Bull. Chem. Soc. Jpn., **74**, 1847–1855 (2001)

Theoretical Conformational Analysis of the Tandem Repeat Sequence in RNA Polymerase II

Isao Ohiso, Masahito Oka,[†] **and Norio Nishi**[*]

Graduate School of Environmental Earth Science, Hokkaido University, Kita-ku, Sapporo 060-0810

†Research Institute for Advanced Science and Technology, Osaka Prefecture University, Sakai, Osaka 599-8570

(Received March 28, 2001)

The largest subunit of RNA polymerase II has an interesting C-terminal domain, which consists of a multiple tandem repeat of a heptapeptide with the consensus sequence Ser-Pro-Thr-Ser-Pro-Ser-Tyr. However, the functional role of this sequence is unclear. One might assume that the conformation of the tandem repeat is very important for its function. A theoretical conformational analysis based on molecular mechanics using the energy functions of ECEPP was carried out for a heptapeptide in the repeating unit (Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe) and a periodic polyheptapeptide corresponding to the tandem repeat sequence (Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe). The minimum conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe had two turn structures, an open turn at Pro^2-Thr^3 and a γ-turn at Pro^5. The contiguous residues to the turns are extended or semi-extended conformations. The low-energy conformations of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe were long-periodic helices with a large number of residues per turn. Their helical radiuses are approximately 10–13 Å. They mostly have tyrosyl side-chains at the helical interior. They are composed of alternative backbone conformations, which are extended and turn structures, patterned on the feature of the low-energy conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe. It was shown that the long-periodic helices contain turn structures as possible conformations of the C-terminal domain.

In the process of the so-called central dogma, RNA is synthesized by RNA polymerase on a DNA template strand. Eucaryotes have three different RNA polymerases, i.e., RNA polymerase I, II, and III.[1] Among them, RNA polymerase I and III synthesize the large ribosomal RNAs and various small RNAs, respectively. While, RNA polymerase II transcribes the genes that encode for proteins. The largest subunit of RNA polymerase II has a very interesting C-terminus consisting of multiple tandem repeat sequences of seven amino acid residues having the consensus Ser-Pro-Thr-Ser-Pro-Ser-Tyr. The number of tandem repeats can vary; for example, there are 26 units in yeast (*S. cerevisiae*),[2] units in *D. melanogaster*,[3] and 52 units in hamster.[4] The largest subunit of RNA polymerase II is homologous with the largest subunits of RNA polymerase I and III.[2] However, the tandem repeat sequence is not found in the largest subunits of eucaryotic RNA polymerase I[5] and III,[2] or in the procaryotic RNA polymerase.[6] It was shown by deletion analysis that approximately half of the repeating units of this tandem repeat are necessary for the function of this enzyme.[3,4,7,8] Though the tandem repeat plays an essential role in the initiation step of transcription,[2–4,7–12] its functional role is unclear. It has been suggested that the tandem repeat binds to transcription factors to form a transcription pre-initiation complex.[4,8,12–15] On the other hand, it is interesting that the tandem repeat contains two copies of the SPXX motif. The SPXX motif is found in various DNA-binding proteins, such as histones and gene regulatory proteins, and has been suggested to be a DNA-binding motif.[16] By examining the effect of synthetic peptides containing one and two units of the consensus sequence Tyr-Ser-Pro-Thr-Ser-Pro-Ser-Tyr and Tyr-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-Ser-Pro-Thr-Ser-Pro-Ser-Tyr, respectively,[17] Suzuki suggested that the heptapeptide repeat might bind to DNA through the intercalation of its tyrosyl side-chain.

The heptapeptide sequence consists of proline residues and three kinds of residues having side-chain hydroxy groups. The proline residue is a conformationally restricted residue, because its pyrrolidine ring formed with the side-chain limits the dihedral angle ϕ of its backbone and brings about steric interactions between its δ carbon and the preceding residue. Moreover, because the amide proton is replaced by a CH$_2$ group, the proline residue cannot be a hydrogen-bond donor. The tandem repeat in RNA polymerase II is estimated to have a characteristic conformation different from ordinary conformations, such as an α-helix and β-sheet, because it is not only proline-rich, but also a periodic sequence. Its characteristic conformation may play an important role in the function of the C-terminal domain. A detailed investigation of a polypeptide corresponding to the tandem repeat is important to clarify the functional role of this repeat.

NMR studies on the octapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser-Tyr were reported by Harding.[18] She proposed the presence of turn structures in equilibrium with unfolded forms of the octapeptide in water. In addition, it was also shown that peptide bonds at its Pro^3 and Pro^6 positions had trans conformations as the major species ($> 90\%$) in water. In our previous study,[19] model polypeptides were synthesized: the polyheptapeptide corresponding to the tandem repeat found in the RNA polymerase II, poly(Ser-Pro-Thr-Ser-Pro-Ser-Tyr), and the heptapeptide corresponding to the repeating unit, H-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-OH. Our circular dichroism (CD) study[20] using the model polypeptides suggested that the heptapeptide included a turn structure in water, and the turn structure was further stabilized in 2,2,2-trifluoroethanol (TFE) and in acetonitrile. It was also suggested that the polyheptapeptide formed not only a turn structure at the heptapeptide unit, Ser-Pro-Thr-Ser-Pro-Ser-Tyr, but also a kind of super conformation induced by the periodicity of the sequence in TFE or acetonitrile.

In the present study, a theoretical conformational analysis based on molecular mechanics was tried for the heptapeptide in the repeating unit (Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe) and the periodic polyheptapeptide corresponding to the tandem repeat sequence (Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe).

Theoretical

All conformational energies (E) were calculated using the energy functions of ECEPP.[21] The conformational energy difference per whole molecule (ΔE) is defined by $\Delta E = E - E_0$, where E_0 is the conformational energy of the lowest-energy conformation of every particular molecule. The conformational energy difference per residue (ΔE_{res}) is defined by $\Delta E_{\mathrm{res}} = \Delta E/m$, where m is the number of residues of a molecule.

In this study, the dihedral angles of all peptide bonds including X–Pro bonds were fixed to trans (i.e., $\omega = 180°$), though it is known that X–Pro peptide bonds have a relatively high probability to adopt a cis ($\omega = 0°$) conformation. It was because the trans conformation of the X–Pro bond was more stable than the cis approximately 2–3 kcal mol^{-1} in the conformational energy difference, for instance, 2.48 kcal mol^{-1} (Ala–Pro), 2.48 kcal mol^{-1} (Asn–Pro), 1.95 kcal mol^{-1} (Phe–Pro), 2.97 kcal mol^{-1} (Tyr–Pro), 3.78 kcal mol^{-1} (Val–Pro), 2.75 kcal mol^{-1} (Val–Pro–Gly–Gly), and 3.08 kcal mol^{-1} (poly(Val–Pro–Gly–Gly)).[22,23] Moreover, it was shown that two X–Pro peptide bonds of Tyr-Ser–Pro-Thr-Ser–Pro-Ser-Tyr had trans conformations as the major species (> 90%).[18]

A conformational energy minimization of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe was carried out by the following procedures (Fig. 1). During the minimization, ϕ, ψ, and all side-chain dihedral angles of Ser, Thr, and Tyr were allowed to vary. For Pro, ϕ was fixed to $-75°$ and only ψ was allowed to vary, namely, the pyrrolidine ring of Pro was fixed to be puckered down. It is known that the pyrrolidine ring of Pro has three energy minima, which are $\phi = -75°$ (puckered down), $-67.6°$ (puckered up), and $-57.6°$ (puckered flat), and that the puckered down is the most stable conformation among the minima.[24] The first step was minimization of the tripeptides included in the sequence of the heptapeptide, Ac-Ser-Pro-Thr-NHMe, Ac-Pro-Thr-Ser-NHMe, Ac-Thr-Ser-Pro-NHMe, Ac-Ser-Pro-Ser-NHMe, and Ac-Pro-Ser-Tyr-NHMe. All combinations of the minima of every single residue[25] were used as starting conformations for minimizations of the tripeptides. The minima of Ser, Pro, Thr, and Tyr are 39, 4, 25, and 30 in number, respectively. To obtain the minima of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe, three routes of minimization procedures were carried out using the obtained minima of the tripeptides. In route 1, the starting conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe were built by overlapping conformations in the minima of the tripeptides (Ac-Ser-Pro-Thr-NHMe, Ac-Thr-Ser-Pro-NHMe, and Ac-Pro-Ser-Tyr-NHMe). The "overlapping" here means that conformations of common residue(s) among two peptides are almost same; in detail, it is that all dihedral angles of the common residue(s) are different by less than 10°. All combinations of the conformations overlapping with each N- or C-terminal residue were picked up from the minima of the tripeptides in the region $\Delta E < 4.5$ kcal mol^{-1}; then, the minimization of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe was performed from the starting conformations built by them. In routes 2 and 3, tetrapeptides included in the heptapeptide sequence (Ac-Ser-Pro-Thr-Ser-NHMe, Ac-Pro-Thr-Ser-Pro-NHMe, Ac-Thr-Ser-Pro-Ser-NHMe, and Ac-Ser-Pro-Ser-Tyr-NHMe) were minimized, using all combinations

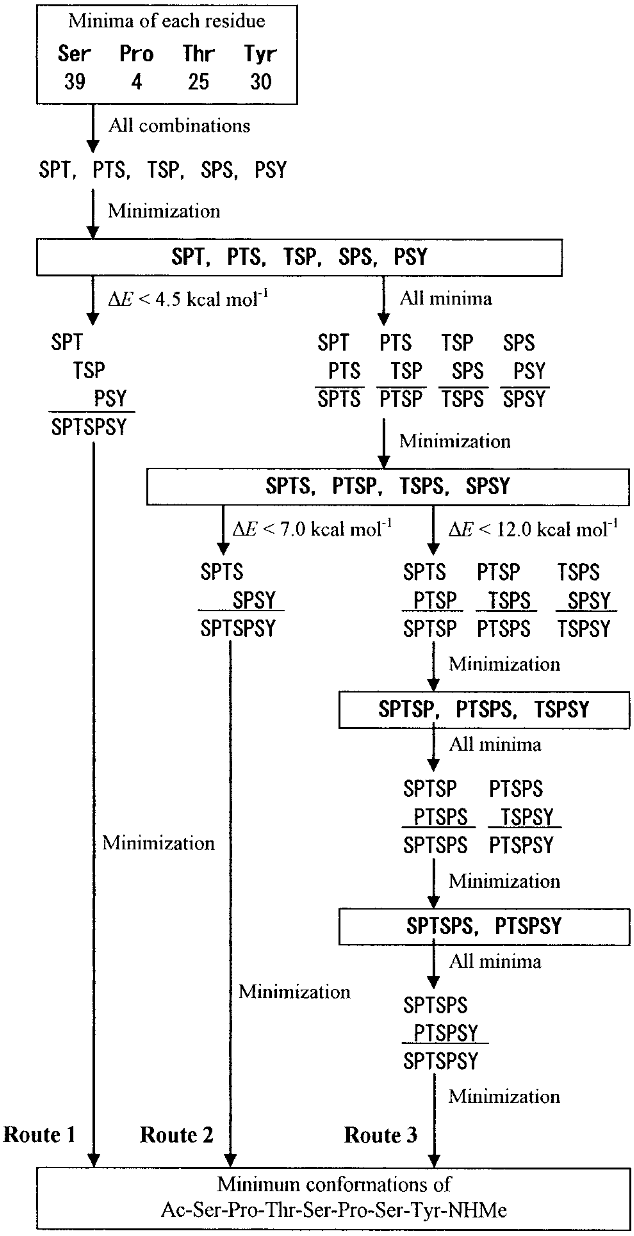

Fig. 1. Conformational energy minimization of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe. Route 1, 2-step minimization; Route 2, 3-step minimization; Route 3, 5-step minimization.

of the overlapping conformations with terminal two residues of all minima of the tripeptides as the starting conformations. In route 2, the starting conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe were built by overlapping of the minima of the tetrapeptides (Ac-Ser-Pro-Thr-Ser-NHMe and Ac-Ser-Pro-Ser-Tyr-NHMe). All combinations of the overlapping conformations with the terminal one residue were picked up from the minima of the tetrapeptides in the region $\Delta E < 7.0$ kcal mol^{-1}; then, the minimization of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe was carried out from the starting conformations built by them. In route 3, pentapeptides included in the heptapeptide sequence (Ac-Ser-Pro-Thr-Ser-Pro-NHMe, Ac-Pro-Thr-Ser-Pro-Ser-NHMe, and Ac-Thr-Ser-Pro-Ser-Tyr-NHMe) were minimized using all combinations of the overlapping conformations with terminal three residues of the minima of the tetrapeptides in the region $\Delta E < 12.0$ kcal mol^{-1} as starting conformations. Then, to obtain the minima of hexapeptides (Ac-Ser-Pro-Thr-Ser-Pro-Ser-NHMe and Ac-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe) the minimization was carried out using all combinations of the overlapping conformations with terminal four residues of all the minima of pentapeptides as starting conformations. The minimization of heptapeptide (Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe) was carried out, using all combinations of the overlapping conformations with terminal five residues of all the minima of hexapeptides as starting conformations. Finally, all of the minima obtained from routes 1, 2, and 3 were sorted according to the conformational energy; the reduplicated minima were excluded.

Minimization of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe as a periodic polypeptide was carried out by a five-step minimization (Fig. 2). During the minimization steps, 24 dihedral angles that were allowed to vary on the minimization of the heptapeptide were adopted as the variables. The first step was the minimization of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_2$-NHMe using all minima of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe as the starting conformations. Then, Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_4$-NHMe, Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_8$-NHMe, Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{12}$-NHMe, and Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe were minimized step-by-step using the minima which were obtained in the previous minimization steps of $\Delta E < 35.0$ kcal mol^{-1}, $\Delta E < 36.7$ kcal mol^{-1}, $\Delta E < 76.0$ kcal mol^{-1}, and all minima, respectively.

The conformational space was divided into 16 regions with the conformational letter codes shown in Fig. 1 of ref. 25. An interatomic distance of less than 2.3 Å between a polar hydrogen atom and an oxygen or a nitrogen atom was regarded to be hydrogen-bonded. In categorizations of helix-type, β^x-helix is defined as a helix that has a spiral structure with x residues per turn,[26] and γ-helix[27] is defined as a helical conformation which cannot be categorized to an α-helix or a β^x-helix. Two helical parameters, n and h, are the number of residues per turn and rise per residue, respectively.

Results

Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe. In the region $\Delta E_{\mathrm{res}} < 1.0$ kcal mol^{-1}, 175 minima of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe were found, and 10 of which are given in Table 1. The lowest-energy conformation is shown in Fig. 3 and its conformational letter code is EACF*CGE. The letter code C is a region containing the C_7 ring, which is a seven-membered ring with a hydrogen bond between CO of the i-1th residue and NH of the $i+1$th residue, i.e., a γ-turn conformation.[28] The lowest-energy conformation has two γ-turns at Thr^3 and Pro^5 with hydrogen bonds, (Pro^2)CO$\cdots$HN(Ser^4) and (Ser^4)CO$\cdots$HN(Ser^6), respectively. The side-chain of Thr^3 has a hydrogen bond to the backbone of, itself, (Thr^3)NH$\cdots$OH-(Thr^3), forming a proline-like five-membered ring. The side-chain hydroxy groups of three serine residues have hydrogen

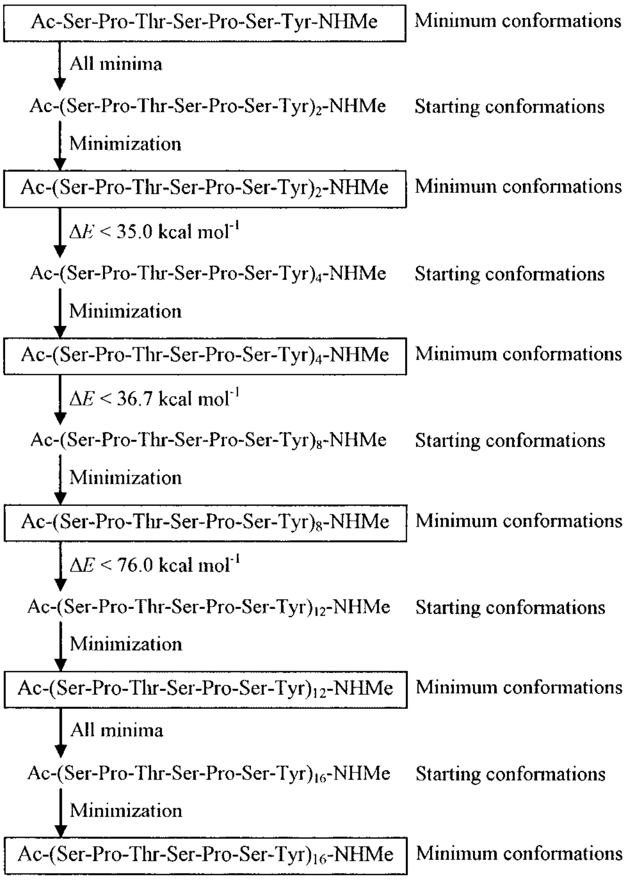

Fig. 2. Conformational energy minimization of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe.

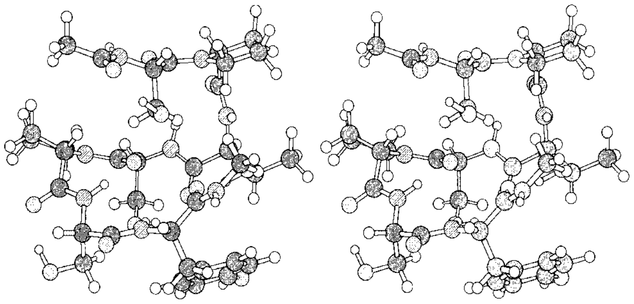

Fig. 3. Stereo view of the lowest-energy conformation of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe. ●, carbon atoms; ◍, oxygen atoms; ▨, nitrogen atoms; ○, hydrogen atoms.

Table 1. Calculated Minimum Energy Conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe[a)]

Conformational letter code	ΔE_{res} b) kcal mol^{-1}	Ser^1	Pro^2	Thr^3	Ser^4	Pro^5	Ser^6	Tyr^7
		ϕ	ϕ	ϕ	ϕ	ϕ	ϕ	ϕ
		ψ	ψ	ψ	ψ	ψ	ψ	ψ
EACF*CGE	0.00	−176	−75	−87	70	−75	−172	−142
		156	−44	64	150	83	−51	152
DFCECD*E	0.03	−171	−75	−86	−140	−75	178	−143
		79	172	94	147	67	−48	138
EAADCGE	0.20	−178	−75	−79	−145	−75	−179	−142
		159	−34	−24	89	76	−45	160
EAADCGE	0.21	−178	−75	−79	−145	−75	−179	−142
		159	−34	−24	89	76	−45	160
EACECD*E	0.32	−175	−75	−71	−155	−75	179	−140
		155	−53	113	151	65	−50	151
EAADCGE	0.32	−177	−75	−72	−140	−75	−179	−142
		158	−39	−28	89	76	−45	160
EAADCGE	0.33	−177	−75	−72	−140	−75	−179	−142
		158	−39	−28	89	76	−45	160
DFCECD*E	0.34	−172	−75	−85	−144	−75	178	−134
		83	167	92	146	65	−51	126
EACDCDE	0.34	−176	−75	−88	−163	−75	−174	−134
		156	−52	67	86	73	105	153
DAAECGE	0.42	−171	−75	−74	−178	−75	−178	−141
		79	−40	−55	157	73	−47	147

a) All 10 minimum-energy conformations with $\Delta E_{\mathrm{res}} < 0.42$ kcal mol^{-1}.

b) $E_0 = -41.87$ kcal mol^{-1}, $\Delta E_{\mathrm{res}} = (E - E_0)/7$.

bonds ((Ser^1)OH···OC(Ac), (Ser^4)OH···OC(Thr^3) and (Ser^6)-OH···OC(Pro^5), respectively), forming seven-membered rings. The 2nd low-energy conformation is DFCECD*E with $\Delta E_{\mathrm{res}} = 0.03$ kcal mol^{-1}, and there is a slight difference in the conformational energy from the lowest. The 3rd and 4th low-energy conformations are EAADCGE with $\Delta E_{\mathrm{res}} = 0.20$ and 0.21 kcal mol^{-1}, respectively, and the conformational difference among them is only the direction of the OH group of the tyrosyl side-chain. The 2nd–10th low-energy conformations have a γ-turn at Pro^5 with a hydrogen bond, (Ser^4)CO···HN(Ser^6). Their conformations at Pro^2–Thr^3 are FC or AC, a γ-turn-like conformation, or AA, a type III β-turn-like (or helical) conformation, without a characteristic hydrogen bond for the turn definition; namely, they are called open turns.

The lowest-energy conformation might be no more than one of the stable conformations for the heptapeptide, because the lowest was not predominantly stable, and many minima with a slight conformational energy difference from the lowest were obtained. The ϕ–ψ plots of each residue of the 175 minima in the region $\Delta E_{\mathrm{res}} < 1.00$ kcal mol^{-1} are shown in Fig. 4. These 175 minimum-energy conformations are not entirely various, but have a common tendency similar to their lower energy classes. There are two kinds of turn structures, the open turn at Pro^2–Thr^3 and the γ-turn at Pro^5, and the contiguous residues with turns are extended or semi-extended conformations (Fig. 5).

Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe. A total of 50 energy minima of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe were found in the region $\Delta E_{\mathrm{res}} < 1.00$ kcal mol^{-1}; 12 of them are given in Table 2. The lowest-energy conformation is a left-handed $\beta^{19.4}$-helix with letter code DCCEACF (Fig. 6). The Pro^{7j+2}–Thr^{7j+3} sequence is a CC conformation with a hydrogen bond, (Ser^{7j+1})CO···HN(Thr^{7j+3}), and its backbone is twisted. This helix contains open turns at Pro^{7j+5}–Ser^{7j+6} (AC conformation) with an interatomic distance of 6.1 Å between C$^{\alpha}$(Ser^{7j+4}) and C$^{\alpha}$(Tyr^{7j+7}), but without a hydrogen bond between CO(Ser^{7j+4}) and NH(Tyr^{7j+7}). The side-chain hydroxy groups of Ser^{7j+1} and Ser^{7j+4} have hydrogen bonds, (Ser^{7j+1})OH···OC(Tyr$^{7(j-1)+7}$) and (Ser^{7j+4})OH···OC(Pro^{7j+5}), respectively, forming seven-membered rings. The side-chain hydroxy hydrogen of Ser^{7j+6} has a hydrogen bond with a carboxyl oxygen of a serine residue, which is 19 residues ahead (approximately one helix turn), (Ser^{7j+6})OH···OC(Ser$^{7(j+3)+4}$). The tyrosyl side-chains are oriented to the inside of the helix

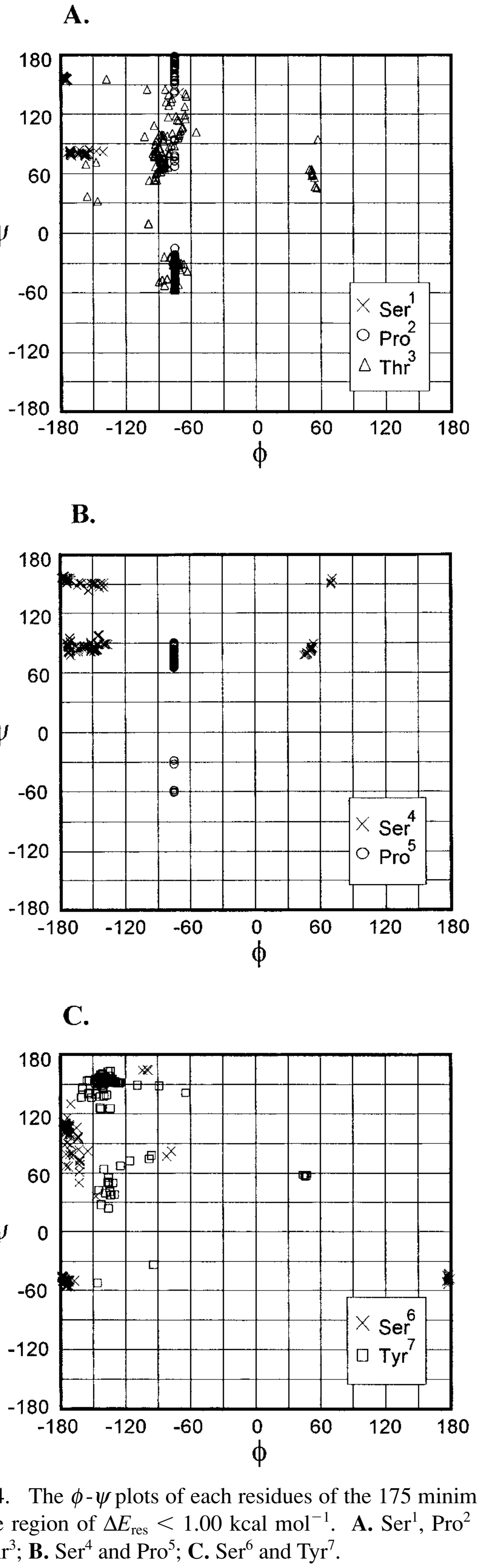

Fig. 4. The ϕ-ψ plots of each residues of the 175 minima in the region of $\Delta E_{\mathrm{res}} < 1.00$ kcal mol^{-1}. **A.** Ser^1, Pro^2 and Thr^3; **B.** Ser^4 and Pro^5; **C.** Ser^6 and Tyr^7.

forming a ranging of hydrogen bonds, (Tyr$^{7(j+1)+7}$)OH$\cdots$OH(Tyr^{7j+7}).

The 2nd low-energy conformation is a left-handed $\beta^{17.3}$-helix (DFA*DACG) with $\Delta E_{\mathrm{res}} = 0.13$ kcal mol^{-1} (Fig. 7). It has γ-turns at Ser^{7j+6} with hydrogen bonds, (Pro^{7j+5})CO$\cdots$HN(Tyr^{7j+7}). The side-chain hydroxy groups of Ser^{7j+1} and Ser^{7j+4} have hydrogen bonds, (Ser^{7j+1})OH$\cdots$OC(Tyr$^{7(j-1)+7}$) and (Ser^{7j+4})OH$\cdots$OC(Thr^{7j+3}), respectively, forming seven-membered rings. The carboxyl oxygen of Tyr^{7j+7} has a hydrogen bond with an amino nitrogen of a threonine residue, which is 17 residues ahead (approximately one helix turn), (Tyr^{7j+7})CO$\cdots$HN(Thr$^{7(j+3)+3}$). The side-chain hydroxy hydrogen of Tyr^{7j+7} has a hydrogen bond with a carboxyl oxygen of a serine residue, which is 8 residues ahead, (Tyr^{7j+7})OH$\cdots$OC(Ser$^{7(j+1)+1}$).

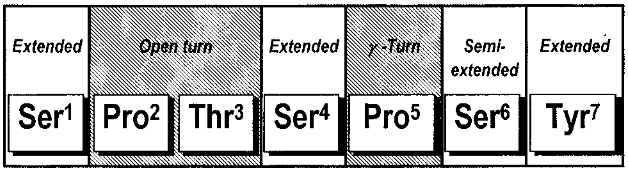

Fig. 5. The feature of low-energy conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe.

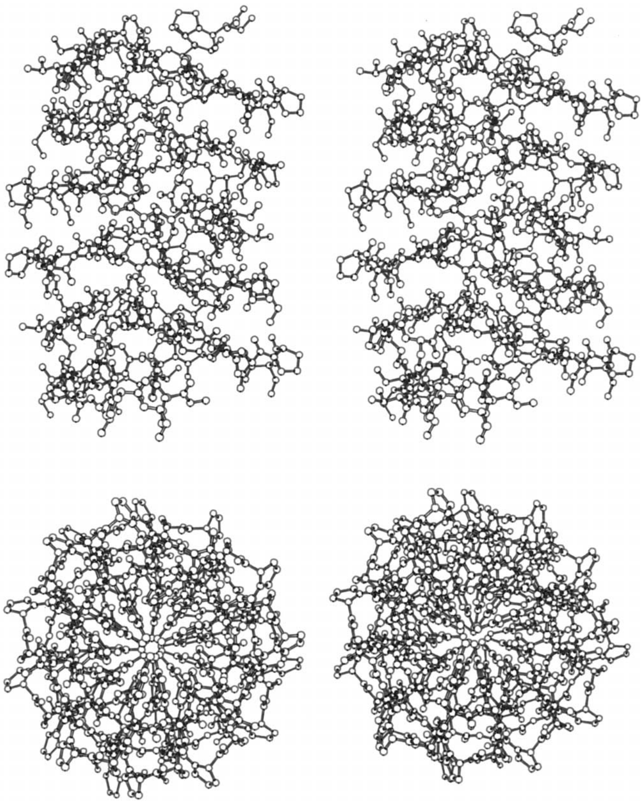

Fig. 6. Stereo views of the left-handed $\beta^{19.4}$-helix of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe with the lowest-energy conformation. Top, view perpendicular to helical axis of downward direction from N- to C-terminus. Bottom, axial view looking down the helix axis from N-terminus. All hydrogen atoms are not shown.

The 3rd low-energy conformation is a left-handed $\beta^{14.4}$-helix (FFA*DCED) with $\Delta E_{\mathrm{res}} = 0.14$ kcal mol^{-1} (Fig. 8). It has γ-turns at Pro^{7j+5} with hydrogen bonds, (Ser^{7j+4})CO$\cdots$HN(Ser^{7j+6}). The side-chain hydroxy groups of Ser^{7j+4} and

Table 2. Calculated Minimum Energy Conformations of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe[a)]

Conformational letter code	ΔE_{res}[b)] kcal mol^{-1}	Helix type[c)]	h[d)]	Ser^{7j+1}	Pro^{7j+2}	Thr^{7j+3}	Ser^{7j+4}	Pro^{7j+5}	Ser^{7j+6}	Tyr^{7j+7}
				ϕ ψ	ϕ ψ	ϕ ψ	ϕ ψ	ϕ ψ	ϕ ψ	ϕ ψ
DCCEACF	0.00	$\beta^{19.4}$-helix (L)	0.30	−168 85	−75 74	−83 87	−169 151	−75 −60	−84 89	−110 133
DFA*DACG	0.13	$\beta^{17.3}$-helix (L)	0.50	−175 84	−75 150	53 53	−173 82	−75 −45	−82 75	−140 −55
FFA*DCED	0.14	$\beta^{14.4}$-helix (L)	0.46	−76 150	−75 163	55 55	−177 78	−75 73	−177 124	−149 105
DFA*DACG	0.17	$\beta^{17.3}$-helix (L)	0.50	−175 83	−75 150	54 51	−173 83	−75 −44	−82 77	−140 −55
ECCF*CEE	0.18	$\beta^{15.7}$-helix (L)	0.40	−173 159	−75 76	−103 63	69 158	−75 76	−163 112	−150 162
DCCEACE	0.28	$\beta^{19.1}$-helix (L)	0.32	−172 85	−75 85	−87 68	−172 154	−75 −48	−70 103	−143 123
DFA*DCGE	0.37	γ-helix	0.66	−160 82	−75 165	55 71	−169 110	−75 76	−175 −53	−158 140
EFA*DCH*E	0.40	$\beta^{14.7}$-helix (L)	0.39	−172 160	−75 −175	60 43	−168 80	−75 66	179 126	−161 172
DFCECGA	0.51	$\beta^{22.8}$-helix (R)	0.58	−152 75	−75 157	−56 99	−161 148	−75 71	−177 −59	−89 −22
DCFFCGE	0.58	$\beta^{16.8}$-helix (R)	0.39	−166 72	−75 74	−84 167	−79 153	−75 81	−173 −50	−144 160
DAAE*CGA	0.62	$\beta^{13.4}$-helix (L)	0.75	−159 77	−75 −30	−73 −69	180 153	−75 70	−177 −55	−71 −37
DAAE*CGA	0.62	$\beta^{13.4}$-helix (L)	0.75	−159 77	−75 −30	−73 −69	180 153	−75 70	−177 −54	−71 −37

a) All 12 minimum-energy conformations with $\Delta E_{\mathrm{res}} < 0.62$ kcal mol^{-1}.

b) $E_0 = -768.93$ kcal mol^{-1}, $\Delta E_{\mathrm{res}} = (E - E_0)/112$.

c) Helix sense is abbreviated as R or L for right- or left-handed, respectively.

d) Rise per residue.

Ser^{7j+6} have hydrogen bonds, (Ser^{7j+4})OH···OC(Thr^{7j+3}) and (Ser^{7j+6})OH···OC(Pro^{7j+5}), respectively, forming seven-membered rings. The side-chain hydroxy hydrogen of Ser^{7j+1} has a hydrogen bond with a carboxyl oxygen of a tyrosine residue which is 14 residues ahead (approximately one helix turn), (Ser^{7j+1})OH···OC(Tyr$^{7(j+1)+7}$). The side-chain hydroxy hydrogen of Tyr^{7j+7} has a hydrogen bond with a carboxyl oxygen of a serine residue, which is 6 residues before, (Tyr^{7j+7})OH···OC(Ser^{7j+1}).

The tyrosyl side-chains of 2nd and 3rd low-energy conformations also locate at a core position in the cylindrical conformation. Conformations whose tyrosyl side-chains were exposed toward the outside of the helix were found as the 11th and 12th low-energy conformations (left-handed $\beta^{13.4}$-helix) with $\Delta E_{\mathrm{res}} = 0.62$. The 11th low-energy conformation contains γ-turns at Pro^{7j+5} with hydrogen bonds, (Ser^{7j+4})CO···HN(Ser^{7j+6}). Its tyrosyl side-chains are outward, while its threonyl hydroxy groups are highly bound in a helical interior with hydrogen bonds, (Thr^{7j+3})NH···OH(Thr^{7j+3}), (Ser^{7j+4})-NH···OH(Thr^{7j+3}) and (Thr^{7j+3})OH···OH(Ser$^{7(j+1)+1}$). It has no long-range hydrogen bonds with more than 5 residues apart. The 12th low-energy conformation only differs from the 11th in the direction of the OH groups of the tyrosyl side-chains.

Although the left-handed $\beta^{19.4}$-helix is the lowest-energy conformation of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe, in the results of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_4$-NHMe, it is the 119th low-energy conformation with $\Delta E_{\mathrm{res}} = 0.85$ kcal mol^{-1}. In the results of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_8$-NHMe, it is the 4th low-energy conformation with $\Delta E_{\mathrm{res}} =$

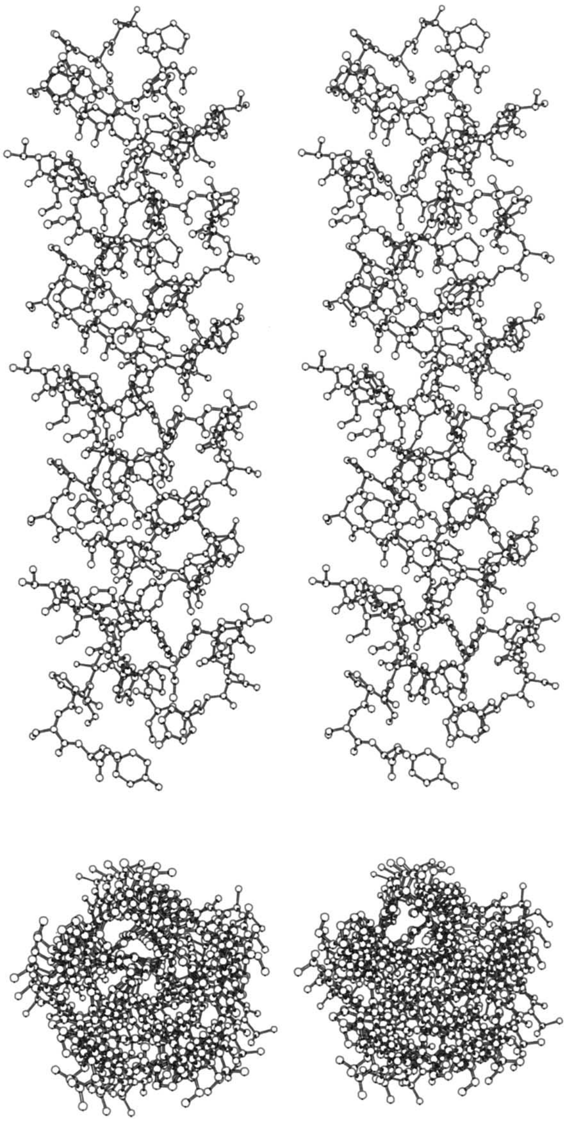

Fig. 7. Stereo views of the left-handed $\beta^{17.3}$-helix of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe with the 2nd low-energy conformation. Top, view perpendicular to helical axis of downward direction from N- to C-terminus. Bottom, axial view looking down the helix axis from N-terminus. All hydrogen atoms are not shown.

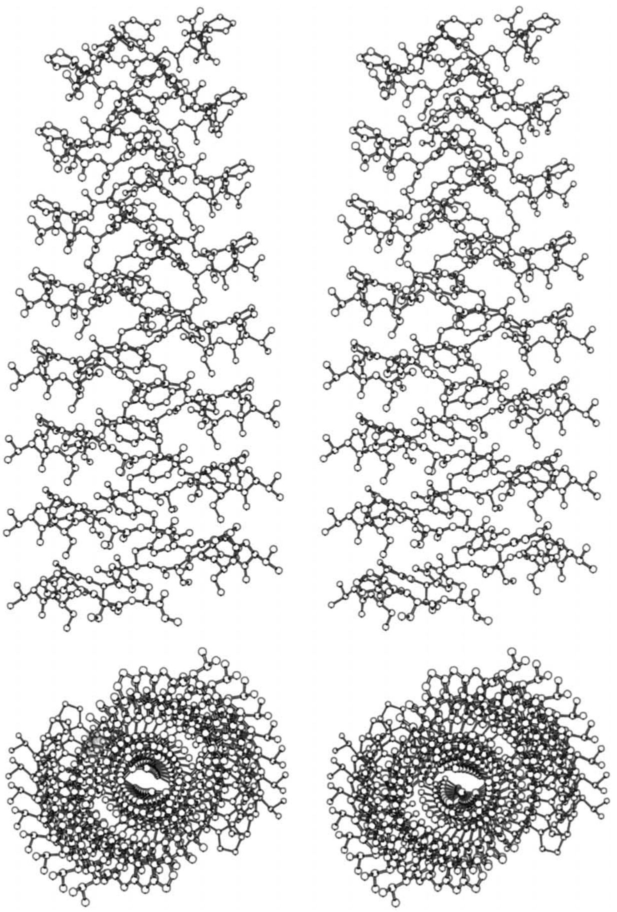

Fig. 8. Stereo views of the left-handed $\beta^{14.4}$-helix of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe with the 3rd low-energy conformation. Top, view perpendicular to helical axis of downward direction from N- to C-terminus. Bottom, axial view looking down the helix axis from N-terminus. All hydrogen atoms are not shown.

0.20 kcal mol^{-1}, and in the results of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{12}$-NHMe and Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe, it is the lowest-energy conformation. Whereas, the lowest-energy conformation is FFA*DCED (left-handed $\beta^{14.1}$-helix) in both results of the minimization of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_4$-NHMe and Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_8$-NHMe. However, in the results of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{12}$-NHMe and Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe, the FFA*DCED conformation is the 2nd and 3rd minimum, left-handed $\beta^{14.2}$ and $\beta^{14.4}$-helix with ΔE_{res} = 0.05 and 0.14 kcal mol^{-1}, respectively. This means that as the number of repeats is increased from 4 to 8, 12, 16, the left-handed $\beta^{19.4}$-helix is relatively stabilized markedly more than the left-handed β^{14}-helix of FFA*DCED. By minimizing Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe based on the results of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{12}$-NHMe, no significant changes in the stability order of the minimum conformations were found. Therefore, if the number of repeats is increased more, the left-handed $\beta^{19.4}$-helix would be the lowest-energy conformation.

Because several minima that were not with very large ΔE_{res} were found around the lowest-energy conformation of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe, only the left-handed $\beta^{19.4}$-helix is not a predominantly stable conformation. The low-energy conformations including the lowest are generally long-periodic helices. The numbers of residues per turn (n) of them are much larger than the α-helix (n = 3.6), and the rises per residue (h) of them are shorter than the α-helix (h = 1.5 Å). Their helical radiuses are approximately 10–13 Å. They mostly have tyrosyl side-chains at the helical interior. They are composed of alternative backbone conformations which are extended and turn structures, patterned on the feature of the low-energy conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe, as illustrated in Fig. 5.

Discussion

In this study, the obtained low-energy conformations of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe contained turn structures. This result corresponded with the NMR study by Harding[18] and our CD study.[20] The NMR study for octapeptide, Tyr-Ser-Pro-Thr-Ser-Pro-Ser-Tyr, suggested the presence of turn structures in equilibrium with unordered forms in water. Our CD results for the heptapeptide, H-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-OH, suggested that the heptapeptide contains a turn structure in water and that the presence of either TFE or acetonitrile favored the turn structure in the heptapeptide. The results of the conformational energy minimization of Ac-Ser-Pro-Thr-Ser-Pro-Ser-Tyr-NHMe support these experimental results. It was suggested that the conformations containing the turn structures observed by the experimental analyses corresponded to the low-energy conformations obtained in this study, as shown as Fig. 5.

Our previous CD results of the polyheptapeptide, poly(Ser-Pro-Thr-Ser-Pro-Ser-Tyr), suggested that the polyheptapeptide includes not only the turn structures caused by the heptapeptide sequence Ser-Pro-Thr-Ser-Pro-Ser-Tyr, but also a kind of super conformation caused by intramolecular hydrogen bond formation within the tandemly repeating sequence in TFE or acetonitrile.[20] The low-energy conformations of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe given in Table 2 were long-periodic helices including repetitive turn structures. There is no evidence that the helices obtained by the minimization correspond with the super conformation suggested by the CD study. We present the long-periodic helices containing turn structures as possible conformations of the polyheptapeptide. A predominantly stable conformation of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe was not found, and there were several minima with small energy differences among the low-energy conformations (Table 2). It was thus suggested that the long-periodic helices could easily be converted to each other or unordered structures, depending upon the environment. Indeed, the CD spectrum of poly(Ser-Pro-Thr-Ser-Pro-Ser-Tyr) in water indicated unordered structures. The CD study of the polyheptapeptide also showed the conformational flexibility between the unordered- and ordered structures under such various conditions favoring hydrogen-bond formation as the various compositions of the TFE/water mixture or the acetonitrile/water mixture. The population of the calculated lowest-energy conformation or other minimum-energy conformations under the unordered- or ordered state will not be known without conducting molecular-dynamics calculations. Because of the proline-rich sequence having multiple pyrrolidine rings, the flexibility of the unordered structures of the polyheptapeptide should be partially restricted. We suggest that the conformational equilibrium includes long periodic helices having repetitive turn structures and partially restricted unordered structures.

It is possible that the tandem repeat sequence having the long-periodic helical conformation may be important for its functional role in RNA polymerase II at the native state. It has been proposed that the tandem repeat is a site interacting with other molecules which participate in a process of transcription initiation, such as transcription factors[4,8,12–15] and/or DNA.[16,17] The CD and NMR analysis reported by our group Imai et al.[29] suggested that the polyheptapeptide poly(Ser-Pro-Thr-Ser-Pro-Ser-Tyr) having a super conformation interacts with DNA in the TFE/water system. If the conformation of the tandem repeat is important for the interaction, and if it is meaningful that the domain is not only proline-rich, but also a periodic tandem repeat sequence, it is possible that the helical conformations proposed here are the essential form of the tandem repeat. It is very interesting that the long-periodic helices of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe have large helical radiuses, and that mostly their tyrosyl side-chains are located at the helical interior, when we estimate the exact functional role of the tandem repeat.

The tandem repeat sequence is known as a phosphorylation site in RNA polymerase II and the phosphorylated form is found when the RNA polymerase becomes transcriptionally active after the formation of a preinitiation complex.[9] It is therefore proposed that the tandem repeat is required for the assembly of the preinitiation complex and its phosphorylation induces a change to the elongation complex disassembling the preinitiation complex.[13] It was shown that the phosphorylation sites were two serine positions at Ser–Pro sequences on the consensus,[30] which were Ser^{7j+1} and Ser^{7j+4} in our model polyheptapeptide. The calculated lowest-energy conformation of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe has hydrogen-bonded hydroxy groups upon Ser^{7j+1} and Ser^{7j+4}, as do other minimum-energy conformations given in Table 2. It was shown that the phosphorylation caused a conformational change in the tandem repeat sequence of RNA polymerase II.[31] For the calculated helical conformations, it is reasonable to assume that a conformational change is induced when the hydrogen bonds are broken by phosphorylation. In addition, considering the role of the tandem repeat in the assembly and disassembly of the preinitiation complex, it is very interesting that the two phosphorylation sites are the first position in the SPXX motif proposed as a DNA-binding motif.[16] On the other hand, phosphorylation of tyrosine residues in the tandem repeat sequence was reported.[32] All 10 minima from the lowest to 10th low-energy conformation obtained in this study have the tyrosyl side-chains located at the helical interior. For these helical conformations, when a tyrosine kinase accesses Tyr to phosphorylate, the conformational change, such as opening the helices or flipping out the tyrosyl side-chain, would be required. Also the Tyr-phosphorylated tandem repeat sequence could not stably adopt the helical conformations.

It was reported that the progressive loss of transcription of the RNA polymerase II was induced by progressive truncation of the tandem repeat sequence.[32] It was shown that approximately half of the repeating units of the tandem repeat are necessary for functioning of the enzyme in yeast,[4,8] *D. melanogaster*,[3] and mouse.[7] According to the lowest energy conformation of Ac-(Ser-Pro-Thr-Ser-Pro-Ser-Tyr)$_{16}$-NHMe obtained in this study, the predictive length of the C-terminal domain is 54.6 Å for 26 repeats (yeast), 92.4 Å for 44 repeats (*D. melanogaster*), and 109.2 Å for 52 repeats (mouse). For the 2nd-low energy conformation, the length is 91.0 Å for 26 repeats, 154.0 Å for 44 repeats, and 182.0 Å for 52 repeats, and for the 3rd-low energy conformation, the length is 83.7 Å for 26 repeats, 141.7 Å for 44 repeats, and 167.4 Å for 52 repeats.

The importance of the length of the C-terminal domain to the function of RNA polymerase II is valid for predicting that the domain has some interactions with transcription factors and/or DNA during the assembly of the preinitiation complex. Furthermore, it is interesting that the left-handed $\beta^{19.4}$-helix, which was the lowest energy conformation of the 12- and 16-times repeated sequences, was markedly stabilized as the number of repeats was increased from 4 to 8, 12 in the minimizations. In yeast, the minimal number of repeats for the cell to be viable was 10–12,[8] which was nearly the same as the minimal number giving the $\beta^{19.4}$-helix as the lowest energy conformation in the minimizations. It is interesting whether the changed relative stability of the helical conformations depending on the number of repeats is related to the defective function to the RNA polymerase II having a truncated C-terminal domain.

References

1 M. K. Lewis and R. R. Burgess, in "The Enzymes," ed by P. D. Boyer, Academic Press, New York (1982), Vol. 15, pp. 109–153.

2 L. A. Allison, M. Moyle, M. Shales, and C. J. Ingles, *Cell*, **42**, 599 (1985).

3 W. A. Zehring, J. M. Lee, J. R. Weeks, R. S. Jokerst, and A. L. Greenleaf, *Proc. Natl. Acad. Sci. U.S.A.*, **85**, 3698 (1988).

4 L. A. Allison, J. K. C. Wong, V. D. Fitzpatrick, M. Moyle, and C. J. Ingles, *Mol. Cell. Biol.*, **8**, 321 (1988).

5 J.-M. Buhler, M. Riva, C. Mann, P. Thuriax, S. Memet, J. Y. Micouin, I. Treich, S. Mariotte, and A. Sentenac, in "RNA polymerase and Regulation of Transcription," ed by W. S. Reznikoff, R. R. Burgess, J. E. Dahlberg, C. A. Gross, M. T. Record, Jr., and M. P. Wickens, Elsevier, New York (1987), pp. 25–36.

6 Y. A. Ovchinnikov, G. S. Monastyrskaya, V. V. Gubanov, S. O. Guryev, I. S. Salomatina, T. M. Shuvaeva, V. M. Lipkin, and E. D. Sverdlov, *Nucleic Acids Res.*, **10**, 4035 (1982).

7 M. S. Bartolomei, N. F. Halden, C. R. Cullen, and J. L. Corden, *Mol. Cell. Biol.*, **8**, 330 (1988).

8 M. Nonet, D. Sweetser, and R. A. Young, *Cell*, **50**, 909 (1987).

9 D. Cadena and M. E. Dahmus, *J. Biol. Chem.*, **262**, 12468 (1987).

10 J. L. Christmann and M. E. Dahmus, *J. Biol. Chem.*, **256**, 11798 (1981).

11 N. E. Thompson, T. H. Steinberg, D. B. Aronson, and R. R. Bourgess, *J. Biol. Chem.*, **264**, 11511 (1989).

12 M. Moyle, J. S. Lee, W. F. Anderson, and C. J. Ingles, *Mol. Cell. Biol.*, **9**, 5750 (1989).

13 P. B. Sigler, *Nature*, **333**, 210 (1988).

14 A. Usheva, E. Maldonado, A. Goldring, H. Lu, C. Houbavi, D. Reinberg, and Y. Aloni, *Cell*, **69**, 871 (1992).

15 A. J. Koleske, S. Buratowski, M. Nonet, and R. A. Young, *Cell*, **69**, 883 (1992).

16 M. Suzuki, *J. Mol. Biol.*, **207**, 61 (1989).

17 M. Suzuki, *Nature*, **344**, 562 (1990).

18 M. M. Harding, *J. Med. Chem.*, **35**, 4658 (1992).

19 N. Nishi, I. Ohiso, N.Sakairi, S. Tokura, M. Tsunemi, and M. Oka, *Biochem. Biophys. Res. Commun.*, **206**, 981 (1995).

20 I. Ohiso, M. Tsunemi, B. Zhao, and N. Nishi, *Bull. Chem. Soc. Jpn.*, **74**, 1139 (2001).

21 F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.*, **79**, 2361 (1975).

22 M. Oka, G. T. Montelione, and H. A. Scheraga, *J. Am. Chem. Soc.*, **106**, 7959 (1984).

23 M. Oka, Y. Baba, A. Kagemoto, and A. Nakajima, *Polym. Bull.*, **25**, 87 (1991).

24 M. Vasquez, G. Némethy, and H. A. Scheraga, *Macromolecules*, **16**, 1043 (1983).

25 S. S. Zimmerman, M. S. Pottle, G. Némethy, and H. A. Scheraga, *Macromolecules*, **10**, 1 (1977).

26 D. W. Urry, *Proc. Natl. Acad. Sci. U.S.A.*, **69**, 1610 (1972).

27 M. Oka, Y. Baba, A. Kagemoto, and A. Nakajima, *Polym. J.*, **22**, 555 (1990).

28 G. Némethy and M. P. Printz, *Macromolecules*, **5**, 755 (1972).

29 K. Imai, I. Ohiso, T. Hiraoki, M. Tsunemi, M. Oka, S. Tokura, and N. Nishi, *Peptide Chem.*, **1996**, 257 (1997).

30 J. Zhang and J. L. Corden, *J. Biol. Chem.*, **266**, 2290 (1991).

31 J. Zhang and J. L. Corden, *J. Biol. Chem.*, **266**, 2297 (1991).

32 S. M. Liao, I. C. Taylor, R. E. Kingston, and R. A. Young, *Genes Dev.*, **5**, 2431 (1991).