Solution Structure of RP 71955, a New 21 Amino Acid Tricyclic Peptide Active against HIV-1 Virus‡

D. Fréchet,* J. D. Guitton, F. Herman, D. Faucher, G. Helynck, B. Monegier du Sorbier, J. P. Ridoux, E. James-Surcouf, and M. Vuilhorgne

Rhône-Poulenc Rorer S.A., 13 Quai Jules Guesde, 94400 Vitry-Sur-Seine, France

Received May 18, 1993; Revised Manuscript Received August 25, 1993®

ABSTRACT: The structure of RP 71955, a new tricyclic 21 amino acid peptide active against human immunodeficiency virus 1, was determined. Its amino acid composition was inferred from the results of fast atom bombardment mass spectrometry, nuclear magnetic resonance, Raman spectroscopy, and amino acid analysis. Its sequence could not be determined classically, using Edman degradation, given the lack of a free terminal NH_2. It was deduced from the interpretation of interresidue nuclear Overhauser effects and confirmed by the sequencing of peptides obtained by limited chemical hydrolysis. It was found to be CLGIGSCNDFAGCGYAVVCFW. An internal amide bond between the NH_2 of C1 and the γ-COOH of D9 was observed, as well as two disulfide bridges, one between C1 and C13 and one between C7 and C19. The three-dimensional structure of RP 71955 was determined from nuclear magnetic resonance derived constraints using distance geometry, restrained molecular dynamics, nuclear Overhauser effect back calculation, and an iterative refinement using a full relaxation matrix approach. Analogies between the structure of RP 71955 and some functional domains of gp41, the transmembrane protein of human immunodeficiency virus 1, suggest hypotheses concerning the mode of action of RP 71955.

The search for new and effective agents in the treatment of HIV[1] infection and AIDS-related complex is a major priority in the pharmaceutical industry. Most drugs presently available are geared toward the inhibition of the two HIV-specific enzymes reverse transcriptase and HIV protease. However, developing strategies aimed at perturbing other steps of the virus life cycle, such as gp120–CD4 interaction, gp120–gp41 interaction, membrane fusion, and so forth, could provide additional means of fighting the viral infection. During screening assays for new anti-HIV agents, a secondary metabolite, RP 71955, isolated in our laboratories from a strain of *Streptomyces*, was found to inhibit HIV replication in cell culture, while its cytotoxicity was very low. The determination of its structure, as well as further biological tests to locate more precisely its molecular target, has been undertaken.

In this paper, we will focus on the determination of the structure of RP 71955. Indeed, RP 71955 proved to be an original 21 amino acid tricyclic peptide. Its primary structure was determined using an ensemble of spectroscopic techniques (FABMS, NMR, Raman) and confirmed by the sequencing of peptide fragments obtained by limited chemical hydrolysis. A detailed conformational analysis of RP 71955 was performed from NMR-derived constraints using distance geometry, restrained molecular dynamics, NOE back calculation, and an iterative refinement using a full relaxation matrix treatment. RP 71955 proved to have some sequence identity with gp41, the transmembrane protein of HIV-1, suggesting hypotheses for structure–function relationships.

MATERIALS AND METHODS

Materials. RP 71955 was isolated from *Streptomyces* strain no. SP9440 as described earlier (Helynck, 1992).

Mass Spectroscopy. Mass spectra were obtained using a VG Autospec spectrometer, in the LSIMS ionization mode (Cs^+ ions accelerated at 35 kV; matrix, NBA).

Raman Spectroscopy. The near-IR FT-Raman spectra were obtained on a Nicolet 60 SXR spectrometer equipped with a Nicolet Raman module. A Nd:YAG laser (1.06 μm) with a 1-W output was used as the excitation source. An NMR tube was used as a cell. The spectral resolution was 2 cm^{-1}.

Amino Acid Analysis. The amino acid analysis was performed on an ABI-420A derivatizer with on-line hydrolysis and a 130A separation system using either standard PITC chemistry or the modified procedure using Marfey's reactant [N^2-(5-fluoro-2,4-dinitrophen-1-yl)-L-alanine amide]. Both techniques were run using the standard programs described by the manufacturer. For the PITC chemistry around 1 μg of product was used to perform the analysis, but 50 μg was necessary when Marfey's reagent was used.

‡ The coordinates of the 20 final MD structures, together with the coordinates of the restrained minimized mean structure of RP 71955, have been deposited in the Brookhaven Protein Data Bank. The accession numbers are 1RPB and 1RPC.

* To whom correspondence should be adressed.

® Abstract published in *Advance ACS Abstracts,* December 15, 1993.

[1] Abbreviations: HIV, human immunodeficiency virus; AIDS, acquired immunodeficiency syndrome; FABMS, fast atom bombardment mass spectrometry; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; LSIMS, liquid secondary ion mass spectrometry; kV, kilovolt; NBA, nitrobenzyl alcohol; near-IR, near infrared; FT, Fourier transform; Nd:YAG, neodymium-doped yttrium aluminium garnet; μm, micrometer; W, watt; cm, centimeter; Å, angstrom; PITC, phenyl isothiocyanate; Tris, tris(hydroxymethyl)aminomethane; μg, microgram; M, molar; mM, millimolar; EDTA, ethylenediaminetetraacetic acid; μL, microliter; mg, milligram; mL, milliliter; N, normal; TFA, trifluoroacetic acid; HPLC, high-pressure liquid chromatography; v/v, volume by volume; MHz, megahertz; DQF-COSY, double quantum filtered bidimensional *J*-correlated spectroscopy; CLEAN-HOHAHA, clean homonuclear Hartman–Hahn bidimensional correlated spectroscopy; Hz, hertz; ppm, part per million; ms, millisecond; CVFF, consistent valence force field; MD, molecular dynamics; kcal, kilocalorie; ps, picosecond; Da, dalton; τ_c, correlation time; rmsd, root mean square deviation; dc, distance constraints; adc, anti-distance constraints; ns, nanosecond. For amino acids, the IUPAC–IUB one-letter symbols have been used.

0006-2960/94/0433-42$04.50/0

Reduction and Carboxymethylation. One hundred micrograms of lyophilized product was solubilized in an 8 M urea buffer containing 360 mM tris and 0.2% EDTA adjusted to pH 9. Ten microliters of 10% β-mercaptoethanol solution was added before incubation for 1 h at 37 °C. The carboxymethylation was obtained by adding 14 μL of a 400 mg/mL solution of iodoacetic acid in 1 N sodium hydroxide. The reaction was stopped after 15 min by addition of 4 μL of β-mercaptoethanol.

Enzymatic Digestion. Trypsin, endoproteinase V8, and chymotrypsin A4 were from Boehringer Mannheim. For all digestions, 95 μL of 40 mM NH_4HCO_3, pH 8.2, containing 2 mM $CaCl_2$ was added to 100 μg of lyophilized product. Then, 5 μL of enzyme at a concentration of 1 mg/mL in the same buffer was added, and the mixture was incubated overnight at 37 °C.

Partial Chemical Hydrolysis. One hundred micrograms of RP 71955 (native or reduced and carboxymethylated) was solubilized in 100 μL of an aqueous 80% TFA solution and incubated for 30 min at 60 °C. The fragments were separated using HPLC: a Brownlee C4 column (4.6 × 30 mm) was eluted using a linear gradient from 0 to 50% of B in 50 min (A, H_2O + 0.07% TFA; B, CH_3CN + 0.07% TFA, v/v). Peptide fragments were sequenced using an ABI 477A microsequencer with an on-line 120 A PTH analyzer according to the standard procedure of the manufacturer.

^1H NMR Sample. The peptide sample was dissolved in a mixture of 50% CD_3OH and 50% H_2O at a concentration of about 10 mM.

NMR Methods. NMR spectra were obtained on Bruker AM 400 and AM 600 spectrometers. Standard homonuclear NMR methods were used to obtain DQF-COSY (Rance et al., 1983), CLEAN-HOHAHA (Griesinger et al., 1988) (mixing times 100 and 300 ms), and NOESY spectra (Jeener et al., 1979) (mixing times 50 and 100 ms). The water resonance was attenuated by low-power irradiation during the relaxation delay (and during the mixing times in the case of the NOESY experiments). NOESY experiments were performed in O1–O2 coherence mode with the phase correction preset. 2D experiments were performed at 2 different temperatures, 10 and 40 °C. 1D NMR spectra were recorded at 3, 10, 20, 30, 40, and 50 °C, in order to measure the chemical shift of amide protons as a function of temperature. Chemical shifts were referenced with respect to the residual methyl resonance of CD_3OH, set at 3.2 ppm. Data processing was performed using the GIFA software (Oxford Molecular) running on a Vaxstation 3100.

Distance Constraint Determination. Internuclear distance constraints were obtained using the program MARDIGRAS (Borgias & James, 1990), starting from NOESY correlation volume integrations measured using GIFA and considering only isotropic overall motion.

Torsion Angle Constraint Determination. Coupling constants were measured from COSY and CLEAN-HOHAHA cross-peak multiplets using the method of Titman and Keeler (1990). Coupling constants smaller than 6 Hz could not be measured precisely because of the width of the ^1H resonances. ϕ angle constraints were determined using the Karplus curve calibrated by Pardi et al. (1984), while χ angle contraints were evaluated using the curve calibrated by Demarco et al. (1978).

Distance Geometry. Distance geometry calculations were performed using the program DIANA (Güntert & Wüthrich, 1991), which minimizes a target function by varying dihedral angles. Standard minimization parameters were used.

Stereospecific Assignments. A first stereospecific assignment of β protons and of the methyl groups of valine was performed as described by Basus (1989). Stereospecific assignments were latter refined using the program GLOMSA (Güntert et al., 1991a). In a first step, distance constraints with pairs of diastereotopic protons were processed so as to allow for both possible stereospecific assignments, and structures were generated using the program DIANA. All distances between the prochiral centers and other protons of the molecules were measured on all of the structures fulfilling a given target function criterion (variable, depending on the stage of the refinement process). These distances were matched to the list of constraints for both possible stereospecific assignments. For a given distance, a particular assignment was accepted if more than 80% of the structures were compatible with this assignment (within a precision of ±0.2 Å). If all distances involving a prochiral center were compatible with the same stereospecific assignment, this assignment was assumed to be accurate and was used for later stages of the refinement process. If the results were inconsistent, distances were examined individually according to the criteria described by Güntert et al. (1991b) and either a stereospecific assignment could be made or the assignment was kept floating. This procedure was repeated at all stages of the refinement process. Whenever chemical shifts of diastereotopic protons were degenerate, pseudoatoms were defined (Wüthrich et al., 1983).

Restrained Molecular Dynamics Calculations. Molecular dynamics simulations were performed with the Discover and Insight II packages (Biosym Technologies Inc.) on a Silicon Graphics SG 340. The consistent valence force field (CVFF) has been used. Force constants on the angles were increased in order to avoid cis–trans inversions during high-temperature MD runs. All calculations were performed in vacuo. Energy minimizations were carried out using the steepest-descent followed by conjugate gradient method (Fletcher & Reeves, 1962) until the maximum derivative was less than 0.1 kcal mol^{-1} $Å^{-1}$.

Interproton distance constraints were imposed by adding a skewed biharmonic potential (Van Gunsteren & Karplus, 1980). The potential was null between the pushing and pulling limits, and the penalty energy increased linearly in the region in which the force had reached its maximum value.

Two different restrained molecular dynamics procedures were used: (1) Restrained molecular dynamics. Molecular dynamics was performed at 500 K for 5 ps and 300 K for 20 ps with K(NOE) = 100 kcal $Å^{-2}$, starting from structures generated by DIANA and submitted to energy minimization. The structures generated were used for back calculation of the NOESY spectra and refinement of the distance constraints. (2) Simulated annealing. Molecular dynamics simulation was performed for 2 ps at 900 K and 2 ps at 300 K followed by minimization, starting from a structure obtained as described previously and using the refined set of NOE constraints. The obtained structure was saved and submitted to another cycle of the same procedure. Fifty cycles were thus performed, generating fifty different structures.

NOESY Back Calculation. Cross-relaxation rates between all proton pairs closer than ~5 Å were calculated using the program CORMA (Borgias et al., 1990), starting from a structure generated by restrained molecular dynamics. This program uses a full relaxation matrix approach. For comparison of experimental and back-calculated NOE intensities, R factors were used: $R = (I_o - I_c)/(I_o)$, where I_o = experimental intensity and I_c = calculated intensity. A

graphical representation of the calculated NOESY was generated using the program GIFA starting from the cross-relaxation rate table and from a chemical shift table and was compared to the experimental NOESY.

Stereochemical Quality Assessment of Peptide Structures. The stereochemical quality of the structures generated was checked using the program PROCHECK V2.1 (Oxford Molecular) (Morris et al., 1992, Laskowski et al., 1993).

RESULTS

Primary Structure Determination. Although the physical properties of RP 71955 show that it is a peptide, RP 71955 could not be sequenced classically, using Edman degradation. Large amounts of material deposited on the sequencer gave rise to weak and nonspecific signals, suggesting that there is no free terminal NH_2. Attempts to generate specific fragments by enzymatic digestion with trypsin, protease V8, or chymotrypsin were unsuccessful, even using a high enzyme-to-substrate ratio. Therefore, the determination of its amino acid composition and sequence had to rely on the results of an ensemble of spectroscopic techniques—FABMS, NMR, and Raman spectroscopy—supported by amino acid analysis and sequencing of peptide fragments obtained by limited chemical hydrolysis.

Mass Spectroscopy. By FAB mass spectrometry, the following ions were observed: $[M + H]^+$ m/z 2164.4; $[M + Na]^+$ m/z 2186.4. Therefore the experimental mass of RP 71955 is 2163.4 Da.

NMR Identification of Spin Systems. Inspection of the relayed signals arising from the amide resonances in a HOHAHA experiment revealed the presence of 21 amino acid spin systems. One L, one I, two Vs, two As, four Gs, and one S could be unambiguously identified, due to their unique coupling pattern (Basus, 1989). The remaining 10 residues showed AMX spin systems ($H\alpha - H\beta,\beta'$). One W, one Y, and two Fs were identified after inspection of the aromatic region of the spectrum. Their β protons were assigned using the large NOE effects with the closest ring protons. One N could also be identified from the observation of NOEs between β protons and δ NH_2. The last five residues remained to be identified.

Raman Spectroscopy. The Raman spectrum of RP 71955 showed the presence of two signals, at 497 and 514 cm^{-1}, characteristic of two different S–S bond stretching vibrations. This observation suggests that four of the unidentified amino acids are cysteines forming two disulfide bridges. The presence of four sulfurs in RP 71955 was confirmed by microanalysis.

Determination of the Molecular Formula. The remaining amino acid was identified as D by the difference between the molecular weight of RP 71955 determined by mass spectroscopy and the sum of the molecular weights of all other previously identified amino acids (taking into account that RP 71955 is cyclic and that the cysteines form disulfide bridges). The molecular formula of RP 71955 is therefore $C_{97}H_{131}O_{26}N_{23}S_4$.

Amino Acid Analysis. The amino acid analysis is in perfect agreement with these results. It does not, however, allow us to detect C and W and to discriminate between D and N. Using Marfey's reactant, it was possible to show that all detectable residues are L isomers.

Sequence Determination. As RP 71955 could not be sequenced classically, we attempted to determine its sequence by NMR spectroscopy, taking into account the fact that, for all secondary structure motifs, sequential $H\alpha_i - HN_{i+1}$ NOEs are observed (Wüthrich et al., 1984). In order to resolve some ambiguities due to spectral overlap, NOESY experiments were carried out at two different temperatures, 10 and 40 °C. As some Hαs showed Overhauser effects to several HNs, other NOEs, in particular $H\alpha - H\alpha$ and HN – HN, were considered in order to identify possible elements of secondary structure. For example, strong NOEs between the Hα of N and Y suggest that these two residues are facing each other on antiparallel β strands. Sequential HN – HN NOEs, characteristic of turns, helped to identify three stretches: XLG, SXN and AVV.

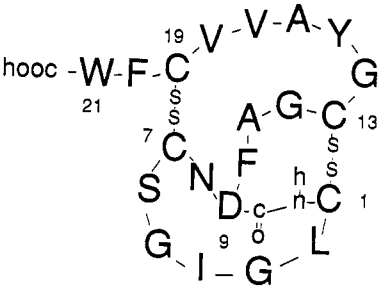
On the basis of this information, the following sequence was proposed :

X L G I G S X N X F A G X G Y A V V X F W
1 5 10 15 20

(X = C or D)

We know that four of the X residues are C and that one is D, that there is no free terminal NH_2, and that the two pairs of C are linked by disulfide bridges. As W21 did not show NOEs to any residue other than F20 and V18, it is probably very flexible and its COOH is not involved in a peptide bond. Therefore, the NH of X1 has to be linked to the γ -CO of D. Indeed, a strong NOE between HN of X1 and the β protons of X9 suggests that residue 9 is D. Therefore, residues 1, 7, 13, and 19 are C. At 40 °C, NOEs between Hα of C1 and the 2 Hβs of C13 and between the Hβ of C7 and the Hβ of C19 show that the disulfide bridges are linking C1 to C13 and C7 to C19 (at 10 °C the resonances of the β protons of C7 and C13 are overlapping).

This leads to the following structure:



Assignments of proton resonances of RP 71955 are listed in Table 1.

Nuclear Overhauser effects and coupling constants supporting the assignments are schematized in Figure 1. A representation of the solvent accessibility of amide protons, as determined from the variation of their chemical shifts with temperature, is also shown in Figure 1.

Furthermore, the sequence of peptides obtained after limited chemical hydrolysis and separation by HPLC of either native or reduced and carboxymethylated RP 71955 confirms the proposed structure.

Tertiary Structure Determination. (A) Structure Generation and Refinement. In a first step, using a limited set of distance constraints (120) determined using MARDIGRAS, one hundred 3D structures of RP 71955 were generated using DIANA. Stereospecific assignments were refined using the program GLOMSA. Correlations presenting systematic violations were examined individually for possible contributions of long-range interactions not taken into account at first. A list of protons having chemical shifts identical to those of the implicated protons (within a tolerance depending on the resolution of the experiment) was generated, using the chemical shift file. A list of proton pairs within a distance of less than

Table 1: Chemical Shifts (ppm) for RP 71955 at 10 and 40 °C

residue	T (°C)	HN	Hα[a]	Hβ[a]	other[b]
C1	10	8.83	4.60	3.49, 2.43	
	40	8.59	4.69	3.53, 2.49	
L2	10	9.77	4.42	1.48, 1.95	Hγ = 1.48, Mδ = 0.95, 0.78
	40	9.58	4.49	1.59, 2.08	Hγ = 1.59, Mδ = 0.94, 0.85
G3	10	8.98	4.07, 3.60		
	40	8.81	4.18, 3.63		
I4	10	6.98	4.07	1.49	Hγ1 = 0.77, 1.20, Mγ2 = 0.77, Mδ1 = 0.62
	40	6.92	4.18	1.59	Hγ1 = 0.88, 1.28, Mγ2 = 0.98, Mδ1 = 0.70
G5	10	7.15	3.49, 4.31		
	40	6.98	3.52, 4.42		
S6	10	7.90	4.54	3.84, 3.66	
	40	7.86	4.54	3.91, 3.79	
C7	10	8.12	4.84	3.31, 2.95	
	40	8.10	4.85	3.36, 3.09	
N8	10	8.96	5.13	2.33, 2.29	Hδ = 6.90, 7.23
	40	8.79	5.21	2.38, 2.36	
D9	10	7.97	4.31	2.66, 2.78	
	40	7.96	4.44	2.80, 2.82	
F10	10	8.51	4.19	2.78, 2.49	Hδ = 7.14, Hε = 7.35
	40	8.40	4.30	2.84, 2.63	Hδ = 7.16, Hε = 7.36
A11	10	8.81	3.78	1.13	
	40	8.56	3.95	1.18	
G12	10	8.78	3.96, 3.49		
	40	8.55	4.03, 3.52		
C13	10	8.32	4.19	2.95, 3.31	
	40	8.22	4.34	3.04, 3.36	
G14	10	7.40	3.02, 3.84		
	40	7.39	3.16, 3.87		
Y15	10	9.39	5.31	1.90, 2.66	Hδ = 6.70, Hε = 6.52
	40	9.34	5.41	2.04, 2.73	Hδ = 6.76, Hε = 6.56
A16	10	8.39	5.60	1.13	
	40	8.38	5.66	1.22	
V17	10	8.20	4.19	1.96	Mγ = 0.95, 0.78
	40	8.06	4.26	2.06	Mγ = 0.94, 0.85
V18	10	7.28	3.84	1.78	Mγ = 0.60, 0.72
	40	7.13	3.95	1.82	Mγ = 0.68, 0.75
C19	10	8.02	4.43	2.13, 2.31	
	40	7.82	4.46	2.21, 2.37	
F20	10	8.12	4.43	2.49, 2.72	Hδ = 6.89, Hε = 7.05
	40	8.04	4.46	2.61, 2.82	Hδ = 6.86, Hε = 7.06
W21	10	7.45	4.49	3.17, 3.02	Hδ1 = 6.90, Hε1 = 10.15, Hε3 = 7.49, Hζ2 = 7.09, Hζ3 = 6.90
	40	7.45	4.54	3.21, 3.06	Hδ1 = 6.97, Hε1 = 10.01, Hε3 = 7.29, Hζ2 = 7.25, Hζ3 = 6.71

[a] For diastereotopic protons, the first number is the chemical shift of the proton with the lower branch number (e.g., Hβ1, Hβ2). All stereospecifically assigned protons are underlined. [b] M = methyl.

5 Å was generated, starting from the rough structures determined previously. Proton pairs possibly contributing to the correlations could thus be identified, on the basis of their chemical shifts and spatial proximity. Additional correlations could also be assigned. Refined distance constraints were used to generate a new set of structures. The best structures were selected for energy minimization, followed by restrained MD calculations. The MD structure showing the best fit with the experimental data was used for back calculation of the NOESY spectrum and comparison with the experimental data. Optimization of the correlation time was performed by monitoring the R factor as a function of τ_c. A value of τ_c of 3 ns was found to give the best agreement between calculated and experimental values. The whole refinement procedure was repeated several times, until the agreement between experimental and calculated data could not be further improved.

During the last steps of the refinement process, the number of constraints was increased, using correlations involving overlapping proton pairs. Proton pairs contributing to the correlation were identified as previously described. The total intensity of the correlation was distributed among these different proton pairs proportionally to their cross-relaxation coefficients, determined by back calculation using CORMA. These coefficients were iteratively adjusted until the best possible agreement between observed and calculated data was obtained. As some correlations were observed on the calculated NOESY, but not on the experimental data, anti-distance constraints (adc) (Brüschweiler et al., 1991) were also cautiously introduced: indeed, wherever no cross peak could be observed at a position corresponding to the intersection of the chemical shifts of a proton pair, a lower distance limit of 4 Å was imposed (given the absence of t_1 noise or other experimental artifacts). The final set of distance constraints comprised 141 distances between isolated proton pairs, 69 distances extracted from the interpretation of overlapping proton pairs, and 1738 anti-distance constraints.

(B) Analysis of the Structures Generated Using DIANA. One hundred structures were generated using DIANA with adc only, dc only, or both dc and adc. Conformers with randomized dihedral angles were used for the initial input. In all cases, the presence of the internal amide bond and of the two disulfide bridges was fixed by imposing explicit distance limits between the CO of D9 and the HN of C1, as well as on the S–S and S–C distances of the cysteines. The values of the rmsd for structures generated with adc only (data not shown) indicate that despite the higher number of adc, compared to the number of dc, adc only do not provide a

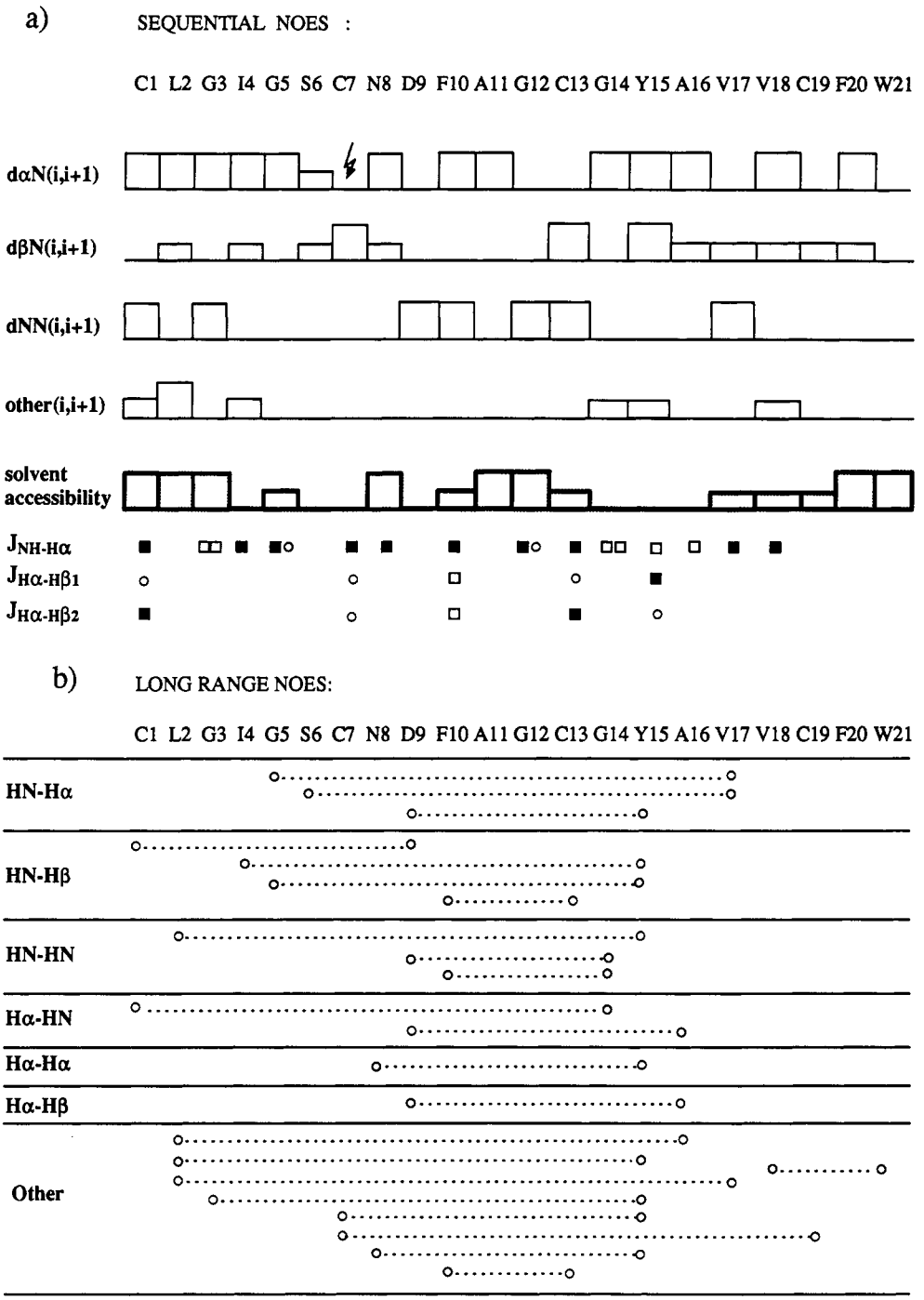


FIGURE 1: Diagrammatic representation qualitatively showing the sequential (a) and long-range (b) interresidue NOE connectivities along the sequence of RP 71955 (mixing time 100 msec; T = 10 °C). The thickness of the lines representing the sequential connectivities is proportional to the cross-peak intensity evaluated by volume integration. The solvent accessibility of amide protons, as determined from the variation of their chemical shift with temperature, is qualitatively represented as a histogram. High accessibility corresponds to $\Delta\delta/\Delta T$ values > 6; intermediate, to $\Delta\delta/\Delta T$ value between 6 and 3.5; low, to $\Delta\delta/\Delta T < 3.5$ ppm/°C . For coupling constants, a black square indicates that $J >$ 10 Hz; a white square, that $7 \leq J \leq 10$ Hz; a white circle, that $J < 7$ Hz. The arrow indicates Hα presaturated at this temperature.

significant restriction of the conformational space. Indeed, adc are much less stringent than dc, since basically they are only lower limit distances, the upper limit being restricted only by the size of the molecule. However, a comparison of the rmsd values for structures generated with dc only and with both dc and adc shows that the inclusion of the latter leads to a significant improvement in the definition of the structures (Figure 2).

All structures were energy minimized, and their stereochemical quality was assessed using the program PROCHECK. It could be noted that 36% of the conformers generated with dc only and 100% of the conformers generated with both dc and adc had ϕ and ψ angles in forbidden regions of the Ramachandran plot for either or all of the following residues: C7, F10, and A11. Figure 3a shows a stereoview of the backbone of the 74 acceptable structures generated with dc only.

A systematic analysis of these conformers shows that several regions of this peptide are fairly rigid: residues 1–5, 8, 9, and 13–18 have ϕ and ψ angle distributions in the same region of the conformational space for all conformers. In the case of residues 6 and 7, there is an equilibrium between two major conformations: the first with S6 in the α region and C7 in the β region, and the second with S6 in the β region and C7 in the ϵ region, of the Ramachandran plot (Laskowski et al.,

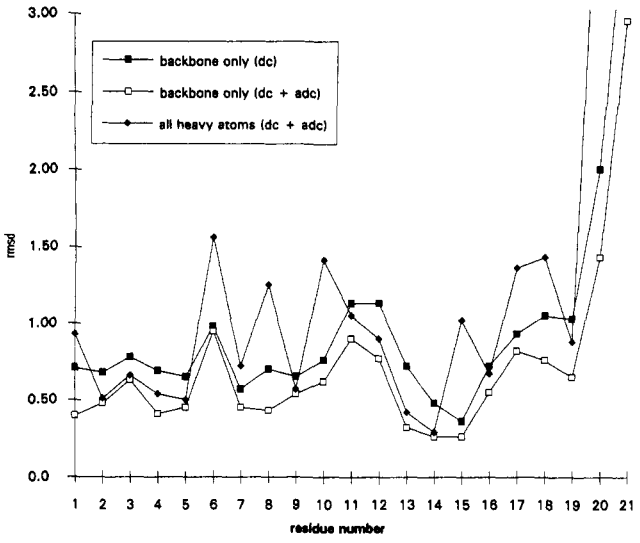


FIGURE 2: Average root mean square deviation (rmsd) in angstroms of the structures generated with DIANA: (■) backbone only (dc); (□) backbone only (dc + adc); (◆) all heavy atoms (dc + adc).

1993). The loop including residues F10, A11, G12, and C13 is also very flexible, and several families of structures could be defined in this region. In particular, the conformation of residue F10 goes from the α region to the β region, and that of A11, between left handed α (l) and ϵ. Finally, the conformation of the C-terminal residues C19, F20, and W21 is very flexible. Conformers having ϕ and ψ angles in forbidden regions of the conformational space are intermediate positions between two (or more) sterically allowed conformations.

Introducing adc at the beginning of the refinement process drives the conformational equilibrium toward these intermediate positions and produces conformers in sterically forbidden areas of the conformational space. A second set of calculations was thus performed using as initial input a set of conformers generated at earlier steps of the refinement process. These conformers were selected from among the ones with the smallest target function value. In this case, using both dc and adc, 31 acceptable structures were generated.

A stereoview of the main chain of these 31 conformers is shown in Figure 3b. In this case, the S6α–C7β ↔ S6β–C7ϵ equilibrium was still present but was shifted toward the former family (70–30%) with regard to calculations including only dc (30–70%). The F10–A11 conformation went from 57% to 97% F10β–A11l. Back calculation of the NOESY intensities was performed using CORMA for all acceptable conformers generated with dc only and with both dc and adc. R factors were calculated using only experimental intensities corresponding to dc. The average R factor was 0.46 for structures calculated with dc only and 0.45 for structures calculated with both dc and adc. Therefore, introducing adc does not lead to additional violations of dc. It does, however, allow one to obtain a better representation of the experimental results by elimination of the conformations giving rise to unobserved signals.

Selected structures, chosen according to their good target function values and good stereochemical quality, were energy minimized and used for MD simulation and back calculation of the NOESY. The conformer showing the best agreement with the experimental data (R = 0.39) belonged to the S6α–C7β, F10β–A11l family. A ribbon representation of this structure is shown in Figure 4.

Simulated Annealing. In order to get further insight into the mobility of the structure, a 50-cycle simulated annealing run was performed. The average rmsd per residue for these 50 structures is shown in Figure 5. As for structures generated with DIANA, the position of the C-terminal residues is not

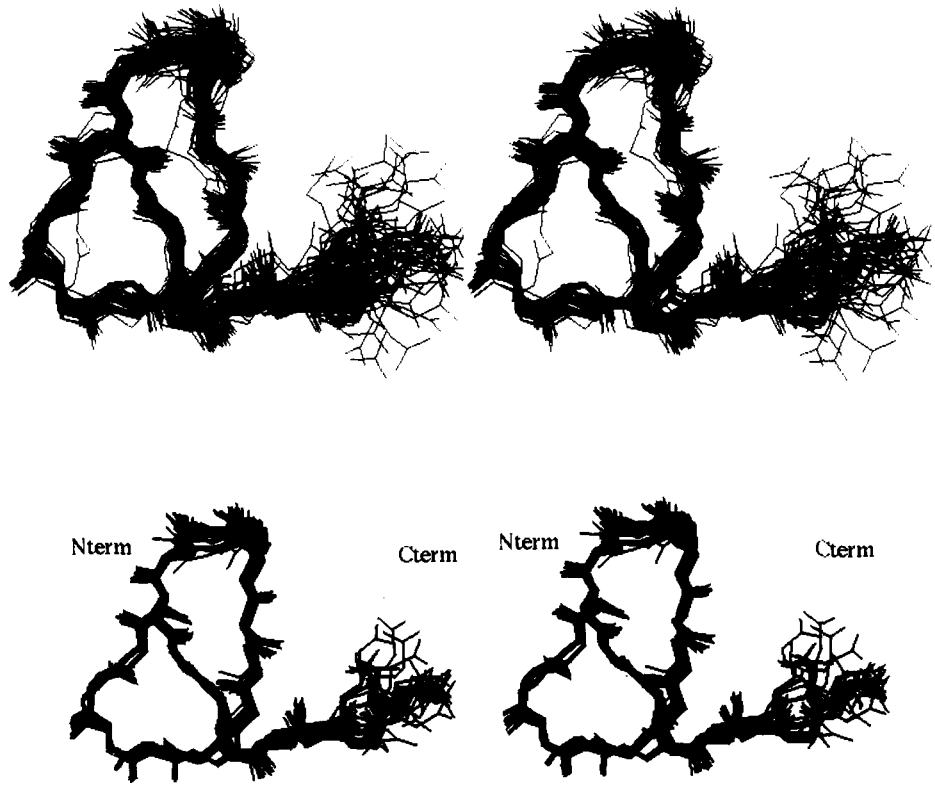


FIGURE 3: Stereo drawings of overlays of the peptide backbones of conformers generated with DIANA. The structures are positioned to optimize the match of the positions of the backbone atoms of residues 1–17: (a) 74 conformers generated using dc only; (b) 31 conformers generated using dc and adc.

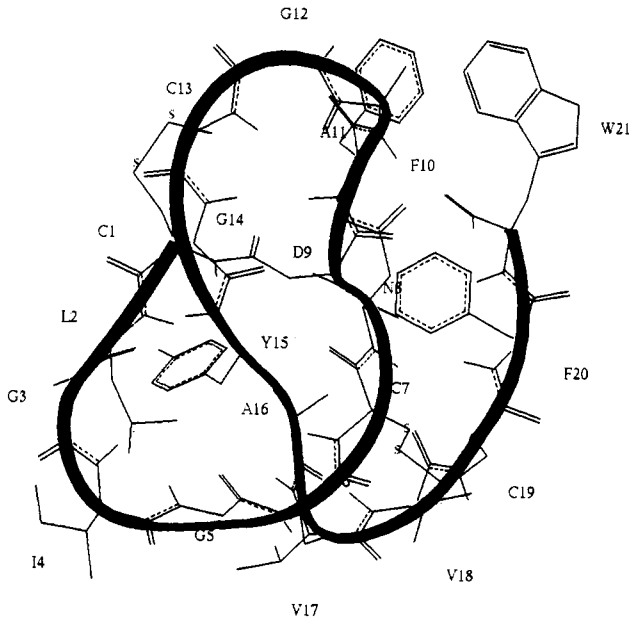


FIGURE 4: Ribbon representation of RP 71955. Shown is the best structure obtained after refinement using back calculation and restrained MD.

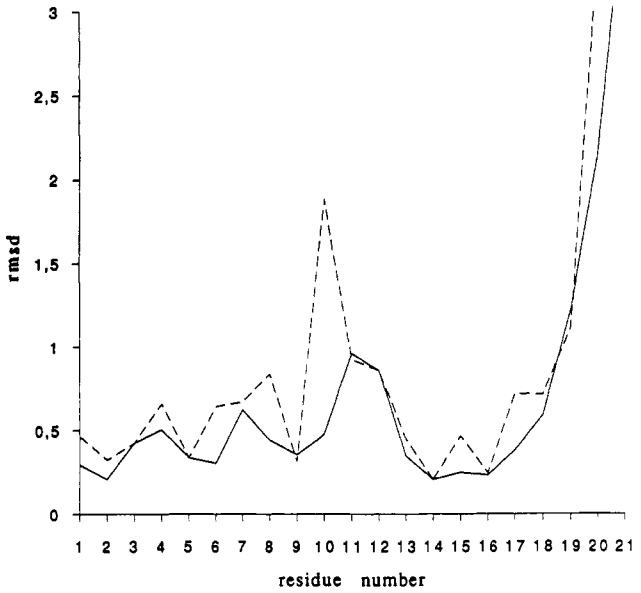


FIGURE 5: Average root mean square deviations (rmsd) in angstroms of the 50 structures generated by simulated annealing: (—) backbone heavy atoms; (– –) all heavy atoms.

very well defined. The loop between F10 and C13 and to a lesser extent residue C7 also have a higher than average rmsd.

Assessing the stereochemical quality of these structures using the program PROCHECK allowed us to identify 30 conformers having ϕ and ψ angles in forbidden regions of the Ramachandran plot, corresponding to intermediate positions between two stable conformations for either C7, F10, or A11.

Analysis of the 20 acceptable conformers (Figure 6) showed that the local conformation for residues in these flexible areas was different from that of conformers generated with DIANA. Indeed, the S6–C7 region showed 30% S6α–C7β and 70% S6β–C7β. The latter family of conformers was never observed in structures generated with DIANA. For F10 and A11, 70% of the structures belonged to the F10β–A11l family and 40% to the F10β–A11β family. The latter type of conformation was seldom observed in DIANA conformers. All residues located in rigid areas of the molecule belonged to the same family (Table 2) whether the conformers were generated with DIANA or simulated annealing. Slight differences observed between the local conformations generated by the two methods in mobile regions could be due to the higher conformational freedom allowed by the simulated annealing protocol.

An estimate of the side-chain mobility of the different amino acids can be obtained by inspection the χ_1 angle population values listed in Table 2. Residues having well-defined side-chain conformations are usually not exposed to the solvent, whereas mobile side chains are. There is a good correlation between the solvent accessibility, as evaluated from the $\Delta\delta/\Delta T$ of amide protons (Figure 1), and the side-chain mobility for all residues except L2 and S6. L2 had a high $\Delta\delta/\Delta T$ but is very rigid. Inspection of the structures shows that the amide proton of L2 is indeed at the surface of the molecule, but its side chain is not accessible to the solvent. The opposite situation is observed for S6.

The right-handed disulfide bridge conformation was observed for all conformers in the case of C7–C19 and for 55% of the conformers in the case of C1–C13. Back calculation of the NOESY intensities was performed using CORMA for the 20 structures having ϕ and ψ angles in allowed regions of the Ramachandran plot. The average R factor for these 20 structures was 0.40. This value is slightly lower than the one observed for structures generated with DIANA.

The energy of the structure was fairly high (between 462 and 500 kcal), given the large number of internal bonds. Nevertheless, the energy due to the constraints represented only 2–5% of the total energy (10–24 kcal). Upon minimization of the energy of the mean structure without constraints, the energy could be lowered to 298 kcal, while the main-chain rmsd between the minimized and the mean structure was only 0.73 Å for residues 1–19. An energy minimization and a 50 ps MD simulation at 300 K have been performed on the mean structure without constraints. The resulting structure has been energy minimized and compared to one of the structures obtained under constraints. The main-chain rmsd between the two structures was 1.42 Å.

DISCUSSION

Methodological Considerations. In this paper, we have determined both the primary structure and the solution conformation of RP 71955 using NMR methods. The solution conformation of this peptide was determined using DIANA and simulated annealing protocols using both dc and adc. The practice of using adc in NMR structure determination is somewhat controversial. Indeed, it has been argued that some cross peaks might be present but unobserved in the NOESY experiment (Sherman & Johnson, 1993). The nonobservation of NOEs might occur either because of the presence of experimental artifacts—t_1 noise, overlap with other cross peaks, proximity to the diagonal, broad resonances, and others—or because of the occurrence of internal motion. In this work, we have used adc only if no signals could be detected above noise level at the expected position for the cross peak, even for experiments performed at long mixing times. Moreover, adc were not used for broad resonances, since in this case the sensitivity threshold is lower. Therefore, the first argument can be ruled out. However, in the presence of a dynamic equilibrium between two (or more) conformational substates, minor conformers might not give rise to detectable signals and the observed cross peaks are averaged between the contributions of these different conformers. The use of adc, coupled with stereochemical analysis and NOE back calcu-

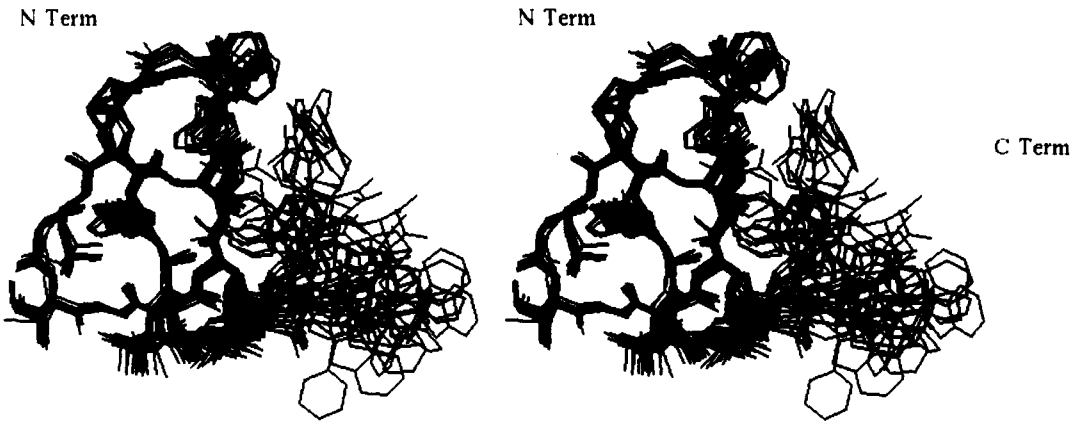


FIGURE 6: Stereo drawing of an overlay of 20 structures of RP 71955, generated by simulated annealing (all heavy atoms). The structures are positioned to optimize the match of the positions of the backbone atoms of residues 1–17.

Table 2: Percentages of Conformers Generated by Simulated Annealing in All Families of Conformers for the Backbone (Region of Ramachandran Plot) and the Side Chains (χ_1 Angle)[a]

residue	region of Ramachandran plot (%)	χ_1 angle (%)			χ_3 angle (%)	
		g^-	t	g^+	rt	lf
C1			80	20	55	45
L2	100 β	100				
G3						
I4	100 α					
G5						
S6	30 α/70 β	60	40			
C7	100 β	100				100
N8	100 β	5	85	10		
D9	100 α	100				
F10	100 β		65	35		
A11	70 l/30 β					
G12						
C13	100 α		100		55	45
G14						
Y15	100 β			100		
A16	100 β					
V17	100 α	10	75	15		
V18	100 β	5	95			
C19	100 β			100		100
F20	90 β/10 α	45	20	35		
W21		65	30	5		

[a] For cysteines the percentage of right-handed (rt) and left-handed (lf) helix is indicated.

lation, is a good tool for the identification of these flexible regions.

In this case, the use of adc at an early stage of the refinement process generates conformers which are metastable intermediates between two stable conformational substates. However, using adc at a later stage enables one to select, from among the conformers, the ones that describe more adequately the experimental results and that are therefore more likely to be major conformers. Indeed, if adc are not introduced, the visualization of a graphical representation of the calculated NOESY shows the presence of spurious peaks, not observed on the experimental NOESY. A more refined description of the local conformation in mobile regions would require taking into account the entire set of dynamic conformations, using approaches such as MEDUSA (Brüschweiler et al., 1991). However, this is beyond the scope of this work.

Hypotheses for the Structure–Function Relationship. The structure of RP 71955 is fairly unusual. Given the presence of an internal amide bond and of two disulfide bridges, it is very rigid. The linear peptide, chemically synthesized, is completely inactive. However, at present, we do not know if the lack of activity of this peptide is due to solubility problems or to a lack of tertiary structure.

RP 71955 is also very resistant to hydrolysis, thus making it a suitable candidate for oral absorption. Another striking feature of RP 71955 is its high hydrophobicity. Indeed, it is only very slightly soluble in water. Although the precise molecular target of RP 71955 is not yet determined, its hydrophobic character makes it a likely candidate for interaction with membranes.

Moreover, a 33% sequence identity was observed between RP 71955 and residues 608–628 of gp41 [amino acid numeration is given for the gp160 precursor; the sequence is from Andeweg et al. (1992)]:

```
          608
gp41      L L G I WG C S G K L I C T T A V P WN A
            | | |     |           |     | |
RP 71955  C L G I G S C N D F A G C G Y A V V C F W
          1
```

This sequence is part of the ectodomain of gp41 (residues 527–655). More precisely, it overlaps with several antigenic epitopes, which have been shown to be crucial for the activity of gp41: residues 608–620 (Gnann et al., 1987; Oldstone et al., 1991), 595–614 (Wang et al., 1986), 598–613 (Norrby et al., 1989), 581–614 (Sattentau & Moore, 1991), and 587–620 (Neurath et al., 1992).

It was suggested that these sequences might be of critical importance in the assembly of *env* protein oligomers (Earl et al., 1990; Doms et al., 1990). Oligomerization is a feature common to many viral membrane proteins, and it is essential for their function. Monoclonal antibodies raised against peptides derived from one of these antigenic regions showed weak or no reactivity with the intact gp41 protein. This observation is consistent with the fact that residues making up its epitope are not accessible in the assembled protein and thus might be located at or near the subunit interaction site (Norrby et al., 1989). Recent studies provided evidence that peptides having a sequence identity with residues 569–606 of gp41 (adjacent to the region having sequence identity with RP 71955) exhibit a coiled coil or leucine zipper type of structure in solution and are stabilized by self-association, suggesting that these residues are directly involved in the oligomerization of gp41 (Wild et al., 1992).

Other studies (Sattentau & Moore, 1991) have shown that monoclonal antibodies having epitopes corresponding to residues 581–614 (Mab 50–69, human) have increased binding to gp41 upon complexing of s-CD4 with gp120 at 4 °C, *i.e.*, without dissociation of the gp120–s-CD4 complex from gp41. It has been proposed that CD4 binding induces conformational changes in gp41, increasing the exposure of this antigenic

region, and that this event is an essential step in inducing fusion. It could, for example, involve binding to a membrane receptor other than CD4 (Qureshi et al., 1990).

This sequence could also be involved in the interaction between gp41 and gp120. Indeed, among synthetic peptides from gp41 tested, the peptide having a sequence corresponding to residues 587–620 bound most effectively to gp120 (Neurath et al., 1992).

The importance of the intramolecular disulfide bond between C residues in positions 614 and 620 for the recognition of some peptide antigens by sera from individuals infected with HIV (Gnann et al., 1987) and by monoclonal antibodies raised against HIV-1 gp41 immunodominant domain (Oldstone et al., 1991) has been emphasized. In RP 71955, the pairing is between C1 and C13 and between C7 and C19, and not between C7 and C13, the two C's showing sequence identity to gp41. However, examination of the 3D structure of RP 71955 shows that given a reorientation of their side chains, C7 and C13 could be paired, without major changes in the conformation of the main chain. The role of C residues, both in gp41 and in RP 71955, is probably mostly structural. The rigid structure maintained in RP 71955 by the two disulfide bridges and the internal amide bond could be conformationaly close to that of gp41.

Another interesting feature is the occurrence of the FXG motif (residues 536–538 in gp41) in one of the β turns of RP 71955 (residues 10–12). This particular sequence is present in the fusion domains of all paramyxoviruses and of HIV (Gallaher, 1987). In RP 71955, this sequence is mobile (local rmsd higher than average) and exposed to the solvent.

Although the fusogenic and immunodominant domains of gp41 are distant in the primary structure, they might be close in the folded protein. The examination of the 3D structure of RP 71955 shows that residues having sequence identity to the immunodominant domain are located on one face of the molecule, while the FXG motif is on the opposite face.

Preliminary results support the hypothesis, suggested by these sequence homologies, that RP 71955 interferes with the molecular function of gp41. Indeed, RP 71955 has been found to inhibit the formation of HIV-1-induced syncytia but not the interaction of gp120 with CD4 (A. Bousseau, personal communication). The formation of giant cells (syncytia) is considered to involve a mechanism analogous to cell fusion (Sattentau & Moore, 1991). It could be hampered by the inhibition of any of the steps mentioned above: either the oligomerization of gp41, its interaction with gp120 or with a putative membrane receptor, or the insertion of the fusion peptide in the membrane. The structure of RP 71955 could provide leads for the design of more potent drugs aimed at interfering with gp41 function. Blocking this step of the virus life cycle could provide an efficient way of preventing virus penetration in T4 lymphocytes, thus complementing anti-HIV therapeutic agents already available.

ACKNOWLEDGMENT

As a part of the BioAvenir project, this work has been partly supported by the use of the 600-MHz NMR spectrometer of Ecole Polytechnique, Palaiseau. We are indebted to J. Y. Lallemand for this fruitful collaboration.

REFERENCES

Andeweg, A. C., Groening, M., Leeflang, P., de Goede, R. E. Y., Osterhaus, A. D. M. E., Tersmette, M., & Bosch, M. L. (1992) *AIDS Res. Hum. Retroviruses 8*, 1803–1813.
Basus, V. (1989) *Methods Enzymol. 117B*, 132–149.
Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475–487.
Borgias, B. A., Gochin, M., Kerwood, D. J., & James, T. L. (1990) *Prog. Nucl. Magn. Reson. Spectrosc. 22*, 83–100.
Brüschweiler, R., Blackledge, M., & Ernst, R. R. (1991) *J. Biomol. NMR 1*, 3–11.
Demarco, A., Llinas, M., & Wüthrich, K. (1978) *Biopolymers 17*, 617–636.
Doms, R. W., Earl, P. L., Chakrabarti, S., & Moss, B. (1990) *J. Virol. 64*, 3537–3540.
Earl, P. L., Doms, R. W., & Moss, B. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 648–652.
Fletcher, R., & Reeves, C. M. (1962) *Comput. J. 7*, 149–154.
Gallaher, W. R. (1987) *Cell 50*, 327–328.
Gnann, J. W., Jr., Nelson, J. A., & Oldstone, M. B. A. (1987) *J. Virol. 61*, 2639–2641.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870–7872.
Güntert, P., & Wüthrich, K. (1991) *J. Biomol. NMR 1*, 447–456.
Güntert, P., Braun, W., & Wüthrich, K. (1991a) *J. Mol. Biol. 217*, 517–530.
Güntert, P., Quian, Y. Q., Otting, G., Müller, M., Gehring, W., & Wüthrich, K. (1991b) *J. Mol. Biol 217*, 531–541.
Helynck, G. (1992) French Patent ST 92041 (request no. 92-08145).
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.
Laskowski, R. A., MacArthur, M. W., Moss, D. S., & Thornton, J. M. (1993) *J. Appl. Crystallogr. 26*, 283–291.
Morris, A. L., Mac Arthur, M. W., Hutchinson, E. G., & Thornton, J. M. (1992) *Proteins 12*, 345–364.
Neurath, A. R., Strick, N., & Jiang, S. (1992) *Virology 188*, 1–13.
Norrby, E., Elliot-Parks, D., Utter, G., Houghten, R. A., & Lerner, R. A. (1989) *J. Immunol. 143*, 3602–3608.
Oldstone, M. B. A., Tishon, A., Lewicki, H., Dyson, H. J., Feher, V. A., Assa-Munt, N., & Wright, P. E. (1991) *J. Virol. 65*, 1727–1734.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741–751.
Qureshi, N. M., Coy, D. H., Garry, R. F., & Henderson, L. A. (1990) *AIDS 4*, 553–558.
Rance, M., Sorensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. P., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.
Sattentau, Q. J., & Moore, J. P. (1991) *J. Exp. Med. 174*, 407–415.
Sherman, S. A., & Johnson, M. E. (1993) *Prog. Biophys. Mol. Biol. 59*, 285–339.
Thomas, P. D., Basus, V. J., & James, T. L. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1237–1241.
Titman, J., & Keeler, J. (1990) *J. Magn. Reson. 89*, 640–646.
Van Gunsteren, W. F., & Karplus, M. (1980) *J. Comput. Chem. 1*, 266–274.
Wang, J. G., Steel, S., Wisniewolski, R., & Wang, C. W. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 6159–6163.
Wild, C., Oas, T., McDanal, C., Bolognesi, D., & Matthews, T. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 10537–10541.
Wüthrich, K., Billeter, M., & Braun, W. (1983) *J. Mol. Biol. 169*, 949–996.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715–740.