Analysis on conformational stability of C-peptide of ribonuclease A in water using the reference interaction site model theory and Monte Carlo simulated annealing

Masahiro Kinoshita, Yuko Okamoto, and Fumio Hirata

Citation: J. Chem. Phys. **110**, 4090 (1999); doi: 10.1063/1.478290
View online: https://doi.org/10.1063/1.478290
View Table of Contents: http://aip.scitation.org/toc/jcp/110/8
Published by the American Institute of Physics

Analysis on conformational stability of C-peptide of ribonuclease A in water using the reference interaction site model theory and Monte Carlo simulated annealing

Masahiro Kinoshita[a)]
Advanced Energy Utilization Division, Institute of Advanced Energy, Kyoto University, Uji, Kyoto 611-0011, Japan

Yuko Okamoto and Fumio Hirata
Department of Theoretical Studies, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan

(Received 21 August 1998; accepted 17 November 1998)

Solvation structure and conformational stability of the C-peptide fragment of ribonuclease A in pure water have been analyzed using the full reference interaction site model (RISM) theory. The charged groups in the side chains of Lys-1^+, Glu-2^-, Lys-7^+, Arg-10^+, and His-12^+ (in particular, the four like-charged groups) play substantial roles in stabilizing the conformations. The solvation free energy and the conformational energy are governed by the contribution from the electrostatic interaction with water and the intramolecular Coulombic energy, respectively, and the conformational stability is determined by competition of these two factors. The contributions from the hydrophobic hydration and the van der Waals and torsion terms in the conformational energy are less important, which is in contrast to the result for Met-enkephalin. The Monte Carlo simulated annealing combined with the RISM theory has been applied to the C-peptide using an almost fully extended conformation as the initial one. The conformation first changes in the direction that the charged groups in the side chains are more exposed to water, and in particular, the positively charged groups are closer together. Thus, the solvation free energy decreases greatly in the initial stage. Although this leads to a significant increase in the intramolecular Coulombic repulsion energy, the decrease in the solvation free energy dominates. In the later stage, however, a further decrease in the solvation free energy gives rise to an even larger increase in the intramolecular Coulombic repulsion energy, and the conformational change is greatly decelerated. The conformations thus stabilized in four different runs of the combined program are quite similar. The peptide conformation in water is stabilized far more rapidly than in the gas phase. © *1999 American Institute of Physics.* [S0021-9606(99)50808-8]

I. INTRODUCTION

The prediction of tertiary structures of proteins from their primary structures, which is one of the most challenging problems in molecular biology, amounts to finding the conformation of a protein with the lowest energy from among possible conformations. However, there are two major stumbling blocks in achieving this goal. One of them is that the number of possible conformations is astronomically large. The other is that the solvent effects can be essential and need to be taken into account in an explicit manner. To overcome the first one, promising simulation methods such as the simulated annealing technique[1] and the generalized-ensemble algorithms[2,3] were developed, and their usefulness was demonstrated for problems of polypeptide conformation prediction in the gas phase.[4–8] As for the second one, we employ the dielectrically consistent reference interaction site model (RISM) theory,[9,10] a statistical-mechanical theory for molecular fluids: we recently developed an algorithm[11,12] for solving the full RISM equations for the system of a solute molecule with many atoms in water and salt solution that is more robust and orders of magnitude faster than the conventional one.

In our earlier work,[13] we analyzed solvation structure and conformational stability of Met-enkephalin (Tyr-Gly-Gly-Phe-Met; the total number of atoms is 75) in the extended simple point charge (SPC/E) water[14] using the RISM theory. The relation between the solvation free energy of the peptide and its conformation was discussed on the atomic level. It was shown that the peptide conformations are greatly sensitive to the microscopic solvent environment and any naive treatment of the solvent, such as the continuum model, will end in failure.

Later, we reported results of the first attempt[15] to combine the Monte Carlo (MC) simulated annealing and the RISM theory. In the combined approach, the energy function is the total energy defined as the sum of the conformational energy and the solvation free energy, and the RISM theory is employed to calculate the latter. The effectiveness of the combined program was demonstrated for Met-enkephalin in water. It was found that most of the conformations with larger solvation free energies are strongly rejected by water and the number of probable conformations is drastically re-

a)Author to whom correspondence should be addressed.

duced. Met-enkephalin is forced to take conformations favored by water much more rapidly than in the gas phase. The set of stable conformations obtained is characterized by almost fully extended backbone structure with large fluctuations in side-chain structure, and is in qualitatively good agreement with those determined by a recent nuclear magnetic resonance (NMR) experiment.[16]

In the present article, we extend the studies for Met-enkephalin to a larger peptide, the C-peptide fragment of ribonuclease A (RNase A) with 221 atoms. A feature of the C-peptide, which is not inherent in Met-enkephalin, is that five of the residues (i.e., Lys-1, Glu-2, Lys-7, Arg-10, and His-12) have charged groups in their side chains. Four representative conformations of the C-peptide in the SPC/E water are considered, and the relation between the conformation and the solvation properties is discussed in detail. The computer program, where the MC simulated annealing is combined with the RISM theory, is applied to the C-peptide. The results obtained are compared with those for Met-enkephalin.

II. METHODS

A. RISM theory

In the present article the subscripts ''V'' and ''S'' denote ''water'' and ''solute,'' respectively. It is assumed that the solute (peptide) is present at the infinite-dilution limit. The calculation process is then split into two steps where bulk water (step 1) and water near a solute molecule (step 2) are successively treated. The calculation in step 1 is performed using the dielectrically consistent RISM theory developed by Perkyns and Pettitt.[9,10] The site–site intermolecular total correlation functions calculated in step 1 are used as input variables for step 2. We consider step 2 hereafter.

It is assumed that the solute molecule and a water molecule have m and 3 interaction sites, respectively. The site–site Ornstein–Zernike (SSOZ) equations are expressed as

$$\tilde{\eta}_{sv}=\tilde{w}_{ss}\tilde{c}_{sv}\tilde{H}_{vv}-\tilde{c}_{sv}, \tag{1a}$$

$$\tilde{\eta}_{sv}=\tilde{h}_{sv}-\tilde{c}_{sv}, \tag{1b}$$

$$\tilde{H}_{vv}=\tilde{w}_{vv}+\rho_v\tilde{h}_{vv}, \tag{2}$$

where $\tilde{H}_{vv}$, $\tilde{\eta}_{sv}$, and $\tilde{w}_{ss}$, for example, are 3×3, $m\times3$, and $m\times m$ matrices, respectively. ρ_v is the number-density matrix for water molecules in the bulk, h is the matrix of site–site intermolecular total correlation functions, c is the matrix of site–site intermolecular direct correlation functions, w is the intramolecular correlation matrix, and ''~'' represents Fourier transforms. $\tilde{H}_{vv}$ is dependent on properties of the bulk water alone and is part of the input data for step 2. More detailed information is given in Ref. 11.

The closure equation employed is of the hypernetted-chain (HNC) type given by

$$c_{AB}(r)=\exp\{-u_{AB}(r)/(k_BT)+\eta_{AB}(r)\}-\eta_{AB}(r)-1, \quad A=1,...,m; \quad B=\mathrm{H,O}, \tag{3}$$

where $u_{AH}(u_{AO})$ is the pair potential between site A of the solute molecule and the water-hydrogen (the water-oxygen), k_B is the Boltzmann constant, and T is the absolute temperature. For instance, $c_{AH}(c_{AO})$ is the site–site direct correlation function between site A of the solute molecule and the water-hydrogen (the water-oxygen).

The solvation free energy for the solute molecule $\Delta\mu_s$ is calculated from[17]

$$\Delta\mu_s/(k_BT)=4\pi\int_0^\infty F(r)dr, \tag{4a}$$

$$F(r)=\sum_{A=1}^{m}\sum_{B=\mathrm{H,O}}\rho_B r^2[\{h_{AB}(r)\}^2/2-c_{AB}(r)-h_{AB}(r)c_{AB}(r)/2], \tag{4b}$$

$$\rho_{\mathrm{H}}=2\rho_v, \quad \rho_{\mathrm{O}}=\rho_v. \tag{4c}$$

The site–site correlation functions $h_{AB}(r)$ and $c_{AB}(r)$ are calculated by solving Eqs. (1) and (3). We also consider the special case where all the site charges of the solute molecule are set to zero. The values of the solvation free energy in this case are denoted by $\Delta\mu_{S0}$. $\Delta\mu_{S0}$ can be regarded as the contribution from the hydrophobic hydration to $\Delta\mu_S$. The difference,

$$\Delta\mu_{SE}=\Delta\mu_S-\Delta\mu_{S0}, \tag{5}$$

implies the contribution from the electrostatic interaction with water.

B. Model

The sequence of the C-peptide considered is Lys-Glu-Thr-Ala-Ala-Ala-Lys-Phe-Leu-Arg-Gln-His-Met. This peptide, which was previously treated in theoretical[18,19] and experimental[20–22] works, is slightly different from residues 1–13 of the native enzyme: the ninth residue Glu is replaced by Leu. A feature of the peptide is that five of the residues (Lys-1, Glu-2, Lys-7, Arg-10, and His-12) have charged groups in their side chains. To make this feature clear, we represent Lys-1, Glu-2, Lys-7, Arg-10, and His-12 by Lys-1^+, Glu-2^-, Lys-7^+, Arg-10^+, and His-12^+, respectively. In the present analysis, it is assumed that the peptide is present in the unionized form.

The model of a water molecule is the SPC/E model.[14] The temperature is set at 273 K. $u_{AB}(r)$ has the form

$$u_{AB}(r)=q_Aq_B/r+4\varepsilon_{AB}\{(\sigma_{AB}/r)^{12}-(\sigma_{AB}/r)^6\}, \quad A=1,...,m; \quad B=\mathrm{H,O}, \tag{6}$$

where q_A is the partial charge on site A of the solute molecule and the standard combination rule

$$\varepsilon_{AB}=(\varepsilon_A\varepsilon_B)^{1/2}, \quad \sigma_{AB}=(\sigma_A+\sigma_B)/2 \tag{7}$$

is employed for calculating the Lennard-Jones potential parameters. The potential energy functions and parameters are adopted from KONF90 (Refs. 18 and 19), which is based on ECEPP/2 (Refs. 23–25) (the dielectric constant is set to

TABLE I. Values of q_A and σ_A used for the C-peptide. The site charges are scaled by the magnitude of the electronic charge.

A			$q_A(-)$	σ_A (nm)
1	1H	Lys-1^+	0.176	0.239
2	N	Lys-1^+	−0.356	0.313
3	2H	Lys-1^+	0.176	0.239
4	CB	Lys-1^+	−0.030	0.367
5	1HB	Lys-1^+	0.015	0.260
6	2HB	Lys-1^+	0.015	0.260
22	C	Lys-1^+	0.450	0.333
23	O	Lys-1^+	−0.384	0.278
24	N	Glu-2^-	−0.356	0.313
25	H	Glu-2^-	0.176	0.239
32	CD	Glu-2^-	0.500	0.333
33	OE1	Glu-2^-	−0.570	0.278
34	OE2	Glu-2^-	−0.570	0.278
97	NZ	Lys-7^+	−0.320	0.313
98	1HZ	Lys-7^+	0.320	0.239
99	2HZ	Lys-7^+	0.320	0.239
100	3HZ	Lys-7^+	0.320	0.239
119	CD2	Leu-9	−0.015	0.330
120	HD2	Leu-9	0.010	0.261
158	NH1	Arg-10^+	−0.390	0.313
159	NH2	Arg-10^+	−0.390	0.313
160	1HH1	Arg-10^+	0.280	0.239
161	2HH1	Arg-10^+	0.280	0.239
162	1HH2	Arg-10^+	0.280	0.239
163	2HH2	Arg-10^+	0.280	0.239
191	ND1	His-12^+	−0.200	0.313
193	HD1	His-12^+	0.275	0.239
195	NE2	His-12^+	−0.200	0.313
196	HE2	His-12^+	0.265	0.239
218	C	Met-13	0.450	0.333
219	O	Met-13	−0.384	0.278
220	OXT	Met-13	−0.380	0.289
221	HXT	Met-13	0.204	0.252

unity). The values of q_A and σ_A of some representative atoms are given in Table I (the charges are scaled by the magnitude of the electronic charge). For the SPC/E water, we have $q_{\mathrm{H}}=0.4238$, $q_{\mathrm{O}}=-0.8476$, $\varepsilon_{\mathrm{H}}=0.046$ kcal/mol, $\varepsilon_{\mathrm{O}}=0.156$ kcal/mol, $\sigma_{\mathrm{H}}=0.040$ nm, and $\sigma_{\mathrm{O}}=0.316$ nm. The dimensionless number density of water $\rho_v d^3 (d=0.28\,\mathrm{nm})$ is 0.7317.

We consider four representative conformations illustrated in Figs. 1 and 2. Conformation 1 is obtained using the torsion angles from the x-ray diffraction data for the corresponding segment of the native enzyme, by minimizing the conformational energy with the MC simulated annealing. It has an α-helix structure[26] and intramolecular hydrogen bonding between ''33 OE1 Glu-2^-'' and ''162 1HH2 Arg-10^+,'' between ''33 OE1 Glu-2^-'' and ''99 2HZ Lys-7^+,'' and between ''34 OE2 Glu-2'' and ''156 HE1 Arg-10^+.'' Conformation 2 is an almost fully extended one. Conformation 3 is one of those with very low solvation free energies found in the course of the present study. A feature of conformation 3 is that the four positively charged groups in the side chains of Lys-1^+, Lys-7^+, Arg-10^+, and His-12^+ are relatively close together. Conformation 4 is the one with the minimum value of the energy function (i.e., the total energy defined as the sum of the conformational energy and the solvation free energy) obtained in the present study. As observed in Fig. 2, the backbone structure of conformations 3 and 4 is rather extended. These conformations will be discussed in more detail in Sec. III.

C. Numerical method

A sufficiently long-range r_L is divided into N mesh points ($r_i=i\,\delta r$, $i=0,1,...,N-1$; $\delta r=r_L/N$), and all the functions are represented by their values on these points. The long-range Coulomb potentials are handled in a special manner so that r_L can be minimized. The details of the algorithm for solving the full RISM equations are described in Ref. 11. The algorithm is a hybrid of the Newton–Raphson and Picard methods, but the Jacobian matrix is treated as part of the input data and need not be recalculated at all.

We now briefly summarize the MC simulated annealing combined with the RISM theory.[15] The scheme for the MC simulated annealing is the same as that employed in the work in the gas phase,[18,19] except that the conformational energy E_C is replaced by the total energy E_T defined by

$$E_T=E_C+\Delta\mu_S. \tag{8}$$

With the RISM theory, we can mimic a peptide immersed in an essentially infinite number of solvent molecules. Moreover, the ensemble-averaged structure of the solvent, which is in equilibrium with the peptide in a particular conformation, is theoretically calculated. Thus, our method is much more efficient than the usual computer simulations treating the peptide and the solvent molecules simultaneously. We note that the important quantities are not the absolute values of E_C, $\Delta\mu_S$, and E_T but the relative values among different conformations of the peptide. The validity of using the RISM theory in the calculation of $\Delta\mu_S$ was already demonstrated for Met-enkephalin in water.[15]

The computer program for solving the RISM equations was incorporated in the program for the MC simulated annealing as a subroutine, and the combined program[15] was thus developed. Starting from an initial conformation of the peptide given, the combined program samples many conformations in accordance with the simulated annealing scheme, and then finds the conformation with the minimum value of E_T. One MC sweep updates all the torsion angles (there are 64 torsion angles for the C-peptide) of the peptide once. The annealing schedule is as follows. The temperature is lowered exponentially from T_I to T_F. The temperature for the nth MC sweep T_n is given by

$$T_n=T_I\gamma^{n-1},\quad \gamma=\exp[\{\ln(T_F/T_I)\}/(M-1)], \tag{9}$$

where M is the total number of MC sweeps. Typical values of the initial and final temperatures are $T_I=473$ and $T_F=273$ K, respectively, but different values are also tested.

The RISM equations are fully solved to calculate $\Delta\mu_S$ for each conformation sampled. Our algorithm for solving the RISM equations is particularly amenable to the construction of the combined program because it is far more robust and orders of magnitude faster than the conventional one.[11,12,15] Further, it has been found that the same matrix can be used for a large set of different conformations and the peptide–water interactions:[11,15] in all the calculations pre-

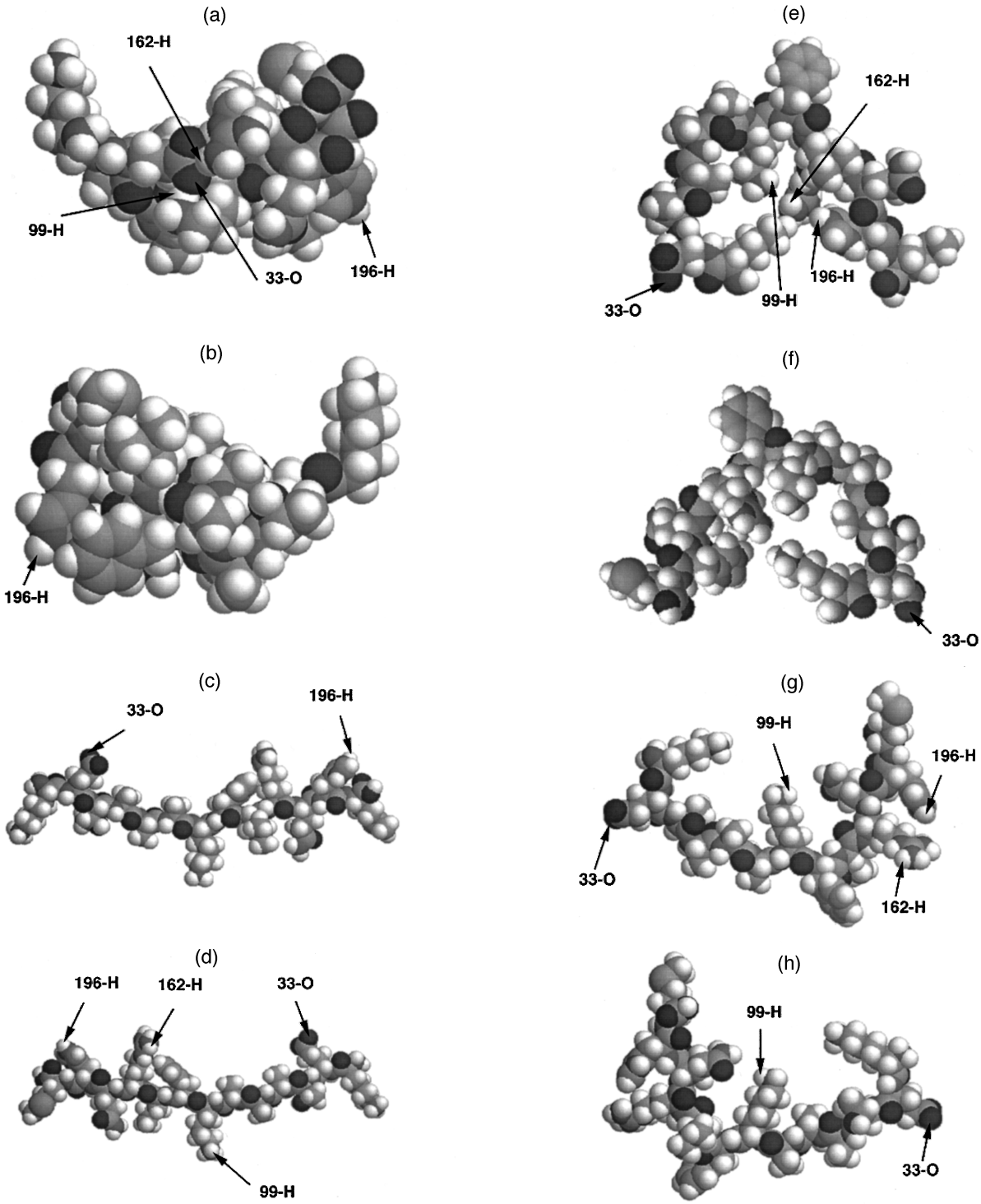

FIG. 1. Four different conformations of the C-peptide considered. ''33-O,'' ''99-H,'' ''162-H,'' and ''196-H'' represent ''33 OE1 Glu-2$^-$,'' ''99 2HZ Lys-7$^+$,'' ''162 1HH2 Arg-10$^+$,'' and ''196 HE2 His-12$^+$,'' respectively. This figure was prepared by RASMOL. (a) Conformation 1. (b) Conformation 1 viewed from another angle. (c) Conformation 2. (d) Conformation 2 viewed from another angle. (e) Conformation 3. (f) Conformation 3 viewed from another angle. (g) Conformation 4. (h) Conformation 4 viewed from another angle.

sented in this article the Jacobian matrix calculated for conformation 2 is always used with no modification.

III. RESULTS AND DISCUSSION

A. Solvation free energy, conformational energy, and total energy

The solvation free energy $\Delta\mu_S$, conformational energy E_C, and total energy E_T calculated for the four conformations are given in Table II. The solvation free energy is split into two terms, the contribution from the hydrophobic hydration $\Delta\mu_{S0}$ and that from the electrostatic interaction with water $\Delta\mu_{SE}$. The conformational energy is also split into two terms, the intramolecular Coulombic (electrostatic) energy E_{CE} and the intramolecular van der Waals energy plus the torsion energy E_{C0}. The contributions from the 13 residues to $\Delta\mu_S$ are given in Table III. We are particularly

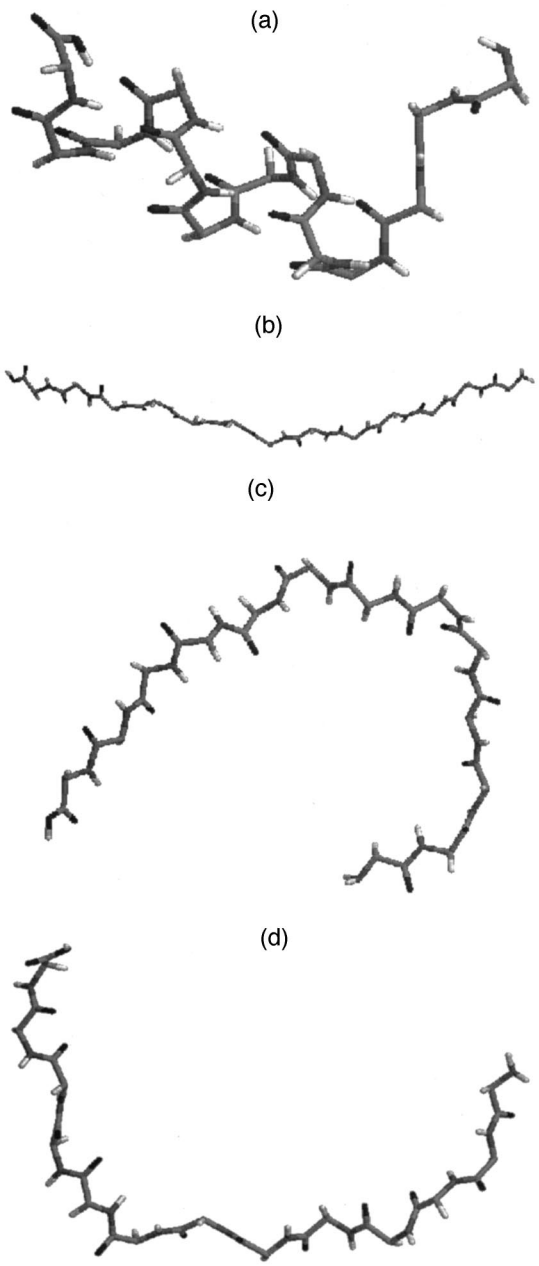

FIG. 2. Backbone structure of conformations 1–4. This figure was prepared by RASMOL. (a) Conformation 1. (b) Conformation 2. (c) Conformation 3. (d) Conformation 4.

concerned with the values for Lys-1^+, Glu-2^-, Lys-7^+, Arg-10^+, and His-12^+ in the table.

In conformation 1, the charged groups in the side chains of Glu-2^-, Lys-7^+, and Arg-10^+ are not well exposed to water, as observed in Fig. 1. Intramolecular hydrogen bonding is present between ''33 OE1 Glu-2^-'' and ''162 1HH2 Arg-10^+,'' between ''33 OE1 Glu-2^-'' and ''99 2HZ Lys-7^+,'' and between ''34 OE2 Glu-2^-'' and ''156 HE1 Arg-10^+.'' For these reasons, it is not likely for water molecules to form hydrogen bonding with the charged groups of Glu-2^-, Lys-7^+, and Arg-10^+. As a result, the contributions from these residues to $\Delta\mu_S$ are positive and relatively large. Conformation 1 has the highest $\Delta\mu_S$ among the four conformations.

TABLE II. Solvation free energy $\Delta\mu_s$ (kcal/mol), conformational energy E_c (kcal/mol), and total energy E_T $(=\Delta\mu_s+E_c)$ (kcal/mol). Comparison among the four different conformations of the C-peptide. $\Delta\mu_{s0}$ (kcal/mol)=solvation free energy in the case where all the site charges of the C-peptide are set to zero, contribution from the hydrophobic hydration. $\Delta\mu_{sE}=\Delta\mu_s-\Delta\mu_{s0}$=contribution from electrostatic interaction with water. E_{cE} (kcal/mol)=intramolecular Coulombic energy. $E_{c0}=E_c-E_{cE}$ =intramolecular van der Waals energy plus torsion energy.

	Conf. 1	Conf. 2	Conf. 3	Conf. 4
$\Delta\mu_s$	457	218	−2	67
E_c	48	200	466	286
E_T	505	418	464	353
$\Delta\mu_{s0}$	709	541	558	541
E_{c0}	−97	−48	5	−37
$\Delta\mu_{sE}$	−252	−323	−560	−474
E_{cE}	145	248	461	323

In our earlier work[13] for Met-enkephalin, we found that water molecules form strong, electrostatically stabilized hydrogen bonding with carbonyl oxygens (atoms with relatively large, negative site charges) and the solvation free energy decreases greatly when multiple carbonyl oxygens get close together. A similar tendency is conspicuous for conformation 3. In this conformation, the charged groups in the side chains of Lys-1^+, Lys-7^+, Arg-10^+, and His-12^+ are relatively close together. Hence, the contributions from these four residues to $\Delta\mu_S$ are negative and very large in magnitude. E_C, on the other hand, is very high because of the strong, repulsive Coulomb interaction among these positively charged groups.

In conformation 4, all of the charged groups in the side chains of Lys-1^+, Glu-2^-, Lys-7^+, Arg-10^+, and His-12^+ are generally well exposed to water. The charged groups of Lys-1^+ and Lys-7^+ are relatively close to each other. These features are reflected on the significantly large, negative contributions from the five residues to $\Delta\mu_S$. The value of $\Delta\mu_S$ is very low, though it is higher than in conformation 3. Conformation 4 has the lowest E_T among the four conformations.

TABLE III. Contributions from the 13 residues of the C-peptide to the solvation free energy (kcal/mol). Comparison among the four different conformations.

	Conf. 1	Conf. 2	Conf. 3	Conf. 4
Lys-1^+	−24.6	11.3	−76.2	−26.5
Glu-2^-	69.0	−30.0	−12.7	−52.6
Thr-3	37.5	28.4	28.8	28.8
Ala-4	33.3	19.6	19.2	20.0
Ala-5	35.3	17.9	18.3	17.7
Ala-6	38.5	20.4	21.0	20.8
Lys-7^+	5.0	−18.3	−64.5	−29.0
Phe-8	68.1	50.0	43.2	44.4
Leu-9	63.1	48.7	51.6	43.5
Arg-10^+	21.1	1.0	−60.2	−36.0
Gln-11	53.2	36.0	37.5	33.1
His-12^+	1.0	−18.3	−59.2	−46.5
Met-13	55.9	50.8	51.4	49.4
Total	457	218	−2	67

TABLE IV. Solvation free energies $\Delta\mu_s$ (kcal/mol) for simple charged hard spheres. The temperature is 273 K. The charges of sphere 1 and sphere 2 are 0.5 and −0.5 (these are scaled by the magnitude of the electronic charge), respectively. The diameter of spheres 1 and 2 is 0.28 nm, and the distance between the centers of the pair of spheres is 0.40 nm.

Solute	$\Delta\mu_s$
Isolated sphere 1	−8.6
Isolated sphere 2	−22.3
Pair of sphere 1-sphere 1	−36.2
Pair of sphere 1-sphere 2	−11.9
Pair of sphere 2-sphere 2	−63.2

The solvation free energy becomes considerably lower when the like-charged groups get closer together. This feature is not found in conformation 2, with the result that it has a much higher solvation free energy than conformations 3 and 4. We note that the values of $\Delta\mu_{S0}$ for conformations 2, 3, and 4 are almost the same. Thus, the difference in $\Delta\mu_S$ comes from that in $\Delta\mu_{SE}$.

To verify the idea that the solvation free energy decreases greatly when like-charged atoms or groups get close together, we consider a model calculation summarized in Table IV. The solvation free energies are calculated for a pair of like-charged and unlike-charged spheres and compared with the values for isolated spheres. It is observed in the table that a great decrease in the solvation free energy occurs when like-charged spheres get close together (e.g., $-36.2<-8.6\times2=-17.2$). On the contrary, when unlike-charged spheres are put close together, the solvation free energy increases significantly (i.e., $-11.9>-8.6-22.3=-30.9$).

B. Structure of water near a peptide

The site–site pair distribution functions $g_{AB}(r)$, which are dependent on the peptide conformation, are shown in Figs. 3–6. We are interested in $g_{AB}(r)$ for the charged groups in the side chains and water molecules. Atoms A in the peptide are ''33 OE1 Glu-2^-,'' ''99 2HZ Lys-7^+,'' ''162 1HH2 Arg-10^+,'' and ''196 HE2 His-12^+'' in Figs. 3, 4, 5, and 6, respectively, and B is the water-hydrogen in Fig.

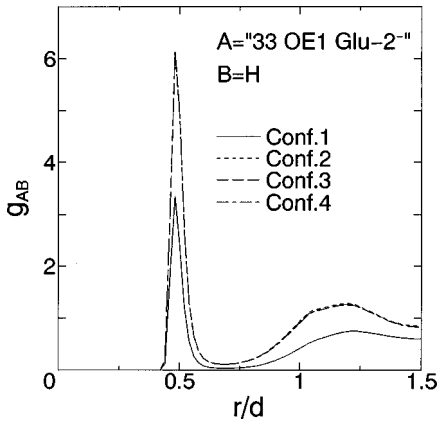

FIG. 3. Site–site pair distribution function $g_{AB}(r)$ for $A=$''33 OE1 Glu-2^-'' in the four conformations ($d=0.28$ nm). B is a water-hydrogen.

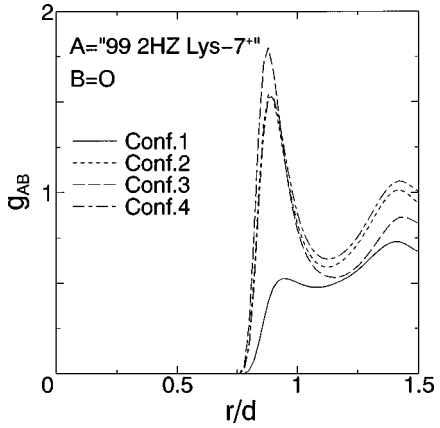

FIG. 4. Site–site pair distribution function $g_{AB}(r)$ for $A=$''99 2HZ Lys-7^+'' in the four conformations ($d=0.28$ nm). B is a water-oxygen.

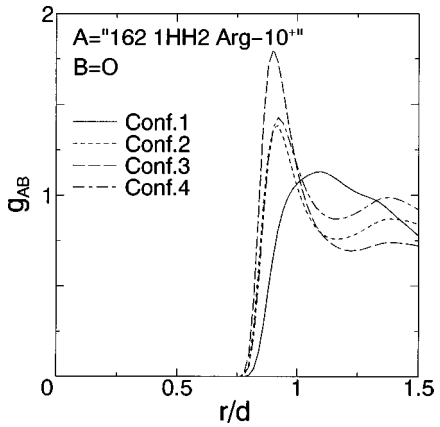

FIG. 5. Site–site pair distribution function $g_{AB}(r)$ for $A=$''162 1HH2 Arg-10^+'' in the four conformations ($d=0.28$ nm). B is a water-oxygen.

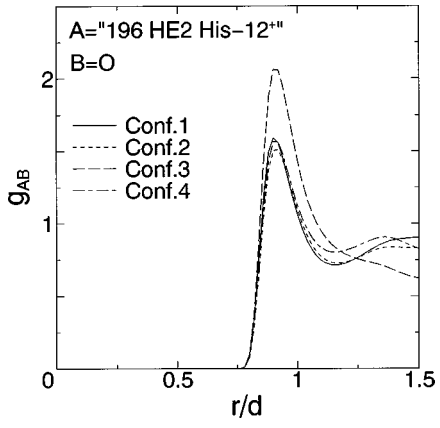

FIG. 6. Site–site pair distribution function $g_{AB}(r)$ for $A=$''196 HE2 His-12^+'' in the four conformations ($d=0.28$ nm). B is a water-oxygen.

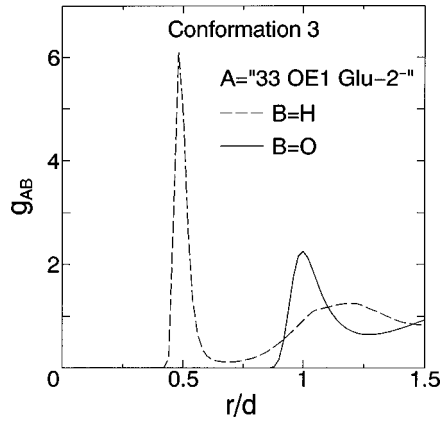

FIG. 7. Site–site pair distribution functions $g_{AB}(r)$ for $A=$''33 OE1 Glu-2^-'' in conformation 3 ($d=0.28$ nm).

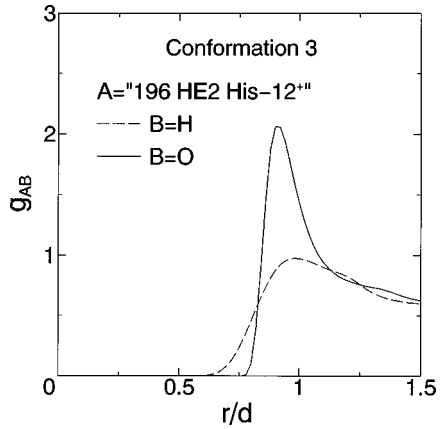

FIG. 8. Site–site pair distribution functions $g_{AB}(r)$ for $A=$''196 HE2 His-12^+'' in conformation 3 ($d=0.28$ nm).

3 and the water-oxygen in the other three figures. Figures 7 and 8 show $g_{AB}(r)$ ($A=$''33 OE1 Glu-2^-'' in Fig. 7 and $A=$''196 HE2 His-12^+'' in Fig. 8, and B is either the water-hydrogen or the water-oxygen) for the peptide in conformation 3.

In Fig. 3 the three curves for conformations 2–4 are almost indistinguishable, while only the curve for conformation 1 is clearly different. In conformation 1, water-hydrogens are not likely to form bonding with ''33 OE1 Glu-2^-'' due to the intramolecular hydrogen bonding between this atom and ''162 1HH2 Arg-10^+'' or ''99 2HZ Lys-7^+.'' The first-peak value of the curve for conformation 1 is much lower. This trend is also observed in Figs. 4 and 5. Due to the intramolecular hydrogen bonding in conformation 1, water-oxygens do not form strong bonding with ''162 1HH2 Arg-10^+'' or ''99 HZ2 Lys-7^+.'' (In conformation 1, ''33 OE1 Glu-2^-,'' ''99 2HZ Lys-7^+,'' and ''162 1HH2 Arg-10^+'' are not well exposed to water.)

In conformation 3, the charged groups in the side chains of Lys-1^+, Lys-7^+, Arg-10^+, and His-12^+ are in proximity to one another. When multiple like-charged atoms are close together, water molecules form strong bonding with these atoms due to the high local charge density, as illustrated in Figs. 4–6. The first-peak values of $g_{AB}(r)$ in conformation 3 are significantly higher than in the others. This feature leads to the large, negative contributions from these residues to $\Delta\mu_S$ and the very low total solvation free energy.

The formation of hydrogen bonding with water molecules is illustrated in Figs. 7 and 8 where conformation 3 is chosen as an example. The sharp first peaks of the two curves in Fig. 7 represent the hydrogen-bonding ''33-O···H—O.'' In Fig. 7, the hydrogen bonding formed is ''196-H···O'': two water-hydrogens can take various positions with the result that the sharp first peak is found only for $g_{AO}(r)$. The first peak of $g_{AH}(r)$ in Fig. 7 is sharper than that of $g_{AO}(r)$ in Fig. 8, since the size of the water-hydrogen is considerably smaller than a hydrogen atom in the peptide.

C. Comparison between Met-enkephalin and C-peptide

In our earlier work,[13] we examined a number of probable, different conformations of Met-enkephalin and calculated the solvation free energy and the conformational energy. Most of those conformations were obtained during the MC simulated annealing combined with the RISM theory. This examination has also been performed for the C-peptide. These energies are dependent on the peptide conformation and thus are variable. The degree of the variation in these energies, which is observed for the C-peptide, is compared

TABLE V. Comparison between Met-enkephalin and the C-peptide in terms of the degree of variation in the solvation free energy $\Delta\mu_s$ (kcal/mol) and the conformational energy E_c (kcal/mol). $\Delta\mu_{sO}$ (kcal/mol)=solvation free energy in the case where all the site charges of the peptide are set to zero, contribution from the hydrophobic hydration. $\Delta\mu_{sE}=\Delta\mu_s-\Delta\mu_{sO}$=contribution from electrostatic interaction with water. E_{cE} (kcal/mol)=intramolecular Coulombic energy. $E_{cO}=E_c-E_{cE}$=intramolecular van der Waals energy plus torsion energy. ''$X\rightarrow Y$'' represents that the value is in the range from X to Y, and the value in parentheses denotes the difference ''$Y-X$.'' The total number of atoms in the C-peptide divided by that in Met-enkephalin is 221/75~3. ''Met-enkephalin×3'' denotes the value for Met-enkephalin multiplied by 3.

	Met-enkephalin	Met-enkephalin×3	C-peptide
$\Delta\mu_s$	150→205 (55)	450→615 (165)	−10→460 (470)
E_c	5→45 (35)	15→135 (120)	20→470 (450)
$\Delta\mu_{sO}$	180→230 (50)	540→690 (150)	535→710 (175)
E_{cO}	−35→−5 (30)	−105→−30 (90)	−95→5 (100)
$\Delta\mu_{sE}$	−30→−20(10)	−90→−60 (30)	−560→−250(310)
E_{cE}	45→60 (15)	135→180 (45)	90→460 (370)

TABLE VI. Four runs of the combined program. M=total number of MC sweeps. T_I(K)=initial temperature. T_F(K)=final temperature. The total number of conformations sampled is 64 M. θ (rad) denotes the maximum value allowed for a change in a torsion angle for the backbone (that for the side chains is fixed at $\pm\pi/2$). In run 4, the conformation with the minimum value of E_T obtained in run 2 (i.e., conformation 4) is used as the initial one.

Run	Initial conformation	M	T_I	T_F	θ
1	Conformation 2	200	673	273	$\pm\pi$
2	Conformation 2	200	473	273	$\pm\pi/2$
3	Conformation 2	400	473	273	$\pm 3\pi/4$
4	Conformation 4	200	473	273	$\pm\pi/2$

with that for Met-enkephalin in Table V. Since the total number of atoms of the C-peptide is about three times larger than that of Met-enkephalin, the values for Met-enkephalin are multiplied by three and compared with those for the C-peptide. The temperatures chosen for Met-enkephalin and for the C-peptide are 298 and 273 K, respectively, but this temperature difference does not alter our conclusions.

The degree of the variation in $\Delta\mu_{S0}$ and E_{C0} (nonelectrostatic terms) among different conformations is about the same for the two peptides. In Met-enkephalin, $\Delta\mu_{S0}$ and E_{C0} are more variable than $\Delta\mu_{SE}$ and E_{CE}. In contrast, in the C-peptide, $\Delta\mu_{SE}$ and E_{CE} vary to a much larger extent than $\Delta\mu_{S0}$ and E_{C0}. The degree of the variation in $\Delta\mu_{SE}$ and E_{CE} (electrostatic terms) for the C-peptide is far larger than that for Met-enkephalin. The conformational stability of the C-peptide is governed by the electrostatic terms. As observed in Table II, lower $\Delta\mu_{SE}$ usually leads to higher E_{CE}, and the conformational stability of the C-peptide in water is determined by a marginal balance of these two opposing terms. Unlike for Met-enkephalin, the nonelectrostatic terms are less important in stabilizing the conformations for the C-peptide.

In the C-peptide, the charged groups in the side chains play essential roles in determining the conformation. Compared with atoms having relatively high site charges in the backbone (e.g., carbonyl oxygens), the charged atoms in the side chains can be arranged in much more variable ways: they can get fairly close together or far apart, resulting in great variation in E_{CE} and $\Delta\mu_{SE}$.

D. Test of combined program

Our combined program has been applied to the C-peptide. As summarized in Table VI, four different runs are tested by changing the initial and final temperatures and the range of a trial change in a torsion angle for the backbone. The total number of MC sweeps is 400 (i.e., 25 600 conformations are sampled and the RISM equations are fully solved 25 600 times) in run 3 and 200 in the other runs. The initial conformation in runs 1–3 is conformation 2, a fully extended conformation. Since fully extended conformations are most stable for Met-enkephalin,[15] we have chosen one of such conformations as the initial one and examined the conformational change. The conformation with the lowest value of E_T (conformation 4), which is obtained in run 2, is used as the initial one in run 4.

TABLE VII. Solvation free energy $\Delta\mu_s$ (kcal/mol), conformational energy E_c (kcal/mol), and total energy E_T ($=\Delta\mu_s+E_c$) (kcal/mol) of the conformation with the minimum value of E_T chosen in each run of the combined program. $\Delta\mu_{sO}$ (kcal/mol)=solvation free energy in the case where all the site charges of the C-peptide are set to zero, contribution from the hydrophobic hydration. $\Delta\mu_{sE}=\Delta\mu_s-\Delta\mu_{sO}$=contribution from electrostatic interaction with water. E_{cE} (kcal/mol)=intramolecular Coulombic energy. $E_{cO}=E_c-E_{cE}$=intramolecular van der Waals energy plus torsion energy. In run 4, the conformation with the minimum value of E_T is the initial conformation (i.e., conformation 4). Conformation 4 gives the lowest value of the total energy.

Run	$\Delta\mu_s$	E_c	E_T	$\Delta\mu_{sO}$	E_{cO}	$\Delta\mu_{sE}$	E_{cE}
1	93	267	360	541	−27	−448	294
2	67	286	353	541	−37	−474	323
3	86	272	359	539	−31	−453	303
4	67	286	353	541	−37	−474	323

The conformation first changes in the direction that the charged groups in the side chains are more exposed to water, and in particular, the positively charged groups are closer together. The solvation free energy thus decreases greatly in the initial stage. This leads to significant increase in the intramolecular Coulombic repulsion energy, but the decrease in the solvation free energy dominates. In the later stage, however, further decrease in the solvation free energy gives rise to an even larger increase in the intramolecular Coulombic repulsion energy, and the conformational change is greatly decelerated. The values of $\Delta\mu_s$, E_c, and E_T of the conformation obtained as the one with the minimum value of E_T in each run are given in Table VII. In run 4, a conformation whose E_T is lower than the initial conformation (conformation 4) is not found. There is no guarantee that conformation 4 is the true lowest-energy conformation, but it is surely one of the most stable conformations. Conformations like conformation 3 have much lower solvation free energies but are less stable due to the high Coulombic repulsion energy among the charged groups.

The conformations with the minimum value of E_T obtained in the four runs are similar to one another, as shown in Fig. 9: the charged groups of Lys-1^+ and Lys-7^+ are relatively close together, and the charged group of Glu-2^- is very well exposed to water. The values of E_T are also well converged. For comparison, we have performed the MC simulated annealing for the C-peptide in the gas phase. We have tested runs 2 and 3. The two conformations obtained as those with the lowest value of $E_T(=E_c)$ are rather different as shown in Figs. 10(a) and 10(b). A different set of M and θ leads to qualitatively different conformational change. It was already found that in the gas phase, about 10^5 MC sweeps are required to obtain sufficiently stable conformations of the C-peptide [one of such conformations is shown in Fig. 10(c)]. This result is suggestive that the conformation of the C-peptide in water is stabilized far more rapidly than in the gas phase, which is in qualitative accord with the result for Met-enkephalin.

The backbone structure of conformation 4 is less extended than that of conformation 2, as shown in Figs. 2(b) and 2(d). Conformation 4 is more stabilized than fully extended conformations like conformation 2 because of the re-

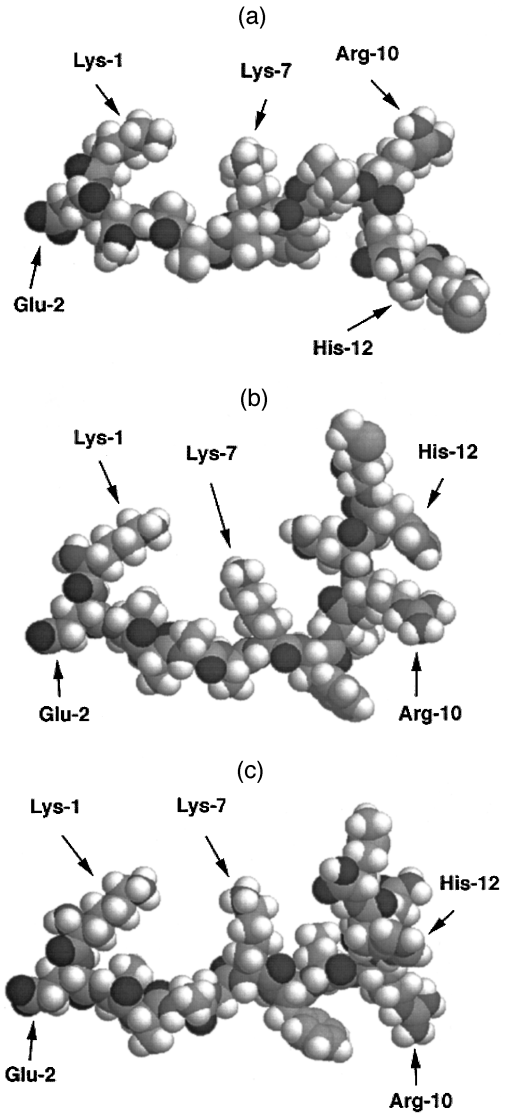

FIG. 9. Conformation in water obtained as the one with the minimum value of E_T in each run of the combined program. The conformation shown in (b) is conformation 4. The charged groups in the side chains are indicated by arrows. The total energies of the conformations in (a), (b), and (c) are 360, 353, and 359 kcal/mol, respectively. This figure was prepared by RASMOL. (a) Run 1. (b) Runs 2 and 4. (c) Run 3.

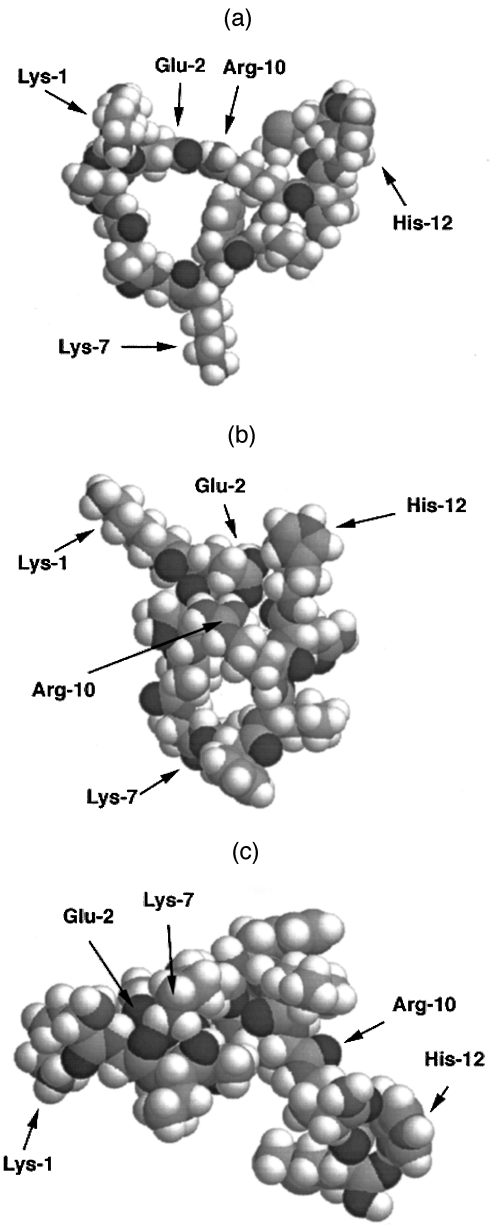

FIG. 10. Conformation in the gas phase obtained as the one with the minimum value of $E_T(=E_c)$. The charged groups in the side chains are indicated by arrows. The total energies of the conformations in (a), (b), and (c) are 99, 125, and 24 kcal/mol, respectively. This figure was prepared by RASMOL. (a) Run 2. (b) Run 3. (c) The lowest-energy conformation obtained by 10^5 MC sweeps ($T_I=973$, $T_F=273$ K, and the dielectric constant is set to unity).

arrangement of charged groups of the side chains. As observed from Tables II and VII, the nonelectrostatic terms for conformation 2 are not significantly different from those for the more stabilized conformations. Thus, the stabilization occurs due to the change in the electrostatic terms. In contrast, for Met-enkephalin,[13] the side-chain structure of stable conformations fluctuates to a large extent, which implies that the conformational stability is not largely influenced by the side-chain structure.

E. Comparison with experimental results

According to the x-ray diffraction data for the native enzyme,[13] the segment from Ala-4 to Gln-11 has an α-helix structure as in conformation 1. Conformation 1 has intramolecular hydrogen bonding between ''33 OE1 Glu-2^-'' and ''162 1HH2 Arg-10^+,'' between ''33 OE1 Glu-2^-'' and ''99 2HZ Lys-7^+,'' and between ''34 OE2 Glu-2'' and ''156 HE1 Arg-10^+.'' Our result shows that in pure water this bonding is broken and the α-helix structure is destabilized. We note that the segment in the native enzyme is in a rather hydrophobic environment (i.e., not well exposed to water). If any segment of a protein or a peptide is isolated from the others, it can take a very different structure.

It is experimentally known that for the isolated C-peptide the α-helix partially ($\sim$30%) remains in an aqueous solution.[24–26] When the ninth residue Glu-9^- is replaced by Leu, the α-helix is more stabilized ($\sim$50%). We note, however, that the formation of the α-helix can be observed

only under a very narrow range of specific solution conditions: 0.1 M-NaCl solution buffered to pH~5.0 near 0 °C. When the temperature is raised or the pH value is changed, the α-helix is destabilized (at 45 °C, for example, the α-helix is melted out). Further, the salt NaCl might play important roles in stabilizing the α-helix. The solvent environment in which the α-helix can partially remain should be quite different from pure water.

In our previous work,[12] we concluded that addition of salts to Met-enkephalin in solution does not significantly influence the peptide conformation. For the C-peptide, however, this is probably not true. Since the charged groups in the side chains, which govern the conformational stability of the C-peptide, will be screened by counterions, the addition of salts could change the peptide conformation to a great extent. Also, the behavior of the zwitterionic form of the C-peptide could be different from that of the unionized form, because the interaction between the *N*- or *C*-terminus and a charged group comes into play (we note that for Met-enkephalin the conformational stability is not significantly influenced by the ionization[12]).

IV. CONCLUSION

Solvation structure and conformational stability of the C-peptide fragment of RNase A with 13 residues, have been analyzed using the dielectrically consistent RISM theory.[9,10] Four representative peptide conformations in the SPC/E water have been considered. The conformation, water structure near a peptide molecule, and details of the solvation properties are quite consistent with one another, which demonstrates the effectiveness of the RISM theory for a peptide with a few hundred atoms in water.

When the like-charged groups in the side chains of Lys-1^+, Lys-7^+, Arg-10^+, and His-12^+ get close together, the solvation free energy decreases to a great extent. This is because water molecules form rather strong electrostatic bonding with these groups due to the increased local charge density. Such a conformation, however, gives rise to higher intramolecular Coulombic repulsion energy. The conformational stability of the C-peptide is determined by competition of the two opposing factors, the contribution from the electrostatic interaction with water to the solvation free energy and the intramolecular Coulombic energy. This result is in contrast with that for Met-enkephalin,[13,15] whose side chains have no charged groups. We note that for Met-enkephalin the hydrophobic hydration and the nonelectrostatic terms in the conformational energy are more important.

The MC simulated annealing combined with the RISM theory has been applied to the C-peptide. From an almost fully extended conformation as the initial one, the conformation first changes in the direction that the charged groups in the side chains are more exposed to water, and in particular, the positively charged groups are closer together: the solvation free energy decreases greatly in the initial stage. This leads to significant increase in the intramolecular Coulombic repulsion energy, but the decrease in the solvation free energy dominates. In the later stage, however, a further decrease in the solvation free energy gives rise to an even larger increase in the intramolecular Coulombic repulsion energy, and the conformational change is greatly decelerated. The conformations thus stabilized in four different runs of the combined program look quite similar, and the total energies are well converged. The conformation of the C-peptide in water is stabilized far more rapidly than in the gas phase, which is in qualitative accord with the result for Met-enkephalin, despite the differences in details of the stabilization process between these two peptides.

The thermodynamics of conformational equilibria for the alanine dipeptide in water was previously reported in several papers:[27–30] In general, water appears to ''flatten'' the free energy surface, decreasing the free energy difference between conformations that differed by large energies on the vacuum surface and lowering the barriers separating those conformations. This feature is likely to be relevant to the marked difference in the speed of conformational stabilization between the peptide behaviors in the gas phase and water, which we have observed for Met-enkephalin and the C-peptide.

Although the stable conformations of the C-peptide in pure water are less extended than those of Met-enkephalin, they have no secondary structure such as the α-helix. For the C-peptide, addition of salts could change the peptide conformation to a great extent, because the charged groups in the side chains will be screened by counterions. The stable conformations in pure water, which are studied in the present article, should be quite different from those in salt solution or in more sophisticated solution used in experiments.[20–22] Effects due to addition of salts must be examined in future studies. We will also try to extend our calculations to larger proteins.

ACKNOWLEDGMENTS

This study was supported by a grant from Research for the Future Program (Project No. JSPS-RFTF98P01101) of the Japan Society for the Promotion of Science, and by the Joint Studies Program (1997 and 1998) of the IMS, ''First-Principle Prediction of Tertiary Structure of Protein in Aqueous Solution.''

[1] S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).
[2] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991); Phys. Rev. Lett. **68**, 9 (1992).
[3] U. H. E. Hansmann and Y. Okamoto, Phys. Rev. E **56**, 2228 (1997).
[4] S. R. Wilson, W. Cui, J. W. Moskowitz, and K. E. Schmidt, Tetrahedron Lett. **29**, 4373 (1988).
[5] H. Kawai, T. Kikuchi, and Y. Okamoto, Protein Eng. **3**, 85 (1989).
[6] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **14**, 1333 (1993).
[7] M. H. Hao and H. A. Scheraga, J. Phys. Chem. **98**, 4940 (1994).
[8] U. H. E. Hansmann and Y. Okamoto, J. Comput. Chem. **18**, 920 (1997).
[9] J. Perkyns and B. M. Pettitt, Chem. Phys. Lett. **190**, 626 (1992).
[10] J. Perkyns and B. M. Pettitt, J. Chem. Phys. **97**, 7656 (1992).
[11] M. Kinoshita, Y. Okamoto, and F. Hirata, J. Comput. Chem. **18**, 1320 (1997).
[12] M. Kinoshita, Y. Okamoto, and F. Hirata, J. Comput. Chem. **19**, 1724 (1998).
[13] M. Kinoshita, Y. Okamoto, and F. Hirata, J. Chem. Phys. **107**, 1586 (1997).
[14] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[15] M. Kinoshita, Y. Okamoto, and F. Hirata, J. Am. Chem. Soc. **120**, 1855 (1998).

[16] W. H. Graham, E. S. Carter II, and R. P. Hickes, Biopolymers **32**, 1755 (1992).
[17] S. J. Singer and D. Chandler, Mol. Phys. **55**, 621 (1985).
[18] H. Kawai, Y. Okamoto, M. Fukugita, T. Nakazawa, and T. Kikuchi, Chem. Lett. **1991**, 213 (1991).
[19] Y. Okamoto, M. Fukugita, T. Nakazawa, and H. Kawai, Protein Eng. **4**, 639 (1991).
[20] K. R. Shoemaker, P. S. Kim, D. N. Brems, S. Marqusee, E. J. York, I. M. Chaiken, J. M. Stewart, and R. L. Baldwin, Proc. Natl. Acad. Sci. USA **82**, 2349 (1985).
[21] K. R. Shoemaker, P. S. Kim, E. J. York, J. M. Stewart, and R. L. Baldwin, Nature (London) **326**, 563 (1987).
[22] J. J. Osterhout, R. L. Baldwin, E. J. York, J. M. Stewart, H. J. Dyson, and P. E. Wright, Biochemistry **28**, 7059 (1989).
[23] F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, J. Phys. Chem. **79**, 2361 (1975).
[24] G. Némethy, M. S. Pottle, and H. A. Scheraga, J. Phys. Chem. **87**, 1883 (1983).
[25] M. J. Sippl, G. Némethy, and H. A. Scheraga, J. Phys. Chem. **88**, 6231 (1984).
[26] H. W. Wychoff, D. Tsernoglou, A. W. Hanson, J. R. Knox, B. Lee, and F. M. Richards, J. Biol. Chem. **245**, 305 (1970).
[27] M. Mezei, P. K. Mehrotra, and D. L. Beveridge, J. Am. Chem. Soc. **107**, 2239 (1985).
[28] B. M. Pettitt and M. Karplus, J. Phys. Chem. **92**, 3994 (1988).
[29] A. Anderson and J. Hermans, Proteins: Struct., Funct., Genet. **3**, 262 (1988).
[30] D. J. Tobias and C. L. Brooks III, J. Phys. Chem. **96**, 3864 (1992).