Secondary Structure of Fibronectin Type 1 and Epidermal Growth Factor Modules from Tissue-Type Plasminogen Activator by Nuclear Magnetic Resonance†

Brian O. Smith,‡ A. Kristina Downing,‡ Timothy J. Dudgeon,§ Mark Cunningham,§ Paul C. Driscoll,‡ and Iain D. Campbell*,‡

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, U.K., and British Biotechnology Ltd., Watlington Road, Cowley, Oxford OX4 5LY, U.K.

Received October 27, 1993; Revised Manuscript Received December 15, 1993®

ABSTRACT: A segment of human tissue-type plasminogen activator (t-PA) corresponding to the fibronectin type 1 (F1) and epidermal growth factor-like (G) pair of modules, residues 1–91, has been produced as a recombinant protein in *Saccharomyces cerevisiae*, with a single conservative Cys to Ser substitution. The sequence-specific assignment of the ^{1}H and ^{15}N nuclear magnetic resonances from the pair of modules has been completed using 2D ^{1}H nuclear magnetic resonance (NMR) spectra in conjunction with 3D, ^{15}N-edited, ^{1}H and 2D ^{15}N–^{1}H NMR spectra. Slowly exchanging amide protons have been identified, and estimates of a number of backbone $^3J_{NH-C\alpha H}$ coupling constants were obtained by line-shape-fitting. The secondary structure of the F1 module in the pair closely matches that previously determined for the isolated F1 module from t-PA, and that of the G module conforms to the "consensus" G module structure determined previously from several isolated G modules. In the module pair, the residues linking the two modules appear to form an extended β-strand, the carboxy-terminal end of which makes up a third strand of the major β-sheet of the G module. The intermodule interface is defined by NOEs between residues in the ranges 22–24 in the F1 module and 65–72 in the G module. The NMR data indicate that there is little or no reorientation of the two modules with respect to one another but rather that they combine with a fixed hydrophobic contact dominated by the side chain of leucine-22.

Tissue-type plasminogen activator (t-PA)[1] is a blood plasma glycoprotein with a critical role in fibrinolysis. Specifically, it is a serine protease which catalyzes the conversion of the zymogen Glu-plasminogen to its active form, plasmin, which is the enzyme responsible for degrading the fibrin network of a blood clot. Unlike other thrombolytic agents such as streptokinase or urokinase, the action of t-PA is localized to the site of the clot, forming a ternary complex with plasminogen and fibrin, thus avoiding systemic fibrinolytic activation (Weimar et al., 1981; Van de Werf et al., 1984). Despite its favorable localization of action, t-PA has not yet been demonstrated to have a markedly better clinical profile than other thrombolytics (Hunt et al., 1992). This is in large part due to its short half-life in the circulation [see Krause (1988) and references cited therein]. Consequently, considerable attention has focused on the structural details of the interaction between t-PA and fibrin and of the mechanisms by which it is cleared from the bloodstream.

Like many other blood plasma proteins, t-PA has a mosaic gene structure (Patthy, 1985) consistent with a modular domain pattern in the expressed protein. As a result, different parts of t-PA share considerable homology with regions of other proteins (Ny et al., 1984). The mature 527 amino acid polypeptide folds into 5 discrete structural domains which show sequence homology to 4 previously identified modules: the first domain resembles the type 1 module of fibronectin (Bányai et al., 1983); the second, epidermal growth factor (Campbell & Bork, 1993); the third and fourth, the kringle (K) modules of plasminogen; and the fifth, a chymotrypsin-like serine protease.

It has been suggested that each of the domains performs autonomous functions in t-PA (van Zonneveld et al., 1986a,b), a hypothesis that has been tested by domain deletion and addition experiments in several groups (van Zonneveld et al., 1986c; Verheijen et al., 1986; Gething et al., 1988; Browne et al., 1988). However, a number of more detailed mutagenesis studies have shown that there is overlap of functionality between the various domains. t-PA displays biphasic fibrin binding kinetics, which led van Zonneveld et al. (1986c) to propose that the F1 domain governs the initial interaction between t-PA and fibrin with the lysine-specific site in the K2 domain mediating the second phase. However, high-resolution structure–function analysis has shown fibrin binding to be affected by mutations in the F1, G, first K, and serine protease domains but not the second K domain (Ahern et al., 1990; Bennett et al., 1991). Three mechanisms have been proposed for hepatic t-PA clearance: one involving N-glycosylation at asparagine-448 in the serine protease module (Lau et al., 1987) or at asparagine-117 in kringle 1 (Hotchkiss et al., 1988), another involving the serpin plasminogen activator inhibitor 1 (PAI-1) (Owensby et al., 1988), and a third involving a determinant in the G module (Bassel-Duby et al., 1992; Nguyen et al., 1992). Site-directed mutagenesis studies have shown that the glycosylation and PAI-1-independent mechanism can be abolished by substitutions to residues toward

† This work is a contribution from the Oxford Centre for Molecular Sciences supported by SERC. B.O.S. was supported by British Biotechnology Ltd. through a CASE award. P.C.D. is a Royal Society University Research Fellow.

‡ University of Oxford.

§ British Biotechnology Ltd.

® Abstract published in *Advance ACS Abstracts,* February 1, 1994.

[1] Abbreviations: F1, fibronectin type 1 module; G, epidermal growth factor module; K, kringle module; t-PA, tissue-type plasminogen activator; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser spectroscopy; HOHAHA, homonuclear Hartmann–Hahn; COSY, correlation spectroscopy; HMQC, heteronuclear multiple quantum correlation; HSQC, heteronuclear single quantum correlation.

0006-2960/94/0433-2422$04.50/0

the end of the F1 domain (Ahern et al., 1990) and residues 66–69 in the G module (Edwards et al., 1989; Bassel-Duby et al., 1992; Browne et al., 1990).

The solution structures of the F1 module from t-PA (Downing et al., 1992) and three F1 modules from fibronectin (Baron et al., 1990; Williams et al., 1993) have previously been determined by ^{1}H NMR spectroscopy. Several G module structures have also been determined by ^{1}H NMR spectroscopy [see Campbell and Bork (1993) for a review] and X-ray crystallography (Padmanabhan et al., 1993). The goals of this work were to compare the structures of single modules with those in the pair and to observe how the two modules combine.

EXPERIMENTAL PROCEDURES

Expression of t-PA F1–G. t-PA F1–G was produced by an expression system modified from the yeast α-factor secretion system (Brake et al., 1984). *Saccharomyces cerevisiae* were transformed with the yeast/*Escherichia coli* shuttle plasmid pBS4 containing the yeast α-factor prepro sequence fused in-frame with a DNA fragment coding for the first 91 residues of t-PA. mRNA synthesis was under the control of a galactose-inducible promoter and the *S. cerevisiae* phosphoglycerate kinase terminator. The 91-residue polypeptide was directed into the culture supernatant by the α-factor leader (Kurjan & Herskowitz, 1982).

Initially, expression was carried out using the *S. cerevisiae* strain MC2 (a leu2–3, 112 ura3–52 trp1–1 pep4–3 prb1–1122 prc1–407). Briefly, transformed cultures were grown in shake flasks for 48 h in YNB medium (6.7 g/L yeast nitrogen base supplemented with a cocktail of nucleotide precursors and amino acids from which leucine was excluded to maintain selection for cells transformed with the plasmid) containing 10 g/L glucose, to bulk up the biomass of the culture. Following this, the cells were transferred to YNB containing 2 g/L glucose and 10 g/L galactose to induce expression of t-PA F1–G into the culture supernatant. Although capable of producing small quantities of protein, this process was not suitable for the production of the larger quantities required for high-resolution studies. Therefore, a 5-L high cell density fed–batch fermentation was developed using the *S. cerevisiae* strain MC101 (a leu2-3,112) transformed with the vector pBS4. The process used conditions optimized for this strain and consisted of two stages. Initially, the culture was grown in a defined salts medium (containing no additional amino acids) while being fed glucose at a predetermined rate. This growth phase was used to produce a culture with a cell density of $\sim$150 A_{600} units. At this point, the feed was stopped, and galactose (100 g) was batched into the fermenter to induce protein secretion into the culture supernatant. The culture was harvested 24-h post-galactose addition and the supernatant recovered by continuous-flow centrifugation.

Construction of Vectors. The plasmid pBS4 is a derivative of pSW6 (Pascall et al., 1991) with the *Hin*dIII–*Bam*HI fragment replaced by a fragment containing the portion of the α-factor prepro sequence distal to the *Hin*dIII site fused to the coding sequence for the F1 and G pair of modules. The appropriate *Hin*dIIII and *Bam*HI restriction sites were introduced into a M13mp19 cDNA clone of t-PA by oligonucleotide site-directed mutagenesis as described by Kunkel et al. (1987). Oligonucleotide 5′-GATCACTTGGTAAGATCTTTTATCCAAGCTTAATCGGGCATGGAT introduced the *Hin*dIII site and the portion of the prepro sequence distal to it so that the codon for Ser-36 of the t-PA precursor, which corresponds to the first residue of the mature protein (Pennica et al., 1983), immediately follows the Lys-Arg codons which make up the KEX 2 protease recognition site at the end of the prepro sequence. Oligonucleotide 5′-GATACCAGGGCCACGTAATAAGGATCCCAGGGCATCAGCTAC introduced two stop codons followed by the *Bam*HI site. The stop codons were placed to terminate translation after Thr-91 of the mature protein, the residue immediately preceding the first cysteine of the kringle 1 module and four residues after the site of the fifth intron of t-PA (Friezner Degen et al., 1986). The locations of the *Hin*dIII and *Bam*HI sites in the primers are underlined. Expression of the native t-PA F1–G polypeptide failed probably due to aggregation through disulfide bonding caused by reduction of the free cysteine (Cys-83); therefore, a point mutation was introduced using oligonucleotide 5′-TATTTCACAGGACTTCCCAG to cause the substitution Cys to Ser at position 83, a substitution known not to affect the biological activity of t-PA (Yuzuriha et al., 1989; Suzuki et al., 1989). The base used to introduce the point mutation is underlined. Correctly altered clones were identified by [^{35}S]dATP-labeled dideoxy sequencing using Sequenase (United States Biochemicals) according to the manufacturer's instructions.

Sample Preparation. The protein was initially concentrated and partially purified by cation-exchange chromatography. Yeast fermentation supernatants were adjusted to pH 5.4 by addition of $^1/_{40}$th volume of 1 M sodium acetate, pH 5.4, and passed through a 10-cm-diameter column packed with fast-flow S-Sepharose (Pharmacia), preequilibrated at pH 5.4 and 4 °C. The column was washed with approximately 10 column volumes of 25 mM sodium acetate, pH 5.4, followed by 3 column volumes of 25 mM sodium acetate (pH 5.4)/0.15 M NaCl. The protein was finally eluted with 25 mM sodium acetate (pH 5.4)/0.5 M NaCl. The effect of each buffer change was monitored by following the A_{280} of the flow-through. The protein was further purified by high-performance liquid chromatography (HPLC) on a reverse-phase column (C18, 250 mm × 10 mm). The eluate from the cation-exchange step was made up to 10% acetonitrile/0.1% trifluoroacetic acid and loaded onto the column, a gradient was run from 20% to 40% acetonitrile in 0.1% aqueous trifluoroacetic acid, and the protein was observed to elute at approximately 35% acetonitrile as detected by A_{280}. Fractions containing the protein were collected and subjected to a further round of reverse-phase HPLC before lyophilization. The purified protein was positively identified by a combination of N-terminal sequence analysis and electrospray mass spectrometry.

NMR Analysis. Samples for NMR spectroscopy contained approximately 2–4 mM protein in D_2O or 90% H_2O/10% D_2O, pH 2.95. Spectra were recorded at 25, 30, and 35 °C on a Bruker AM600 (600.1 MHz for ^{1}H) spectrometer or on spectrometers built in-house here at the Oxford Centre for Molecular Sciences (500.1 MHz for ^{1}H, 50.7 MHz for ^{15}N) interfaced with GE/Nicolet or GE/Omega computers. Multidimensional experiments were recorded with time-proportional phase incrementation (Bruker) or States-TPPI for quadrature detection in the indirectly detected dimensions. Suppression of the solvent resonance was achieved either by presaturation or by use of a jump–return read pulse (Driscoll et al., 1989). ^{1}H 2D nuclear Overhauser effect spectroscopy (NOESY) (Kumar et al., 1980), homonuclear Hartmann–Hahn (HOHAHA) (Braunschweiler & Ernst, 1983; Davis & Bax, 1985), and double quantum filtered correlated spectroscopy (DQF-COSY) (Rance et al., 1983) experiments, with sweep widths of 7042.25 Hz (Bruker) or 6024.1 Hz in both dimensions, were acquired as 2048 real points (Bruker) or 1024 complex points in f_2 and up to 1024 real points (Bruker)

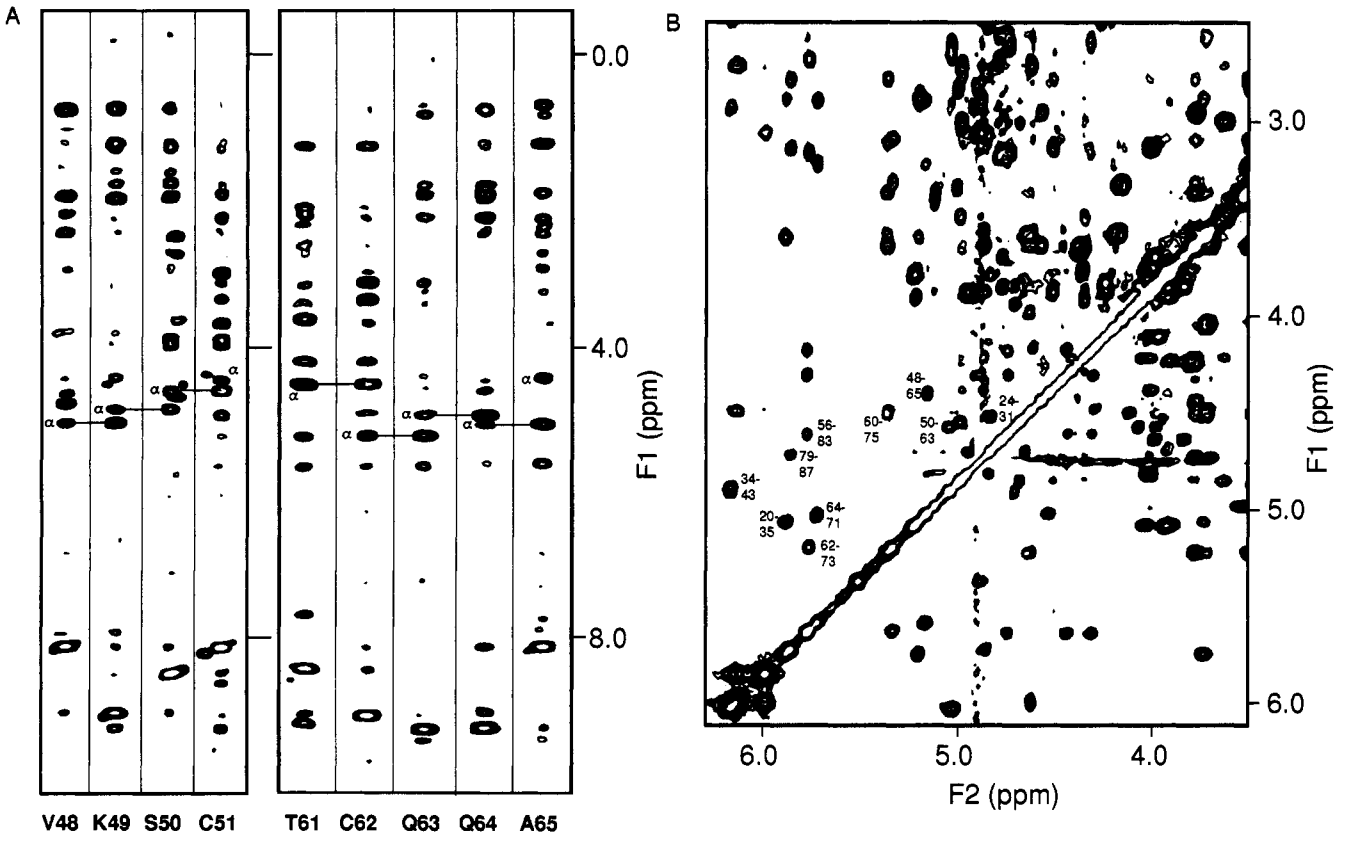

FIGURE 1: Examples of the NMR spectra obtained from the F1–G pair of modules from t-PA. (A) Strips showing cross-peaks to the NH resonance of the residues indicated, extracted from the ^1H–^1H planes of a ^{15}N-edited 3D NOESY recorded with 150-ms mixing time. CαH cross-peaks are labeled, and CαH_i to NH_{i+1} sequential NOEs are indicated by horizontal lines. (B) A portion of a 2D D_2O NOESY recorded with 150-ms mixing time. A number of CαH to CαH cross-peaks are labeled.

or 1024 complex points in f_1. Mixing times between 60 and 200 ms were used in the 2D NOESY spectra. A WALTZ-17y spin lock sequence (Bax et al., 1987) of approximately 35-ms duration was used in the 2D HOHAHA experiments. ^{15}N–^1H 2D heteronuclear single quantum correlation (HSQC) (Bodenhausen & Ruben, 1980) and heteronuclear multiple quantum correlation J (HMQCJ) (Kay & Bax, 1990) experiments with sweep widths of 1014.2 and 7042.25 Hz in f_1 and f_2, respectively, were acquired as 256 (t_1) × 512 (t_2) complex points. ^1H, ^{15}N 3D NOESY–HMQC and HOHA-HA–HMQC (Messerle et al., 1989; Driscoll et al., 1990) experiments with sweep widths of 5376.34, 1014.2, and 7042.25 Hz in f_1, f_2, and f_3, respectively, were acquired as 128 (t_1) × 32 (t_2) × 512 (t_3) complex points. Each 3D spectrum was recorded with 8 scans per increment over a period of about 3 days. A mixing time of 150 ms was used in the NOESY–HMQC experiment, and a DIPSI-3 (Shaka et al., 1988) composite pulse sequence of approximately 32-ms duration was used in the HOHAHA–HMQC experiment. Data processing was performed using the FELIX 1.1 and 2.05 software packages (Hare Research Inc.) on Sun workstations.

RESULTS

An expression system, using a defined growth medium suitable for isotopic labeling and giving adequate yields, was successfully developed for producing the t-PA F1–G pair of modules in *S. cerevisiae*. In conjunction with a simple and efficient purification protocol, yields of approximately 20 mg of pure protein per liter of culture medium were obtained.

The sequence-specific assignment of the F1–G pair of modules from t-PA was initially attempted using only 2D data; however, even with the assistance of the assignments of the F1 module made previously (Downing et al., 1992), overlap of the ^1H resonances prevented a complete assignment. Two problems in particular contributed to these difficulties. First, the part of the fingerprint region between 8.3 and 8.6 ppm in f_2 was particularly difficult to interpret, containing cross-peaks to 21 NH resonances. For nine of these residues, the intraresidue NH to CαHH cross-peaks were partly or wholly overlapped with the CαH resonances lying between 4.5 and 4.8 ppm. Second, many cross-peaks to the NH protons of residues in the F1 module showed evidence of significant exchange-broadening due to the effect of an unidentified conformational equilibrium. Therefore, a strategy using 3D heteronuclear experiments for resonance assignment (Driscoll et al., 1990) was adopted involving the production of uniformly ^{15}N-labeled protein. Despite the problems mentioned with the fingerprint region of the 2D spectra, generally good quality 2D D_2O and 3D spectra were obtained (Figure 1).

The heteronuclear experiments were used to correlate CαH resonances with ^{15}N and ^1H resonances of the backbone amide group of the same residue. The 3D spectra permitted the unravelling of most of the overlap problems seen in the 2D spectra, with the cross-peaks to only a few pairs of NH resonances remaining unresolved in the ^{15}N dimension. Where possible, spin system assignments were extended to the side-chain resonances through homonuclear correlations in 2D and 3D spectra. The sequence-specific assignment of the backbone resonances of all residues was completed by the analysis of strong sequential NOEs observed between the CαH and CβH resonances of one residue and the amide proton resonance of the following residue (Wüthrich, 1986). In addition, almost all side-chain resonances were assigned unambiguously. A list of proton chemical shiftt assignments is given in Table 1.

Wherever possible, assignments were made for cross-peaks in the various spectra, and, from the shorter mixing time

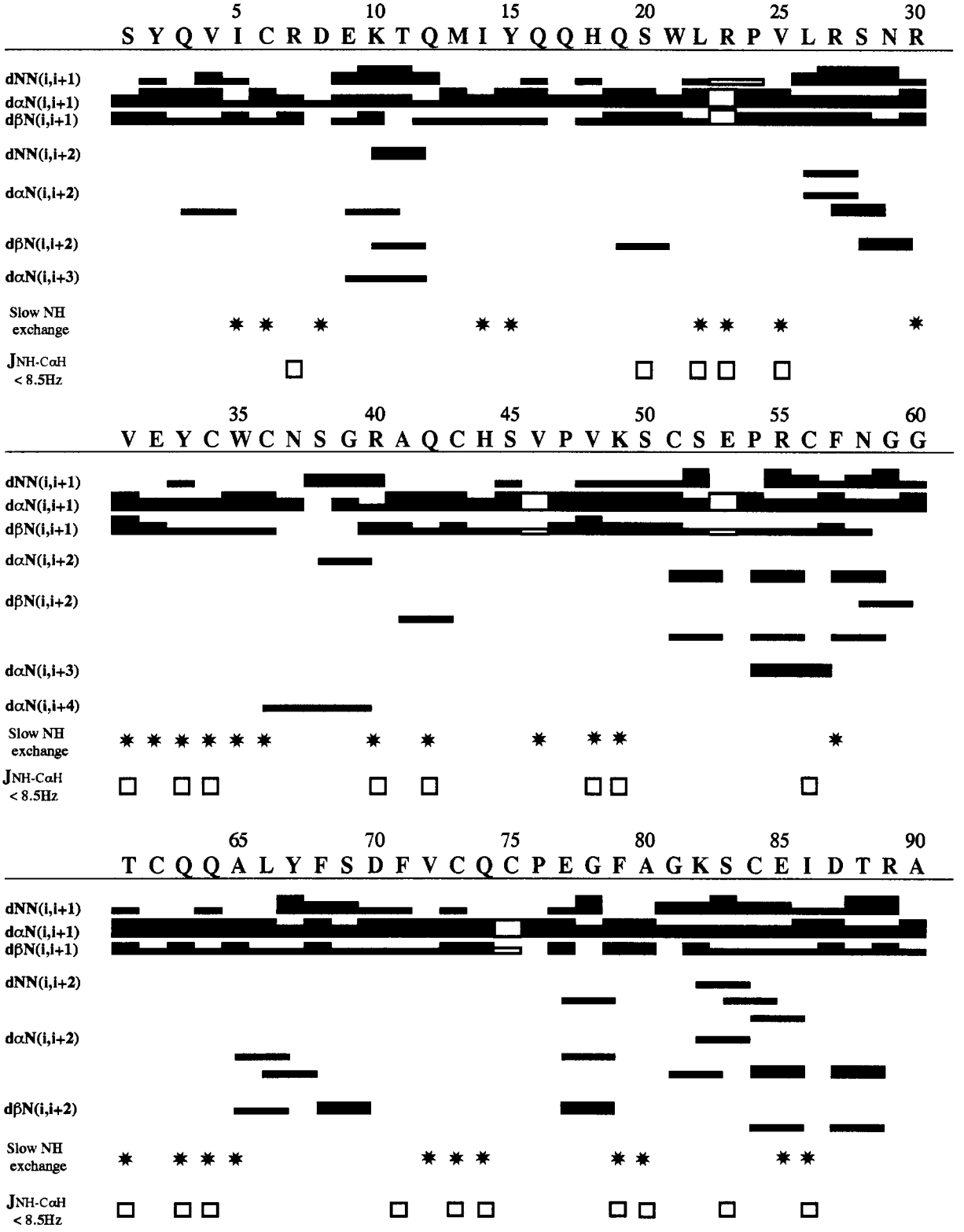

FIGURE 2: Summary of sequential- and medium-range backbone NOEs, slowly exchanging amide protons, and measurable high (>8.5Hz) $^3J_{NH-C\alpha H}$ values. The bars join residues i to $i+n$ as indicated with the width of the bar corresponding to either a strong, medium, or weak NOE, categorized by the relative cross-peak intensity. An unfilled bar is used to denote an NOE to the CδH protons of a proline residue.

NOESY spectra, interproton distances were derived and classified as short, medium, or long depending on the relative cross-peak intensities. This gave rise to 14 CαH to CαH NOEs, 38 NH_i to CαH_j (where $|i-j|>4$) NOEs, and 17 NH_i to NH_j NOEs (where $|i-j|>4$).

Estimates of a number of backbone coupling constants ($^3J_{NH-C\alpha H}$) were made by fitting simulated line shapes to cross sections of peaks taken from MHQCJ spectra (Kay & Bax, 1990). Those giving coupling constants greater than 8.5 Hz are shown in Figure 2. In addition, 33 slowly exchanging amide protons were identified by lyophilizing a sample from aqueous solution, redissolving it in D_2O, and immediately acquiring a 2D HOHAHA experiment.

DISCUSSION

The elements of regular secondary structure of a protein can be identified from the pattern of sequential-, medium-, and long-range NOEs along with backbone $^3J_{NH-C\alpha H}$ values

Table 1: Proton Resonance Assignments of t-PA F1–GF at 298 K and pH 2.95[a]

residue	NH	CαH	CβH	other
Ser-1	–	4.47	4.06, 3.92	
Tyr-2	8.58	4.60	3.00, 2.94	2,6H 7.08, 3,5H 6.78
Gln-3	8.23	4.30	1.96, 1.84	CγH 2.24*, NH –
Val-4	8.16	4.02	1.94	$C\gamma H_3$ 0.84, 0.74
Ile-5	7.57	4.70	1.82*	CγH 1.22, 1.00, $C\gamma H_3$ 0.74, $C\delta H_3$ 0.68
Cys-6	8.97	4.80	3.00, 2.68	
Arg-7	8.57	4.72	1.60, 1.43	CγH 1.24, 1.14, CδH 3.04, 3.00, NH 7.16
Asp-8	8.54	4.73	3.03, 2.25	
Glu-9	8.95	4.04	2.13, 2.06	CγH –
Lys-10	8.61	4.32	2.05, 2.00	CγH 1.49*, CδH 1.70*, CεH 3.00*
Thr-11	7.87	4.33	4.36	$C\gamma H_3$ 1.28
Gln-12	8.10	3.74	2.16*	CγH –, NH –
Met-13	7.42	4.28	1.56, 1.32	
Ile-14	7.53	4.27	1.42	CγH 0.80*, $C\gamma H_3$ 0.61, $C\delta H_3$ 0.66
Tyr-15	8.55	4.45	2.68, 1.80	2,6H 5.96, 3,5H 5.82
Gln-16	8.52	4.36	1.75*	CγH 2.38, 2.34, NH 7.55, 6.71
Gln-17	8.28	4.16	1.94, 1.78	CγH 2.21*, NH –
His-18	9.13	4.13	–	2H –, 4H –
Gln-19	8.43	4.62	2.42*	CγH 2.65*, NH –
Ser-20	8.42	5.70	3.55*	
Trp-21	8.60	4.70	3.00, 2.86	2H 6.54, 4H 6.21, 5H 6.38, 6H 6.25 7H 6.83 NH 9.76
Leu-22	8.14	5.32	1.51, 1.06	CγH 1.58, $C\delta H_3$ 0.73, 0.62
Arg-23	9.09	5.04	1.79*	CγH 1.66, 1.62, CδH 3.32, 3.19, NH –?
Pro-24		4.60	2.13*	CγH 2.24*, CδH 3.74*
Val-25	8.28	4.04	1.88	$C\gamma H_3$ 0.84, 0.73
Leu-26	8.37	4.12	1.65*	CγH 1.53, $C\delta H_3$ 0.92, 0.88
Arg-27	8.69	4.04	1.95, 1.88	CγH 1.57*, CδH 3.20*, NH 7.22
Ser-28	7.89	4.59	3.81, 3.67	
Ash-29	8.38	4.73	2.89*	NH 6.81, 2.90
Arg-30	8.06	4.61	1.73, 1.59	CγH 1.48*, CδH 3.11*, NH 7.15
Val-31	8.85	4.49	1.84	$C\gamma H_3$ 0.72, 0.66
Gln-32	8.57	4.70	1.76, 1.61	CγH –, NH –
Tyr-33	8.79	4.71	2.44, 2.38	2,6H 6.57, 3,5H 6.31
Cys-34	8.41	4.87	0.91, 0.04	
Trp-35	7.88	5.02	2.85*	2H 7.03, 4H 7.19, 5H 6.85, 6H 6.96, 7H 7.06 NH 9.90
Cys-36	8.37	4.82	3.45, 2.96	
Ash-37	8.63	4.73	2.79*	NH 7.45, 6.97
Ser-38	9.62	3.84	4.07, 3.79	
Gly-39	6.92/6.86	4.20, 3.60		
Arg-40	7.46	4.58	1.74*	CγH 1.53*, CδH 3.18, 3.09, NH –
Ala-41	8.76	4.51	$C\beta H_3$ 1.34	
Gln-42	8.76	4.50	1.74, 1.66	CγH 2.15*, NH –
Cys-43	8.77	5.97	2.89, 2.44	
His-44	9.01	4.94	3.35*	2H 8.47, 4H 7.32
Ser-45	9.01	4.78	3.83*	
Val-46	8.62	5.04	2.24	$C\gamma H_3$ 0.94, 0.77
Pro-47		4.72	2.35, 1.85	CγH 2.08*, CδH 3.86, 3.73
Val-48	8.06	4.97	1.91	$C\gamma H_3$ 0.70, 0.64
Lys-49	8.97	4.77	1.86, 1.68	CγH 1.22, 1.13, CδH 1.53* CεH –
Ser-50	8.45	4.53	3.90, 3.81	
Cys-51	8.07	4.39	3.59, 2.91	
Ser-52	8.55	4.18	3.88, 3.83	
Glu-53	7.36	4.69	1.93, 1.60	CγH 2.20*
Pro-54		4.29	2.01*	CγH 1.89*, CδH 3.60, 3.52
Arg-55	7.20	4.01	1.08*	CγH 1.19, 0.84, CδH 2.36, 2.23, NH 6.85
Cys-56	7.72	4.57	2.54*	
Phe-57	8.57	4.57	3.13, 2.59	
Asn-58	9.04	3.98	2.02, 1.39	NH –
Gly-59	8.42	4.00, 3.54		
Gly-60	7.62	4.45, 3.54		
Thr-61	8.38	4.43	4.11	$C\gamma H_3$ 1.17
Cys-62	9.03	5.14	3.27, 3.05	
Gln-63	9.20	4.87	1.86, 1.71	CγH 2.13*, NH 7.42, 6.59
Gln-64	9.19	4.98	2.15, 1.82	CγH 2.38, 2.28, NH 7.79, 6.79
Ala-65	8.07	4.36	$C\beta H_3$ 1.16	
Leu-66	7.87	3.74	0.87, 0.78	CγH 0.97, $C\delta H_3$ 0.66, 0.60
Tyr-67	7.91	4.60	3.14, 2.53	2,6H 7.02, 3,5H 6.74
Phe-68	7.25	4.82	3.23, 2.81	2,6H 7.00, 3.5H 7.04, 4H –
Ser-69	8.62	4.45	3.94*	
Asp-70	7.69	4.85	2.63, 2.56	
Phe-71	8.27	5.54	3.17, 2.85	2,6H 7.12, 3,5H 7.34, 4H 7.25
Val-72	9.36	4.48	1.98	$C\gamma H_3$ 0.97, 0.75
Cys-73	8.80	5.58	3.12, 2.65	
Gln-74	9.13	4.60	2.13, 2.03	CγH 2.25*, NH 7.51, 6.79
Cys-75	9.01	5.18	3.06, 2.75	
Pro-76		4.66	2.46, 2.10	CγH 2.22, 2.00, CδH 3.33*
Glu-77	8.44	4.09	2.06, 1.97	CγH 2.45*
Gly-78	8.64	3.99, 3.29		

Table 1 (Continued)

residue	NH	CαH	CβH	other
Phe-79	7.92	5.67	3.11, 2.76	2,6H 7.02, 3,5H 7.18, 4H –
Ala-80	9.47	4.81	$C\beta H_3$ 1.33	
Gly-81	8.09	4.76, 3.84		
Lys-82	9.12	4.06	2.01*	CγH 1.51, 1.43, CδH 0.77*, CεH 1.67*
Ser-83	8.52	5.59	4.27, 4.11	
Cys-84	7.72	4.33	3.83, 3.09	
Glu-85	9.62	4.14	2.24, 2.02	CγH 2.56*
Ile-86	8.66	4.15	1.43	CγH 1.04*, $C\gamma H_3$ 0.71, $C\delta H_3$ 0.78
Asp-87	8.85	4.67	3.03, 2.73	
Thr-88	8.39	4.18	4.24	$C\gamma H_3$ 0.97
Arg-89	7.84	4.15	1.78, 1.73	CγH 1.52*, CδH 3.09*, NH 7.17
Ala-90	7.94	4.32	$C\beta H_3$ 1.34	
Thr-91	7.86	4.34		$C\gamma H_3$ –

[a] Chemical shifts are given in ppm with respect to TSP; (–) indicates unassigned or unobserved resonances; an asterisk indicates protons with degenerate chemical shifts.

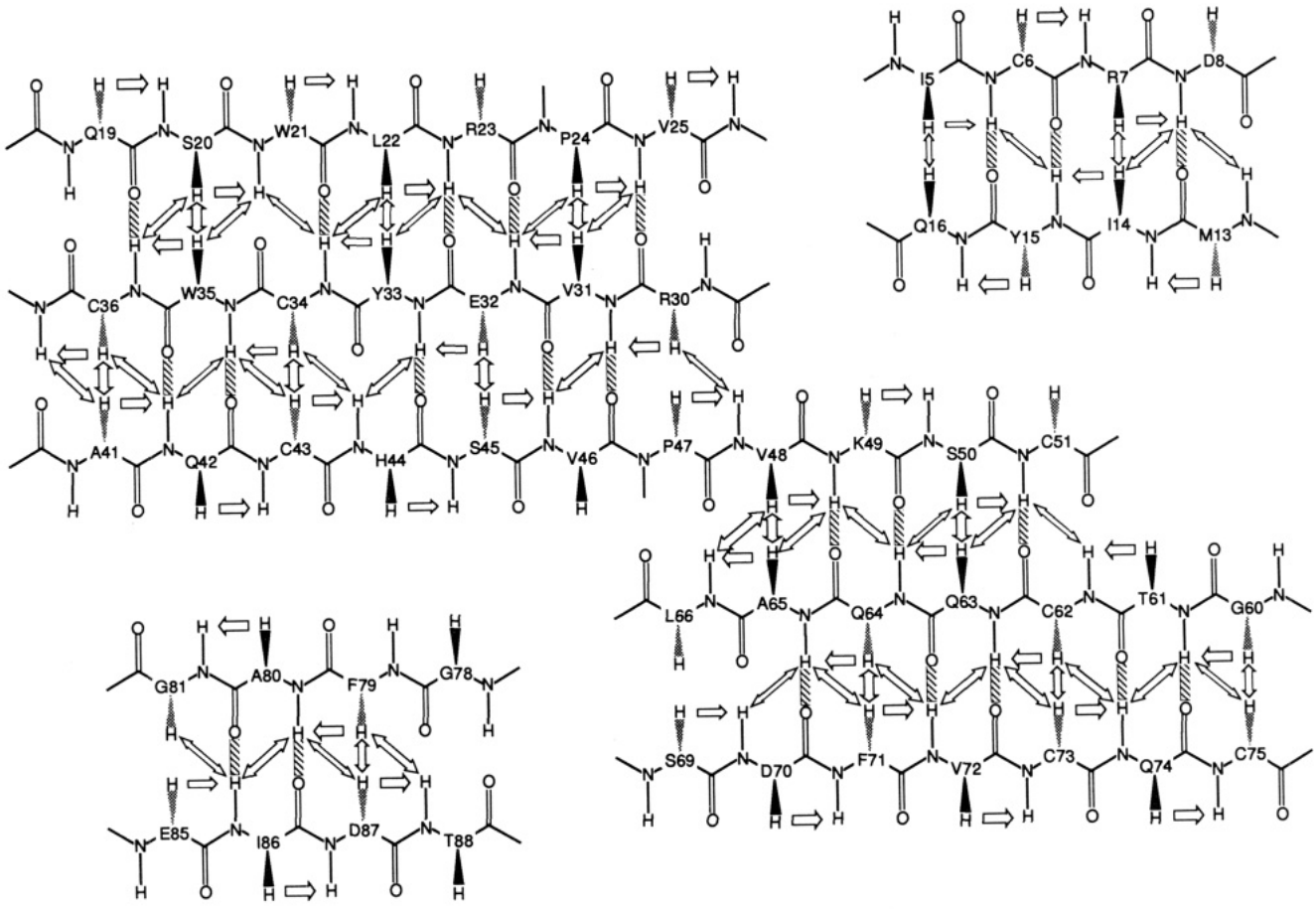

FIGURE 3: Representation of the backbone atoms of β-strands of F1–G deduced from backbone to backbone NOE and slowly exchanging amide proton data. Sequential and interstrand connectivities are denoted by double-headed arrrows whose width corresponds to either weak, medium, or strong NOEs. Putative hydrogen-bonding patterns for slowly exchanging backbone amide protons are denoted by grey boxes.

(Wüthrich, 1986). Figure 2 summarizes the sequential- and medium-range NOE data observed for the F1–G pair of modules from t-PA. The data exhibit a large number of strong sequential CαH to NH NOEs and an absence of NH to NH NOEs as expected for a structure that is largely composed of β-sheets. The features of the F1 module consensus structure (Baron et al., 1991) are, from the N-terminus, a short double-stranded antiparallel β-sheet followed by a longer triple-stranded antiparallel β-sheet. The two sheets are constrained to fold together, enclosing a conserved hydrophobic core by a single disulfide bond, with another disulfide linking the second and third strands of the major β-sheet. The dominant features of the G module consensus structure (Campbell & Bork, 1993) are, from the N-terminus, two loops and two regions of double-stranded antiparallel β-sheet. The structure is constrained by three disulfide bonds, two of which anchor the start of each of the loops (residues 7–17 and 15–17 in EGF) to the first and second strands of the major β-sheet, respectively, while the third causes the minor β-sheet to fold back on the major β-sheet. The pattern of strands and turns derived from the NMR data is entirely consistent with the consensus G and F1 module structures. Figure 3 shows the pattern of sequential and interstrand NOEs and putative hydrogen bonds, deduced from CαH to CαH NOEs and slowly exchanging amides, which define the sheets.

Figure 4 shows two representations of the chemical shift data for the t-PA F1–G pair of domains which illustrate that the structure contains F1 and G modules with consensus secondary structure. Figure 4A shows that the CαH chemical shift values for the F1 module in the pair of domains match

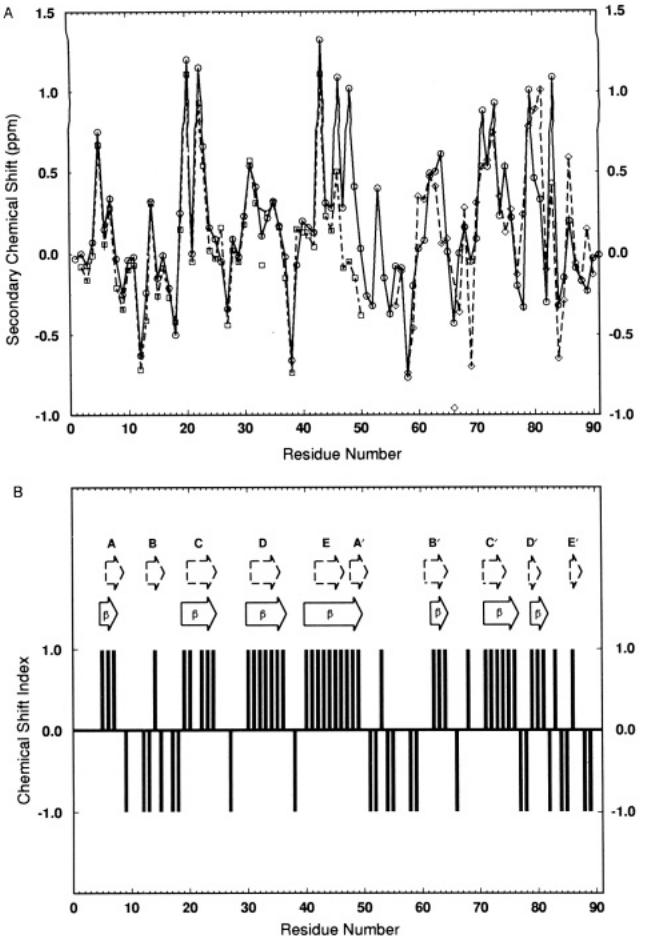

FIGURE 4: Secondary shifts of CαH protons. (A) Difference between the observed and random-coil chemical shifts (Wishart et al., 1992) of CαH protons plotted for F1–G (circles), for F1 single (1–50) (Downing *et al.*, 1992) (squares), and for an aligned subset of residues from hEGF (Cooke *et al.*, 1990) (diamonds). Note the change in secondary shift for residues 47–50 (see text). (B) Chemical shift index for F1–G according to the method of Wishart et al. (1992) with the predicted β-strands shown by solid arrows and the β-strands for which interstrand NOEs and slowly exchanging amides are observed shown by dashed arrows.

those of the isolated t-PA F1 module very closely while the CαH secondary chemical shift values (defined as their difference between the observed CαH shift and the random-coil value for that residue) for the t-PA G module follow similar trends to those of an aligned subset of residues from EGF. Wishart *et al.* (1992) have shown that, for Cα protons, the secondary chemical shift correlates strongly with the secondary structure in which a residue is found. Figure 4B shows an analysis of the chemical shift index. The profile clearly shows the β-strand structure consistent with the consensus module structures.

The NMR spectra exhibit a number of interesting features which indicate that a number of dynamic processes may be at work in the F1–G pair of modules. Most noticeably, two distinct chemical shifts are observed for the amide proton of Gly-39 which displays resonances which are broadened in both ^{15}N and ^{1}H dimensions. The residues in the minor β-sheet of the F1 module also have anomalously broadened ^{15}N resonances. A fuller treatment of the dynamics of the molecule will be presented shortly.

It is interesting to consider the way the modules of a mosaic protein intteract with one another. The "phase 1" class of modules (Patthy, 1991) have largely been shown to be autonomously folding domains with a correspondence between exon structure and units of secondary and tertiary structure

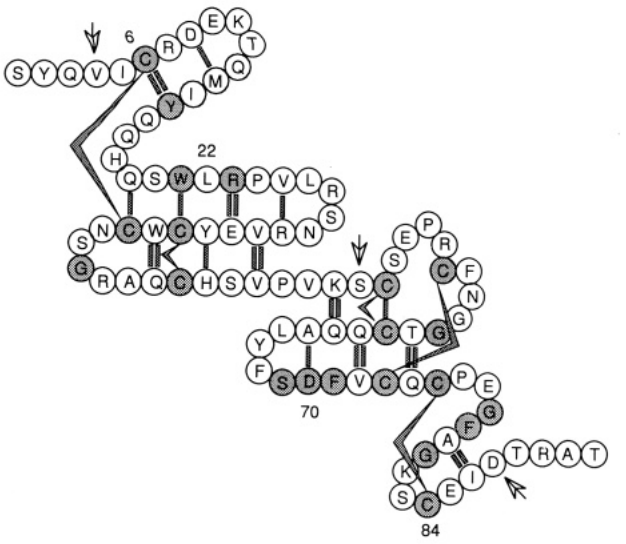

FIGURE 5: Schematic representation of the F1–G pair of modules from t-PA. Residues that form the consensus sequences for each module are shaded. Disulfide bonds are indicated by chevrons. The locations of the phase 1 introns are indicated by large arrows. Putative hydrogen bonds are represented by shaded boxes. Note that the last three residues encoded by the F1 exon appear to form part of the secondary structure of the G module.

(Patthy, 1991; Baron, 1991). With a sequential pair of modules from a mosaic protein, there are two sets of interactions which may govern their relative orientation: the first involves the sequence of residues that fall between the carboxy-terminal end of one module and the amino-terminal end of the next; the second involves residues which are distant from the linker in the primary sequence.

In the F1–G pair of modules from t-PA, the intron between the F1 and G modules falls in Ser-50 (see Figure 5), making the linker between the modules a difficult one to define. The last residue of the isolated F1 module to be structurally well-defined is Val-46 (Downing et al., 1992), so we might expect residues 47–50 to form a flexible linker similar to that seen between C modules in the complement protein factor H (Barlow et al., 1993). G modules usually have four to five residues between the intron and the first Cys (residue 51 in t-PA); in t-PA, this G module leader is absent because the intron falls in Ser-50. Our data suggest that Pro-47 to Ser-50 forms a structured, extended strand which adds a short third β-strand to the G module major β-sheet. This interaction is defined by several NOEs including CαH to CαH NOEs between Val-48 and Ala-65 and by putative hydrogen bonds between Lys-49 and Gln-64.

The change of structure for residues 47–50, from random coil in the isolated F1 module to β-sheet in the F1–G pair of modules, is further exemplified by the change in CαH chemical shifts (see Figure 4). The CαH chemical shift is a sensitive indicator of secondary structure (Wishart et al., 1991), with resonances shifted upfield from random-coil values indicative of α-helix and downfield shifts indicative of β-sheet conformations. In the F1–G pair of modules from t-PA, the CαH chemical shifts of residues 47–50 move downfield from 4.35, 3.90, 4.21, and 4.12 ppm, respectively, in the isolated F1 to 4.72, 4.97, 4.77, and 4.53 ppm, taking them from secondary shift values typical of random-coil (47, 48) or α-helical (49, 50) conformations to secondary shift values indicative of β-sheet (47–49) or random-coil (50) conformations.

The intermodule interface is implicitly defined in part by the shortness of the linker between the two modules; only Pro-47 appears to be unconstrained by involvement in the secondary structure of one or the other module. This alone

would not greatly restrict the relative reorientation of the modules. The observation of a number of NOEs between residues 22–24 and 65–72 suggests that the relative orientation of the modules is fixed. In order to accommodate these long-range NOE restraints, it is likely that the pair of modules folds together to form a compact globular structure.

From the secondary structure data presented here, we can see that residues 65–69, which have been postulated to be involved in binding to the t-PA-specific hepatocyte receptor (Basel-Duby et al., 1992; Browne et al., 1990), do indeed lie in the turn between the two strands of the G module major β-sheet. Additionally, the observed secondary structure places residues 44–49, mutations to which also affect t-PA clearance (Ahern et al., 1990), close to this turn. Further insight into those residues responsible for fibrin binding and the determinants of the hepatocyte receptor binding site should come from knowledge of the full solution structure of the F1–G pair of modules, work which is in progress at present. The solution of the structure of the F1–G pair of modules should also contribute to the debate over the conformation of intact t-PA.

ACKNOWLEDGMENT

We thank Deborah Brotherton, Dr. Jonathan Boyd, and Arthur Crawford for their assistance.

REFERENCES

Ahern, T. J., Morris, G. E., Barone, K. M., Horgan, P. G., Timony, G. A., Angus, L. B., Henson, K. S., Stoudemire, J. B., Pennina, R. L.-S., & Larsen, G. R. (1990) *J. Biol. Chem. 265*, 5540–5545.

Bányai, L., Váradi, A., & Patthy, L. (1983) *FEBS Lett. 163*, 37–41.

Barlow, P. N., Steinkasserer, A., Norman, D. G., Kieffer, B., Wiles, A. P., Sim, R. B., & Campbell, I. D. (1993) *J. Mol. Biol. 232*, 268–284.

Baron, M., Norman, D., Willis, A., & Campbell, I. D. (1990) *Nature 345*, 642–646.

Baron, M., Norman, D. G., & Campbell, I. D. (1991) *Trends Biochem. Sci. (Pers. Ed.) 16*, 13–17.

Bassel-Duby, R., Jiang, N. Y., Bittick, T., Madison, E., McGookey, D., Orth, K., Shohet, R., Sambrook, J. & Gething, M. J. (1992) *J. Biol. Chem. 267*, 9668–9677.

Bax, A., Sklenar, V., Clore, G. M., & Gronenborn, A. M. (1987) *J. Am. Chem. Soc. 109*, 6511–6513.

Bennett, W. F., Paoni, N. F., Keyt, B. A., Botstein, D., Jones, A. J. S., Presta, L., Wurm, F. M., & Zoller, M. J. (1991) *J. Biol. Chem. 266*, 5191–5201.

Bodenhauseen, G., & Ruben, D. J. (1980) *Chem. Phys. Lett. 69*, 185.

Brake, A. J., Merryweather, J. P., Coit, D. G., Heberlein, U. A., Masiarz, F. R., Mullenbach, G. T., Urdea, M. S., Valenzuela, P., & Barr, P. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4642–4646.

Braunschweiler, L. R., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521–528.

Browne, M. J., Carey, J. E., Chapman, C. G., Tyrell, A. W. R., Entwisle, C., Lawrence, G. M. P., Reavy, B., Dodd, I., Esmail, A. & Robinson, J. H. (1988) *J. Biol. Chem. 263*, 1599–1602.

Browne, M. J., Chapman, C. G., Dodd, I., Esmail, A. F., & Robinson, J. H. (1990) *Thromb. Res. 59*, 687–692.

Campbell, I. D., & Bork, P. (1993) *Curr. Opin. Struct. Biol. 3*, 385–392.

Clore, G. M., Kimber, B. J., & Gronenborn, A. M. (1983) *J. Magn. Reson. 54*, 170–173.

Cooke, R. M., Tappin, M. J., Campbell, I. D., Kohda, D., Miyake, T., Fuwa, T., Miyazawa, T., & Inangaki, F. (1990) *Eur. J. Biochem. 193*, 807–815.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2820–2821.

Downing, A. K., Driscoll, P. C., Harvey, T. S., Dudgeon, T. J., Smith, B. O., Baron, M., & Campbell, I. D. (1992) *J. Mol. Biol. 225*, 821–833.

Driscoll, P. C., Clore, G. M., Beress, L., & Gronenborn, A. M. (1989) *Biochemistry 28*, 2178–2187.

Driscoll, P. C., Clore, G. M., Marion, D., Wingfield, P. T., & Gronenborn, A. M. (1990) *Biochemistry 29*, 3542–3556.

Edwards, R. M., Dawson, K. M., Fallon, A., & Craig, S. (1993) U.S. Patent 5 232 847.

Friezner Degen, S. J., Rajput, B., & Reich, E. (1986) *J. Biol. Chem. 261*, 6972–6985.

Gething, M.-J., Adler, B., Boose, J.-A., Gerard, R. D., Madison, E. L., McGookey, D., Meidell, R. S., Roman, L. M., & Sambrook, J. (1988) *EMBO J. 7*, 2731–2740.

Hotchkiss, A., Refino, C. J., Leonard, C. K., O'Connor, J. V., Crowley, C., McCabe, J., Tate, K., Nakamura, G.; Powers, D., Levinson, A., Mohler, M., & Spellman, M. W. (1988) *Thromb. Haemostasis 39*, 511–521.

Hunt, D., Varigos, J., Dienstl, F., Lechleitner, P., Debacker, G., Kornitzer, M., Cairns, J., Turpie, A., Fritzhansen, P., Skagen, K., et al. (1992) *Lancet 339*, 753–770.

Kay, L. E., & Bax, A. (1990) *J. Magn. Reson. 86*, 110–126.

Krause, J. (1988) *Fibrinolysis 2*, 133–142.

Kumar, A., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1–6.

Kunkel, T. A., Roberts, J. D., & Zakour, R. A. (1987) *Methods Enzymol. 154*, 367–382.

Kurjan, J., & Herskowitz, I. (1982) *Cell 30*, 933–943.

Lau, D., Kuzma, G., Wei, C.-M., Livingston, D. J., & Hsiung, N. (1987) *BioTechnology 5*, 953–958.

Messerle, B. A., Wider, G., Otting, G., Weber, C., & Wüthrich, K. (1989) *J. Magn. Reson. 85*, 608–613.

Nguyen, G., Self, S. J., Camani, C., & Kruithof, E. K. (1992) *J. Biol. Chem. 267*, 6249–6256.

Ny, T., Elgh, F., & Lund, B. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 5355–5359.

Owensby, D. A., Sobel, B. E., & Schwartz, A. L. (1988) *J. Biol. Chem. 263*, 10587–10594.

Padmanabhan, K., Padmanabhan, K. P., Tulinsky, A., Bode, W., Huber, R., Blankenship, D. T., Cardin, A. D., & Kisiel, W. (1993) *J. Mol. Biol. 232*, 947–966.

Pascall, J. C., Jones, D. S. C., Doel, S. M., Clements, J. M., Hunter, M., Fallon, T., Edwards, M., & Brown, K. D. (1991) *J. Mol. Endocrinol. 6*, 63–70.

Patthy, L. (1985) *Cell 41*, 657–663.

Patthy, L. (1991) *Curr. Opin. Struct. Biol. 1*, 351–361.

Pennica, D., Holmes, W. E., Kohr, W. J., Harkins, R. N., Vehar, G. A., Ward, C. A., Bennett, W. F., Yelverton, E., Seeburg, P. H., Heyneker, H. L., Goeddel, D. V., & Collen, D. (1983) *Nature (London) 301*, 214–221.

Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wütrich, K. (1983) *Biochem. Biophys. Res. Commun. 117*, 479–485.

Shaka, A. J., Lee, C. J., & Pines, A. (1988) *J. Magn. Reson. 77*, 274–293.

Suzuki, S., Nagaoka, N., Suzuki, N., Fujimori, T., & Yoshitake, S. (1989) *Thromb. Haemostasis 62*, 543 Abstr. 1724.

Van de Werf, F., Ludbrook, P. A., Bergmann, S. R., Tiefenbrunn, A. J., Fox, K. A. A., de Geest, H., Verstraete, M., Collen, D., & Sobel, B. E. (1984) *N. Engl. J. Med. 310*, 609–613.

van Zonneveld, A.-J., Veerman, H., & Pannekoek, H. (1986a) *Proc. Natl. Acad. Sci. U.S.A. 83*, 4670–4674.

van Zonneveld, A.-J., Veerman, H., MacDonald, M. E., van Mourik, J. A., & Pannekoek, H. (1986b) *J. Cell. Biochem. 32*, 169–178.

van Zonneveld, A.-J., Veerman, H., & Pannekoek, H. (1986c) *J. Biol. Chem. 261*, 14214–14218.

Verheijen, J. H., Caspers, M. P. M., Chang, T. T. G., de Munk, G. A. W., Pouwels, P. H., & Enger-Valk, B. E. (1986) *EMBO J. 5*, 3525–3530.