Polypeptide Chain Fold of Human Transforming Growth Factor α Analogous to Those of Mouse and Human Epidermal Growth Factors As Studied by Two-Dimensional ^{1}H NMR

Daisuke Kohda,[‡] Ichio Shimada,[‡] Tetsuo Miyake,[§] Toru Fuwa,[§] and Fuyuhiko Inagaki*[,‡]

Department of Molecular Physiology, Tokyo Metropolitan Institute of Medical Science, Bunkyo-ku, Honkomagome, Tokyo 113, Japan, and Central Research Laboratories, Wakunaga Pharmaceuticals Company, Shimokotachi, Koda-cho, Takata-gun, Hiroshima 729-64, Japan

Received June 14, 1988; Revised Manuscript Received August 15, 1988

ABSTRACT: The ^{1}H NMR spectrum of human transforming growth factor α (TGF-α) was analyzed almost completely by the sequential assignment method using two-dimensional NMR techniques. On the basis of the nearly complete sequence-specific resonance assignment, secondary and tertiary structures of human TGF-α in solution (pH 4.9, 28 °C) were determined to satisfy the upper limits of proton–proton distances derived from nuclear Overhauser effect experiments. Although human TGF-α and mouse epidermal growth factor (EGF) share 27% homology in amino acid sequence, the backbone chain folds in the two growth factors are quite similar. The structure and function of TGF-α is well characterized by the "mitten model" previously proposed for mouse EGF. The gross shape of the TGF-α molecule resembles a mitten. TGF-α interacts with the receptor as a mitten would grasp an object. However, there is an appreciable structural difference between the two growth factors in the back of the mitten that is formed by the N-terminal polypeptide segment. This is consistent with the evidence that the backs of these molecules are not involved in the receptor binding.

Transforming growth factor α (TGF-α)[1] is a mitogenic polypeptide hormone produced by a variety of retrovirus-transformed cells, human tumors, and embryonic cells (Ozanne et al., 1980; Todaro et al., 1980; Twardzik et al., 1982). Human TGF-α is a single-chain polypeptide with 50 amino acid residues (Derynck et al., 1984). The sequence of human TGF-α can be aligned with those of mouse epidermal growth factor (EGF) and human EGF so that six cysteine residues are located on homologous positions (Figure 3). Locations of the disulfide bridges of human TGF-α were found to be analogous to those of EGFs (Winkler et al., 1986). Although human TGF-α shares only 27% and 37% homologies in amino acid sequence with mouse EGF and human EGF, respectively, they compete for the EGF receptor with similar affinities (0.55 mg of EGF equiv/mg of TGF-α) (Winkler et al., 1986) and activate the tyrosine phosphokinase activity of the EGF receptor in a similar manner (Reynolds et al., 1981; Pike et al., 1982). TGF-α shows a variety of biological actions similar to EGF, but is found to be more potent than EGF on several biological functions (Stern et al., 1985; Ibbotson et al., 1986).

Since TGF-α and EGF bind to the same receptor competitively (Carpenter et al., 1983; Massagué, 1983), the two growth factors are expected to be similar in tertiary structures. The backbone chain fold of mouse EGF has been determined through NMR by us (Kohda et al., 1988) and another group (Montelione et al., 1986, 1987), independently, on the basis of sequence-specific resonance assignments (Kohda & Inagaki, 1988; Montelione et al., 1988). The backbone chain fold of human EGF has also been reported (Carver et al., 1986; Cooke et al., 1987). In the present study, we completed the sequence-specific resonance assignment of human TGF-α and constructed a mechanical molecular model using the inter-proton distance constraints (ca. 4 Å) derived from NOESY spectra. The backbone chain fold of human TGF-α was found to be similar to that of mouse EGF except for the N-terminal segment, in spite of a single residue deletion next to residue 18 and a substitution of proline residue at residue 30 as compared with mouse EGF.

Materials and Methods

Human Recombinant TGF-α Preparation. Recombinant human TGF-α was produced as a fused protein with human growth hormone. After treatment with CNBr, human TGF-α was isolated by using gel filtration and reversed-phase HPLC. The amino acid sequence and the arrangement of three disulfide linkages of the recombinant TGF-α were in accord with those of native TGF-α (Derynck et al., 1984; Winkler et al., 1986). The recombinant TGF-α competed with EGF for the EGF receptor and enabled normal rat kidney cells to grow in soft agar in a manner similar to that of native TGF-α (Marquardt & Todaro, 1982).

NMR Measurements. Human TGF-α samples were dissolved at 5.6 mM in 99.8% ^2H_2O or in 90% ^1H_2O/10% ^2H_2O. The sample pH was 4.9 for human TGF-α, 2.0 for mouse EGF, and 2.7 for human EGF. The pH values given are direct pH-meter readings measured at 25 °C with a Radiometer PHM84. ^{1}H NMR spectra at 500 MHz were recorded on a JEOL JNM-GX500 spectrometer at a probe temperature of 28 °C. Chemical shifts were measured from the internal standard of sodium 2,2-dimethyl-2-silapentane-5-sulfonate. DQF-COSY (Rance et al., 1983), HOHAHA (Bax & Davies, 1985), and NOESY (Jeener et al., 1979; Macura et al., 1981) spectra were recorded in the phase-sensitive mode as described in States et al. (1982). All two-dimensional spectra were recorded with 512 × 2048 data points and with a spectral

* Author to whom correspondence should be addressed.
‡ Tokyo Metropolitan Institute of Medical Science.
§ Wakunaga Pharmaceuticals Co.

[1] Abbreviations: DQF-COSY, double quantum filtered correlated spectroscopy; EGF, epidermal growth factor; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; TGF-α, transforming growth factor α.

0006-2960/89/0428-0953$01.50/0

width of 6000 Hz. The HOHAHA spectra were recorded at mixing times of 45 and 95 ms. The NOESY spectra were recorded with a mixing time of 150 ms. Typically, 128 scans were accumulated for each t_1 with a relaxation delay of 1.0 s. The digital resolution was about 6 Hz/point in both dimensions by zero-filling in the t_1 dimension. A phase-shifted sine-bell function was applied for both t_1 and t_2 dimensions. After Fourier transformation, t_1 noise was reduced by the substraction method (Klevit, 1985).

NMR Spectra Analyses. The sequence-specific resonance assignment of a protein ^{1}H NMR spectrum was obtained by the sequential assignment procedure reviewed by Wüthrich (1986). Briefly, the amino acid spin systems are identified through spin–spin couplings observed on DQF-COSY and RCT-COSY (relayed coherence transfer spectroscopy) spectra, and then the spin systems are aligned along the primary structure through nuclear Overhauser effects (NOEs) observed on NOESY spectra. HOHAHA is also very useful in identifying the amino acid types, especially for long side chain bearing amino acid residues (Clore et al., 1986; Klevit & Drobny, 1986; Weber et al., 1987; Kohda & Inagaki, 1988).

NOEs were classified into three categories: intraresidue, sequential, and tertiary NOEs. Intraresidue NOEs were those within an amino acid residue, sequential NOEs were those used in the sequential assignment, and tertiary NOEs were all other NOEs. The distance between protons A and B located on residues i and j, respectively, in the sequence is denoted $d_{AB}(i,j)$ (Wüthrich, 1986). The indices i and j are omitted for the sequential distances; for example, $d_{\alpha N}(i,i+1) = d_{\alpha N}$. Making reference to the proton resonance assignments of human TGF-α, NOEs were assigned to individual proton pairs. The three-dimensional molecular model of human TGF-α was constructed with an HGS mechanical model (Maruzen, Tokyo) using upper interproton distance limits (ca. 4 Å) derived from these NOEs.

It is generally argued that NOESY spectra recorded with a long mixing time (150 ms) are not free from spin diffusion effects, which obscure the quantitative distance information (Kumar et al., 1981), and that manual model building is not sufficient to explore the conformations consistent with the distance constraints derived from NOEs. However, in the case of mouse EGF, we investigated these points in detail to compare the molecular model constructed by using 150-ms NOESY and manual model building (Kohda et al., 1988) with that constructed by using 80-ms NOESY and distance geometry calculation (unpublished results). The two models of mouse EGF were found to be very similar with regard to the gross three-dimensional structure. The extra NOE cross-peaks observed in 150-ms NOESY were at most 10% of the total number of the NOE cross-peaks in 80-ms NOESY. These extra NOEs may arise from spin diffusion. However, spin diffusion posed no severe problems because it provides qualitative information on spatial proximity of protons in manual model building. Even in the preliminary distance geometry calculation, it is also useful information since the interproton distances derived from tertiary NOEs are just used qualitatively as fixed distance constraints, usually 5 Å. Thus, we consider that the combination of 150-ms NOESY and manual model building as was used in the present study can provide reliable gross three-dimensional structure of proteins.

Results

Sequence-Specific Resonance Assignment. Nearly complete sequence-specific resonance assignment of human TGF-α was obtained in a manner similar to that described in detail for mouse EGF (Kohda & Inagaki, 1988). Human TGF-α

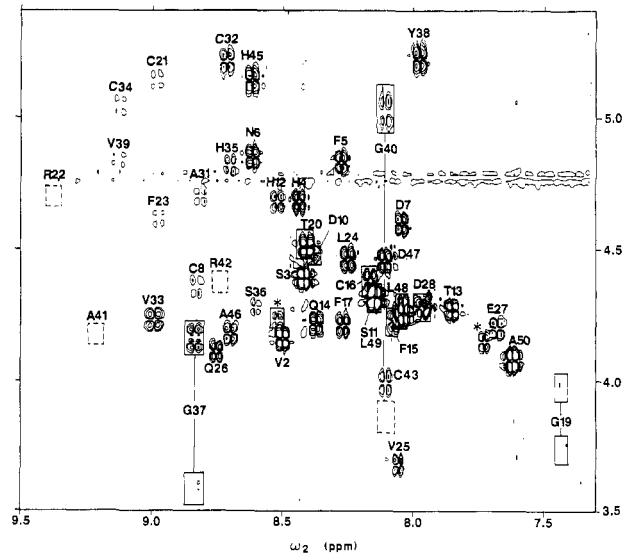

FIGURE 1: Fingerprint region of the DQF-COSY spectrum of TGF-α at pH 4.9 and 28 °C in ^1H_2O. Several cross-peaks are enclosed by solid boxes for clarity. The assignments are shown with the one-letter code for the amino acids. A broken box indicates the location of missing cross-peaks (R22, G40, A41, R42) due to the low intensity. The cross-peak of E44 is located outside the region of the figure and is not shown. The cross-peaks of V1, P9, H18, K29, and P30 are not shown because the NH resonances were not observed. The cross-peaks of S11 and L49 are almost overlapped. The cross-peaks from impurities are indicated by asterisks.

consists of 50 amino acid residues: Gly (3), Thr (2), Ala (4), Ser (3), Val (5), Leu (3), Tyr (1), Phe (4), His (5), Asp (4), Asn (1), Cys (6), Glu (2), Gln (2), Pro (2), Arg (2), and Lys (1). This protein contains no Ile, Met, or Trp residues (Derynck et al., 1984). These spin systems were identified by using DQF-COSY recorded in ^1H_2O (Figure 1) and HOHAHA in ^1H_2O and in ^2H_2O. The ring protons of aromatic residues were connected to $C^\beta H_2$ by the NOE connectivities observed on NOESY in ^1H_2O. The side-chain amide protons of Asn and Gln residues were also connected to $C^\beta H_2$ (Asn) or $C^\gamma H_2$ (Gln) by NOEs on NOESY in ^1H_2O. The 19 spin systems of Gly (3), Thr (2), Ala (3), Ser (3), Tyr (1), Phe (1), His (1), Asn (1), Gln (2), Glu (1), and Pro (1) were completely identified, and the 24 spin systems were grouped into [Leu (3), Val (5)] and [Phe (3), His (3), Asp (4), Cys (6)]. Since human TGF-α contains single Tyr and Asn residues, Tyr38 and Asn6 were immediately sequence-specific assigned. The remaining seven spin systems were identified, but incompletely as their NH and/or $C^\alpha H$ could not be found in the DQF-COSY and the HOHAHA spectra. They were Ala (1), His (1), Glu (1), Pro (1), and a group of [Arg (2), Lys (1)]. Most of these missing chemical shifts were found directly from the position of an intraresidue NH–$C^\alpha H$ NOE cross-peak or indirectly from the position of a sequential NOE cross-peak in the NOESY spectra after the sequential assignment was accomplished.

The spin systems identified as above were aligned along the primary structure of human TGF-α by the NOE connectivities observed on the NOESY spectra in ^1H_2O (Figure 2) and in ^2H_2O. A summary of sequential NOE connectivities used in the sequential assignment of human TGF-α is shown in Figure 3. The $d_{\alpha N}$ connectivities were used mainly to connect neighboring amino acid residues with the help of the $d_{\beta N}$ and the d_{NN} connectivities if necessary. The $d_{\alpha N}$ walk from Tyr38 was extended in both directions and gave the sequence His35-Ser36-Gly37-Tyr38*-Val39-Gly40-Ala41, where the asterisk indicates the starting point of the sequential walk (Figure 2, solid lines). The sequence Asn6*-Asp7-Cys8-Pro9

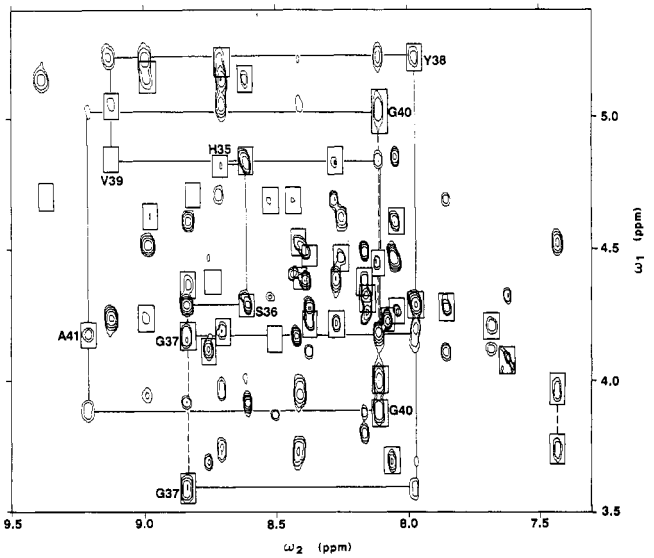

FIGURE 2: Fingerprint region of the NOESY spectrum of TGF-α under the same conditions and in the same spectral region as in Figure 1. The positions of the COSY cross-peaks are indicated by boxes. The two cross-peaks of a Gly residue are connected with a vertical broken line. The $d_{\alpha N}$ walk, His35-Ser36-Gly37-Tyr38-Val39-Gly40-Ala41, is shown by horizontal and vertical solid lines, for example.

was obtained by starting from Asn6. The assignment of the two Gly residues (Gly37 and Gly40) immediately led to the assignment of the remaining Gly residue to Gly19, which subsequently gave the sequence Gly19*-Thr20-Cys21-Arg22. In a similar manner, the following sequences were obtained one after another: Pro9-Asp10-Ser11-His12-Thr13*-Gln14; Val1-Val2-Ser3*; Arg22-Phe23-Leu24-Val25-Gln26*-Glu27-Asp28; His45-Ala46-Asp47-Leu48-Leu49-Ala50*; Pro30-Ala31*-Cys32-Val33-Cys34-His35; Ala41-Arg42-Cys43-Glu44-His45*; Gln14*-Phe15-Cys16-Phe17; Ser3*-His4-Phe5-Asn6. There remained three breaks of the NOE connectivities (Figure 3). A break at Phe17-His18 occurred because the NH resonance of His18 was missing, while breaks at Asp28-Lys29 and Lys29-Pro30 were ascribed to missing resonances of NH, $C^{\alpha}H$, and $C^{\beta}H_2$ protons of Lys29. The assignments of (NH, $C^{\alpha}H$) cross-peaks in the fingerprint region of the DQF-COSY spectrum are shown in Figure 1. The chemical shifts of the assigned proton resonances of human TGF-α are summarized in Table I.

Secondary Structure. On the basis of the sequential NOE connectivities presented in Figure 3 and tertiary NOEs between backbone protons, secondary structure elements were identified in human TGF-α. TGF-α includes a short antiparallel β-sheet involving the segments Cys21-Arg22 and Ala31-Cys32, which was recognized from the interstrand NOE connectivities $d_{\alpha\alpha}$(21,32), d_{NN}(22,31), and $d_{N\alpha}$(22,32). This β-sheet element may be extended irregularly to involve Phe23 and Pro30 because of the many NOEs reflecting short contacts between the side chains of Phe23 and Pro30. A turn involving the segment Val25-Asp28 was indicated by the sequential d_{NN} connectivities and the $d_{\alpha N}$(24,26) and $d_{\alpha N}$(25,28) connectivities. A second short antiparallel β-sheet involves the segments Tyr38-Val39 and His45-Ala46 shown by the interstrand NOE connectivities $d_{\alpha\alpha}$(38,46), $d_{\alpha N}$(38,47), and d_{NN}(39,45). The isolated strong d_{NN} connectivity at Gly37-Tyr38 and many NOEs between the side chains of His35 and Tyr38 suggest that the segment His35-Tyr38 forms a chain reversal of type II β-turn. Strong $d_{\alpha\delta}$ connectivity shows that the Cys8-Pro9 peptide bond takes a trans form, while no NOE connectivities were obtained to elucidate the configuration of the Lys29-Pro30 peptide bond.

Tertiary Structure. Seventy-one tertiary NOEs were assigned unambiguously to individual proton pairs, making reference to the chemical shifts of the amino acid residues (Table I). The mechanical model of the human TGF-α molecule was constructed to satisfy distance constraints (ca. 4 Å) derived from these tertiary NOEs. With the help of the constructed mechanical model, the unassigned tertiary NOEs were rexamined, and 62 NOEs were subsequently assigned. The total 133 tertiary NOEs are summarized in a distance map (above diagonal, Figure 4), and the backbone topology of human TGF-α obtained from the final mechanical model is shown in Figure 5.

Discussion

Sequence-specific resonance assignment of TGF-α was made (Table I) and the backbone chain fold was determined by model building (Figure 5) using the sequential NOEs (Figure 3) and the tertiary NOEs (Figure 4). Since TGF-α and EGF bind competitively to the EGF receptor (Carpenter et al., 1983; Massagué, 1983), the tertiary structures of the receptor binding site of both molecules should be expected to be similar. The tertiary structures of mouse and human EGFs have already been determined by means of NMR and distance geometry calculation (Montelione et al., 1987; Cooke et al., 1987; Kohda et al., 1988), so that we can now discuss the structure of human

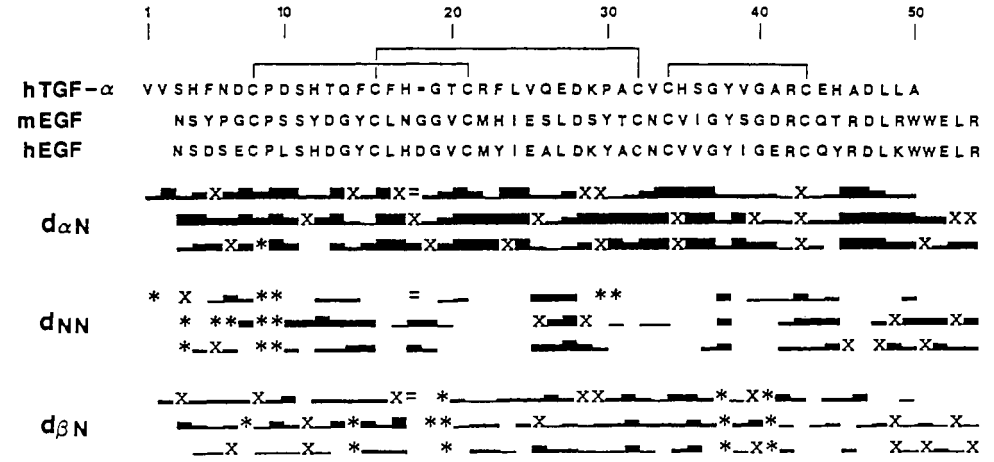

FIGURE 3: Sequence of human TGF-α and summary of the sequential NOE connectivities. For comparison, sequences and the NOE data of mouse EGF and human EGF are also shown. The $d_{\alpha\delta}$ and $d_{\beta\delta}$ connectivities involving proline residues are shown as the $d_{\alpha N}$ and $d_{\beta N}$ connectivities. The heights of the bars indicate the approximate magnitude of the NOESY cross-peaks. (=) A gap has been inserted in the TGF-α sequence for optimal alignment; (×) connectivity could not be determined; (*) connectivity cannot be defined.

Table I: Chemical Shifts of the Assigned Proton Resonances of Human TGF-α at 28 °C and pH 4.9[a]

residue	chemical shift (ppm) NH	$C^{\alpha}H$	$C^{\beta}H$	others
Val1	–	3.88	2.19	$C^{\gamma}H$ 1.00, 1.00
Val2	8.50	4.16	2.01	$C^{\gamma}H$ 0.93, 0.86
Ser3	8.42	4.39	3.70, 3.70	
His4	8.44	4.69	3.16, 3.02	C2H 8.49; C4H 7.14
Phe5	8.28	4.84	3.12, 2.94	C2,6H 7.18; C3,5H 7.31; C4H (7.3)
Asn6	8.62	4.86	2.87, 2.70	$N^{\delta}H$ 7.54, 6.92
Asp7	8.05	4.60	2.64, 2.50	
Cys8	8.83	4.36	3.17, 3.06	
Pro9	–	4.39	2.17, 1.90	$C^{\gamma}H$ 1.72, 1.72; $C^{\delta}H$ 3.30, 3.30
Asp10	8.38	4.49	2.69, 2.69	
Ser11	8.16	4.31	3.89, 3.79	
His12	8.52	4.68	3.39, 3.09	C2H 8.40; C4H 7.08
Thr13	7.86	4.27	4.11	$C^{\gamma}H$ 1.16
Gln14	8.37	4.22	1.95, 1.81	$C^{\gamma}H$ 2.15, 2.10; $N^{\epsilon}H$ 7.30, 6.73
Phe15	8.08	4.23	3.05, 2.86	C2,6H 7.13; C3,5H 7.19; C4H (7.2)
Cys16	8.17	4.38	2.49, 2.10	
Phe17	8.27	4.21	2.72, 2.72	C2,6H 7.11; C3,5H 7.17; C4H (7.2)
His18	*b*	4.53	2.74, 2.31	C2H 8.43; C4H 6.89
Gly19	7.44	3.97, 3.73		
Thr20	8.41	4.51	3.93	$C^{\gamma}H$ 1.26
Cys21	8.99	5.15	3.25, 3.17	
Arg22	9.38	4.72	1.68, 1.64	$C^{\gamma}H$ 1.49, 1.49; $C^{\delta}H$ 2.95, 2.95 $N^{\epsilon}H$ 7.01
Phe23	8.98	4.62	2.95, 2.77	C2,6H 6.87; C3,5H 7.14; C4H 7.17
Leu24	8.26	4.47	1.76, 1.76	$C^{\gamma}H$ 1.52; $C^{\delta}H$ 0.75, 0.75
Val25	8.06	3.68	2.04	$C^{\gamma}H$ 1.08, 1.00
Gln26	8.76	4.12	2.09, 2.09	$C^{\gamma}H$ 2.40, 2.40; $N^{\epsilon}H$ 7.59, 6.84
Glu27	7.68	4.20	1.73, 1.73	$C^{\gamma}H$ 2.23, 2.19
Asp28	7.98	4.28	3.11, 2.47	
Lys29	*b*	*b*	*b*	$C^{\gamma}H^{b}$; $C^{\delta}H$ 1.51, 1.51; $C^{\epsilon}H$ 3.04, 3.04 $N^{\zeta}H$ 7.08
Pro30	–	4.55	1.84, 1.64	$C^{\gamma}H$ 1.51, 1.51; $C^{\delta}H$ 3.55, 3.55
Ala31	8.82	4.70	1.37	
Cys32	8.72	5.22	(2.79), 2.66	
Val33	9.00	4.23	2.03	$C^{\gamma}H$ 0.94, 0.88
Cys34	9.12	5.05	3.44, 2.75	
His35	8.70	4.82	3.37, 2.97	C2H 8.21; C4H 7.18
Ser36	8.61	4.28	3.91, 3.91	
Gly37	8.84	4.17, 3.58		
Tyr38	7.97	5.23	2.94, 2.75	C2,6H 6.66; C3,5H 6.39
Val39	9.13	4.85	2.27	$C^{\gamma}H$ 0.86, 0.80
Gly40	8.11	5.02, 3.89		
Ala41	9.21	4.18	1.56	
Arg42	8.74	4.36	1.64, 1.26	$C^{\gamma}H$ 1.45, 1.04; $C^{\delta}H$ 2.99, 2.28 $N^{\epsilon}H$ 7.05
Cys43	8.11	3.99	2.92, 2.54	
Glu44	10.04	4.32	1.74, 1.74	$C^{\gamma}H$ (2.0, 2.0)
His45	8.62	5.15	3.22, 3.11	C2H 8.57; C4H 7.17
Ala46	8.70	4.19	1.15	
Asp47	8.12	4.46	2.50, 2.27	
Leu48	8.04	4.27	(2.00, 2.00)	$C^{\gamma}H$ 1.57; $C^{\delta}H$ 0.88, 0.82
Leu49	8.15	4.33	(1.59, 1.59)	$C^{\gamma}H$ (1.63); $C^{\delta}H$ 0.90, 0.83
Ala50	7.62	4.08	1.30	

[a] Chemical shifts are measured to ±0.01 ppm. The chemical shift values in parentheses are approximate because of spectral overlaps. [b] Resonance assignments were not obtained because no cross-peaks were observed on NMR spectra.

TGF-α in comparison with those of mouse and human EGFs. Sequential NOE data of mouse EGF (Kohda & Inagaki, 1988) and human EGF (unpublished data) are also presented in Figure 3, and tertiary NOE data of mouse EGF are shown in the lower left half of Figure 4 (Kohda et al., 1988) for comparison.

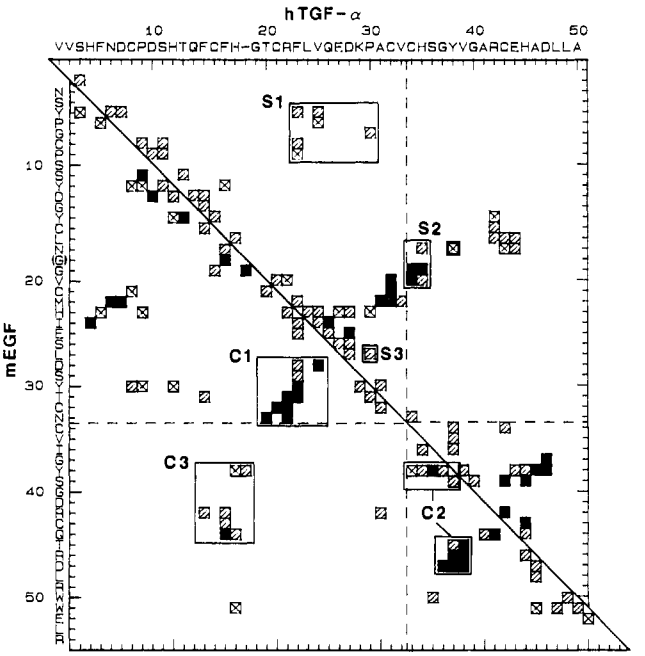

FIGURE 4: Distance maps of tertiary NOEs in human TGF-α (above diagonal, pH 4.9, 28 °C) and in mouse EGF (below diagonal, pH 2.0, 28 °C) (Kohda et al., 1988). Both axes are calibrated with the sequences of human TGF-α and mouse EGF. A gap (marked with a dash) was inserted in the TGF-α sequence between residues 18 and 19. The squares connect pairs of residues linked by one or more tertiary NOEs. Filled, hatched, and crossed squares indicate backbone–backbone, backbone–side chain, and side chain–side chain NOEs, respectively. When two or three kinds of NOEs were observed at a pair of residues, the NOE involving more backbone protons is shown. The broken lines separate the N-terminal domain region (1–33 for human TGF-α and 1–32 for mouse EGF) and the C-terminal domain region (34–50 for human TGF-α and 33–53 for mouse EGF). The regions where the tertiary NOEs are conserved between TGF-α and EGF are enclosed by boxes below the diagonal and labeled C1, C2, and C3. The regions containing the tertiary NOEs specific to human TGF-α are enclosed by boxes above the diagonal and labeled S1, S2, and S3.

In consequence of the high homology (74%) in amino acid sequence between mouse and human EGFs, the patterns of the sequential NOE connectivities are very similar to each other, showing that mouse and human EGFs take identical secondary structures except for the N-terminal segment (Cooke et al., 1987). On the other hand, in spite of the low homology in the sequence of human TGF-α to mouse EGF (27%) or human EGF (35%), the pattern of sequential NOE connectivities of human TGF-α is surprisingly similar to those of mouse and human EGFs. This implies that the secondary structure of human TGF-α is basically the same as those of mouse and human EGFs. The main differences in the sequence of TGF-α are the substitution of proline at residue 30 and the deletion next to residue 18 in comparison with EGF. (In the following discussion, we use the residue numbering for EGFs according to TGF-α as shown in the top of Figure 3.) Pro30 in TGF-α corresponds to Tyr30 that is located at the center of the antiparallel β-sheet structure in mouse and human EGFs. As expected, Pro30 distorts the β-sheet structure. The irregularity of the β-sheet structure induced by this substitution can be seen in the interruption of $d_{\alpha N}$ connectivities of TGF-α around Pro30 and Phe23 (Figure 3). On the other hand, the deletion site next to residue 18 in TGF-α is involved in the loop region in EGFs. This deletion reduced the d_{NN} connectivities of TGF-α (Figure 3), indicating that this segment adopts more extended structure compared with those of EGFs.

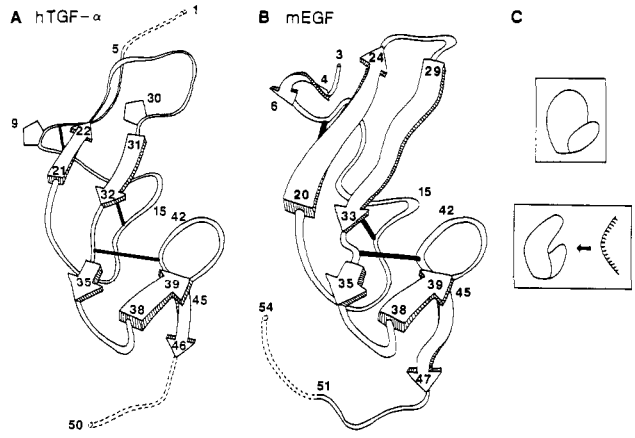

FIGURE 5: (A) Schematic drawing of the backbone topology of human TGF-α derived from the mechanical model constructed with NOE distance constraints. The arrowed ribbons represent the location and direction of the β-sheet strands. The side chains of the two proline residues (Pro9 and Pro30) are indicated by pentagons on the backbone rope. The broken ropes indicate the N-terminal segment, Val1–His4, and the C-terminal segment, Asp47–Ala50, whose conformations could not be determined for lack of tertiary NOEs. (B) Backbone topology of mouse EGF previously reported by Kohda et al. (1988). Residue numbering according to TGF-α is used for comparison. (C) Schematic front view of human TGF-α (top). The overall structure of a human TGF-α molecule resembles a mitten. A side view is shown at the bottom. Human TGF-α interacts with the EGF receptor as a mitten would grasp an object.

The tertiary NOEs for human TGF-α and mouse EGF are shown in the upper right half and the lower left half of the distance map, respectively, in Figure 4. The distributions of the tertiary NOEs for both molecules are similar, indicating that the gross structures are the same. Tertiary NOEs conserved for the two growth factors are enclosed by boxes below the diagonal and labeled C1, C2, and C3 (Figure 4). They demonstrate that human TGF-α can be divided into two structural domains, the N-terminal domain (Val1-Val33) and the C-terminal domain (Cys34-Ala50), as can be mouse and human EGFs. The NOEs in the C1 region indicate that the N-terminal domain of both growth factors contains an antiparallel β-sheet structure. The β-sheet of human TGF-α is, however, shorter than that of mouse EGF. For example, interstrand NOE $d_{\alpha\alpha}$(23,30) was not observed and was instead replaced by side chain–side chain NOEs between Phe23 and Pro30. No NMR data were available to determine the configuration around the Lys29-Pro30 peptide bond and the five chemical shifts of the Lys29 proton resonances (Table I). This may be attributed to the irregular β-sheet of human TGF-α. The NOEs in the C2 region indicate that the C-terminal domain contains a double-hairpin structure that consists of a reverse turn and an antiparallel β-sheet as described in Montelione et al. (1986). The Cys33-Cys42 disulfide bond forces the turn and the β-sheet in the C-terminal domain to be packed into highly restricted structure, so that the portions of the distance maps corresponding to the C-terminal domains are particularly similar between the two growth factors. The NOEs in the C3 region reflecting interdomain NOEs indicate that the relative orientations of the two domains in mouse EGF are also maintained in human TGF-α.

The tertiary NOEs specific to human TGF-α are enclosed by boxes above the diagonal and labeled S1, S2, and S3 (Figure 4). The NOEs in the S2 region indicate the influence of the deletion next to residue 18. Many backbone–backbone NOEs, $d_{\alpha\alpha}$(19,34), $d_{\alpha N}$(19,35), and $d_{N\alpha}$(20,34), were present in human TGF-α but absent in mouse EGF. These specific NOEs demonstrate that the shorter segment, Phe17-Thr20 of TGF-α, runs straighter than the counterpart of mouse EGF (Figure 5). The NOE $d_{N\gamma}$(27,30) in the S3 region shows the direct effect of the proline substitution at residue 30. The fixed torsion angle φ for a planar proline ring changes the direction of the main-chain flow. The N-terminal segment, Phe5-Phe15, adopts a different conformation compared with mouse EGF as shown in the S1 region of Figure 5. The N-terminal segment of mouse EGF associates with one strand of the β-sheet, while the N-terminal segment of human TGF-α runs along the center of the two β-sheet strands, which is well recognized from the NOE $d_{N\gamma}$(7,30) in the S1 region. For the N-terminal segment, Val1-His4, and the C-terminal segment, Leu48-Ala50, no tertiary NOEs were observed, which implies the flexibility of both tail segments.

As proposed in Kohda et al. (1988), the structure and function of the mouse EGF molecule are well characterized by a "mitten model". This mitten model is also applicable to human TGF-α (Figure 5C). The TGF-α molecule consists of a palm (N-terminal domain) and a thumb (C-terminal domain). Due to the deletion next to residue 18, the loop comprising Phe17-Thr20 in TGF-α is shortened and runs straighter than the corresponding loop of EGF. Although the substitution of a proline at residue 30 disrupts the regular β-sheet in the N-terminal domain, the secondary structure near Pro30 can be still regarded as a distorted β-sheet-like structure. Thus, the deletion and substitution little affect the structure of the hollow formed by the two domains in the TGF-α mitten. Since EGF is considered to interact with the receptor through this hollow from several lines of biochemical evidence (Kohda et al., 1988), this result is in good agreement with the fact that TGF-α and EGF bind to the same site on the EGF receptor (Carpenter et al., 1983; Massagué, 1983).

However, TGF-α shows biological effects distinct from those of EGF (DeLarco & Todaro, 1978; Stern et al., 1985; Ibbotson et al., 1986). It should be noted that there are no slowly exchangeable amide protons in TGF-α, in contrast to mouse and human EGFs, where more than 10 slowly exchangeable amide protons were observed (Mayo, 1985; Carver et al., 1986), suggesting that TGF-α is more flexible than EGF. We infer that Pro30 makes the TGF-α molecule more flexible than EGF. Such a difference in dynamic behavior may be responsible for the difference in the biological activities of the two growth factors. In this respect, the site-directed mutagenesis of Tyr30 → Pro30 in the EGF sequence and/or Pro30 → Tyr30 in the TGF-α sequence may elucidate the mechanism of the biological activities specific to TGF-α.

Registry No. TGF-α, 93585-89-8; EGF, 62229-50-9.

REFERENCES

Bax, A., & Davies, D. G. (1985) *J. Magn. Reson. 65*, 393–402.
Carpenter, G., Stoscheck, C. M., Preston, Y. A., & DeLarco, J. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 5627–5630.
Carver, J. A., Cooke, R. M., Esposito, G., Campbell, I. D., Gregory, H., & Sheard, B. (1986) *FEBS Lett. 205*, 77–81.
Clore, G. M., Martin, S. R., & Gronenborn, A. M. (1986) *J. Mol. Biol. 191*, 553–561.
Cooke, R. M., Wilkinson, A. J., Baron, M., Pastore, A., Tappin, M. J., Campbell, I. D., Gregory, H., & Sheard, B. (1987) *Nature 327*, 339–341.
DeLarco, J. E., & Todaro, G. J. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4001–4005.
Derynck, R., Roberts, A. B., Winkler, M. E., Chen, E. Y., & Goeddel, D. V. (1984) *Cell 38*, 287–297.
Ibbotson, K. J., Harrod, J., Gowen, M., D'Souza, S., Smith, D. D., Winkler, M. E., Derynck, R., & Mundy, G. R. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 2228–2232.

Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546-4553.
Klevit, R. E. (1985) *J. Magn. Reson. 62*, 551-555.
Klevit, R. E., & Drobny, G. P. (1986) *Biochemistry 25*, 7770-7773.
Kohda, D., & Inagaki, F. (1988) *J. Biochem. (Tokyo) 103*, 554-571.
Kohda, D., Go, N., Hayashi, K., & Inagaki, F. (1988) *J. Biochem. (Tokyo) 103*, 741-743.
Kumar, A., Wagner, G., Ernst, R. R., & Wüthrich, K. (1981) *J. Am. Chem. Soc. 103*, 3654-3658.
Macura, S., Huang, Y., Suter, D., & Ernst, R. R. (1981) *J. Magn. Reson. 43*, 259-281.
Marquardt, H., & Todaro, G. J. (1982) *J. Biol. Chem. 257*, 5220-5225.
Massaguê, J. (1983) *J. Biol. Chem. 258*, 13614-13620.
Mayo, K. H. (1985) *Biochemistry 24*, 3783-3794.
Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 8594-8598.
Montelione, G. T., Wüthrich, K., Nice, E. C., Burgess, A. W., & Scheraga, H. A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 5226-5230.
Montelione, G. T., Wüthrich, K., & Scheraga, H. A. (1988) *Biochemistry 27*, 2235-2243.
Ozanne, B., Fulton, R. J., & Kaplan, P. L. (1980) *J. Cell. Physiol. 105*, 163-180.
Pike, L. J., Marquardt, H., Todaro, G. J., Gallis, B., Casnellie, J. E., Bornstein, P., & Krebs, E. G. (1982) *J. Biol. Chem. 257*, 14628-14631.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479-485.
Reynolds, F. H., Jr., Todaro, G. J., Fryling, C., & Stephenson, J. R. (1981) *Nature 292*, 259-262.
States, D. J., Haberkorn, R. A., & Reuben, D. J. (1982) *J. Magn. Reson. 48*, 286-292.
Stern, P. H., Krieger, N. S., Nissenson, R. A., Williams, R. D., Winkler, M. E., Derynck, R., & Strewler, G. J. (1985) *J. Clin. Invest. 76*, 2016-2019.
Todaro, G. J., Fryling, C., & DeLarco, J. E. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 5258-5262.
Twardzik, D. R., Ranchalis, J. E., & Todaro, G. J. (1982) *Cancer Res. 42*, 590-593.
Weber, P. L., Sieker, L. C., Samy, A., Reid, B. R., & Drobny, G. P. (1987) *J. Am. Chem. Soc. 109*, 5842-5844.
Winkler, M. E., Bringman, T., & Marks, B. J. (1986) *J. Biol. Chem. 261*, 13838-13843.
Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, pp 117-161, Wiley, New York.

Ubiquitinated Histone H2B Is Preferentially Located in Transcriptionally Active Chromatin†

Barbara E. Nickel,‡ C. David Allis,§ and James R. Davie*,‡

Department of Biochemistry, Faculty of Medicine, University of Manitoba, Winnipeg, Manitoba, Canada R3E 0W3, and Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

Received April 18, 1988; Revised Manuscript Received July 12, 1988

ABSTRACT: Using an anti-ubiquitin antibody in Western blotting experiments, we detected polyubiquitinated species of histones H2A, H2A.Z, and H2B in histone preparations of bovine thymus, chicken erythrocyte, and *Tetrahymena* macro- and micronuclei. Histone H2A had the greatest level of polyubiquitinated species, with tetra- to hexaubiquitinated forms of this histone being observed. The fraction of bovine thymus and chicken erythrocyte chromatin enriched in transcriptionally active gene sequences was enriched in mono- and polyubiquitinated species of histones H2A, H2B, and H2A.Z, especially in the ubiquitinated forms of histone H2B. Histones H2A and H2B were ubiquitinated in the transcriptionally active *Tetrahymena* macronucleus, with monoubiquitinated (u) H2B being the predominant ubiquitinated histone species. Ubiquitinated forms of histones H2A and H2B were found in transcriptionally inert micronuclei, but at lower levels than seen in macronuclear histones. Also, the level of micronuclear uH2A was greater than that of uH2B which may be from macronuclei that contaminate the preparation. These results indicate that the mono- and polyubiquitinated species of histone H2B are preferentially located in transcriptionally active chromatin regions. Ubiquitinated histone H2A is located in both expressed and repressed chromatin domains, but expressed chromatin is enriched in mono- and polyubiquitinated forms of this histone. These observations are consistent with the hypothesis that ubiquitinated histones have a role maintaining the structure of transcriptionally active chromatin.

Ubiquitin, a small protein of 76 amino acids, can be covalently attached to histones H2A and H2B and their variants via an isopeptide bond between the ubiquitin C-terminal glycine and the ϵ-amino groups of lysine-119 in H2A (Goldknopf & Busch, 1977) and lysine-120 in H2B (Thorne et al., 1987). In multicellular eukaryotes, ubiquitinated (u) histone H2A replaces approximately 10% of the nucleosomal H2A (Busch & Goldknopf, 1981), and uH2B replaces approximately 1-2% of the nucleosomal H2B (West & Bonner, 1980). In the cytoplasm, ubiquitin is used to "tag" proteins for protease digestion with the rate of proteolysis increasing with the number of attached ubiquitins (Ciechanover et al., 1981). It

† This work was supported by the Medical Research Council of Canada (to J.R.D.), National Institutes of Health Grant HD16259 (to C. D.A), a Medical Research Council of Canada scholarship (to J.R.D), and a Manitoba Health Research Council studentship (to B.E.N.).
‡ University of Manitoba.
§ Baylor College of Medicine.

0006-2960/89/0428-0958$01.50/0