Journal of Peptide Science
J. Peptide Sci. **7**: 588–597 (2001)
DOI: 10.1002/psc.356

N-Terminal Domain of HTLV-I Integrase. Complexation and Conformational Studies of the Zinc Finger

FRANÇOISE BERTOLA,[a] CLAUDE MANIGAND,[a] PHILIPPE PICARD,[a]* MICHAEL GOETZ,[b] JEAN-MARIE SCHMITTER[c] and GILLES PRECIGOUX[a]

[a] Unité de Biophysique Structurale, UMR 5471 CNRS, Université Bordeaux I, Talence, France
[b] Institut Européen de Chimie Biologie, FRE 2247 CNRS, ENSCPB, Talence, France
[c] LPTC, UMR 5472 CNRS, Université Bordeaux I, Talence, France

Received 25 June 2001
Accepted 20 August 2001

Abstract: The HTLV-I integrase *N*-terminal domain [50-residue peptide (IN50)], and a 35-residue truncated peptide formed by residues 9–43 (IN35) have been synthesized by solid-phase peptide synthesis. Formation of the 50-residue zinc finger type structure through a HHCC motif has been proved by UV-visible absorption spectroscopy. Its stability was demonstrated by an original method using RP-HPLC. Similar experiments performed on the 35-residue peptide showed that the truncation does not prevent zinc complex formation but rather that it significantly influences its stability. As evidenced by CD spectroscopy, the 50-residue zinc finger is unordered in aqueous solution but adopts a partially helical conformation when trifluoroethanol is added. These results are in agreement with our secondary structure predictions and demonstrate that the HTLV-I integrase *N*-terminal domain is likely to be composed of an helical region (residues 28–42) and a β-strand (residues 20–23), associated with a HHCC zinc-binding motif. Size-exclusion chromatography showed that the structured zinc finger dimerizes through the helical region.

Keywords: HTLV-I; integrase; peptide synthesis; solution structure; zinc finger

INTRODUCTION

Human T-cell leukaemia virus type I or HTLV-I, belonging to the retroviral family, was found to be the causative agent of two severe human diseases: adult T-cell leukaemia (ATL) [1] and HTLV-I associated myelopathy or tropical spastic paraparesis (HAM/TSP) [2,3], with an estimate of 40 million people infected with the virus. Diseases break out a few years (HAM/TSP) or about 20 years (ATL) after the infection and affect less than 5% of HTLV-I infected people. To date, there is no really efficient therapy to fight the HTLV-I infection and to check the development of associated diseases. However, research focuses mainly on the different functional proteins of this retrovirus [4,5] in order subsequently to develop new therapeutic strategies. This paper is devoted to the study of integrase (IN), an essential HTLV-I protein. This protein (297 residues) mediates the insertion of viral DNA into the genome of the human host cell, an essential step in the viral life cycle [6,7]. IN appears thus to be a privileged target for the development of an anti-retroviral therapy.

Abbreviations: Boc, *tert*-butyloxycarbonyl; BPTI, bovine pancreatic trypsin inhibitor; DCC, 1,3-dicyclohexylcarbodiimide; DCM, dichloromethane; DIEA, diisopropylethylamine; DMF, dimethylformamide; DMSO, dimethylsulphoxide; Fmoc, 9-fluorenylmethyloxycarbonyl; HBTU, 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HMP, 4-hydroxymethylphenoxymethyl; HOBt, 1-hydroxy-benzotriazole; HTLV-I, human T-cell leukaemia virus type I; IN, integrase protein; PAM, phenylacetamidomethyl; TFE, 2,2,2-trifluoroethanol; TFA, trifluoroacetic acid.

*Correspondence to: Dr P. Picard, UBS, Bât B8, avenue des Facultés, Université Bordeaux I, 33405 Talence Cedex, France; e-mail: P.Picard@ubs.u-bordeaux.fr
Contract/grant sponsor: Comité Départemental de la Dordogne de la Ligue Nationale Contre le Cancer.

To date, as far the HTLV-I IN protein is concerned, only its expression in *E. coli* and a few *in vitro* activity studies have been published [8]. The HIV-1 IN protein has been the most widely investigated, and from the results, a general model for the integration reaction has been proposed [7,9,10]. Nevertheless, an increasing number of structural characterizations of different integrases is requested for a better knowledge of the real mechanism of retroviral integration. Three distinct IN functional domains can be distinguished: the *N*-terminal and *C*-terminal domains and the catalytic core. The structure and function of the central and the *C*-terminal domains are well established for HIV-1 [11–15], ASV [16] and SIV [17]. Studies would suggest that separate domains resemble those of domains in the full-length protein [14,15]. The *N*-terminal domain, characterized by a His-X_{3-7}-His-X_{22-32}-Cys-X_2-Cys sequence well conserved in all retroviral IN proteins, that is required for the integration process, has a mode of action which still remains unclear [18].

In the case of HIV-1 and HIV-2 IN *N*-terminal domains, which have 60% of sequence similarity, this HHCC motif was shown to be able to bind a Zn^{2+} ion and to form an atypical 'zinc finger' type structure with a 1 : 1 domain/zinc stoichiometry [19–23].

The HTLV-I *N*-terminal domain (50 residues) presents only 14% and 18% of sequence similarity [24] with the HIV-1 and HIV-2 corresponding domains, respectively.

The present work is aimed at the investigation of a zinc finger in the IN *N*-terminus of HTLV-I.

MATERIALS AND METHODS

Peptide Synthesis

All solvents, reagents, Fmoc- and Boc-protected amino acids were purchased from SDS (France), Sigma-Aldrich (France) and Novabiochem (Switzerland).

IN50 (C_{16} acetamidomethyl protected-50-mer) and IN35 (C_{16}S-35-mer and C_{16}S-D_1N-35-mer) peptides were obtained by solid-phase synthesis using Fmoc and Boc strategies on Applied Biosystems 433A and 431A Peptide Synthesizers, respectively. For the IN50 Fmoc synthesis, the Fmoc-Met-HMP-resin (0.70 mmol/g), which provides a peptide carboxylic acid after cleavage, was used. Amino acids with reactive side-chain groups were protected as follows: Arg: 2,2,5,7,8-pentamethylchroman-6-sulfonyl; Asn, Cys, Gln, His: trityl; Asp, Glu: O-*tert*-butyl; Cys_{16}: acetamidomethyl; Lys: Boc; Ser, Thr: *tert*-butyl. The Fmoc protecting group was removed at the beginning of every cycle by treatment with 20% (v/v) piperidine in NMP and the deprotection was controlled using an UV absorbance (301 nm) monitoring. The deprotection time was increased in comparison with the classical synthesizer conditions. Additional NMP and piperidine were added to the peptide-resin and allowed to react for 30 min more. The synthesizer programming was also executed in order to induce a new deprotection when the yield difference between two consecutive deprotections was more than 5%. A maximum of four deprotection cycles was necessary. After deprotection, the peptide-resin was washed with NMP to remove the piperidine. Activation of amino acids with a solution of HBTU/HOBt (0.50/0.45 M) in DMF was performed before each coupling step. A six-fold excess of each amino acid was employed. Coupling time was 50% increased (22 min) in comparison with the classical protocol. After each coupling and a resin washing with NMP, a capping step with a solution of 4.75% acetic anhydride, 2.25% DIEA (v/v) and 0.2% HOBt (w/v) in NMP was included in order to block the unreacted amino groups. Cleavage of the peptide from the resin and simultaneous deprotection of side-chain groups were performed with a solution of TFA (10 ml), phenol (0.75 g), 1,2-ethanedithiol (0.25 ml), thioanisole (0.50 ml) and deionized H_2O (0.50 ml). This solution was cooled on ice before being added to the cooled peptide-resin. The mixture was stirred for 2 h at ambient temperature and filtered. The resin was washed with TFA (1 ml) and DCM (2 × 5 ml). The peptide was precipitated by addition of chilled diethyl ether (50 ml), then filtered and washed with diethyl ether (20 ml). The crude peptide was dissolved in aqueous 0.1% TFA, freeze-dried and stored under N_2 at 0 °C.

For the IN35 Boc syntheses, the Boc-Lys(2-chlorobenzyloxycarbonyl)-PAM-resin (0.59 mmol/g), which ensures a free *C*-terminal carboxylic acid after HF treatment, was used. The protecting groups of the reactive side chains were: Arg: tosyl; Asp, Glu: O-cyclohexyl; Cys: p-methoxybenzyl; His: 2,4-dinitrophenyl; Lys: 2-chlorobenzyloxycarbonyl; Ser, Thr: benzyl. The Boc protecting group was removed by 50% TFA in DCM (v/v). Then, the peptide-resin was washed with DCM, neutralized with 10% DIEA in DCM (v/v) and washed with NMP. Activation of protected amino acids was performed with 1 M

HOBt in NMP. Coupling with 1 M DCC in NMP used four equivalents of the activated amino acids per one equivalent of the growing peptide chain. During this step, 15% DMSO in NMP (v/v) and DIEA were added. Then, a capping step with a solution of 10% acetic anhydride, 5% DIEA in NMP (v/v) was performed. After the automated synthesis, the Boc-α-amino peptide-resin (corresponding to C_{16}S-35-mer) or the acetylated α-amino peptide-resin (corresponding to C_{16}S-D_1N-35-mer) was mixed with 20% 2-mercaptoethanol and 10% DIEA in DMF (v/v), for 90 min at ambient temperature, to remove the His protecting group. The Boc-α-amino terminal protecting group was removed by treatment with 50% TFA in DCM (v/v), for 15 min, followed by DCM and methanol washings. Cleavage of the dried peptide from the resin and simultaneous deprotection of side-chain groups were performed by treatment with liquid HF in the presence of 10% anisole and 0.1% 1,2-ethanedithiol (v/v) at 0 °C for 90 min. After HF removal, the mixture was poured into a filter, washed with cooled diethyl ether and subsequently with 10% acetic acid (v/v), three times. The combined water filtrates were freeze-dried and the crude peptide was stored under N_2 at 0 °C.

RP-HPLC Analysis

Analytical RP-HPLC was performed on a C_{18} Vydac column (4.6 × 250 mm, 300 Å, 5 μm) with a Waters 600E system controller and 996 photodiode array detector. Elution was achieved with a linear gradient from 10% to 60% acetonitrile in water −0.08% (v/v) TFA, for 30 min, at a flow rate of 1 ml/min with detection at 220 nm. An identical column, with the same elution conditions, was used for the stability studies. Semi-preparative RP-HPLC was performed on a C_{18} Vydac column (10 × 250 mm, 300 Å, 5 μm) with a linear gradient from 25% to 40% acetonitrile in water −0.08% (v/v) TFA, for 30 min, at a flow rate of 2.5 ml/min, with a Waters 510 apparatus equipped with a 481 Lambdamax detector adjusted to 220 nm. To avoid possible oxidation of Cys residues, the RP-HPLC-purified peptide samples were immediately freeze-dried after their elution. Their identity and purity were confirmed by amino acid analysis on a Millipore-Waters Pico-Tag Work Station (Table 1) and mass spectrometry.

Table 1 Amino acid analysis of synthetic peptides

Amino acid	Peptide		
	IN50	IN35	
		C_{16}S-35-mer	C_{16}S-D_1N-35-mer
Pro	3.1 (3)		
Val	1.0 (1)		
Leu	6.2 (6)	3.9 (4)	4.0 (4)
Glu/Gln	6.4 (6)	2.8 (3)	2.8 (3)
Ser	3.7 (4)	3.9 (4)	3.8 (4)
Ala	5.2 (5)	3.9 (4)	3.8 (4)
Asn/Asp	4.2 (4)	1.9 (2)	2.1 (2)
His	4.0 (4)	2.7 (3)	2.8 (3)
Phe	1.1 (1)	1.0 (1)	1.1 (1)
Thr	5.9 (6)	5.8 (6)	6.0 (6)
Cys	3.3 (3)	2.0 (2)	2.0 (2)
Gly	2.1 (2)	1.8 (2)	1.9 (2)
Ile	1.1 (1)	1.0 (1)	1.1 (1)
Arg	2.2 (2)	2.1 (2)	2.1 (2)
Lys	1.2 (1)	1.2 (1)	1.2 (1)
Met	1.0 (1)		

Mass Spectrometry Analysis

In ESI mode, analyses were performed on a Micromass LCT instrument equipped with an electrospray source and a time-of-flight analyser. Calibration was made with horse heart apomyoglobin. Samples were introduced direct via a syringe pump (Harvard) at a constant flow rate of 5 μl/min, at a 10 μM concentration in a water/methanol (1 : 1, v/v) mixture containing 1% (v/v) acetic acid. The IN50 and IN35 peptides were identified by this method: average mass M_{IN50} exp.: 5500.7 ± 0.7 Da, calc. mass: 5500.2 Da; monoisotopic mass $M_{IN35(C16S-35-mer)}$ exp.: 3755.8 ± 0.7 Da, calc. mass: 3755.8 Da; $M_{IN35(C16S-D1N-35-mer)}$ exp.: 3796.4 ± 0.7 Da, calc. mass: 3796.8 Da. The IN50-Zn complex was analysed at 20 μM concentration in 20 mM ammonium acetate, pH 6.5. Average mass was measured: 5562.0 ± 2.0 Da; calc. mass: 5563.6 Da. For LC/MS experiments, a C_{18} column (1 × 150 mm, 300 Å, 5 μm) was used with the same eluant as for analytical HPLC, except that the TFA concentration was decreased to 0.05% (v/v). The flow rate was 50 μl/min and a splitter was inserted before the mass spectrometer, allowing a 20 μl/min flow rate to reach the electrospray source. The average mass for the oxidized IN50 peptide was measured: 5498.7 ± 0.6 Da; calc. mass: 5498.2 Da.

Metal Complexes Preparation

To a solution of IN50 or IN35 peptides in 20 mM phosphate buffer pH 7.4, 1.2 equivalents of $CoCl_2$ or $ZnCl_2$ solution was rapidly added.

UV-visible Measurements

UV-visible absorption spectra were recorded from 250 to 800 nm on a Safas model 190 DES spectrometer with quartz cells of 0.5 cm pathlength. The concentration of solutions was 0.2 mM. The spectra were corrected for background and free peptide interferences.

CD Spectroscopy

The CD spectra of the IN 50 peptide and the corresponding zinc complex (35 μM) in 20 mM phosphate buffer, pH 7.4, were recorded at room temperature on a Mark VI Jobin Yvon dichrograph, between 184 and 270 nm, with a cell of 1 mm path length (increment 0.5 nm; bandwidth 2 nm; integration time 2 s). Spectra were also recorded in the presence of increasing amounts of TFE. For each sample, three spectra were collected and averaged, and the background interference was corrected. The spectrum deconvolution was calculated by the method of Johnson [25].

Size Exclusion Chromatography

Analytical separation was performed on a Superdex peptide HR 10/30 column, using a Waters 2690 separation module and a 486 tunable absorbance detector adjusted to 220 nm. Elution was achieved with a 20 mM phosphate buffer, pH 7.2, containing 0.25 M NaCl at a flow rate of 0.5 ml/min. A mixture of horse heart cytochrome c (MW 12.4 kDa) and BPTI (MW 6.5 kDa) was used as a molecular weight marker.

RESULTS

Peptide Synthesis

The sequence of the HTLV-I IN *N*-terminal domain [24] corresponds to a 50-residue fragment (IN50) (Figure 1). In order to avoid the potential formation of the non-native $H_{11}C_{16}C_{38}C_{41}$ (HTLV-I numbering) zinc finger, the cysteine in position 16 was protected in a permanent way with an acetamidomethyl group. At first, the synthesis of the IN50 peptide with the Boc strategy (Boc-Met-PAM-resin, 0.60 mmol/g) in a simple coupling was tried and it led to a too complex mixture of truncated peptides. Use of a double coupling protocol proved also to be unsuccessful. Eventually, the IN50 peptide synthesis was achieved successfully using the Fmoc strategy (FastMoc method: activation with HBTU/HOBt) [26,27]. In this basic chemistry, the side chains of Gln and Asn residues are protected by the stable trityl group.

As some difficulty was encountered in solubilizing the IN50 purified peptide, a shorter compound was synthesized. As described in the literature, a shorter peptide must contain at least two amino acids before the *N*-terminal and after the *C*-terminal Cys and/or His residues implicated in the metal chelation to allow zinc finger formation [28,29]. Therefore, in our case, the minimal peptide (35-residue) results from the truncation of 1–8 and 44–50 segments. Such a peptide, missing the hydrophobic *N*-terminus, was expected to be more water-soluble. Moreover, in order to avoid undesirable side effects, the Cys_{16} residue, not conserved among the retroviral integrase proteins, was substituted with Ser_{16} (C_{16}S-35-mer). Another 35-residue peptide (C_{16}S-D_1N-35-mer), which is different in the first residue from the previous IN35, was also synthesized. A D_1N substitution was introduced and the *N*-terminus was acetylated. The purpose of this modification was to suppress the possible contribution of electrostatic forces that could disturb metal finger formation. The IN35 peptides (C_{16}S-35-mer and C_{16}S-D_1N-35-mer) (Figure 1) were synthesized by the Boc strategy [27]. ESI-MS spectra (not shown) confirmed the success of these syntheses.

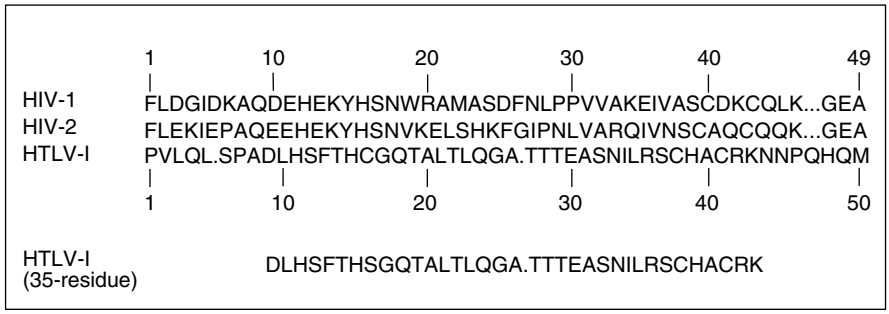

Figure 1 Alignment of amino acid sequences of HIV-1, HIV-2 and HTLV-I IN *N*-terminal domains. Amino acid sequence corresponding to the HTLV-I IN35 *N*-terminal domain is also indicated.

Characterization of Peptide-zinc Finger

According to UV-visible absorption spectroscopy a previous coordination of a metallic ion, which absorbs into the 200–800 nm range such as Co^{2+} [30,31] is required for the control of an effective zinc complexation. The UV-visible absorption spectrum of IN50 cobalt complex showed a *d-d* electronic transition band of the Co^{2+} ion at 627 nm with a shoulder at 583 nm (Figure 2). Furthermore, $-S^- \rightarrow Co^{2+}$ charge transfer bands appeared at 289 nm and 316 nm. These results are in accordance with a tetrahedral metal coordination by nitrogen and sulphur atoms of two His and two Cys residues respectively [31–34]. Spectra recorded at different pH values showed that His protonation, below pH 5.8, prevents metal coordination. Substitution of the cobalt ion in the tetrahedral complex by addition of a $ZnCl_2$ solution resulted in the disappearance of the characteristic *d-d* transition bands in the visible spectrum.

In ESI-MS, by infusion of IN50 peptide treated with 1.2 equivalents of $ZnCl_2$ in ammonium acetate solution, the IN50–Zn complex was observed.

These results confirm the effective formation of a zinc finger with the synthetic HTLV-I IN50 peptide.

The UV-visible absorption spectrum of the IN35 cobalt complexes displayed an unusual broad band at 312 nm and a very tiny band at 627 nm (Figure 2). This result suggests that these complexes are highly unstable.

Stability Studies

Kinetics studies of the IN50 cobalt finger, using UV-visible absorption spectroscopy, showed a very fast degradation of the complex in the presence of air; in particular, it showed a gradual decrease of the *d-d* transition bands ($K_d^{Co} = 4 \times 10^{-4}$ M). However, under anaerobic conditions (under nitrogen atmosphere), the cobalt complex proved to be stable.

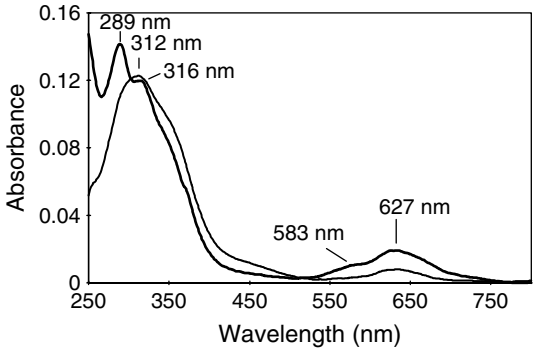

Figure 2 UV-visible absorption spectrum of the HTLV-I IN50 (thick line) and IN35 (thin line) in the presence of 1.2 eq. of $CoCl_2$.

In order to control the IN50 zinc finger stability, that cannot be tested by UV-visible spectroscopy, we optimized an original method using RP-HPLC chromatography. Since the His residues are protonated at the very low pH used for RP-HPLC elution, zinc finger dissociation occurs on the column and the IN50 peptide chromatogram is obtained. Moreover, we assumed that zinc finger instability must translate into a peptide release. But, at the neutral pH of the solution, oxidation of the Cys residues can occur and the resulting oxidized peptide (OxIN50) would be eluted at a retention time different from that of the IN50 peptide. Therefore, the absence of an oxidation product in the RP-HPLC chromatogram of a zinc finger solution at neutral pH would be a good proof of its stability.

Experimentally, we studied first the chromatographic behaviour of the IN50 peptide dissolved in phosphate buffer at pH 7.4. The RP-HPLC analysis, performed 10 min after dissolution, showed the presence of 80% IN50 peptide and 20% of a secondary product (Figure 3a). After 5 h, this secondary product increased to 75% (Figure 3a). It should be noted that this product did not appear when the peptide was dissolved in an acidic medium. LC/MS experiments showed that this compound is an intramolecular oxidation product (OxIN50) corresponding to a disulphide bridge formation between two Cys residues, i.e. a 2 Da mass difference between IN50 and OxIN50 peptides. Secondly, we analysed the IN50 zinc finger under the same solvent conditions (phosphate buffer at pH 7.4). Analysis of a 24 h old solution showed a single chromatographic peak at the same retention time as that observed for the IN50 peptide dissolved into acidic medium (Figure 3b). Thus, the IN50 zinc finger proved to be highly stable.

The IN50 cobalt finger instability, observed in the UV-visible study, was confirmed by RP-HPLC. After 5 h, the chromatogram profile was similar to that obtained for the IN50 peptide dissolved at pH 7.4 and analysed after the same time (Figure 3a).

The IN35 metal complexes were also studied. Experiments revealed an important oxidation phenomenon for both the IN35 peptides (Figure 3c) and the IN35 cobalt fingers (Figure 3d). This shows the very high instability of the IN35 cobalt fingers and confirms the result obtained in the UV-visible absorption characterization study. As far as the IN35 zinc fingers are concerned, an oxidative process was

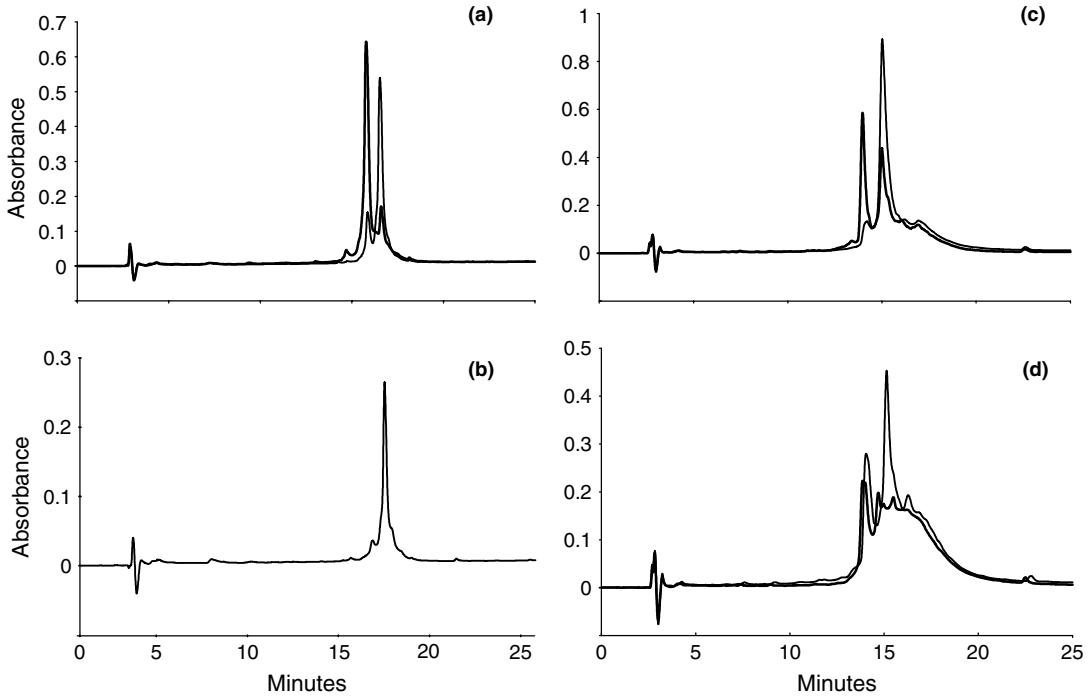

Figure 3 Time evolution of analytical RP-HPLC chromatograms (220 nm): (a) IN50 peptide solution after 10 min (thin line) and after 5 h (thick line); (b) IN50 zinc finger solution after 24 h; (c) IN35 peptide solution after 10 min (thin line) and after 5 h (thick line); (d) IN35 zinc finger solution after 24 h (thin line) and IN35 cobalt finger solution after 10 min (thick line).

also observed but this phenomenon is less important that the one observed for the IN35 cobalt fingers.

Conformational Studies

Experimental CD studies, performed on the IN50 zinc finger in a 20 mM phosphate buffer, pH 7.4, with addition of 0 to 50% TFE (v/v), allowed the quantification of the percentage of secondary structure elements in the molecule (Figure 4). Without TFE, the IN50 peptide and the IN50 zinc finger spectra mainly presented a negative band at 201 nm that is characteristic of a random-coil structure. Both the IN50 peptide and its corresponding complex would be unstructured in aqueous solution. Above 20% TFE, the spectrum began to show some significant modifications. Whereas the band at 201 nm was shifted to 206 nm, the magnitude of the shoulder at 222 nm increased and an additional positive signal appeared at 192 nm. All these observations are characteristic of an helical conformation. An increase of the TFE percentage to 50% led to evidence of an increase in the helical structure. A singular value deconvolution method [25] gave a global result of 32% α-helices, 4% β-strands, 33% β-turns and 31% random coil.

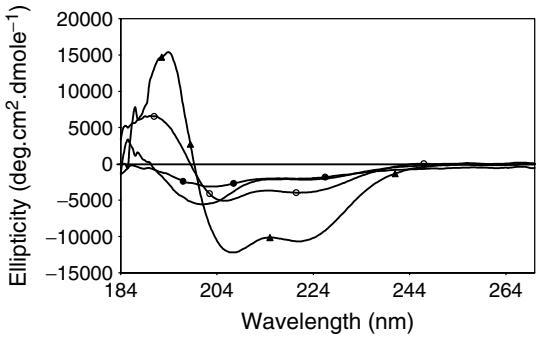

Figure 4 CD spectra of the HTLV-I IN50 zinc finger (thin line) after addition of 10% (full circles), 20% (empty circles) and 50% (triangles) TFE.

Classical algorithms for the prediction of protein secondary structures [35–49] were used to discuss the possible solution structures. The predicted conformations of the HIV-1 and HTLV-I IN *N*-terminal domains are reported in Figure 5, together with the NMR derived structure obtained for the isolated HIV-1 *N*-terminal domain [22]. As far as HIV-1 is concerned, there is a reasonable correlation between experimentally observed and predicted structures. For HTLV-I, the predicted conformation appears to be specific. Only a helical region, encompassing residues 28–42 and representing

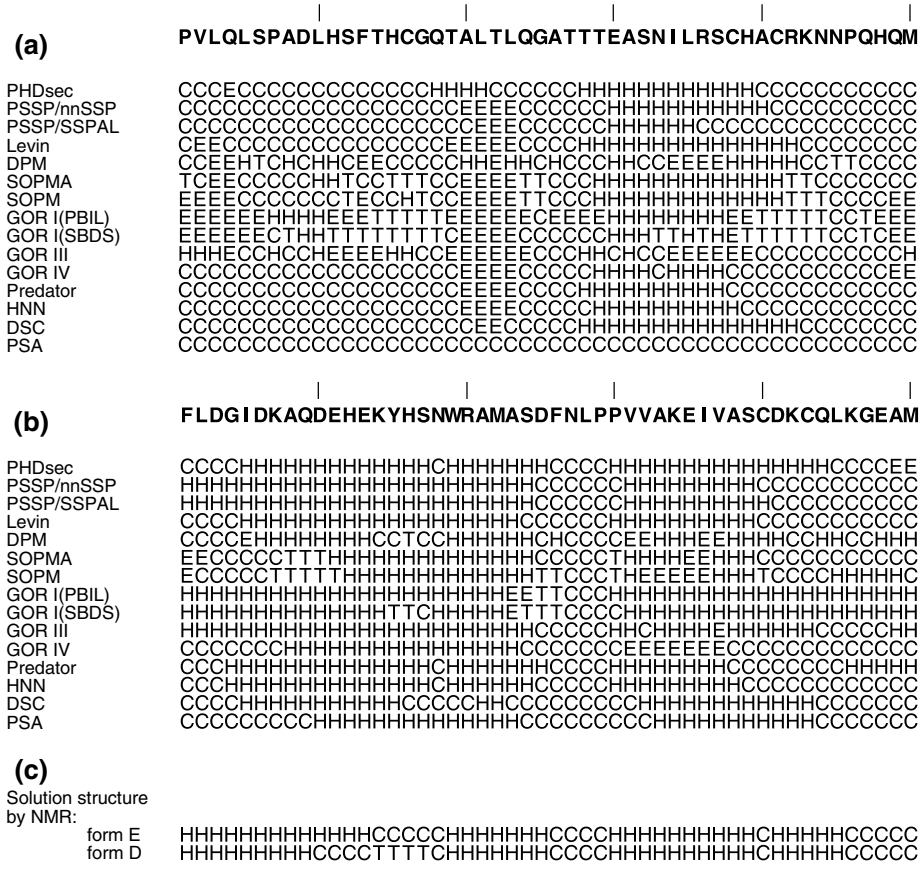

Figure 5 Secondary structure predictions of HTLV-I (a) and HIV-1 (b) IN *N*-terminal domains using different prediction algorithms: PHDsec [35,36], PSSP/nnSSP [37], PSSP/SSPAL [38], Levin [39], DPM [40], SOPM [41], SOPMA [42], GOR I [43], GOR III [44], GOR IV [45], Predator [46], HNN [47], DSC [48], PSA [49]. (c) Solution structure of HIV-1 IN *N*-terminal domain determined by NMR is indicated. (H, helix; E, extended; T, turn; C, coil).

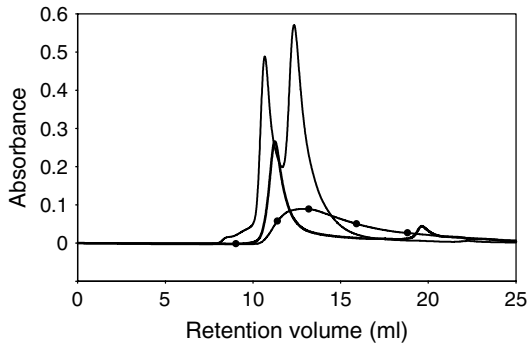

Figure 6 Size exclusion chromatography profile of the HTLV-I IN50 zinc finger in the absence (spotted line) and in the presence (thick line) of 50% TFE. Molecular weight markers (thin line): cytochrome c (10.7 ml) and BPTI (12.3 ml).

30% of the sequence, would be conserved. A β-strand structure would correspond to the residues 20–23 (8%). The *N*-terminal and the *C*-terminal parts would mainly present a randomly coil type structure including approximately 20% of turns.

Whereas size exclusion chromatography of the IN50 zinc finger solution showed a broad signal at a retention volume close to that of BPTI (6.5 kDa) in the absence of TFE, indicating a monomeric structure, a well defined peak was observed at 11.2 ml retention volume in the presence of 50% TFE (v/v) (Figure 6). This latter retention volume is close to the value observed for cytochrome c (12.4 kDa), indicating the formation of a dimeric HTLV-1 IN zinc finger (M.W. 11 kDa).

DISCUSSION

Concerning the failure of the IN50 peptide synthesis using the Boc strategy, one explanation could be the presence of two Gln ($Q_{47,49}$) and two Asn ($N_{44,45}$) residues at the *C*-terminus of the peptide. These residues are usually not protected in acidic

chemistry, and their amide-type side chains are known to be able to interact with neighbouring main chains through the formation of hydrogen bonds. The resulting network can induce a steric hindrance that leads to a weak α-NH_2 function accessibility for the addition of the next bulky Lys_{43} residue. The DIEA addition during the coupling step did not allow the effective disruption of peptide–peptide hydrogen bonds. The xanthanyl or the 4,4′-dimethoxybenzhydryl groups may be used to protect the side chain of these residues. However, neither is stable to the TFA treatments used for Boc group cleavage and their progressive removal during the synthesis is known. Thus, when Gln and Asn residues are present from the outset of a synthesis, the use of these protecting groups is not effective. Therefore, the peptide was synthesized using the Fmoc chemistry using Gln and Asn protected residues. Because the deprotection yield proved to be low from residues 25 to 34 (from G_{25} to C_{16}), the deprotection time was prolonged during synthesis. Introduction of side-chain protecting groups for Gln and Asn is probably the main reason for the success of the synthesis.

The IN35 peptides which do not contain the two Gln and two Asn incriminated residues were synthesized using the Boc strategy without specific difficulties, thereby confirming our aforementioned explanations of the failure of Boc synthesis of the IN50 peptide.

The HTLV-I IN50 zinc finger solubility was proved to be even lower than that of peptide IN50. Different organic solvents, suitable for peptide structural studies, mixed with a small amount of water in order to ensure a neutral pH and the solubility of zinc salts, were tested. The most successful solvent medium was DMSO/H_2O (80/20 v/v), pH 7. However, as DMSO is known to be a mild oxidizing agent of the Cys residues [50], the stability of the IN50 zinc finger in the presence of such a solvent has been monitored using RP-HPLC. Unfortunately, after a few hours, an oxidation product (OxIN50) was observed (data not shown). This progressive oxidative phenomenon is incompatible with NMR experiments which require a long time for data acquisition. In buffer, at neutral pH, the solubility was also weak. Attempts at peptide refolding using slow dialysis [51,52] or the addition of aggregation-limiting compounds, such as glycerol and sodium chloride [22], did not allow us to overcome the problem. The maximal concentration (0.4 mM) was obtained in phosphate buffer, pH 7.4, allowing the mass spectrometry, spectroscopic (UV-visible, CD) and chromatographic studies previously described to be performed, which require low peptide concentrations.

Using the scale of hydropathicity [53], the HIV-1 and HIV-2 IN *N*-terminal domains exhibit a hydrophilic region (residues 3–20 and 3–25 in HIV-1 and HIV-2 respectively), but the HTLV-I IN50 domain does not present a highly hydrophilic or hydrophobic character. This difference could explain the insolubility of HTLV-I IN50 in contrast with the *N*-terminal domain of HIV-1 and HIV-2 integrases.

The truncation of the IN50 peptide at both ends, leading to the IN35 peptide, did not improve significantly the *N*-terminal domain solubility. Furthermore, it has been shown that this truncation did not prevent cobalt and zinc complexation, but significantly decreased the stability of the resulting complexes. Moreover, RP-HPLC analysis of the IN50 zinc finger showed the absence of an oxidation product, attesting a full complexation and a high zinc finger stability.

Knowledge of the *in vitro* stability of the zinc fingers is a *sine qua non* condition before undertaking biological or structural studies. Therefore, our control method by RP-HPLC is an interesting tool.

Some structural features were determined by CD, size exclusion chromatography and structural predictions. CD experiments indicate a random coil structure, for both the IN50 peptide and the corresponding zinc complex, suggesting that the peptide–zinc complexation does not introduce a well-defined 3D structure. The results obtained in the presence of TFE are in agreement with many studies which showed that TFE, used as a cosolvent in aqueous solution, promotes helix formation in peptides that have a propensity to form helices [54–56]. A satisfactory agreement exists between the secondary structure obtained by CD and by 2D predictions, especially for α-helices and β-strands percentages.

Thus, the HTLV-I IN *N*-terminal domain appears to be less structured than that of HIV-1 and of HIV-2, mainly constituted by four and three α-helices, respectively. Both CD and size exclusion chromatography results suggest that the helical structure of the HTLV-I IN50–zinc complex induces a dimerization. In HIV-1 and HIV-2, dimerization of the IN–zinc complex is also observed and in the same helical region (residues 30–45) [22,23].

Acknowledgements

The authors thank E. Loret (CNRS Marseille) and C. Aznar (UBS Bordeaux) for their assistance

in peptide synthesis and CD spectroscopy, respectively.

This work was supported by a grant from the Comité Départemental de la Dordogne de la Ligue Nationale Contre le Cancer.

REFERENCES

1. Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl Acad. Sci. USA* 1980; **77**: 7415–7419.
2. Gessain A, Barin F, Vernant JC, Gout O, Maurs L, De Calender A. Thé G. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* 1985; **2**: 407–410.
3. Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M. HTLV-I associated myelopathy, a new clinical entity. *Lancet* 1986; **1**: 1031–1032.
4. Goëtz M, Geoffre S, Busetta B, Manigand C, Nespoulos C, Londos-Gagliardi D, Guillemain B, Hospital M. Synthesis and CD studies of an 88-residue peptide containing the main receptor binding site of HTLV-I SU-glycoprotein. *J. Peptide Sci.* 1997; **3**: 347–353.
5. Khorasanizadeh S, Campos-Olivas R, Summers MF. Solution structure of the capsid protein from the human T-cell leukemia virus type-I. *J. Mol. Biol.* 1999; **291**: 491–505.
6. Katz RA, Skalka AM. The retroviral enzymes. *Annu. Rev. Biochem.* 1994; **63**: 133–173.
7. Kulkosky J, Skalka AM. Molecular mechanism of retroviral DNA integration. *Pharmac. Ther.* 1994; **61**: 185–203.
8. Muller B, Krausslich HG. Characterization of human T-cell leukemia virus type I integrase expressed in *Escherichia coli*. *Eur. J. Biochem.* 1999; **259**: 79–87.
9. Vincent KA, Ellison V, Chow SA, Brown PO. Characterization of human immunodeficiency virus type 1 integrase expressed in *Escherichia coli* and analysis of variants with amino-terminal mutations. *J. Virol.* 1993; **67**: 425–437.
10. Vink C, Plasterk RHA. The human immunodeficiency virus integrase protein. *TIG* 1993; **9**: 433–437.
11. Dyda F, Hickman AB, Jenkins TM, Engelman A, Craigie R, Davies DR. Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science* 1994; **266**: 1981–1986.
12. Lodi PJ, Ernst JA, Kuszewski J, Hickman AB, Engelman A, Craigie R, Clore GM, Gronenborn AM. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry* 1995; **34**: 9826–9833.
13. Eijkelenboom AP, Puras Lutzke RA, Boelens R, Plasterk RHA, Kaptein R, Hard K. The DNA-binding domain of HIV-1 integrase has an SH3-like fold. *Nat. Struct. Biol.* 1995; **2**: 807–810.
14. Eijkelenboom AP, Sprangers R, Hard K, Puras Lutzke RA, Plasterk RHA, Boelens R, Kaptein R. Refined solution structure of the *C*-terminal DNA-binding domain of human immunovirus-1 integrase. *Proteins* 1999; **36**: 556–564.
15. Chen JC, Krucinski J, Miercke LJ, Finer-Moore JS, Tang AH, Leavitt AD, Stroud RM. Crystal structure of the HIV-1 integrase catalytic core and *C*-terminal domains: a model for viral DNA binding. *Proc. Natl Acad. Sci. USA* 2000; **97**: 8233–8238.
16. Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, Katz RA, Skalka AM. High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J. Mol. Biol.* 1995; **253**: 333–346.
17. Chen Z, Yan Y, Munshi S, Li Y, Zugay-Murphy J, Xu B, Witmer M, Felock P, Wolfe A, Sardana V, Emini EA, Hazuda D, Kuo LC. X-ray structure of simian immunodeficiency virus integrase containing the core and *C*-terminal domain (residues 50–293) — an initial glance of the viral DNA binding platform. *J. Mol. Biol.* 2000; **296**: 521–533.
18. Esposito D, Craigie R. HIV integrase structure and function. *Adv. Virus Res.* 1999; **52**: 319–333.
19. Burke CJ, Sanyal G, Bruner MW, Ryan JA, LaFemina RL, Robbins HL, Zeft AS, Middaugh CR, Cordingley MG. Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase. *J. Biol. Chem.* 1992; **267**: 9639–9644.
20. Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R. Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc. Natl Acad. Sci. USA* 1993; **90**: 3428–3432.
21. Haugan IR, Nilsen BM, Worland S, Olsen L, Helland DE. Characterization of the DNA-binding activity of HIV-1 integrase using a filter binding assay. *Biochem. Biophys. Res. Commun.* 1995; **217**: 802–810.
22. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, Gronenborn AM. Solution structure of the *N*-terminal zinc binding domain of HIV-1 integrase. *Nat. Struct. Biol.* 1997; **4**: 567–577.
23. Eijkelenboom AP, van den Ent FM, Wechselberger R. Refined solution structure of the dimeric *N*-terminal HHCC domain of HIV-2 integrase. *J. Biomol. NMR* 2000; **18**: 119–128.
24. Khan E, Mack JP, Katz RA, Kulkosky J, Skalka AM. Retroviral integrase domains: DNA binding and the recognition of LTR sequences. *Nucleic Acids Res.* 1991; **19**: 851–860.
25. Johnson Jr WC. Protein secondary structure and circular dichroism: a practical guide. *Proteins: Structure, Function Gen.* 1990; **7**: 205–214.
26. Dourtoglou V, Gross B, Lambropoulou V, Zioudrou C. *O*-Benzotriazolyl-*N*,*N*,*N*′,*N*′-tetramethyluronium hexafluorophosphate as coupling reagent for the

synthesis of peptides of biological interest. *Synthesis* 1984; 572–574.
27. Llyod-Williams P, Albericio F, Giralt E. *Chemical Approaches to the Synthesis of Peptides and Proteins*. CRC Press: New York, 1997; 1–75.
28. Green LM, Berg JM. A retroviral Cys-Xaa$_2$-Cys-Xaa$_4$-His-Xaa$_4$-Cys peptide binds metal ions: spectroscopic studies and a proposed three-dimensional structure. *Proc. Natl Acad. Sci. USA* 1989; **86**: 4047–4051.
29. Mely Y, De Rocquigny H, Morellet N, Roques BP, Gerard D. Zinc binding to the HIV-1 nucleocapsid protein: a thermodynamic investigation by fluorescence spectroscopy. *Biochemistry* 1996; **35**: 5175–5182.
30. Fitzgerald DW, Coleman JE. Physicochemical properties of cloned nucleocapsid protein from HIV. Interactions with metal ions. *Biochemistry* 1991; **30**: 5195–5201.
31. Klemba M, Regan L. Characterization of metal binding by a designed protein: single ligand substitutions at a tetrahedral Cys$_2$His$_2$ site. *Biochemistry* 1995; **34**: 10094–10100.
32. Parraga G, Horvath SJ, Hood L, Young ET, Klevit RE. Spectroscopic studies of wild-type and mutant 'zinc finger' peptides: determinants of domain folding and structure. *Proc. Natl Acad. Sci. USA* 1990; **87**: 137–141.
33. Weiss MA, Keutmann HT. Alternating zinc finger motifs in the male-associated protein ZFY: defining architectural rules by mutagenesis and design of an 'aromatic swap' second-site revertant. *Biochemistry* 1990; **29**: 9808–9813.
34. Krizek BA, Amann BT, Kilfoil VJ, Merkle DL, Berg JM. A consensus zinc finger peptide: design, high-affinity metal binding, a pH-dependent structure, and a His to Cys sequence variant. *J. Am. Chem. Soc.* 1991; **113**: 4518–4523.
35. Rost B, Sander C. Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* 1993; **232**: 584–599.
36. Rost B, Sander C. Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 1994; **19**: 55–72.
37. Salamov AA, Solovyev VV. Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments. *J. Mol. Biol.* 1995; **247**: 11–15.
38. Salamov AA, Solovyev VV. Protein secondary structure prediction using local alignments. *J. Mol. Biol.* 1997; **268**: 31–36.
39. Levin JM, Garnier J. Improvements in a secondary structure prediction method based on a search for local sequence homologies and its use as a building tool. *Biochim. Biophys. Acta* 1988; **955**: 283–295.
40. Deléage G, Roux B. An algorithm for protein secondary structure prediction based on class prediction. *Protein Eng.* 1987; **1**: 289–294.
41. Géourjon C, Deléage G. SOPM: a self-optimized method for protein secondary structure prediction. *Protein Eng.* 1994; **7**: 157–164.
42. Géourjon C, Deléage G. SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* 1995; **11**: 681–684.
43. Garnier J, Osguthorpe DJ, Robson B. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. *J. Mol. Biol.* 1978; **120**: 97–120.
44. Gibrat J-F, Garnier J, Robson B. Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs. *J. Mol. Biol.* 1987; **198**: 425–443.
45. Garnier J, Gibrat J-F, Robson B. GOR secondary structure prediction method version IV. *Methods Enzymol.* 1996; **266**: 540–553.
46. Frishman D, Argos P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* 1996; **9**: 133–142.
47. Guermeur Y. *Combinaison de classifieurs statistiques. Application à la prédiction de structure secondaire des protéines*. 1997; PhD Thesis. University of Paris 6.
48. King RD, Sternberg MJ. Identification and application of the concepts important for accurate and reliable protein secondary structure prediction. *Protein Sci.* 1996; **5**: 2298–2310.
49. Stultz CM, White JV, Smith TF. Structural analysis based on state-space modeling. *Protein Sci.* 1993; **2**: 305–314.
50. Tam JP, Shen Z-Y. Efficient approach to synthesis of two-chain asymmetric cysteine analogs of receptor-binding region of transforming growth factor-α. *Int. J. Peptide Protein Res.* 1992; **39**: 464–471.
51. McPhie P. Enzyme purification and related techniques. *Methods Enzymol.* 1971; **22**: 23–32.
52. McCartney RG, Sanderson SJ, Lindsay JG. Refolding and reconstitution studies on the transacetylase-protein X (E2/X) subcomplex of the mammalian pyruvate dehydrogenase complex: evidence for specific binding of the dihydrolipoamide dehydrogenase component to sites on reassembled E2. *Biochemistry* 1997; **36**: 6819–6826.
53. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 1982; **157**: 105–132.
54. Dyson HJ, Wright PE. Peptide conformation and protein folding. *Curr. Opin. Struct. Biol.* 1993; **3**: 60–65.
55. Jasanoff A, Fersht AR. Quantitative determination of helical propensities from trifluoroethanol titration curves. *Biochemistry* 1994; **33**: 2129–2135.
56. Shiraki K, Nishikawa K, Goto Y. Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding. *J. Mol. Biol.* 1995; **245**: 180–194.

