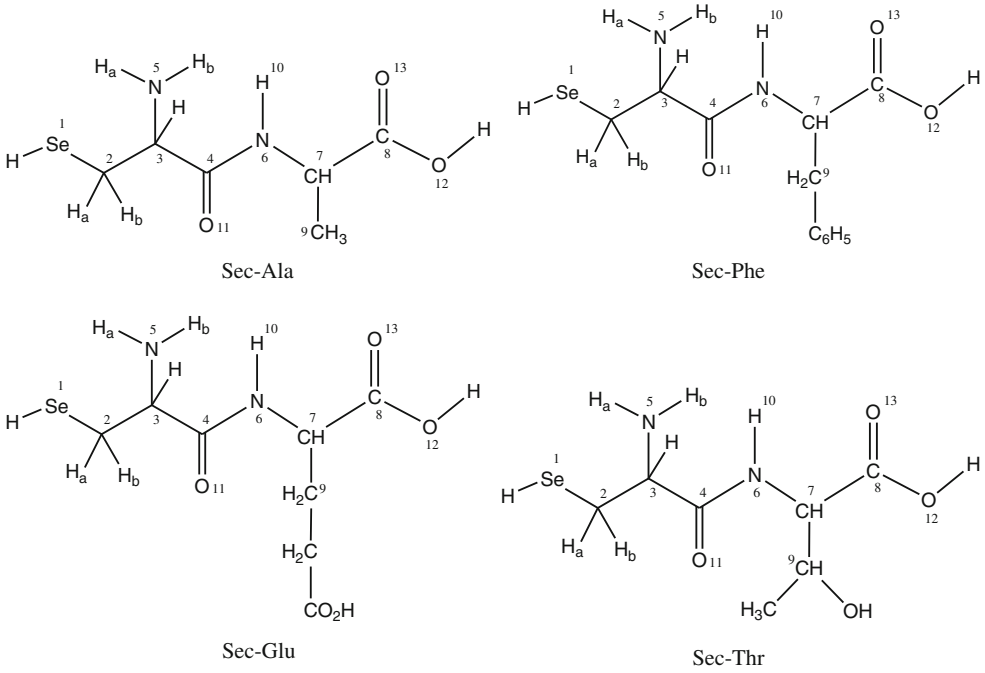


Fig. 1 Schematic representation of chemical structures of Sec-Ala, Sec-Phe, Sec-Glu and Sec-Thr systems

0.9679 is used [42]. Use of diffuse functions is important to take into account the relative diffuseness of lone pair of electrons when a molecule under investigation contains lone pair of electrons [43] while polarization functions are useful in studying the conformational aspects where stereoelectronic effects play an important role [44].

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 protein and energetics of protein folding [45].

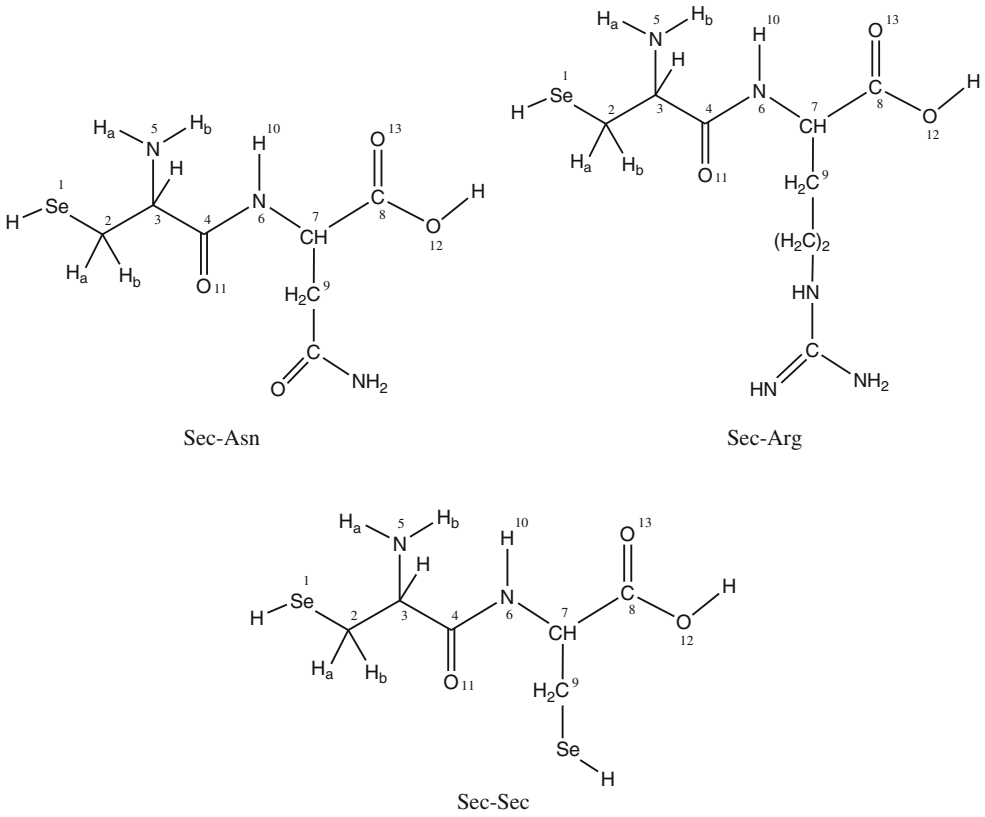


Fig. 2 Schematic representation of chemical structures of Sec-Asn, Sec-Arg and Sec-Sec systems

Table 1 Calculated total energies[a] (kcal mol^{-1}), rotational constants (GHZ) and dipole moments (Debye) of the dipeptides of selenocysteine in gas and aqueous phase using B3LYP/6−311++G(d,p) level of theory

Dipeptides	Phases	Total energies	Rotational constants			Dipole moments
			A	B	C	
Sec-Ala	Aqueous	−1865356.26	1.21042	0.29516	0.25763	8.729
	Gas	−1865343.62	1.25601	0.27852	0.25345	5.581
Sec-Phe	Aqueous	−2010327.82	0.38823	0.17476	0.14113	8.345
	Gas	−2010314.36	0.38548	0.17881	0.14284	5.477
Sec-Glu	Aqueous	−2008376.99	0.62830	0.14853	0.13552	8.976
	Gas	−2008360.64	0.58872	0.15383	0.14016	6.503
Sec-Thr	Aqueous	−1937228.86	0.92053	0.20334	0.19842	10.793
	Gas	−1937213.41	0.91880	0.20296	0.20015	6.966
Sec-Asn	Aqueous	−1971240.59	0.86287	0.16588	0.16434	10.912
	Gas	−1971222.17	0.77377	0.17601	0.16280	4.805
Sec-Arg	Aqueous	−2042801.06	0.86795	0.08254	0.07921	11.072
	Gas	−2042780.72	0.42573	0.10192	0.08866	7.141
Sec-Sec	Aqueous	−3372349.43	0.80193	0.13838	0.13245	9.764
	Gas	−3372334.93	0.78253	0.14315	0.13367	6.667

[a]ZPVE corrected; Scaled with a correction factor 0.9877

The geometrical parameters that have been 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. Table 1 presents the gas and aqueous phase data on total energies, rotational constants and dipole moments of the dipeptides calculated at B3LYP/6−311++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_3–C_4–N_6–C_7, O_{11}–C_4–N_6–H_{10}, C_3–C_4–N_6–H_{10} and O_{11}–C_4–N_6–C_7, are listed in Table 4. Table 4 also lists the two well known Ramachandran backbone dihedral angles ψ (N_5–C_3–C_4–N_6) and ϕ (C_4–N_6–C_7–C_8) 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/6−311++G(d,p) level of theory. Figures 3, 4, 5 and 6 represent the theoretical IR spectra of the seven dipeptides both in gas and aqueous phase (scaled with a correction factor 0.9679).

Dipeptide structure

As listed in Table 1 all seven dipeptide geometries exhibit large values of total dipole moments, ranging from 4.805 to 7.141 D in gas phase and 8.345 to 11.072 D in aqueous phase, indicating that they have greater polar character and consequently possess greater affinity to polar solvents.

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

Dipeptides	C_3–C_4	C_4 = O_{11}	C_4–N_6	N_6–H_{10}	N_6–C_7
Sec-Ala	1.538 (1.541)	1.233 (1.223)	1.348 (1.358)	1.013 (1.012)	1.450 (1.447)
Sec-Phe	1.539 (1.540)	1.232 (1.223)	1.350 (1.358)	1.014 (1.012)	1.446 (1.444)
Sec-Glu	1.539 (1.541)	1.231 (1.223)	1.351 (1.358)	1.014 (1.011)	1.449 (1.452)
Sec-Thr	1.540 (1.543)	1.230 (1.221)	1.353 (1.361)	1.015 (1.013)	1.451 (1.454)
Sec-Asn	1.540 (1.542)	1.230 (1.223)	1.352 (1.357)	1.015 (1.014)	1.452 (1.460)
Sec-Arg	1.538 (1.542)	1.233 (1.224)	1.348 (1.356)	1.014 (1.012)	1.452 (1.455)
Sec-Sec	1.541 (1.542)	1.230 (1.222)	1.354 (1.360)	1.015 (1.013)	1.457 (1.456)
Average	1.539 (1.542)	1.231 (1.223)	1.351 (1.358)	1.014 (1.012)	1.451 (1.453)
MD[a]	0.002 (0.002)	0.002 (0.002)	0.003 (0.003)	0.001 (0.002)	0.006 (0.009)

[a]Maximum deviation from average values

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

Dipeptides	C_3–C_4–O_{11}	C_3–C_4–N_6	O_{11}–C_4–N_6	C_4–N_6–C_7	C_4–N_6–H_{10}	H_{10}–N_6–C_7
Sec-Ala	121.6 (121.4)	115.3 (114.7)	123.1 (123.9)	122.6 (122.7)	116.4 (116.2)	121.0 (120.8)
Sec-Phe	121.4 (121.6)	115.0 (114.6)	123.6 (123.8)	123.1 (122.6)	115.8 (116.3)	121.0 (121.1)
Sec-Glu	121.3 (121.7)	114.8 (114.7)	123.9 (123.6)	123.0 (121.7)	115.3 (116.6)	121.7 (121.4)
Sec-Thr	121.4 (121.5)	114.5 (114.3)	124.1 (124.2)	123.7 (123.3)	115.6 (116.5)	120.7 (120.2)
Sec-Asn	121.5 (121.7)	114.5 (114.6)	124.0 (123.7)	123.5 (122.1)	115.5 (116.9)	120.9 (120.4)
Sec-Arg	121.7 (121.7)	115.3 (114.7)	123.0 (123.6)	122.6 (121.6)	116.4 (117.4)	121.0 (120.8)
Sec-Sec	121.5 (121.9)	114.5 (114.3)	124.0 (123.8)	123.5 (122.2)	115.6 (117.0)	120.7 (120.4)
Average	121.5 (121.6)	114.8 (114.6)	123.7 (123.8)	123.1 (122.3)	115.8 (116.7)	121.0 (120.7)
MD[a]	0.2 (0.3)	0.5 (0.3)	0.7 (0.4)	0.6 (1.0)	0.6 (0.7)	0.7 (0.7)

[a] Maximum deviation from average values

Thus, the data on the total energy 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 12.64 to 20.34 kcal mol^{-1}. The accuracy of DFT method in predicting the rotational constants of conformers of some aliphatic amino acids has been discussed in the literature [46, 47]. 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 seven 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_3–C_4, $C_4 = O_{11}$, C_4–N_6, N_6–H_{10} and N_6–C_7, that very little variance in the bond length values of the amide plane results as the identity of the C-terminal residue of a given dipeptide changes. Maximum deviations of 0.009 Å in gas phase and 0.006 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 exposed $C_4 = O_{11}$ and N_6–H_{10} bonds are elongated up to 0.01 and 0.003 Å respectively; whereas the buried C_4–N_6 bonds are shortened by a range of 0.005 to 0.01 Å for all the systems. Table 3 lists the values of the six bond angles of the amide planes, i.e., C_3–C_4–O_{11}, C_3–C_4–N_6, O_{11}–C_4–N_6, C_4–N_6–C_7, C_4–N_6–H_{10} and H_{10}–N_6–C_7; and the data in both phases indicates very little changes in the bond angle values as the individuality of the C-terminal residue of the dipeptides changes. Maximum deviations of 1.0° in gas phase and 0.7° in aqueous phase indicate that the bond angles are also essentially fixed. The solvent effects on these bond angles are quite apparent when their aqueous phase data is compared with the corresponding gas phase values; a maximum deviation up to 1.4° is observed for C_4–N_6–C_7 and C_4–N_6–H_{10} angles.

Investigating the four dihedral angles of the dipeptides viz. C_3–C_4–N_6–C_7, O_{11}–C_4–N_6–H_{10}, C_3–C_4–N_6–H_{10} and O_{11}–C_4–N_6–C_7, listed in Table 4, can provide valuable information regarding the planarity of the peptide planes. The values of the two dihedral angles C_3–C_4–N_6–C_7 and O_{11}–C_4–N_6–H_{10} should be close to 180° and those for the other two, i.e., C_3–C_4–N_6–H_{10} and O_{11}–C_4–N_6–C_7 should

Table 4 Calculated dihedral angles (in degrees) for the peptide planes of the dipeptides of selenocysteine at B3LYP/6-311++G(d,p) level of theory; the gas phase values are given in brackets

Dipeptides	-SC Groups	C_3–C_4–N_6–C_7	O_{11}–C_4–N_6–H_{10}	C_3–C_4–N_6–H_{10}	O_{11}–C_4–N_6–C_7	ψ	ϕ
Sec-Ala	$-CH_3$	178.0 (173.6)	−179.0 (−179.2)	−0.6 (−0.4)	−0.3 (−5.2)	19.6 (25.9)	−72.6 (−94.5)
Sec-Phe	$-CH_2C_6H_5$	176.6 (177.2)	−178.8 (179.6)	−0.4 (−1.6)	−1.8 (−1.6)	19.9 (26.5)	−89.3 (−97.1)
Sec-Glu	$-CH_2CH_2CO_2H$	176.6 (173.7)	−179.8 (−179.2)	−1.6 (−1.0)	−1.6 (−4.6)	20.1 (28.2)	−119.6 (−137.8)
Sec-Thr	$-CH(CH_3)(OH)$	176.7 (173.3)	179.5 (177.2)	−2.2 (−4.3)	−1.6 (−5.2)	20.4 (28.0)	−108.6 (−125.9)
Sec-Asn	$-CH_2CONH_2$	178.2 (178.3)	177.3 (171.3)	−4.5 (−10.5)	−0.0 (0.0)	22.5 (29.6)	−113.7 (−152.5)
Sec-Arg	$-(CH_2)_3NHC(NH)(NH_2)$	178.0 (177.0)	−179.1 (174.7)	−1.0 (−7.1)	−0.2 (−1.1)	20.0 (28.9)	−67.4 (−149.8)
Sec-Sec	$-CH_2SeH$	179.4 (178.1)	176.9 (173.0)	−4.9 (−8.8)	1.2 (−0.1)	20.9 (30.6)	−121.0 (−141.7)
MD[a]		3.4 (6.7)	3.1 (8.7)	4.9 (10.5)	1.8 (5.2)		

[a] Maximum deviation from expected values

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

Dipeptides	α-carbon atoms C_3			α-carbon atoms C_7		
	N_5–C_3–C_2	N_5–C_3–C_4	C_2–C_3–C_4	N_6–C_7–C_8	N_6–C_7–C_9	C_9–C_7–C_8
Sec-Ala	113.4 (114.3)	110.8 (110.5)	111.5 (110.9)	113.6 (114.1)	110.4 (111.6)	109.6 (109.9)
Sec-Phe	113.5 (114.4)	110.9 (110.4)	111.5 (110.8)	114.0 (114.3)	111.7 (112.1)	109.2 (109.0)
Sec-Glu	113.4 (114.5)	110.7 (110.3)	111.6 (110.9)	113.1 (112.8)	113.4 (113.6)	111.8 (110.6)
Sec-Thr	113.4 (114.4)	110.8 (110.4)	111.5 (110.9)	110.4 (109.2)	111.4 (112.0)	111.0 (111.0)
Sec-Asn	113.5 (114.3)	110.7 (110.2)	111.5 (110.9)	110.5 (106.6)	111.9 (111.3)	111.7 (111.4)
Sec-Arg	113.5 (114.4)	110.7 (110.2)	111.5 (110.9)	112.6 (111.0)	109.4 (112.0)	111.0 (110.5)
Sec-Sec	113.4 (114.5)	110.8 (110.1)	111.4 (110.8)	108.7 (109.8)	110.4 (110.6)	113.3 (111.9)
Average	113.4 (114.4)	110.8 (110.3)	111.5 (110.9)	111.8 (111.1)	111.2 (111.9)	111.1 (110.6)
MD[a]	0.1 (0.1)	0.1 (0.2)	0.1 (0.1)	3.1 (4.5)	2.2 (1.7)	2.2 (1.6)

[a] Maximum deviation from average values

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 up to a maximum value of 4.9° from the expected value whereas in gas phase the maximum deviation observed is 10.5°. Thus, these dihedral angles do not deviate dramatically from their expected values in both phases, however, the extent of deviations observed in the values of the four dihedral angles obviously suggests 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 [28] is evident in the cases of Sec-Ala, Sec-Thr, Sec-Asn and Sec-Arg out of the seven systems considered in this paper. However, in Sec-Phe and Sec-Sec systems the solvent effects could not enhance the planarity of the amide planes. We expect that the conformations of the seven dipeptides predicted at B3LYP/6−311++G(d,p) level are reliable since it has been pointed out that full geometry optimization of gaseous tryptophan conformers at B3LYP/6−311 G(d) and MP2/6−311++G(d,p) levels do not produce any noticeable structural changes, only the conformer energies change by small amounts [48]. Therefore, it is reasonable to assume that solvation effects 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 6 H-bond distances[a] (in angstrom) of the intramolecular H-bond interactions detected in the dipeptides of selenocysteine in gas and aqueous phase

Dipeptides	Phases	N_5…H_{10}-N_6	O_{11}…H-C_7	O_{12}…H-C_7	O_{13}…H-C_7
Sec-Ala	Aqueous	2.179	2.622	*abs*	2.553
	Gas	2.214	2.401	*abs*	2.688
Sec-Phe	Aqueous	2.166	2.466	*abs*	2.615
	Gas	2.209	2.392	*abs*	2.665
Sec-Glu	Aqueous	2.146	2.343	*abs*	*abs*
	Gas	2.236	2.404	*abs*	*abs*
Sec-Thr	Aqueous	2.149	2.411	*abs*	2.532
	Gas	2.214	2.427	*abs*	2.554
Sec-Asn	Aqueous	2.150	2.377	2.597	*abs*
	Gas	2.218	2.530	2.431	*abs*
Sec-Arg	Aqueous	2.180	2.683	*abs*	2.530
	Gas	2.232	2.496	*abs*	2.588
Sec-Sec	Aqueous	2.144	2.395	*abs*	2.536
	Gas	2.229	2.454	*abs*	2.555

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

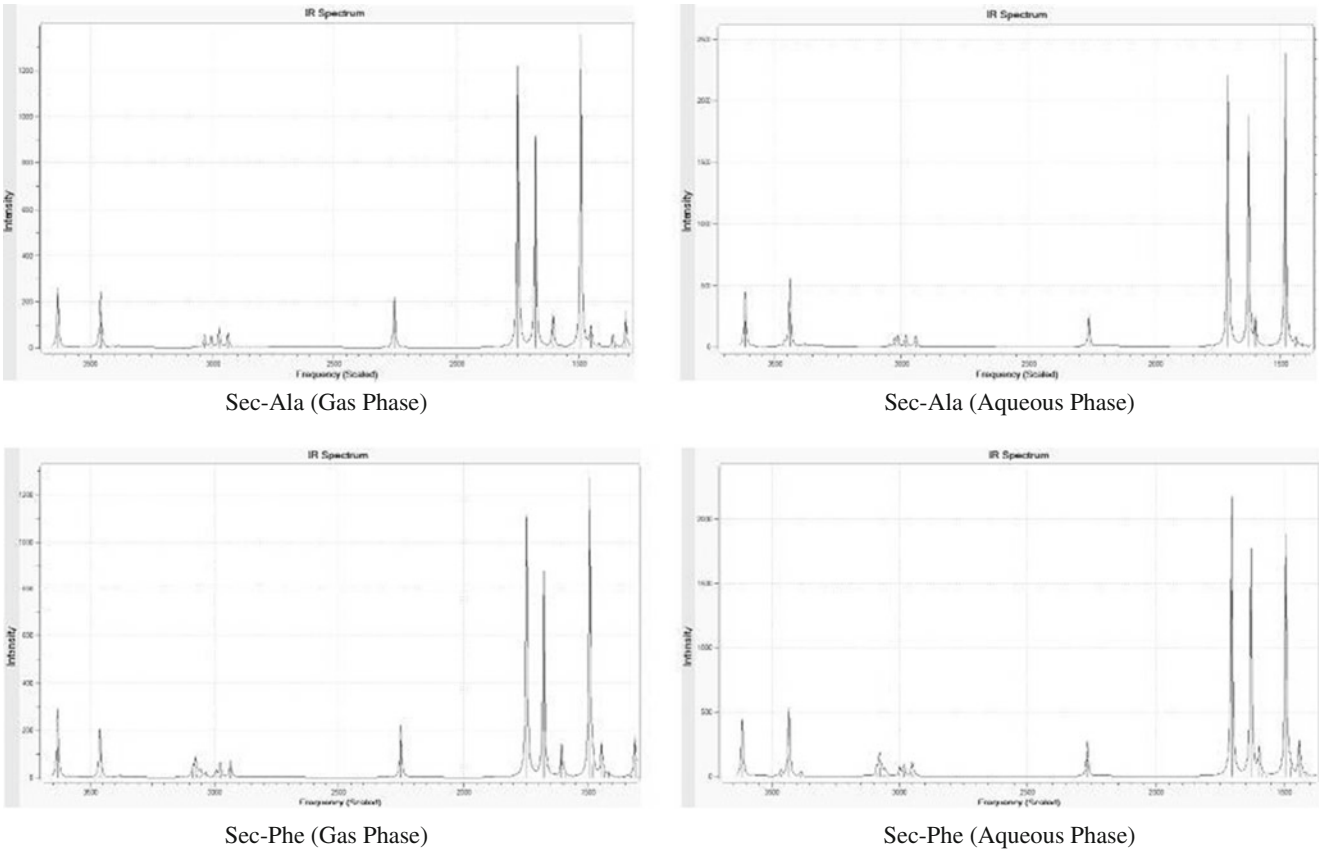
Table 7 Frequencies[a] (in cm^{-1}) and IR intensities (in km mol^{-1}) of various vibrational modes[b] obtained from the theoretical vibrational spectra of the selenocysteine dipeptides in gas and aqueous phase. Intensities are given in brackets

Dipeptides	Phases	$\nu(C_4 = O_{11})$	$\nu(N_6\text{-}H_{10})$	$\nu(C_4\text{-}N_6)$	$\nu_s(N_5\text{-}H)$	$\nu_{as}(N_5\text{-}H)$	$Sis(N_5\text{-}H)$	$\nu(C_7\text{-}H)$	$\nu(C_3\text{-}H)$
Sec-Ala	Aqueous	1696 (509)	3439 (161)	1546 (587)	3380 (9)	3460 (12)	1669 (60)	2981 (7)	2941 (6)
	Gas	1750 (279)	3457 (69)	1557 (374)	3382 (3)	3464 (3)	1675 (42)	2973 (9)	2939 (2)
Sec-Phe	Aqueous	1699 (544)	3431 (157)	1558 (544)	3384 (10)	3463 (13)	1667 (60)	2985 (1)	2941 (6)
	Gas	1750 (251)	3348 (61)	1557 (352)	3385 (3)	3464 (2)	1674 (39)	2977 (4)	2938 (2)
Sec-Glu	Aqueous	1700 (574)	3426 (174)	1550 (607)	3383 (10)	3463 (13)	1669 (57)	2954 (35)	2942 (6)
	Gas	1749 (252)	3471 (69)	1550 (423)	3387 (3)	3463 (7)	1676 (42)	2959 (15)	2938 (2)
Sec-Thr	Aqueous	1705 (506)	3418 (180)	1549 (566)	3384 (12)	3465 (13)	1668 (58)	3020 (12)	2942 (6)
	Gas	1754 (259)	3450 (77)	1549 (379)	3386 (3)	3466 (3)	1676 (43)	3021 (7)	2939 (2)
Sec-Asn	Aqueous	1700 (110)	3410 (173)	1554 (490)	3384 (13)	3464 (15)	1670 (59)	2986 (22)	2944 (6)
	Gas	1744 (339)	3443 (80)	1545 (343)	3385 (3)	3466 (4)	1677 (41)	3022 (1)	2942 (2)
Sec-Arg	Aqueous	1697 (493)	3436 (162)	1549 (607)	3382 (10)	3462 (12)	1669 (60)	2978 (17)	2941 (6)
	Gas	1743 (239)	3463 (81)	1545 (444)	3382 (3)	3465 (10)	1676 (43)	2996 (5)	2939 (2)
Sec-Sec	Aqueous	1702 (544)	3414 (190)	1540 (677)	3384 (14)	3464 (14)	1670 (58)	3028 (7)	2944 (6)
	Gas	1752 (264)	3454 (85)	1543 (444)	3386 (3)	3465 (3)	1677 (43)	3015 (4)	2940 (2)

[a] The frequencies below 1800 cm^{-1} are scaled with 1.01 and for those above 1800 cm^{-1} a correction factor 0.9679 is used

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

Table 4 also lists the –SC groups of the C-terminal residues of the dipeptides as well as the gas and aqueous phase values of the ψ and ϕ angles. 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



Sec-Ala (Gas Phase) Sec-Ala (Aqueous Phase)

Sec-Phe (Gas Phase) Sec-Phe (Aqueous Phase)

Fig. 3 Vibrational spectra of Sec-Ala and Sec-Phe in gas and aqueous phase

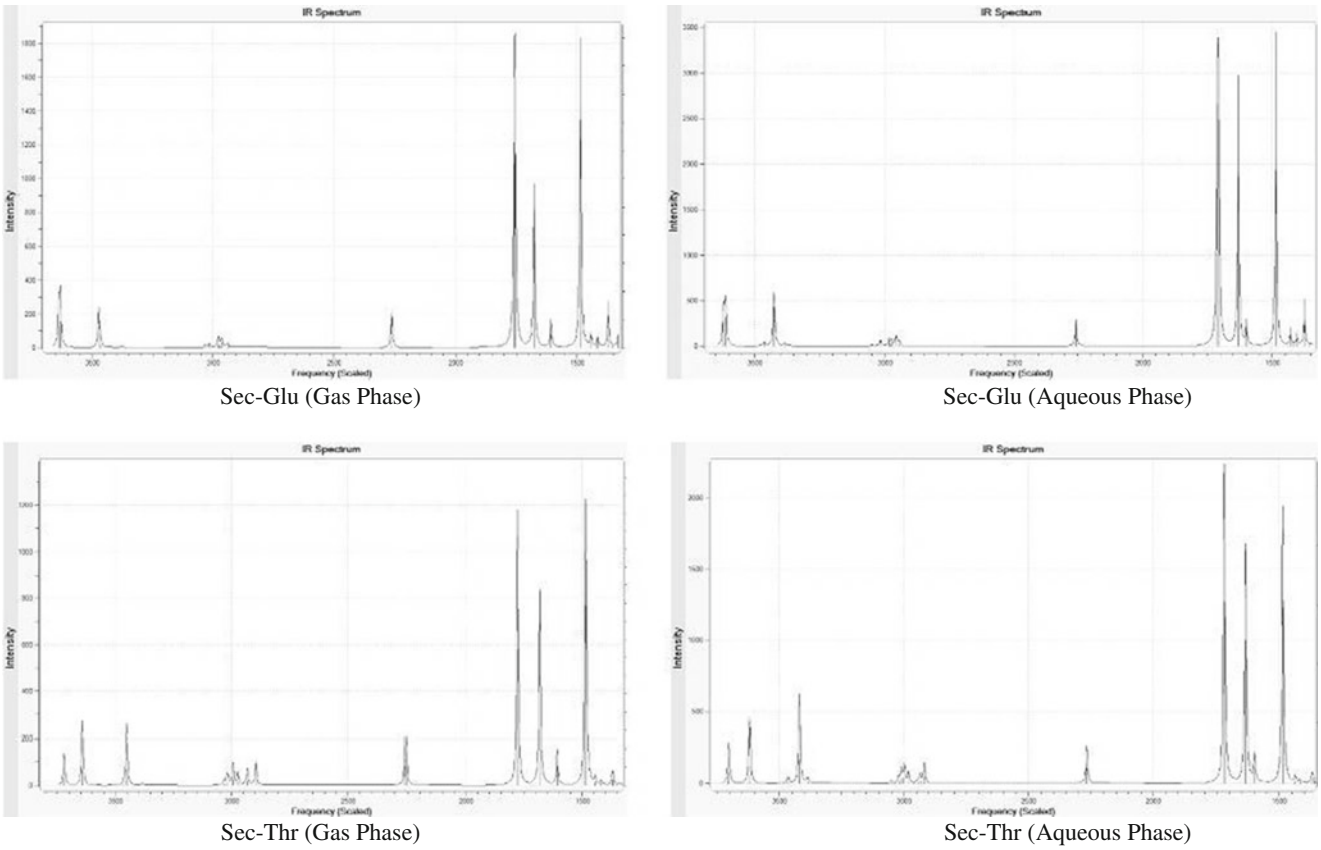
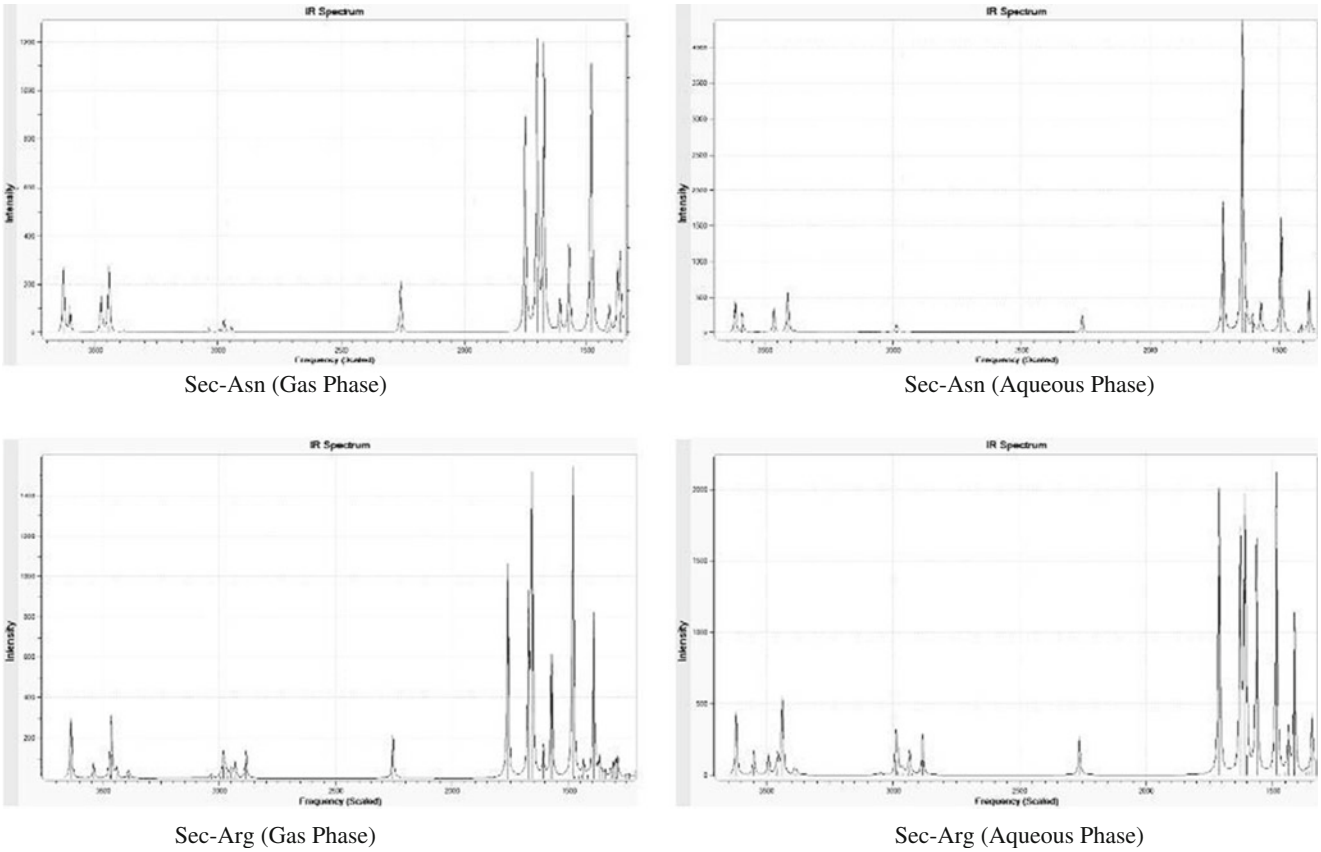


Fig. 4 Vibrational spectra of Sec-Glu and Sec-Thr in gas and aqueous phase

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 phase values of the ϕ angles reveal that the value of ϕ increases as the size of a given –SC group increases. This point has been well discussed in various literature [15, 16]. Among the seven dipeptides studied in this paper the ϕ value of Sec-Ala in gas phase is −94.5° while in the other six systems, which have much bigger sized –SC groups compared to that of Sec-Ala, the ϕ values increase up to −152.5°. 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 ϕ as well as planarity of the amide planes. The aqueous phase values of the ϕ angles of the dipeptides reveal that in solvent phase the type of functional groups present in the –SC groups is more important in influencing the ϕ values of the dipeptides than the size of the -SC groups. For example, in aqueous phase the ϕ value of Sec-Arg is smaller than that of Sec-Ala even though the –SC group of Arg is much bigger in size than Ala. As shown in Fig. 7, in aqueous phase the $-(CH_2)_3NHC(NH)(NH_2)$ group of Arg residue adopts a different orientation from that in gas phase allocating more physical space to the carboxylic group of the Arg residue. It has been suggested that polar solvents remarkably influence the conformational properties of dipeptides, by weakening the intra-residue hydrogen bonds and leading to the appearance of new energy minima [30–32]. Thus, as a result of this new orientation of the $-(CH_2)_3NHC(NH)(NH_2)$ group of Arg residue adopted in aqueous phase the ϕ value is reduced to −67.4° (−149.8° in gas phase).

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 role in deciding the overall structure of the proteins. The three bond angles considered to monitor the geometry around the C_3 α-carbon atoms of the dipeptides are N_5–C_3–C_2, N_5–C_3–C_4 and C_2–C_3–C_4 while N_6–C_7–C_8, N_6–C_7–C_9 and C_9–C_7–C_8 are the same for the C_7 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 residue can affect the geometries

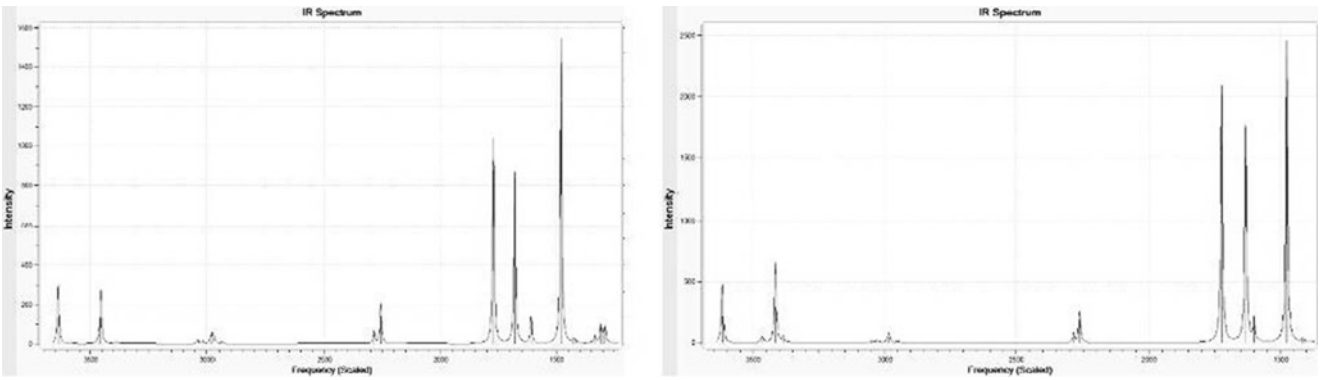


Sec-Asn (Gas Phase) Sec-Asn (Aqueous Phase)

Sec-Arg (Gas Phase) Sec-Arg (Aqueous Phase)

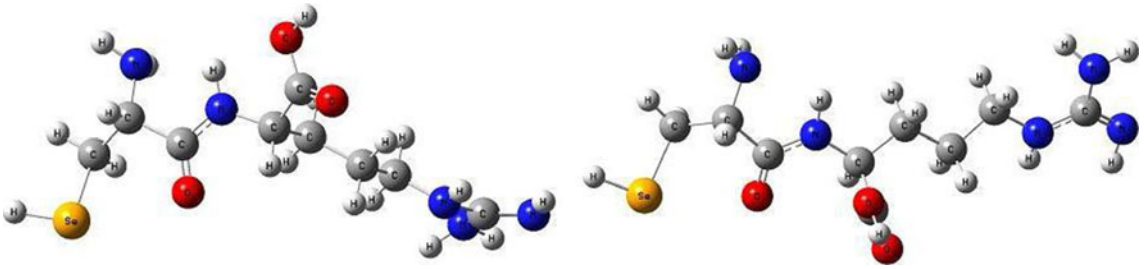
Fig. 5 Vibrational spectra of Sec-Asn and Sec-Arg in gas and aqueous phase

about these α-carbon atoms. This DFT study also provides us 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 0.1° in aqueous and 0.2° in gas phase from their respective average values suggest that the geometries about the C_3 atoms do not change much with the change in the identity of the C-terminal residues. On the other hand, with maximum deviations up to 3.1° in aqueous and 4.5° in gas phase from their respective average values, the bond angles around the C_7 change appreciably with the change in identity of the C-terminal residue of the dipeptides. These observations can be justified by invoking the 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_3 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_7 atoms the geometry



Sec-Sec (Gas Phase) Sec-Sec (Aqueous Phase)

Fig. 6 Vibrational spectra of Sec-Sec in gas and aqueous phase



Sec-Arg (Gas Phase) Sec-Arg (Aqueous Phase)

Fig. 7 Optimized structures of Sec-Arg system in gas and aqueous phase

around them are affected by the changing identity of the –SC groups. The solvation effects are also more prominent on the geometry of the C_7 atoms (a maximum deviation up to 3.9° is observed for the angle N_6–C_7–C_8 in Sec-Asn system) than that on the C_3 atoms where the maximum deviation predicted is only up to 1.1° for the N_5–C_3–C_2 angle.

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 [16]. The strength of these H-bonds depends on two factors, (a) is the distance A–H….B shorter than the sum of their van der Waals radii and (b) closer the angle A–H….B to 180° [14], where A–H is H-bond donor and B is H-bond acceptor. Table 6 lists two types of intramolecular H-bonds, namely 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 selenocysteine containing dipeptides. The gas phase intramolecular H-bond combinations of the dipeptides are similar to those in the aqueous phase. In gas phase the B….H distances of the two H-bonds N_5…H_{10}-N_6 and O_{11}…H-C_7 range from 2.209 to 2.53 Å while in aqueous phase they range from 2.146 to 2.683 Å. On the other hand, the gas and solvent phase data on the two H-bonds O_{12}…H-C_7 and O_{13}…H-C_7 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 absence of O_{13}…H-C_7 bonds only in the cases of Sec-Glu and Sec-Asn systems can be explained on the basis of the identity of their –SC groups belonging to the C-terminal residues. Similarly, the presence of O_{12}…H-C_7 only in the case of Sec-Asn also depends on the identity of the –SC group of the Asn residue.

Vibrational spectra

The theoretically predicted vibrational spectra of the seven selenocysteine dipeptides in both 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 shortening of N_5…H_{10}-N_6 bonds in aqueous phase structures is well reflected by the lowering in the frequency values of the $\nu(N_6$-$H_{10})$ stretching by a range of 17 to 45 cm^{-1} than those in the gas phase. Solvent effects also lower the frequency values of the $\nu(C_4 = O_{11})$, $\nu_s(N_5$-$H)$, $\nu_{as}(N_5$-$H)$ and $Sis(N_5$-$H)$ modes by a magnitude up to 54 cm^{-1} in the aqueous phase which can be due to elongation in the bond length values in solvent phase (the $C_4 = O_{11}$ bonds are elongated up to 0.01 Å in the aqueous phase). The variations in the $\nu(C_7$-$H)$ stretching values can be attributed to the effects of the changing identity of the –SC groups of the C-terminal residues.

Conclusions

This DFT study on dipeptides containing selenocysteine as a fixed component at their N-terminal positions predicts large values of total dipole moments for the dipeptides, 4.805 to 7.141 D in gas phase and 8.345 to 11.072 D in aqueous phase, and as a consequence the aqueous phase structures show more thermodynamic stabilities by a range of 12.64 to 20.34 kcal mol^{-1} than those in the gas phase. The identity of the varying C-terminal residue influences the values of ϕ, planarity of the peptide planes and geometry around the C_7 α-carbon atoms while the solvation effects are evident on the values of bond lengths and bond angles of the amide planes. The geometry of the amide planes are not perfectly planar regardless of whether the systems are in gas or in strong polar solvents like 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. In gas phase the ϕ values depend on the size of a given –SC group which is evident from the fact that the ϕ value of Sec-Ala in gas phase is −94.5° while in the other six systems, that have much bigger sized –SC groups than Sec-Ala, the ϕ values increase up to −152.5°. However, in solvent phase the type of functional group present in the –SC groups is more important in influencing the ϕ values of the dipeptides than the size of the–SC groups which is evident from the fact that the ϕ value of Sec-Arg is smaller than that of Sec-Ala even though the –SC group of Arg is much bigger in size than Ala. The presence or absence of two types of intramolecular H-bonds, namely 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 selenocysteine dipeptides. The variations in the values of $\nu(C_7\text{-H})$ stretching frequencies of the dipeptides reflect the effects of the changing –SC groups on the geometry around the C_7 atoms.

Acknowledgments Financial assistance from the Special Assistance Program of the University Grants Commission to the Department of Chemistry, NEHU, is gratefully acknowledged. The authors gratefully acknowledge Dr. AK Chandra for useful discussions and suggestions. SM is also grateful to the University Grants Commission, Government of India, New Delhi, for financial assistance through a research fellowship.

References

1. Campbell NA (1996) In Biology, 4th edn. Cummings, New York
2. Garrett RH (1999) In Biochemistry. Saunders College, New York, Ch 6
3. Rother M, Krzycki JA (2010) Archaea 1–14
4. Ibba M, Soll D (2004) Genes Dev 18:731–738
5. Lukashenko NP (2010) Russ J Genet 46:899–916
6. Chambers I, Frampton J, Goldfarb P, Affara N, McBain W, Harrison PR (1986) EMBO J 5:1221–1227
7. Zinoni F, Birkmann A, Stadtman TC, Bock A (1986) Proc Natl Acad Sci U S A 83:4650–4654
8. Bock A, Forchhammer K, Heider J, Leinfelder W, Sawers G, Veprek B, Zinoni F (1991) Mol Microbiol 5:515–520
9. Stadtman TC (1996) Annu Rev Biochem 65:83–100
10. Hatfield DL, Gladyshev VN (2002) Mol Cell Biol 11:3565–3576
11. Bock A, Forchhammer K, Heider J, Baron C (1991) Trends Biochem Sci 16:463–467
12. Dudev T, Lim C (2009) J Phys Chem B 113:11754–11764
13. Spezia R, Tournois G, Cartailler T, Tortajada J, Jeanvoine YJ (2006) Phys Chem A 110:9727–9735
14. Kaur D, Sharma P, Bharatam PV, Kaur M (2008) Int J Quant Chem 108:983–991
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) Comput Theor Chem 969:76–82
18. Vargas R, Garza J, Hay BP, Dixon DA (2002) J Phys Chem A 106:3213–3218
19. Wormald MR, Petrescu AJ, Pao YL, Glithero A, Elliott T, Dwek RA (2002) Chem Rev 102:371–386
20. Foloppe N, Hartmann B, Nilsson L, MacKerell AD Jr (2002) Biophys J 82:1554–1569
21. Sponer J, Zgarbova M, Jurecka P, Riley KE, Sponer JE, Hobza P (2009) J Chem Theory Comput 5:1166–1179
22. Antohi O, Naider F, Sapse AM (1996) J Mol Struct (THEOCHEM) 360:99–108
23. Kapota C, Ohanessian G (2005) Phys Chem Chem Phys 7:3744–3755
24. Kolev T, Koleva BB, Spiteller M (2007) Amino Acids 33:719–725
25. Koleva BB, Kolev TM, Todorov S (2007) Chem Papers 61:490–496
26. Clavaguera C, Piuzzi F, Dognon JP (2009) J Phys Chem B 113:16443–16448
27. Tobiast DJ, Brooks CL III (1992) J Phys Chem 96:3864–3870
28. Wang ZX, Duan Y (2004) J Comput Chem 25:1699–1716
29. García-Prieto FF, Galvan IF, Aguilar MA, Martín ME (2011) J Chem Phys 135:194502–194509
30. Gould IR, Cornell WD, Hillier IH (1994) J Am Chem Soc 116:9250–9256
31. Head-Gordon T, Head-Gordon M, Frisch MJ, Brooks CL III, Pople J (1991) J Am Chem Soc 113:5989–5997
32. Adamo C, Dillet V, Barone V (1996) Chem Phys Lett 263:113–118
33. Ramachandran GN (1963) Biopolymers 6:1494–1495
34. Ramachandran GN, Ramakrishnan C, Sasisekharan V (1963) J Mol Biol 7:95–99
35. Osawa S, Jukes TH, Watanabe K, Muto A (1992) Microbiological reviews 229–264
36. Becke AD (1993) J Chem Phys 98:5648–5652
37. Lee C, Yang W, Parr RG (1988) Phys Rev B 37:785–789
38. 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, Pople JA (2004) Gaussian 03, Revision D.01. Gaussian Inc, Wallingford, CT
39. Tehrani ZA, Tavasoli E, Fattahi A (2010) J Mol Struct (THEOCHEM) 960:73–85
40. Miertus S, Scrocco E, Tomasi J (1981) Chem Phys 55:117–129
41. Gould IR, Hillier IH (1993) J Chem Soc, Chem Commun 951–952
42. Andersson MP, Uvdal P (2005) J Phys Chem A 109:2937–2941
43. Foresman JB, Frisch A (1996) Exploring Chemistry with Electronic Structure Methods, 2nd edn. Gaussian, Inc, Pittsburgh
44. Freeman F, Le KT (2003) J Phys Chem A 107:2908–2918
45. Chasse GA, Rodriguez AM, Mak ML, Deretey E, Perczel A, Sosa CP, Enriz RD, Csizmadia IG (2001) J Mol Struct (THEOCHEM) 537:319–361
46. Stepanian SG, Reva ID, Radchenko ED, Rosado MTS, Duarte MLTS, Fausto R, Adamowicz L (1998) J Phys Chem A 102:1041–1054
47. Stepanian SG, Reva ID, Radchenko ED, Adamowicz L (1998) J Phys Chem A 102:4623–4629
48. Huang Z, Lin Z (2005) J Phys Chem A 109:2656–2659