Siamycins I and II, New Anti-HIV-1 Peptides:

II. Sequence Analysis and Structure Determination of Siamycin I

D. J. DETLEFSEN*, S. E. HILL, K. J. VOLK, S. E. KLOHR, M. TSUNAKAWA, T. FURUMAI[b], P. F. LIN, M. NISHIO[†,a], K. KAWANO[†,a], T. OKI[†,b] and M. S. LEE[††]

Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 5100, Wallingford, CT 06492, U.S.A.
[†]Bristol-Myers Squibb Research Institute, Ltd., Tokyo Research Center
[††]Bristol-Myers Squibb Pharmaceutical Research Institute, P.O. Box 4000, Princeton, NJ 08543, U.S.A.

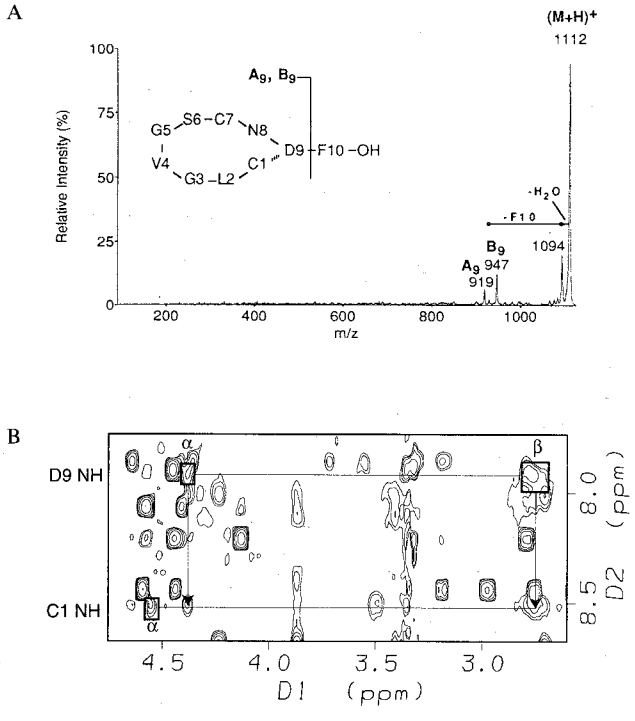
(Received for publication July 18, 1995)

We have previously reported on the preparation and initial characterization of the structure and biological activity of siamycins I and II[1]. Here, we report on the novel strategy used to rapidly determine the primary, secondary and preliminary solution structure of siamycin I using a multidisciplinary approach that utilized primarily MS and NMR data. The presence of three internal covalent cross-linkages found in siamycin I made structure analysis by traditional methods difficult, particularly in the *N*-terminal region.

Low resolution full scan and product ion mass spectra were acquired on a Sciex API *III* tandem quadrupole mass spectrometer equipped with an Ionspray (pneumatically assisted electrospray) interface. Samples were analyzed in the flow injection mode using a 0.1% trifluoroacetic acid in methanol mobile phase at a flow rate of 60 μl/minute. The ionspray tip was operated at +5300 V. Full scan mass spectra were obtained using an orifice voltage of 75 eV while scanning from m/z 800 to 2250 with a step size of 0.4 D and a dwell time of 1.5 msec. Product ion spectra were obtained from collision induced dissociation (CID) with an argon collision gas at 450×10^{12} atoms/cm^2 and a collision energy of 65 eV. NMR data were collected on a Varian Unity 500 spectrometer and processed on a Silicon Graphics Crimson workstation using Felix (Biosym). Spectra were collected on a 3.5 mM sample of siamycin I in 99% DMSO-d_6. Double quantum filtered COSY spectra (DQF-COSY)[2], 2D total correlation spectra (TOCSY)[3] and NOESY[4] two-dimensional ^{1}H NMR datasets were collected with 512 t_1 values and 2K complex t_2 data points using the hyercomplex method for frequency discrimination in $\omega 1$[5]. Crosspeak intensities were measured by counting contours on a 500 ms NOESY spectrum plotted with a contour level multiplier of 1.4. Upper bounds distances were quantified using an internal calibration based on dαN(i, i+1) and dαN(i, i+1) crosspeak intensities and possible distances for these protons[6]. Lower bounds distances were set to van der Waals contact. An ensemble of ten structures was computed based on the quantified NOE data using *DG-II* (Biosym). No additional energy minimization procedure was performed on the ensemble.

Complete acid hydrolysis of siamycin I indicated an amino acid content of: 2 Ala, 2 Val, 1 Ile, 4 Cys, 1 Ser, 4 Gly, 2 Phe, 1 Trp, 1 Leu, 1 Tyr and 2 Asx and the primary sequence of residues 10 through 21 were determined as previously described[1]. Secondary structural information on native siamycin I was obtained by co-analysis of MS and NMR data. Mass spectroscopic data on an isolated, purified proteolytic fragment that corresponded to the *N*-terminal region of siamycin I indicated a molecular weight of 1111 D for the proteolytic fragment and a fragment ion consistent with the *C*-terminal loss of Phe (Fig. 1A). The observed MW was consistent with the amino acid composition of the unassigned residues (2 Cys, 1 Val, 2 Gly, 1 Leu, 1 Ser,

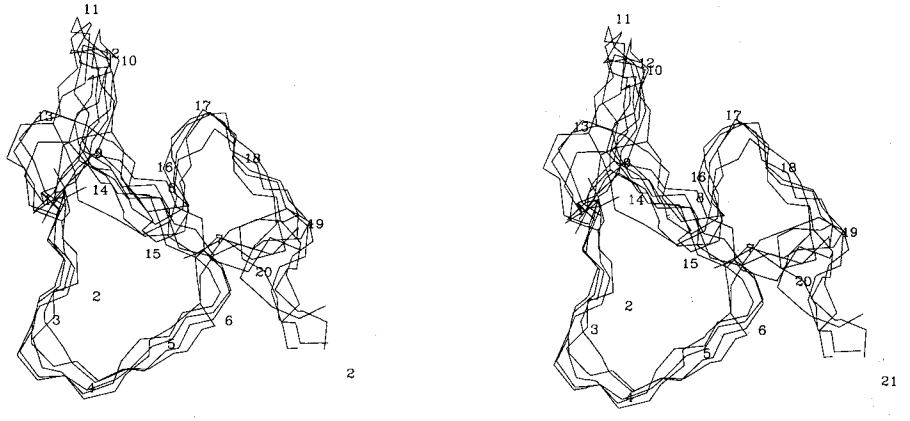
Fig. 1. Co-analyzed NMR and MS data for structure determination of *C*-terminal region of siamycin I.



A) MS/MS product ion spectrum of an *N*-terminal chymotryptic fragment of siamycin I. Fragment ions are noted on the spectrum are labeled according to ROEPSTORFF & FOHLMAN[9]. The proposed structure of the *N*-terminal cyclic region is shown as an inset. B) Expansion of the 500 ms NOESY spectrum of siamycin I. NOE crosspeaks in boxes correspond to intraresidue crosspeaks (see inset in part A). NOE crosspeaks that demonstrate the amide bond between are noted by arrows from the from Asp-9 H_α and H_β protons to the Cys-1 NH.

Present address:
[a] Tanabe Seiyaku Co., Ltd. 2-2-50 Kawagishi, Toda, Saitama 335, Japan.
[b] Biotechnology Research Center, Toyama Prefectural University, Kosugi, Toyama 939-03, Japan.

Fig. 2. Structural ensemble of 5 *DG-II* structures of siamycin I with only backbone atoms shown.
Structures were superimposed to the solution with the lowest residual violations to the experimental NMR data.



2 Asx) plus F10. The NOESY spectrum of intact siamycin I was re-examined with these data and a Leu-Gly-Val-Gly fragment was detected based on clear sequential connectivities between these unique spin systems (*N*-terminus contains LGVG sequence). The Leu residue had one AMX spin system on its *N*-terminus and the *C*-terminal Gly was sequentially connected to four AMX residues with the fourth connected to Phe-10 (*N*-terminus= X1 L2 G3 V4 G5 X6 X7 X8 X9 F10). X8 was identified as an Asn based on NOE crosspeaks between the H_β of this AMX spin system and two non-backbone amide protons (*N*-terminus=X1 L2 G3 V4 G5 X6 X7 N8 X9 F10). Based on the lack of MS/MS fragmentation beyond Phe-10 (stability of the product ion at *m/z* 919 in Fig. 1A) coupled with problems encountered with Edman degradation of this proteolytic fragment, cyclization was inferred at X9. A more detailed examination of the NOESY data on native siamycin I revealed NOE crosspeaks between the backbone amide proton of the X1 and the H_β protons of the X9 (Fig. 1B). This bond could not be a sequential linkage since L2 was already sequentially linked to X1 and G3 while the X9 was sequentially linked N8 and F10. These observations suggested the presence an amide bond between the backbone amide nitrogen of the X1 and the sidechain of X9 suggesting that the ninth residue was an aspartate to accommodate such a linkage. The remaining residues (3 AMXs) were resolved based on an analysis of the chemical shifts of the H_β protons (Ser H_β protons are shifted downfield with respect to other H_β protons)[7] resulting in the determination of the *N*-terminal 9 residues to Cys1 Leu2 Gly3 Val4 Gly5 Ser6 Cys7 Asn8 Asp9 with an amide linkage between C1 backbone nitrogen and D9 side-chain carboxyl.

Combining the previous analysis, additional secondary structural information in the form of disulfide linkages was suggested by the MS data on native siamycin I. A knowledge of the sequence and *N*-terminal cyclization, suggested the presence of two disulfide bonds based on the observed molecular weight of native siamycin I. Again, a re-examination of the NOESY data suggested the presence of a disulfide linkage between C1 and C13 based on NOEs between the H_α of Cys-1 and the H_β protons of Cys-13. The remaining disulfide bond was assigned to Cys-7 and Cys-19. These data, along with the sequential and other secondary structure information, were combined in structural calculations and an ensemble of ten structures was computed based on the quantified NOE data using *DG-II*. No violations in excess of 0.1 Å were observed in any of the structures. Fig. 2 shows a superimposition of five from the ensemble of 10 structures. The average RMSD of the backbone atoms for the five structures shown is 1.83 Å using the structure with the lowest residual error from the *DG-II* calculations as the fixed structure. The backbone fold shows no evidence of regular secondary structure but rather consists of a series of extended stretches and turns. However, siamycin I is highly structured in solution and is closely similar to another syncitia inhibitory peptide RP71955 having Ile-4 and Val-17[8].

Acknowledgments

The authors wish to thank SHIU-LOK HU for initial screening work on siamycin I as well as SADIA ABID and JINPING LIU for helpful discussions.

References

1) TSUNAKAWA, M.; S.-L. HU, Y. HOSHINO, D. J. DETLEFSEN, S. E. HILL, T. FURUMAI, R. WHITE, M. NISHIO, K. KAWANO, S. YAMAMOTO, Y. FUKAGAWA & T. OKI: Siamycins I and II, new anti-HIV peptides: I. Fermentation, isolation, biological activity and initial characterization. J. Antibiotics 48: 433~434, 1995
2) PIANTINI, U.; O. W. SØRENSEN & R. R. ERNST: Multiple quantum filters for elucidating NMR coupling networks. J. Am. Chem. Soc. 104: 6800~6801, 1982
3) BRAUNSCHWEILER, L. & R. R. ERNST: Coherence transfer by isotropic mixing: application to proton correlation

spectroscopy. J. Magn. Reson. 53: 521~528, 1983

4) KUMAR, A.; R. R. ERNST & K. WÜTHRICH: A two-dimensional nuclear overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. 95: 1~6, 1980

5) STATES, D. J.; R. A. HABERKORN & D. J. RUBEN: A two-dimensional nuclear overhauser experiment with pure absorption phase in four quadrants. J. Magn. Reson. 48: 286~492, 1982

6) DETLEFSEN, D. J.; V. THANABAL, V. L. PECORARO & G. WAGNER: Solution Structure of Fe(II) cytochrome *c*551 from *Pseudomonas aeruginosa* as determined by two-dimensional ^{1}H-NMR. Biochemistry 30: 9040~9046, 1991

7) WÜTHRICH, K.: "NMR of Proteins and Nucleic Acids", John Wiley & Sons, New York, 1986

8) FRECHET, D.; J. D. GUITTON, D. HERMAN, G. FAUCHER, G. HELYNCK, DU MONEGIER, B. SORBIER, J. P. RIDOUX, E. JAMES-SURCOUF & M. VUILHORGNE: Solution structure of RP 71955, a new 21 amino acid tricyclic peptide active against HIV-1 virus. Biochemistry 33: 42~50, 1994

9) ROEPSTORFF, P. & J. FOHLMAN: Proposal for a nomenclature for sequence ions in mass spectra of peptides. Biomed. Mass Spec. 11: 60, 1984