Use of Recombinant DNA Derived Human Relaxin To Probe the Structure of the Native Protein

Eleanor Canova-Davis,*,‡ T. Jeremy Kessler,‡ Paul-Jane Lee,‡ David T. W. Fei,‡ Patrick Griffin,§ John T. Stults,§ John D. Wade,∥ and Ernst Rinderknecht⊥

Medicinal and Analytical Chemistry, Protein Chemistry, and Process Recovery Departments, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, and Howard Florey Institute of Experimental Physiology and Medicine, University of Melbourne, Parkville, Victoria, Australia 3052

Received September 6, 1990; Revised Manuscript Received February 27, 1991

ABSTRACT: This report describes the physical, chemical, and biological characterization of recombinant human relaxin (rhRlx) used as a probe to establish the disulfide pairing in native human relaxin. This strategy is necessary since native human relaxin is only available in the nanogram range. The relaxin molecule is composed of two nonidentical peptide chains, an A-chain 24 amino acids in length and a B-chain of 29 amino acids, linked by two disulfide bridges with an additional disulfide linkage in the A-chain. Native relaxin isolated from human corpora lutea was compared to rhRlx by reversed-phase chromatography, partial sequence analysis, mass spectroscopy, and bioassay. The potency of rhRlx was established by its ability to stimulate cAMP from primary human uterine endometrial cells. Native relaxin isolated from human corpora lutea was equipotent to chemically synthesized relaxin, which in turn was equipotent to rhRlx. A tryptic map was developed for rhRlx to confirm the complete amino acid sequence and assignment of the disulfide bonds. The three disulfide bonds (Cys^{A10}–Cys^{A15}, Cys^{A11}–Cys^{B11}, and Cys^{A24}–Cys^{B23}) were assigned by mass spectrometric analysis of the tryptic peptides and by comparison to chemically synthesized peptides disulfide linked in the two most probable configurations. In addition, the observed amino acid composition and sequence of rhRlx was in agreement with that predicted from the cDNA sequence with the exception that the A-chain amino terminal was pyroglutamic acid. The migration of rhRlx upon sodium dodecyl sulfate–polyacrylamide gel electrophoresis was consistent with a monomeric structure, and the identity of the band was demonstrated by immunoblotting.

Hisaw's (1926) observation that a factor in the serum from pregnant guinea pigs or rabbits caused a relaxation of the pubic ligament when administered to virgin guinea pigs shortly after estrus led to the partial purification of relaxin from porcine corpora lutea (Fevold et al., 1930), but not until twenty years later was its ability to inhibit spontaneous contractions of the uterine myometrium in estrogen-primed guinea pigs (Krantz et al., 1950) and to promote cervical ripening in estrogen-primed cows (Graham & Dracy, 1953) reported.

Another 20 years elapsed before the elucidation of the primary structure of porcine relaxin revealed its homology to insulin by its strikingly similar number and distribution of half-cystine residues (James et al., 1977). Both the relaxin and insulin molecules contain two nonidentical peptide chains linked by two disulfide bridges with an intrachain disulfide linkage in the smaller A-chain. As with insulin, relaxin is expressed as a single-chain peptide precursor with the following overall structure: signal peptide/B-chain/C-peptide/A-chain (Hudson et al., 1983). Two human genes have been identified, only one of which, gene 2, is expressed (Hudson et al., 1984). The amino acid sequence for human relaxin was first deduced from the gene sequence due to the lack of available protein from human tissue. Subsequently, the processed protein was isolated from human corpora lutea, pregnancy serum (Winslow et al., 1989), and seminal fluid (Shih et al., 1989). Protein sequence analysis and mass spectrometry confirmed that the human relaxin protein is derived from gene 2 and consists of an A-chain of 24 residues and a B-chain of 29 residues. Synthetic relaxin produced by chemical synthesis of both the A- and B-chains with subsequent formation of interchain disulfide bonds was shown to be identical with that isolated from human sources (unpublished data).

Shortly thereafter, quantities of human relaxin were produced by use of recombinant DNA techniques (unpublished experiments). The separate A- and B-chains were produced in *Escherichia coli*, purified, and combined to form recombinant human relaxin (rhRlx)[1] in a manner similar to that previously described for a chemically synthesized human relaxin analogue (Canova-Davis et al., 1990). Hence, with use of this recombinant DNA derived molecule, it was now possible to characterize the hormone in detail, and especially to assign the disulfide bonds unequivocally. The analytical investigations undertaken to ascertain the homogeneity and disulfide-bonded structure of rhRlx and its identity to relaxin isolated from human corpora lutea are described in this report.

Materials and Methods

Amino Acid Composition. Analyses were performed on a Beckman 6300 amino acid analyzer connected to a Nelson Analytical M6000 data system (Moore et al., 1958). For

‡ Medicinal and Analytical Chemistry Department, Genentech, Inc.
§ Protein Chemistry Department, Genentech, Inc.
∥ University of Melbourne.
⊥ Process Recovery Department, Genentech, Inc.

[1] Abbreviations: Acm, acetamidomethyl; Boc, (*tert*-butyloxy)-carbonyl; Bop, (benzotriazol-1-yloxy)tris(dimethylamino)phosphonium hexafluorophosphate; *t*-Bu, *tert*-butyl; DMAP, 4-(dimethylamino)-pyridine; DTT, dithiothreitol; FAB-MS, fast atom bombardment mass spectrometry; Fmoc-OPfp, [(9-fluorenylmethyl)oxy]carbonyl-substituted pentafluorophenyl ester; HOBt, 1-hydroxybenzotriazole; HPLC, high-performance liquid chromatography; IEF, isoelectric focusing; Npys, 3-nitro-2-pyridinesulfenyl; pGlu, pyroglutamate; rhRlx, recombinant human relaxin; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; S-*t*-Bu, *tert*-butylthio; TFA, trifluoroacetic acid; Trt, triphenylmethyl.

0006-2960/91/0430-6006$02.50/0

cysteine determination, samples were treated with performic acid prior to a 24-h hydrolysis (Hirs, 1967). For tryptophan analysis, samples were hydrolyzed in 6 N HCl containing 7% thioglycolic acid (J. T. Baker) at 110 °C for 24 h (Matsubara & Sasaki, 1969).

Amino-Terminal Sequence Analysis. Recombinant hRlx samples were desalted on a Nucleosil C_{18} column (4.6 × 250 mm, 5-μm particle size, 300-Å pore size) and evacuated to dryness in a Savant Speed-Vac. The sample (11.8 μg) was reconstituted in 64 μL of 40 mM 4-ethylmorpholine (Aldrich) containing 11 mM DTT (Sigma) adjusted to pH 8.5 with hydrochloric acid. The sample was subsequently incubated with 0.23 unit of pyroglutamate aminopeptidase (Boehringer Mannheim Biochemicals) for 5 h at 25 °C. [One unit of enzyme is defined as that amount of enzyme causing an increase of absorbance at 410 nm of 0.001/min at 25 °C in 40 mM 4-ethylmorpholine, pH 8.5, with pyroglutamic acid-*p*-nitroanilide (Boehringer Mannheim Biochemicals) as the substrate in a final volume of 3 mL.] The enzyme-treated sample was then subjected directly to Edman degradation on an Applied Biosystems 477A protein sequencer with on-line phenylthiohydantoin detection with an Applied Biosystems 120A analyzer. Quantitation was performed by the Applied Biosystems 900A data system using peak height comparison to an external standard.

Carboxy-Terminal Sequence Analysis. A sample of rhRlx was dissolved in 4 M guanidine and 0.25 M Tris·HCl, pH 9.2. The solution was brought to 20 mM DTT, incubated at 37 °C for 30 min, and then brought to 50 mM iodoacetic acid (Kodak) and left in the dark at room temperature for 1 h. The carboxymethylated rhRlx was dialyzed against 5% acetic acid for 24 h, dried under vacuum, and then reconstituted in 0.2 M pyridine–acetate, pH 5.5. After removal of 25 μL for a zero-time point, carboxypeptidase Y (Worthington Biochemicals) was added to bring the enzyme to substrate ratio to 1:50 by weight. The digest was analyzed according to the procedure of Hayashi et al. (1973).

Isoelectric Focusing (IEF). A precast LKB PAG plate (pH range 3.5–9.5) was loaded with samples (20 μL) and p*I* standards (Bio-Rad). Focusing was conducted at constant power up to a maximum voltage of 1500 V. The gel was stained according to Holbrook and Leaver (1976) using Coomassie Brilliant Blue G-250.

Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS–PAGE). Samples of human relaxin (with and without 5% β-mercaptoethanol) and Electran molecular weight standards (BDH Limited) were loaded onto a 16% Tricine SDS–PAGE gel (Novex EC1695, 1.0 mm). Electrophoresis was carried out by using the Tricine system of Schägger and von Jagow (1987), and the gel was stained with Coomassie Brilliant Blue R-250.

Electrotransfer and Immunoblotting. Immunoblotting was carried out by a modification of the procedure of Burnette (1981). After SDS–PAGE of the rhRlx and chemically synthesized relaxin samples, electrotransfer to nitrocellulose was effected in a Novex Western transfer apparatus in carbonate transfer buffer (Novex LC3677) at 25 V for 30 min. The nitrocellulose paper was then immersed in 50 mM Tris, pH 7.4, 150 mM NaCl, 4.3 mM EDTA, 0.05% Triton X-100, and 0.25% gelatin buffer to saturate additional binding sites. The nitrocellulose was probed with rabbit anti-relaxin produced by using chemically synthesized human relaxin as the immunogen (Genentech) and subsequently with the indicator antibody, goat anti-rabbit IgG conjugated to horseradish peroxidase. Immunoreactive bands were detected by reaction with diaminobenzidine (Adams, 1981; Anicetti et al., 1986) in the presence of hydrogen peroxide (Aldrich).

Ion-Exchange Chromatography. Cation-exchange chromatography was performed utilizing a Hewlett-Packard liquid chromatograph equipped with a TSK CM-2-SW (4.6 × 250 mm) ion-exchange column. The column was equilibrated in 0.1 M ammonium acetate, pH 6.0. Recombinant hRlx or chemically synthesized human relaxin analogues (15 μg) were loaded, and elution was performed with a linear gradient of sodium acetate in 0.1 M ammonium acetate from 0 to 0.4 M, pH 6.0, in 40 min. The flow rate was held constant at 0.6 mL/min; the column temperature was controlled at 35 °C; and the absorbance was monitored at 280 nm.

Reversed-Phase High-Performance Liquid Chromatography. This analysis was performed by using a reversed-phase column (Vydac C_{18}, 4.6 × 250 mm, 5 μm, 300 Å). Solvent A was 0.1% aqueous trifluoroacetic acid (TFA). Solvent B was 0.085% TFA in 85% acetonitrile (Burdick and Jackson). The column was equilibrated with 65% solvent A and 35% solvent B. The rhRlx sample (5 μg) was loaded, and solvent B was held at 35% for 10 min and then followed by a linear gradient to 38% solvent B in 20 min. The flow rate was 1 mL/min, the column temperature was controlled at 40 °C, and the absorbance was monitored at 214 nm. Native hRlx (100 ng) was analyzed on a microbore reversed-phase column (Synchrom C_4, 1 × 100 mm). The hormone was eluted by the application of a linear gradient of acetonitrile in 0.1% TFA from 15% (held for 15 min) to 60% in 45 min at 100 μL/min.

Bioassay. Relaxin ranging from 0.39 to 25 ng/mL was incubated for 30 min with human uterine endometrial cells (37 500 cells/cm^2 growth area) in the presence of 1 μM forskolin and 50 μM isobutylmethylxanthine. After the used medium was discarded, cellular cyclic AMP was extracted with 0.1 N HCl and measured by radioimmunoassay (Fei et al., 1990). Sample concentrations were determined from the standard curve which was fitted by a nonlinear least-squares (four-parameter fit) program.

Trypsin Digestion. A recombinant hRlx sample (375 μg) was dissolved in 750 μL of 10 mM Tris, pH 7.2, in 10 mM sodium acetate and 2 mM $CaCl_2$. Digestion was conducted with L-1-(tosylamino)-2-phenylethyl chloromethyl ketone treated trypsin (Cooper Biomedical) at 37 °C for 4 h. An aliquot of 1:100 enzyme/substrate by weight was added at zero time and another after 2 h of digestion. The reaction was terminated with 5 μL of 2 N HCl. The resulting peptide mixture was separated by reversed-phase HPLC and monitored at 214 nm. The column was packed with Nucleosil C_{18} resin (4.6 × 150 mm, 5 μm, 300 Å). Elution was effected with a linear gradient from 0% to 50% acetonitrile containing 0.1% TFA over 100 min. The flow rate was 1 mL/min, and the column temperature was controlled at 35 °C. The peptide peaks collected manually from the HPLC system were characterized by amino acid analysis after acid hydrolysis, amino-terminal sequencing by Edman degradation, and FAB-MS as described below.

Fast Atom Bombardment Mass Spectrometry of Tryptic Peptides. The peptides were dissolved in 1 μL of 0.1% TFA/30% acetonitrile, transferred to the mass spectrometer probe, dried, and mixed with 1 μL of the liquid matrix, 1-thioglycerol. The spectra were obtained on a JEOL JMS-HX110HF/HX110HF tandem mass spectrometer equipped with a cesium ion source. The cesium ion primary beam energy was 15 keV. The secondary ions were analyzed with an acceleration voltage of 10 kV and a resolution of 1:3000. Data were acquired over a mass range of 100–4000 daltons

using the ACM program of the JEOL DA-5000 data system.

Collisionally activated dissociation daughter spectra were also obtained with the JEOL JMS-HX110HF/HX110HF. Precursor ions, formed by FAB of 0.5–1.0 nmol of peptide, were selected by MS-1 at a resolution of 1:1000. The ^{12}C isotope peak was selected to ensure that the daughter ions were isotopically pure. The selected ions leaving MS-1 were fragmented with helium in a grounded collision cell. The helium pressure was sufficient to reduce the precursor intensity to 25% of its original value. The daughter ions were measured with MS-2, using a linked scan at constant B/E produced by the ACM program, with a resolution of 1:1000.

Solid-Phase Syntheses of T_2-T_7 Disulfide Analogues. (*A*) *T_2 Peptide.* Pepsyn K (0.2 mmol/g) was derivatized by attaching a reference amino acid (Ala) and the acid-labile handle (Atherton & Sheppard, 1989). The carboxy terminal of the A-chain peptide, Lys, was coupled to this handle by using a double 1.5-h coupling of the Fmoc-OPfp active ester (5-fold molar excess) in the presence of 0.1 equiv of the catalyst DMAP. The remaining amino acids were coupled by using a 3-fold molar excess of their respective Fmoc-OPfp active esters. The exceptions, Thr and S-*t*-Bu-Cys, were coupled after Bop activation in the presence of HOBt. Half-hour coupling times were employed and Fmoc deprotection was with 20% piperidine in dimethylformamide. Side-chain protection was as follows: Lys, Boc; Thr, *t*-Bu; His, Trt; Cys[15], Acm; and Cys[10,11], *S*-*t*-Bu or Acm. Prior to the attachment of Cys[11], the deprotected peptide–resin was divided into two equal parts. To allow each Cys residue to be selectively deblocked, the order of attachment of S-*t*-Bu-Cys and Acm-Cys was reversed in the two aliquots. The peptides were cleaved from their respective resins with 95% TFA/5% phenol. The crude peptides were diluted with water and lyophilized. Before purification by reversed-phase HPLC in aqueous TFA-containing acetonitrile gradients, the peptides were reduced by treatment with a 10-fold molar excess of β-mercaptoethanol in 0.5 M Tris, pH 8.0, and 6 M urea for 2 h at room temperature, which removed the S-*t*-Bu group, resulting in a free thiol at the corresponding Cys residue.

(*B*) *T_7 Peptide.* The synthesis was carried out automatically on a 0.5-mmol scale using an ABI 430A synthesizer employing Boc-polystyrene methodology. In this case, the Cys protecting group was Npys. The peptide was cleaved from the resin by treatment with 85% hydrogen fluoride/15% *m*-cresol for 1 h at 0 °C. The crude peptide was then purified by reversed-phase HPLC in a manner similar to that applied to the T_2 peptide.

(*C*) *Combination of T_2 Peptide to the T_7 Peptide.* The respective T_2 peptides and the T_7 peptide was dissolved in 0.1 M potassium phosphate buffer, pH 4.45, separately. The T_7 peptide was added to each of the T_2 peptide containing solutions at a 3:1 ratio, respectively. The mixtures were vortexed and allowed to stand at room temperature for 2.5 h before purification by reversed-phase HPLC. The T_2-T_7 peptides were dissolved in 80% aqueous acetic acid. A solution of iodine was added dropwise with vigorous stirring to effect the intrachain disulfide bond formation. After 2 h the solution was diluted with water, extracted with chloroform, and purified by reversed-phase HPLC.

RESULTS

Characterization of rhRlx. The amino acid composition of a rhRlx preparation is shown in Table I. The residues were calculated on the basis of an assumption of complete recovery of alanine and leucine. These results demonstrate excellent correlation with the theoretical values (Hudson et al., 1984).

Table I: Amino Acid Composition of Recombinant Human Relaxin

amino acid	residues/mol	amino acid	residues/mol
CyA (6)[a]	5.95[b]	Met (2)	1.80
Asx (2)	2.01	Ile (3)	2.75[d]
Thr (2)	2.04[c]	Leu (5)	5.03
Ser (5)	5.13[c]	Tyr (1)	1.01
Glx (5)	5.18	Phe (1)	1.00
Pro + Cys-SH (0)	0[b]	His (1)	1.00
Gly (3)	3.08	Lys (3)	2.88
Ala (5)	4.97	Trp (2)	1.90[e]
$^1/_2$-Cys	0[b]	Arg (4)	3.82
Val (3)	3.04[d]		

[a] Theoretical values are in parentheses. [b] Performic acid oxidized sample results. [c] Extrapolation to zero time of hydrolysis. [d] After a 50-h hydrolysis. [e] Determined in the presence of thioglycolic acid.

Table II: Amino-Terminal Sequence Analysis of Recombinant Human Relaxin and Relaxin Isolated from Human Corpora Lutea

	A-chain[a]		B-chain		
				yield (pmol)	
cycle no.	sequence	yield (pmol) recombinant	sequence	recombinant	native
1	Leu	2143.8	Asp	1038.5	
2	Tyr	2259.1	Ser	836.8	
3	Ser	1160.1	Trp	936.9	0.3
4	Ala	2390.9	Met	1466.9	0.5
5	Leu	1870.9	Glu	672.5	0.8
6	Ala	2138.3	Glu	654.1	1.2
7	Asn	1333.6	Val	846.1	0.4
8	Lys	1558.1	Ile	731.1	0.3
9	Cys[b]	368.0	Lys	794.2	0.5
10	Cys[b]	293.2	Leu	678.3	0.3
11	His	522.5	Cys[b]	232.6	
12	Val	814.2	Gly	532.2	0.4
13	Gly	838.7	Arg	619.9	0.5
14	Cys[b]	137.8	Glu	184.0	0.3
15	Thr	424.2	Leu	418.0	0.4
16	Lys	431.7	Val	360.0	
17	Arg	933.7	Arg	933.7	
18	Ser	125.6	Ala	352.7	
19	Leu	309.8	Gln	257.3	
20	Ala	316.1	Ile	234.8	
21	Arg	217.1	Ala	352.9	
22	Phe	183.4	Ile	213.7	
23	Cys[b]	61.9	Cys[b]	61.9	
24			Gly	150.6	
25			Met	119.1	
26			Ser	60.2	
27			Thr	55.8	
28			Trp	8.2	
29			Ser	24.6	

[a] Sequence after pyroglutamate aminopeptidase treatment. [b] Identified as the DTT adduct of dehydroalanine.

Recombinant hRlx was analyzed by Edman degradation to determine the amino acid sequence at the amino terminal of the two chains present in the molecule. The DNA sequence for the A-chain codes for a glutamine amino terminus. Acid treatment converts this residue to pyroglutamic acid. Hence, sequence data could be obtained only after treatment of the molecule with pyroglutamate aminopeptidase. The enzymatic reaction was followed by reversed-phase HPLC (data not shown) of an aliquot reduced by DTT. The A-chain fractions were further analyzed by FAB-MS. The expected molecular ion of 2656.9 atomic mass units (u) was obtained for the intact A-chain. An earlier eluting peak was identified as the des-pGlu peptide (2546.5 u). The expected entire sequences for the two chains were found (Table II). Since DTT was required to preserve the activity of pyroglutamate aminopeptidase (Podell & Abraham, 1978), the relaxin molecule was partially reduced to its component A- and B-chains. Assignment of the disulfide linkages from the sequence data is therefore impossible. Another preparation was sequenced

without prior treatment with pyroglutamate aminopeptidase (data not shown). No A-chain sequence was detectable, supporting the contention that the amino terminal of the A-chain was indeed blocked.

Digestion of reduced and alkylated rhRlx with carboxypeptidase Y resulted in the release of amino acids which were consistent with the carboxy termini as shown in Table II. After 8 h of digestion the following ratios of amino acids were observed: carboxymethyl-S-Cys, 1.11; Phe, 0.62; Ser, 1.45; Trp, 0.42; Thr, 0.42; Met, 0.47; Gly, 0.47. This procedure was necessary since the carboxy terminal of the A-chain is a disulfide-bonded cysteine residue. The trypsin digestion data (discussed below) support the conclusion from the carboxypeptidase Y analyses.

Four different preparations of rhRlx were analyzed by isoelectric focusing as described under Materials and Methods. This procedure was effective in focusing basic proteins (data not shown). A single band was visible at an estimated p*I* of 9.1 on the basis of standards run on the same gel. Hence, no evidence for deamidation was observed. The theoretical p*I* is 9.8 as calculated by the method of Shire (1983). Briefly, this calculation is made by assuming intrinsic p*K* values for all titratable groups, computing their fractional dissociation at pH values from 2 to 12 at intervals of 0.1, and using these values to calculate the average charge at each pH by subtracting the contribution of all titratable groups from the maximum protein charge. The theoretical p*I* is obtained by determining the pH for which the average charge is zero. This method assumes no interactions among titratable groups.

An SDS–PAGE analysis of relaxin is difficult due to its low molecular weight. The problems are primarily lack of resolution and diffusion out of the gel. On Tricine SDS–PAGE, rhRlx migrates as a monomer with an apparent molecular weight of 6000 (data not shown). This is an appropriate migration position, given the theoretical molecular weight of 5963. The monomer band is the only band detectable by Coomassie staining of a 2-μg load of rhRlx. Samples that are reduced prior to electrophoresis migrate as a broad band in the 2500–3000 molecular weight range. This is due to the inability of this gel system to resolve the M_r 2656 A-chain from the M_r 3313 B-chain. A chemically synthesized analogue of human relaxin (Canova-Davis et al., 1990) was used as a standard to probe the performance of this gel system. This human relaxin analogue, which contains the sequence Lys-Arg-Ser-Leu as an extension of the C-terminal of the B-chain, migrates with an apparent molecular weight that is approximately 500 greater than rhRlx. Upon reduction before electrophoresis this B-chain analogue of M_r 3797 is resolved from the M_r 2656 A-chain. The identity of the 6000 molecular weight band was confirmed to be relaxin by immunoblotting with antibody directed against chemically synthesized human relaxin.

A cation-exchange column was chosen to analyze relaxin since it is a basic protein (Figure 1). Two chemically synthesized analogues of human relaxin were used as probes to demonstrate that the system described under Materials and Methods was indeed capable of resolving differentially charged variants. The chemically synthesized analogues required higher salt concentrations for elution from the column as expected. The analogue containing a longer B-chain as described in the SDS–PAGE section contains two more positive charges than rhRlx. The second analogue contains a lysine for methionine substitution at position 4 and an alanine substitution for the methionine at position 25 in addition to the longer B-chain for a total of three more positive charges than

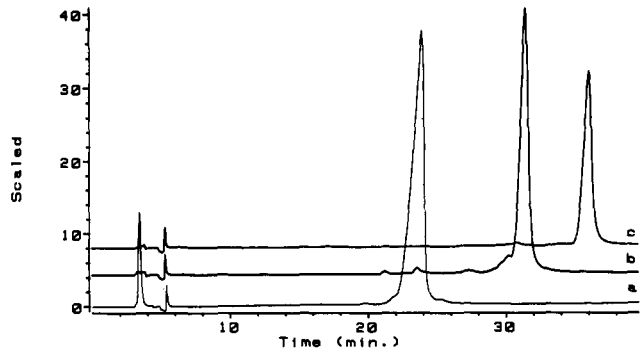

FIGURE 1: Cation-exchange chromatography of a human relaxin preparation and chemically synthesized relaxin analogues. The chromatography was performed on a TSK CM-2-SW column and eluted with a sodium acetate gradient as described under Materials and Methods. (a) Recombinant human relaxin; (b) chemically synthesized human B ($Lys^{30}Arg^{31}Ser^{32}Leu^{33}$) relaxin; and (c) chemically synthesized human B ($Lys^{4}Ala^{25}Lys^{30}Arg^{31}Ser^{32}Leu^{33}$) relaxin.

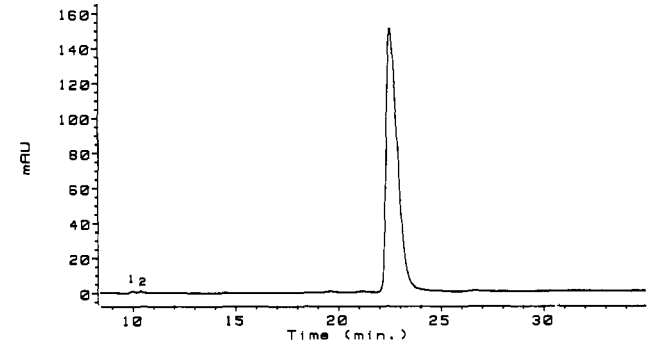

FIGURE 2: Reversed-phase HPLC of recombinant human relaxin. Chromatography was performed on a Vydac C_{18} column and eluted with an acetonitrile gradient as described under Materials and Methods.

rhRlx. The rhRlx preparation contains greater than 99% of the total integrated area in the main peak again, suggesting that no deamidated species are present.

Gradient conditions were utilized for the reversed-phase chromatography as described under Materials and Methods to improve the resolution of minor components of slightly greater hydrophilicity. The analysis of a rhRlx preparation by reversed-phase HPLC is shown in Figure 2. The profile indicates that 98% of the total integrated area is in the main peak. Analyses of peaks 1 and 2 by FAB-MS identified these variants as containing B-chains of mass 3345 and 3329 (i.e., 3313 + 32 and 3313 + 16), respectively, suggesting that oxidation, possibly at the methionine residues, was responsible for generating these small peaks. In fact, this was observed for native human relaxin isolated from corpora lutea also (see below).

Relaxin stimulates the production of cyclic AMP in the mouse pubic symphysis (Braddon, 1978), rat uterus (Sanborn et al., 1980), rat uterine tissue slices (Judson et al., 1980), rat myometrial cells (Hsu et al., 1985), rat anterior pituitary gland (Cronin et al., 1987), and newborn Rhesus monkey uterine cells (Kramer et al., 1990). Hence, a bioassay for relaxin was developed by using primary human uterine endometrial cells (Fei et al., 1990). Recombinant human relaxin produced a dose-dependent stimulation of cyclic AMP accumulation in these cells with an ED_{50} of 3.7 ± 1 ng/mL (n = 49); chemically synthesized human relaxin had an ED_{50} of 4.4 ± 0.7 (n = 5). This represents the first demonstration that recombinant human relaxin is biologically active in a human cell line. Neither rhRlx A-chain, rhRlx B-chain, or A-chain mixed with

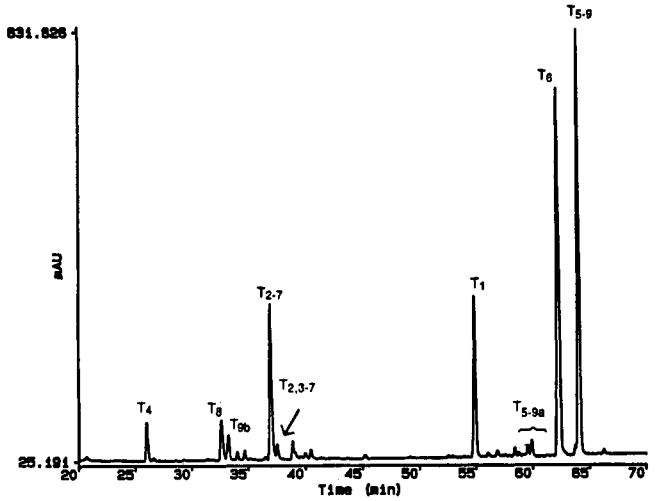

FIGURE 3: Tryptic map of a recombinant human relaxin preparation. Chromatography was performed on a Nucleosil C_{18} column and eluted with an acetonitrile gradient as described under Materials and Methods.

B-chain at a concentration of 1 μg/mL elicits a cyclic AMP response in this bioassay (Fei et al., 1990). Of 18 other peptides and steroids tested (including insulin, insulin-like growth factor I, oxytocin, ACTH, hydrocortisone, and progesterone) none significantly elevated cyclic AMP concentrations in these cells (Fei et al., 1990). Preincubation of chemically synthesized human relaxin with neutralizing monoclonal antibody to chemically synthesized relaxin analogues completely abolished its ability to stimulate cyclic AMP production (Fei et al., 1990). The data indicate that this bioassay is specific to relaxin and can be used to measure the biological activity of relaxin.

Trypsin Digestion and Disulfide Bond Assignments. Digestion of the A-chain of human relaxin with trypsin can theoretically result in the release of five fragments (T_1, A1–9; T_2, A10–17; T_3, A18; T_4, A19–22; T_5, A23–24) while four fragments are released from the digestion of the B-chain (T_6, B1–9; T_7, B10–13; T_8, B14–17; T_9, B18–29). By analogy with the disulfide pairing established for insulin (Ryle et al., 1955), peptide T_2 would be expected to be covalently bonded to T_7 and peptide T_5 to T_9 in human relaxin. Since it had been shown that trypsin digestion at pH 8.2 of the chemically synthesized relaxin analogue containing a B-chain 33 amino acids in length results in scrambling of the disulfide-containing peptides (Canova-Davis et al., 1990), the digestion was conducted at pH 7.2. Also, the method of Ellman (1959) was used to analyze intact rhRlx for free sulfhydryl groups which could catalyze a disulfide exchange reaction. No free sulfhydryl groups were detected. A typical tryptic map of rhRlx is shown in Figure 3. The peptide assignments were made after analyses of the peaks by acid hydrolysis for amino acid composition, amino-terminal sequence analysis (data not shown), and FAB-MS (Table III). A non-tryptic-like cleavage occurred in the T_9 peptide. These peptides were identified as T_{9a}, B18–25, and T_{9b}, B26–29.

The disulfide assignments of T_2-T_7 and T_5-T_9 are confirmed by the tryptic map data. This unequivocally establishes Cys^{A24}–Cys^{B23}. The positions of the two disulfides in T_2-T_7 required further analysis. The three possible arrangements for the two disulfide bridges in this tryptic fragment are illustrated in Figure 4. The peptide structures 1 and 2 were synthesized, and the reversed-phase HPLC profiles of these chemically synthesized peptides were examined. The peptides arranged as shown in structure 1 had an identical retention

Table III: Mass Spectral Analysis of Tryptic Peptides from rhRlx

assignment	theoretical masses[a] (u)	observed (u)
T_4	446.27	446.0
T_8	516.31	516.0
T_{9b}	480.21	479.9
T_2-T_7	1293.57	1293.6
		1295.6[b]
$T_{2,3}$-T_7	1449.67	1449.8
T_1	990.55	990.5
T_5-T_{9a}	1072.49	1072.4
		1088.4[c]
T_6	1136.53	1136.6
T_5-T_9	1533.68	1533.7

[a] Monoisotopic masses. [b] Intrachain disulfide reduced. [c] T_5–T_{9a} Met sulfoxide.

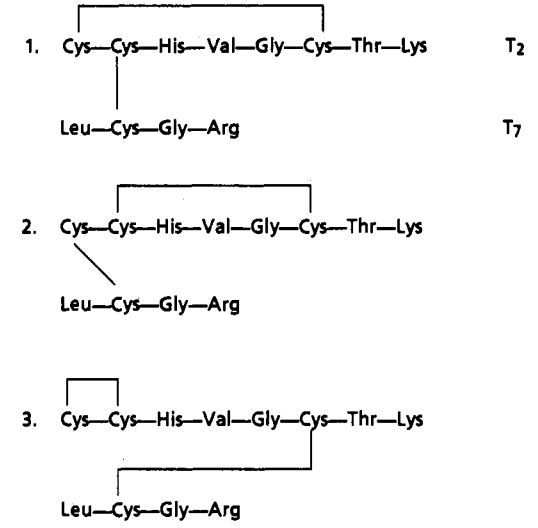

FIGURE 4: Possible disulfide bridge arrangements for the T_2-T_7 fragment from a tryptic digestion of human relaxin.

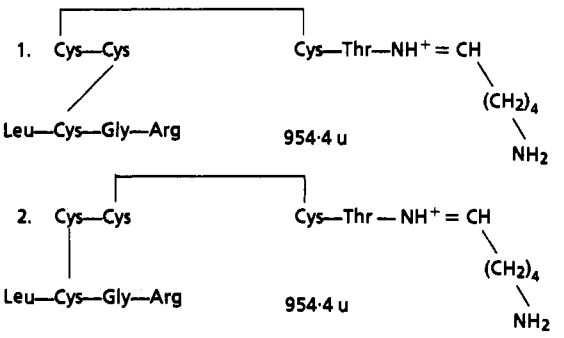

FIGURE 5: Possible structures for the 954.4 fragment molecular ion in the tandem mass spectral analysis of the T_2-T_7 tryptic peptide molecular ion 1293.5.

time to the T_2-T_7 tryptic fragment obtained from the recombinant material (data not shown).

In addition, an FAB-MS analysis of the T_2-T_7 tryptic fragment obtained from the recombinant material generated the molecular ions 1293.5 and 1295.6 (Table III), which were subjected to MS/MS. The third structure drawn in Figure 4 was eliminated by the MS/MS analysis of the 1293.5 molecular ion (data not shown) which generated the fragment molecular ion of 954.4 u whose proposed structure is shown in Figure 5. The absence of this fragment molecular ion (954.4 u) from the MS/MS analysis of the 1295.6 u molecular ion (data not shown), in which the intrachain disulfide bond is reduced, supports the identification of the 954.4 u fragment molecular ion as depicted in Figure 5. This type of fragmentation with the loss of a carboxyl group is a characteristic

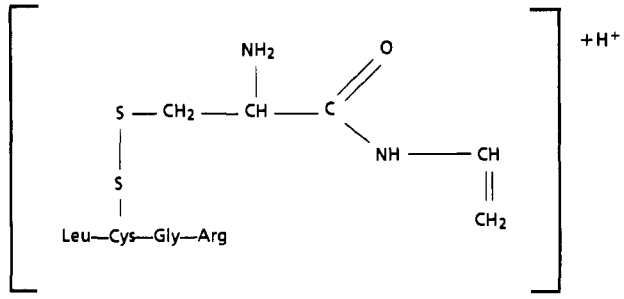

FIGURE 6: Proposed structure of the molecular ion fragment 592.2 observed in the tandem mass spectral analysis of the chemically synthesized T_2-T_7 peptide pair configured as in structure 2 of Figure 4.

"a" ion found in MS/MS spectra (Biemann & Scoble, 1987).

An MS/MS analysis was also performed on the chemically synthesized peptide pairs. The profile of the peptides disulfide bonded as in structure 1 gave an identical spectrogram to that obtained from the recombinant T_2–T_7 tryptic fragment (molecular ion 1293.5) including the 954.4 u fragment. The MS/MS profile obtained from the peptide disulfide bonded as in structure 2 gave a spectrogram containing an additional fragment of 592.2 u, consistent with the structure shown in Figure 6 which can easily be produced from structure 2 in Figure 4. This type of fragmentation, a "d"-type ion, retains the charge at the amino-terminal fragment and corresponds to the loss of the side chain from the carboxy-terminal amino acid of an "a"-type ion (Johnson et al., 1988). This fragment was not seen in the MS/MS spectrum of the recombinant T_2-T_7 tryptic peptide. These data, taken together, confirm that the disulfide arrangement is as shown in Figure 4, structure 1, and is identical with that observed for insulin.

Comparison of Recombinant Human Relaxin to a Human Relaxin Preparation Isolated from Human Corpora Lutea. A sequence was attempted on a preparation of human relaxin isolated from human corpora lutea (Winslow et al., 1989) which was shown to be biologically active as determined by the in vitro cell bioassay using chemically synthesized human relaxin as a reference since the recombinant form was not yet available. No A-chain sequence was evident, suggesting that its amino terminal was blocked. A sequence was detected consistent with the B-chain sequence as shown in Table II.

The reversed-phase HPLC profile obtained from a microbore column of human relaxin isolated from corpora lutea was compared to a chromatogram of 100 ng of rhRlx generated by the same gradient system. The retention times were identical (39.81 and 39.89 min, respectively), indicating that the disulfide arrangements are also the same, especially in view of the observations of Sieber et al. (1978) that the three chemically synthesized disulfide isomers of insulin differ distinctly in their physiochemical properties, and of Iwai et al. (1989) that the disulfide isomers of insulin-like growth factor I which also shares cysteine sequence homology to relaxin can be separated by reversed-phase HPLC. An FAB-MS analysis of the material isolated from human corpora lutea gave an A-chain mass of 2657 and a B-chain mass of 3313 (Winslow et al., 1989). From these data it was deduced that the A-chain amino-terminal glutamic acid is cyclized to the pyroglutamic acid form, substantiating the amino-terminal sequencing data, and that the B-chain is 29 amino acids in length. An early eluting peak was shown by FAB-MS analysis to be native human relaxin containing a B-chain of mass 3329 (3313 + 16) or 3345 (3313 + 32), suggesting that the methionine residues had become oxidized. Since only nanogram quantities of native human relaxin were available, it was not possible to generate a tryptic map.

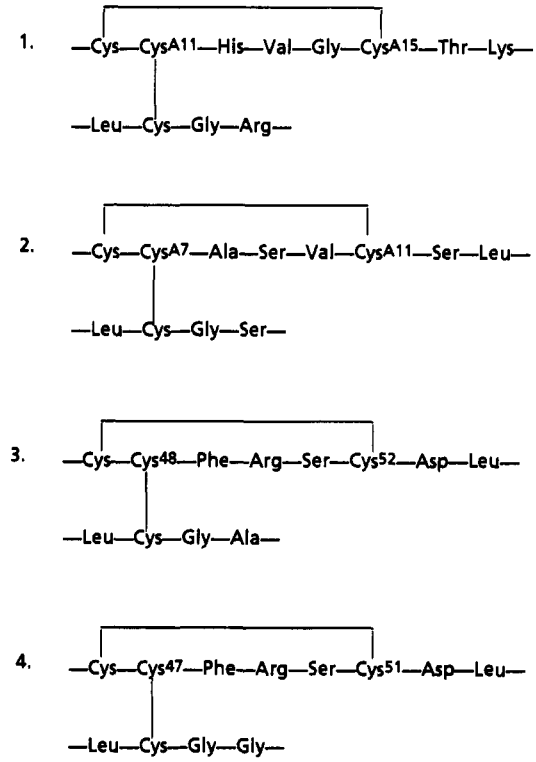

FIGURE 7: Homology of half-cystine residues of human relaxin (1), bovine insulin (2), and human insulin-like growth factors I (3) and II (4).

Discussion

The scarcity of relaxin as isolated from human sources has precluded an extensive characterization of its chemical and biological properties. Even the application of recombinant DNA techniques did not immediately lead to an unlimited supply. Initial investigations relied heavily upon chemical synthesis of the hormone to provide material for study (Tregear et al., 1983; Canova-Davis et al., 1990). Recently, innovative constructions of the human gene have resulted in the synthesis of human relaxin by recombinant DNA techniques (unpublished results).

Even though porcine relaxin has been relatively abundant (Sherwood & O'Byrne, 1974), its disulfide-bonded structure has not been rigorously defined. Schwabe and McDonald (1977) claimed that the disulfide bond distribution of porcine relaxin is homologous to that of insulin. However, an examination of their results reveals that they had only eliminated structure 2 (Figure 4) but not structure 3 as the possible configurations. The present report has definitively assigned the bonding in human relaxin to be that shown in structure 1, which is identical with the pattern found in insulin (Ryle et al., 1955).

A number of researchers have attempted to determine the exact disulfide bond linkage system in insulin-like growth factors I and II (Iwai et al., 1989; Raschdorf et al., 1988; Smith et al., 1989). As with relaxin, these growth factors have sequence homology to insulin (Figure 7), with the half-cystine residues being strictly conserved (Dafgard et al., 1985). The strategy employed by Raschdorf et al. (1988) was a sequential digestion of the substrate by more than one enzyme to produce peptides that were small enough for optimal utilization of tandem mass spectrometry. This approach yielded data from the resultant peptide mixture consistent with a disulfide bond structure identical with that of insulin. Smith et al. (1989)

attempted to obtain more exact data by isolating the individual peptides from an enzymatic digest before analysis by a number of analytical techniques. However, they failed to rigorously distinguish between the disulfide structures 1 and 2 (Figure 4). The Japanese group (Iwai et al., 1989) took an alternative tack. They chose to chemically synthesize the structures 1 and 2 (Figure 4) and compare their retention times upon reversed-phase HPLC to that of the peptide pairs isolated from recombinant insulin-like growth factor I isomers. In this case, structure 3 was ignored as a possibility even though Sieber et al. (1978) reported the chemical synthesis of an insulin isomer with this disulfide bonding. Intrachain disulfide bonds between adjacent Cys residues are both rare and strained. However, the recent kinetic experiments of Zhang and Snyder (1989) conducted using chemically synthesized peptides demonstrated that the formation of a disulfide between sequentially adjacent cysteines is not especially difficult. Documentation does exist for such a disulfide bond in human von Willebrand factor (Marti et al., 1987), the acetylcholine receptor (Kao & Karlin, 1986), the γ subunit of bovine transducin (Ovchinnikov et al., 1985), and the cyclic pentapeptide fungal product malformin A (Bodanszky & Stahl, 1975).

The work presented in this paper includes the more refined approaches initiated in these three reports, i.e., peptide pair isolations from enzyme digests and chemical synthesis of these peptide pairs in addition to the utilization of tandom mass spectrometry for direct structural analyses. This latter approach was necessary due to the differing amino acid sequence in the region between Cys^{A11} and Cys^{A15} of human relaxin as opposed to these sequences in the insulin-like growth factors (Figure 7). The presence of the protease-sensitive sites Phe and Arg in the latter proteins lends itself to a facile elimination of the disulfide isomer shown in structure 3 (Figure 4). Digestion by either trypsin or chymotrypsin would result in the production of two peptides. The sequence of human relaxin in this region contains no such site. Ryle et al. (1955) resorted to partial acid hydrolysis to effect cleavages in this same region in bovine insulin (Figure 7). The present study utilized tandem mass spectrometry to obtain fragments in the corresponding sequence of human relaxin. The structure 2 isomer (Figure 4) was eliminated by the comparison to the chemically synthesized peptide pairs in both reversed-phase HPLC and tandem mass spectral analyses.

In conclusion, a recombinant human relaxin preparation has been characterized with the aim of effecting a direct disulfide bond assignment in addition to detecting potential variants such as deamidated, oxidized, proteolytically degraded, and aggregated species. The comparison to relaxin isolated from human corpora lutea indicates that the disulfide bond configurations of the recombinant DNA derived relaxin are representative of those existing in the native molecule.

Acknowledgments

We thank Karen Wagner for amino acid analyses; Reed Harris for amino acid sequence analysis; Victor Ling and James Bourell for FAB-MS analyses; Dr. John Winslow for the isolation of human relaxin from corpora lutea and its analysis by reversed-phase HPLC, Daniel Drolet for its bioassay, and William Henzel for its partial sequence; and Dr. Geoffrey Tregear for discussions relating to peptide synthesis. We especially thank Dr. Richard Vandlen for his interest and support in making these investigations possible.

Registry No. rhRlx, 9002-69-1.

References

Adams, J. C. (1981) *J. Histochem. Cytochem. 29*, 775.

Anicetti, V. R., Fehskens, E. F., Reed, B. R., Chen, A. B., Moore, P., Geier, M. D., & Jones, A. J. S. (1986) *J. Immunol. Methods 91*, 213–224.

Atherton, E., & Sheppard, R. C. (1989) in *Solid Phase Peptide Synthesis* (Rickwood, D., & Hames, B. D., Eds.) pp 131–148, Oxford University Press, Oxford.

Biemann, K., & Scoble, H. A. (1987) *Science 237*, 992–998.

Bodanszky, M., & Stahl, G. L. (1975) *Bioorg. Chem. 4*, 93–105.

Braddon, S. A. (1978) *Endocrinology 102*, 1292–1299.

Burnette, W. N. (1981) *Anal. Biochem. 112*, 195–203.

Canova-Davis, E., Baldonado, I. P., & Teshima, G. M. (1990) *J. Chromatogr. 508*, 81–96.

Cronin, M. J., Malaska, T., & Bakhit, C. (1987) *Biochem. Biophys. Res. Commun. 148*, 1246–1251.

Dafgard, E., Bajaj, M., Honegger, A. M., Pitts, J., Wood, S., & Blundell, T. (1985) *J. Cell Sci. Suppl. 3*, 53–64.

Ellman, G. L. (1959) *Arch. Biochem. Biophys. 82*, 70–77.

Fei, D. T. W., Gross, M. C., Lofgren, J. L., Mora-Worms, M., & Chen, A. B. (1990) *Biochem. Biophys. Res. Commun. 170*, 214–222.

Fevold, H. L., Hisaw, F. L., & Meyer, R. K. (1930) *J. Am. Chem. Soc. 52*, 3340–3348.

Graham, E. F., & Dracy, A. E. (1953) *J. Dairy Sci. 36*, 772–777.

Hayashi, R., Moore, S., & Stein, W. H. (1973) *J. Biol. Chem. 248*, 2296–2302.

Hirs, C. H. W. (1967) *Methods Enzymol. 11*, 59–62.

Hisaw, F. L. (1926) *Proc. Soc. Exp. Biol. Med. 23*, 661–663.

Holbrook, I. B., & Leaver, A. G. (1976) *Anal. Biochem. 75*, 634–636.

Hsu, C. J., McCormack, S. M., & Sanborn, B. M. (1985) *Endocrinology 116*, 2029–2035.

Hudson, P., Haley, J., John, M., Cronk, M., Crawford, R., Haralambidis, J., Tregear, G., Shine, J., & Niall, H. (1983) *Nature 301*, 628–631.

Hudson, P., John, M., Crawford, R., Haralambidis, J., Scanlon, D., Gorman, J., Tregear, G., Shine, J., & Niall, H. (1984) *EMBO J. 3*, 2333–2339.

Iwai, M., Kobayashi, M., Tamura, K., Ishii, Y., Yamada, H., & Niwa, M. (1989) *J. Biochem. 106*, 949–951.

James, R., Niall, H., Kwok, S., & Bryant-Greenwood, G. (1977) *Nature 267*, 544–546.

Johnson, R. S., Martin, S. A., & Biemann, K. (1988) *Int. J. Mass Spectrom. Ion Processes 86*, 137–154.

Judson, D. G., Pay, S., & Bhoola, K. D. (1980) *J. Endocrinol. 87*, 153–159.

Kao, P. N., & Karlin, A. (1986) *J. Biol. Chem. 261*, 8085–8088.

Kramer, S. M., Gibson, U. E. M., Fendly, B. M., Mohler, M. A., Drolet, D. W., & Johnston, P. D. (1990) *In Vitro Cell. Dev. Biol. 26*, 647–656.

Krantz, J. C., Jr., Bryant, H. H., & Carr, C. J. (1950) *Surg., Gynecol. Obstet. 90*, 372–375.

Marti, T., Rösselet, S. J., Titani, K., & Walsh, K. A. (1987) *Biochemistry 26*, 8099–8109.

Matsubara, H., & Sasaki, R. M. (1969) *Biochem. Biophys. Res. Commun. 35*, 175–181.

Moore, S., Spackman, D. H., & Stein, W. H. (1958) *Anal. Chem. 30*, 1185–1190.

Ovchinnikov, Y. A., Lipkin, V. M., Shuvaeva, T. M., Bogachuk, A. P., & Shemyakin, V. V. (1985) *FEBS Lett. 179*, 107–110.

Podell, D. N., & Abraham, G. N. (1978) *Biochem. Biophys. Res. Commun. 81*, 176–185.

Raschdorf, F., Dahinden, R., Maerki, W., Richter, W. J., & Merryweather, J. P. (1988) *Biomed. Environ. Mass Spectrom. 16*, 3–8.
Ryle, A. P., Sanger, F., Smith, L. F., & Kitai, R. (1955) *Biochem. J. 60*, 541–556.
Sanborn, B. M., Kuo, H. S., Weisbrodt, N. W., & Sherwood, O. D. (1980) *Endocrinology 106*, 1210–1215.
Schägger, H., & von Jagow, G. (1987) *Anal. Biochem. 166*, 368–379.
Schwabe, C., & McDonald, J. K. (1977) *Science 197*, 914–915.
Sherwood, C. D., & O'Byrne, E. M. (1974) *Arch. Biochem. Biophys. 160*, 185–196.
Shih, A., Goldsmith, L. T., Weiss, G., Bourell, J., & Winslow, J. (1989) *Endocrine Soc.*, 1543.
Shire, S. J. (1983) *Biochemistry 22*, 2664–2671.
Sieber, P., Eisler, K., Kamber, B., Riniker, B., Rittel, W., Marki, F., & deGasparo, M. (1978) *Hoppe-Seyler's Z. Physiol. Chem. 359*, 113–123.
Smith, M. C., Cook, J. A., Furman, T. C., & Occolowitz, J. L. (1989) *J. Biol. Chem. 264*, 9314–9321.
Tregear, G. W., Du, Y.-C., Wang, K.-Z., Southwell, C., Jones, P., John, M., Gorman, J., Kemp, B., & Niall, H. D. (1983) in *Biology of Relaxin and Its Role in the Human* (Bigazzi, M., Greenwood, F. C., & Gasparri, F., Eds.) pp 42–55, Excerpta Medica, Amsterdam.
Winslow, J., Shih, A., Laramee, G., Bourell, J., Stults, J., & Johnston, P. (1989) *Endocrine Soc.*, 889.
Zhang, R., & Snyder, G. H. (1989) *J. Biol. Chem. 264*, 18472–18479.

Activating Region of HIV-1 Tat Protein: Vacuum UV Circular Dichroism and Energy Minimization†

Erwann P. Loret,*,‡ Eric Vives,§ Pui Shing Ho,‡ Hervé Rochat,§ Jurphaas Van Rietschoten,§ and W. Curtis Johnson, Jr.‡

Department of Biochemistry and Biophysics, Oregon State University, Corvallis, Oregon 97331-6503, and Laboratoire de Biochimie, CNRS URA 1179, Faculté de Médecine, Secteur Nord, Boulevard Pierre Dramard, 13326 Marseille Cedex 15, France

Received March 21, 1991

ABSTRACT: Tat protein is a trans-acting transcriptional activator of the human immunodeficiency virus type 1 and is essential for viral transcription. By homology with other transcriptional activators, Tat is expected to possess a nucleic acid binding region and a separate adjacent activating region. In order to localize the activating region of Tat, we have synthesized the sequences 2–23 and 38–60 of the protein. These two peptides contain the two candidates for the activating regions proposed from mutation experiments in previous studies: sequence 1–13 and sequence 38–45. The argument advanced to justify the location of the activating region within the sequence 1–13 was the periodicity of acidic, polar, and hydrophobic residues consistent with that of an amphipathic α-helix, similar to the activating region of many eukaryotic transcriptional activators. We have monitored by vacuum UV circular dichroism the ability of each peptide to adopt an α-helical conformation under conditions that strongly favor the formation of secondary structures. Only peptide 38–60 adopts an α-helical conformation in these conditions, in keeping with Chou–Fasman prediction. Energy minimization and molecular dynamics were carried out for several possible conformations of sequences 1–14 and 38–60. Our results indicate that only the sequence 38–45 is able to form an α-helix with amphipathic characteristics.

The protein Tat is a trans activator of the human immunodeficiency virus type 1 (HIV-1)[1] in vivo (Arya et al., 1985; Sodroski et al., 1985). The major action of Tat is to increase the steady-state level of transcription from the HIV-1 long terminal repeat (LTR), although the mechanism of this trans activation remains unclear [see Rosen and Pavlakis (1990) and Cullen and Greene (1990) for reviews]. The Tat gene is composed of two exons predicting a protein of 86 amino acid residues (Figure 1); however, the first 72 N-terminal residues, corresponding to the first exon, are sufficient for the full activity (Cullen, 1986). Moreover, the residues 58–72 can also be deleted without impairing the activation function of Tat (Siegels et al., 1986; Wright et al., 1986).

Mutational analyses of Tat (Garcia et al., 1988, 1989; Sadaie et al., 1989; Hauber et al., 1989; Kuppuswamy et al., 1989; Green et al., 1989; Ruben et al., 1989; Rappaport et al., 1989) made it possible to propose a functional organization of this protein (Figure 1). Two regions have been functionally characterized. Region II, a cysteine-rich cluster containing

† This work was supported by grants from the Association pour la Recherche sur le Cancer, the American Foundation for AIDS Research, the ANRS, the American Cancer Society (NP-740 and JFRA306), the CNRS, and the National Institutes of Health (Grant GM 21479).
* To whom correspondence should be addressed.
‡ Oregon State University.
§ Laboratoire de Biochimie, CNRS URA 1179.

[1] Abbreviations: HIV-1, human immunodeficiency virus type 1; Tat, trans-acting transcriptional activator of HIV-1; LTR, long terminal repeat sequence located at the two extremities of the HIV-1 DNA; C18 HPLC, reverse-phase high-performance liquid chromatography on C18 columns; TFA, trifluoroacetic acid; TFE, 2,2,2-trifluoroethanol; CD, circular dichroism.

0006-2960/91/0430-6013$02.50/0