Sequential ^1H NMR Assignments of Kistrin, a Potent Platelet Aggregation Inhibitor and Glycoprotein IIb–IIIa Antagonist†

Marc Adler and Gerhard Wagner*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115

Received July 18, 1991; Revised Manuscript Received October 14, 1991

ABSTRACT: Sequence-specific nuclear magnetic resonances assignments have been obtained for the protons of kistrin. Kistrin is a small naturally occurring snake venom protein that inhibits platelet aggregation by blocking the interaction of fibrinogen with the membrane-bound glycoprotein IIb–IIIa (GP IIb–IIIa), a receptor from the integrin family. Kistrin has an Arg–Gly–Asp sequence which is believed to form an adhesion recognition sequence that is essential for activity. Therefore, the interaction between kistrin and GP IIb–IIIa may provide important information on the motif used by integrins to recognize their target proteins which bear RGD sequences. Kistrin consists of 68 residues and contains six intramolecular disulfide bonds. Although one-third of the amide protons are protected from exchange with the solvent, there appears to be little or no regular secondary structure. The large number of NOE's between residues separated by two and three positions in the sequence indicates that the protein contains a large number of tightly packed loops. Along with the sequential assignments, this paper also discusses the construction and use of computerized data bases for manipulating NMR results. A strategy for computer-assisted sequential resonance using these data bases is also presented.

The binding of platelets to fibrinogen is an essential step in the formation of platelet-rich blood clots (Stein et al., 1989). This process is mediated by the glycoprotein IIb–IIIa found on the surface of activated platelets. GP IIb–IIIa[1] binds to a variety of extracellular ligands, including fibrinogen, fibronectin, vitronectin, and the von Willebrand factor. It has been proposed that GP IIb–IIIa specifically binds to an Arg–Gly–Asp (RGD) sequence found on each of these ligands (Ruoslahti & Pierschbacher, 1987; Phillips et al., 1988). Potentially, a GP IIb–IIIa antagonist could be used as an antiplatelet agent for the treatment of thrombotic diseases (Ruggeri et al., 1989).

A recently identified protein called kistrin, which is purified from the venom of the Malayan pit viper *Agkistrodon rhodostoma*, contains an RGD sequence and binds to GP IIb–IIIa with nanomolar efficiency (Dennis et al., 1990). In vitro assays indicate that kistrin competitively inhibits the binding of GP IIb–IIIa to fibrinogen and inhibits platelet aggregation in human platelet-rich plasma. In canine models of coronary artery thrombosis, there is measurable improvement of the rate and extent of thrombolysis as well as prevention of reocclusion when kistrin is used in conjunction with recombinant tissue-type plasminogen activator (Yasuda et al., 1991). Kistrin is a 68 amino acid protein that contains six intramolecular disulfide bonds. It has a high degree of sequence homology with other proteins obtained from *Viperidae* venom which show similar antiplatelet activity (Dennis et al., 1990; Gould et al., 1990). There is an absolute conservation of the RGD sequence and high degree of conservation of the cysteine residues. A detailed structure of kistrin may provide important insight into its interaction with GP IIb–IIIa; this knowledge may in turn aid in the design of clinical useful antithrombotic agents. The sequence-specific assignments of proton resonances presented in this paper are the basis for the structure determination in solution. A preliminary structure has already been determined (Adler et al., 1991).

This paper also discusses an alternative approach to the automated sequential assignment programs that have appeared in literature (Billeter et al., 1988; Cieslar et al., 1988; Kleywegt et al., 1990; Nelson et al., 1991; Wand & Nelson, 1991). The computer is used to rapidly organize and display the data in forms that simplify the analysis of the spectra.

Materials and Methods

Kistrin was isolated from the venom of the Malayan pit viper *A. rhodostoma* (Dennis et al., 1990) and was provided by Drs. M. S. Dennis and R. A. Lazarus (Genentech, South San Francisco, CA). The chemical shift of the $C^{\epsilon}H_3$ of M52 was calibrated against sodium 2,2-dimethyl-2-silapentanesulfonate (DSS) in a low concentration sample of the protein (0.5 mM) dissolved in D_2O. This peak was then used as the reference in all subsequent spectra. The sequential assignments were determined using samples that contained 2.5 mM protein at pH 2.2. Measurements were made in either 100% D_2O or in a mixture of 5% D_2O and 95% H_2O. For the D_2O sample, all the amide protons were exchanged by heating the sample in D_2O to 50 °C for roughly an hour prior to lyophilization. Total correlation (TOCSY) spectra (Braunschweiler & Ernst, 1983) were collected using an MLEV-17 (Bax & Davis, 1985) spin lock. Delays were introduced on either side of the 180° pulse (clean TOCSY) so that the development of the negative ROE peaks during the spin lock were offset by a positive NOE

† We thank Dow Chemical Company for a fellowship given to Dr. M. Adler. Partial support was provided by NIH Grant GM38608. Some of the spectra were collected with an GN-500 spectrometer that was acquired with a grant from the NSF (BBS-8615223).

* To whom correspondence should be addressed.

[1] Abbreviations: DQF-COSY, double-quantum filtered correlation spectroscopy; d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$, NOE connectivity between the amide proton of residue i and the amide, α, or β protons, respectively, of the residue $i-1$; GP IIb–IIIa, glycoprotein IIb–IIIa; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; ppm, parts per million; ROE, rotating frame nuclear Overhauser effect; TOCSY, two-dimensional total correlation spectroscopy.

0006-2960/92/0431-1031$03.00/0

effect during the delays (Griesinger et al., 1986). Empirically, the best results were obtained with the delays set equal to 2.1 times the 90° pulse width. Separate data sets were recorded with the length of the spin locked varied between 36 and 95 ms. These data sets were then coadded to produce the final spectrum. Double-quantum filtered correlation (DQF-COSY) spectra (Piantini et al., 1982; Shaka & Freeman, 1983; Rance et al., 1984) were recorded in both D_2O and H_2O at 37 °C. Vicinal $^3J(H^N-H^\alpha)$ coupling constants were estimated from the H^N-H^α cross peaks in the H_2O DQF-COSY spectrum along ω_2 axis, which had been processed with a 10°-shifted sine bell. No correction was made for the line width of the peaks. Similarly, $H^\alpha-H^\beta$ coupling constants were estimated as small or large from the cross peaks in a D_2O DQF-COSY spectrum. NOESY spectra (Jeener et al., 1979; Kumar et al., 1980) were collected with the following mixing times: 200 ms (D_2O and H_2O), 80 ms (D_2O), or 60 ms (H_2O). The NOESY spectra used to prepare Figure 1 were obtained at 20 °C from a second sample of a 5 mM solution of kistrin. Measurements on this second sample were made both in H_2O and D_2O using a 200-ms mixing time. The D_2O sample was not heated prior to lyophilization, and, therefore, it contained some slowly exchanging amide protons. The baseline of the H_2O spectrum was flattened using a dispersive fit (Adler & Wagner, 1991). Phase-sensitive detection along ω_1 was achieved using time-proportional phase incrementation (TPPI) as described by Marion and Wüthrich (1983). Measurements were performed on General Electric GN-500 spectrometers.

Results

The data analysis was facilitated by creation of a data base. The purpose of this data base was (i) to organize the assignments and create lists of assigned resonances, (ii) to automatically annotate newly recorded spectra, (iii) to produce figures documenting sequential assignments, surveys of NOE connectivities in diagonal plots, and (iv) to arrange and plot selections of small portions of spectra for documentation of assignments.

Organization of the Computer Data Base. Careful attention was paid in the beginning of the project to the construction of a computerized data base for the experimental results. Specifically, we wanted to avoid generating parallel copies of the same data that could not be rapidly interconverted. For instance, Figures 1, 2, and 3 contain a large amount of information from both the sequential assignments and the list of NOE cross peaks. Programs were written to generate these figures directly from the computer data base with a minimal amount of human intervention.

The graphics package called PLOT (New Unit, Inc., Ithaca, NY) was selected as the primary programming tool. It was chosen because it had all the graphic tools necessary to produce complex plots. Furthermore, its macros have many features of a programming language, including such useful devices as subroutines, DO loops, and "if ... then ... else" constructions. However, the package does not include any user defined matrices and is only designed to handle two-dimensional data files. Also, the macros are not compiled and, therefore, run quite slowly on a computer workstation. Thus, there are practical limits on the complexity of the programs. Specifically, most programs are executed without consulting a reference file that gives the structure of each amino acid.

A standard format was developed to handle the chemical shift information for the amino acids. The chemical shift file is constructed from pairs of real numbers, consisting of the atom number and then the chemical shift. The atom number consists of the residue number, followed by a decimal point. The two digits after the decimal point are used to signify the location of protons within the residue. The protons are numbered relative to their position to the HN. The HN's are numbered as .01, $C^\alpha H$'s are numbered as .02 and .03, $C^\beta H$'s are .04 and .05, and so on. Obviously only glycines have two $C^\alpha H$'s. The other residues have a dummy proton in the second $C^\alpha H$ position, .03, and its chemical shift is set to −9.00. Other nonexistent protons which are required as placeholders, such as proline HN, are also assigned the chemical shift of −9.00. Unassigned protons are given a chemical shift of −5.00. Isoleucine 44 presents a unique problem in that its has three nondegenerate $C^\gamma H$'s. According to this system, the $C^\gamma H_3$ group is placed in the fifth position and thus treated as the second $C^\beta H$. Within this numbering system, the β-methyl of Ala-16 becomes atom 16.04; the $C^\delta H$'s of Phe-38 are numbered 38.08 and 38.09.

Using this numbering system, it is very simple to write programs that manipulate the data. One program is used to label spectra by putting a box around a cross peak and then labeling the box with the residue numbers of the protons involved in the cross peaks. For example, one subroutine of this program generates a partial set of labels for a NOESY spectrum. It uses the table of chemical shifts to predict where all the intraresidue and sequential NOE's should appear. The program makes use of nested DO loops to create x and y values that match the atom number of both protons. Intraresidue cross peaks have the property that

$$\text{int(atom\_number\_}x\text{)} = \text{int(atom\_number\_}y\text{)}$$

The variables "atom_number_x" and "atom_number_y" are the atom numbers of both protons and the function int(x) truncates the atom numbers to the nearest integer. The x and y variables are incremented until no chemical shift can be found to match the atom number, i.e., when the atom number is too large for the given type of residue. (The use of dummy protons, such as the HN of proline, prevent a premature exit from this DO loop.) The subroutine creates labels for cross peaks for sequential NOE's when

$$\text{atom\_number\_}y - \text{int(atom\_number\_}y\text{)} = 0.01$$

and

$$\text{int(atom\_number\_}x\text{)} = \text{int(atom\_number\_}y\text{)} - 1$$

and

$$\text{atom\_number\_}x - \text{int(atom\_number\_}x\text{)} < 0.055$$

The first Boolean expression specifies the HN of residue y. The second and third expressions select the HN, $C^\alpha H$(s), and $C^\beta H$(s) from the preceding residue.

Another subroutine of the same program is used to simulate a DQF-COSY spectrum. The program predicts a cross peak when

$$\text{abs[int(50*atom\_number\_}x\text{)} - \text{int(50*atom\_number\_}y\text{)]} < 1.5$$

abs[x] obtains the absolute value of x. The function int(50*x) sets the atom numbers for all methylene pairs of the same residue equal to each other. As an example, for the α, β, and γ protons of K20, it becomes 1001, 1002, and 1003, respectively. Therefore, this Boolean argument is true when two protons are separated by no more than three covalent bonds. Unfortunately this simple expression fails for the branched amino acids such as valine or isoleucine. However, it gives a very good approximation of a DQF-COSY without having to check the structure of each amino acid.

The plot of the sequential NOE's (Figure 1) is generated using a PLOT macro and the chemical shift file. The program

has two output files. The first file is a macro for FTNMR (Hare Research, Inc., Woodinville, WA) that specifies the plot limits for a small region of the spectrum whose X-axis is centered at the HN frequency of a given residue. The macro then directs the FTNMR to make a plot of the selected region. The second file contains the rescaling and x and y offset information needed to place the sequential slices from FTNMR together into a single HPGL (Hewlet Packard Graphics Language) file. This file also contains commands, written in HPGL, for drawing the boxes and lines for the sequential NOE's. Prolines are handled by a special subroutine that is called up when the value of the HN frequency equals −9.00.

A second independent data base was generated to keep track of the NOE information. The file has a three-place formate. (Special subroutines were written to handle three-place data.) The first two columns of this file are the atom numbers of the protons responsible for the NOE, and the third place gives an upper bound for the distance constraint. This file is used to generate Figures 2 and 3. More importantly, a separate PLOT macro converts this NOE information into a constraint file for structure calculations. A subroutine of the program converts the atom number for each proton into the residue name and the atom type. The program automatically handles both assigned and unassigned stereopairs and figures out the shortest possible constraint from the available information.

Strategy for Computer-Assisted Sequential Assignments. The proton assignments for kistrin were obtained using the same procedures as described by Dubs et al. (1978), Nagayama and Wüthrich (1981), Wagner et al. (1981), Wüthrich et al. (1982), Wagner and Wüthrich (1982), and Wider et al. (1982). Initially, the protons were grouped into spin systems. Spin systems represent a subset of protons that all belong to a single residue. Most of the information at this stage came from TOCSY spectra collected in H_2O at 20 and 37 °C. A D_2O DQF-COSY (37 °C) was used to determine the relative positions of the protons in the side chain. The chemical shifts of all protons were entered into a data file (see previous section), and each spin system was assigned a temporary residue number starting at 100. Since kistrin has 68 amino acids, there could be no overlap between assigned and unassigned spin systems. The unassigned spin systems were broken down into four groups; unique spin systems, such as glycine, leucine, etc., were numbered starting at 100; the AMX spin systems started at 200; 300–399 were reserved for the long spin systems; and numbers above 400 were reserved for unclassified spin systems.

The classification of the spin systems according to amino acid type was relatively straightforward. Since kistrin has only three aromatic residues, most of the protons have chemical shifts (Table I) that are within 1 ppm of the random-coil values (Bundi & Wüthrich, 1987). Furthermore, most long-chain residues could be identified by the passive coupling patterns observed in the $C^{\alpha}H$–$C^{\beta}H$ cross peak in the DQF-COSY spectrum. The initial classification was correct for 62 of the amino acids, and only one error in the assignment interfered with analysis of the spectra (see discussion of S61 below). The three aromatic residues were rapidly identified by the large number of intense intraresidue NOE's from ring protons to the other protons in the AMX spin systems (data not shown).

All the intraresidue cross peaks in the fingerprint region of the NOESY were then labeled using the information from the spin system assignments. The temporary residue numbers were used as labels. For most spin systems, the $d_{\alpha N}$ cross peak could be distinguished by its strong intensity. Once this peak was identified, the next step was to identify potential d_{NN} and $d_{\beta N}$ cross peaks from spin systems that have the $C^{\alpha}H$ which matched the $d_{\alpha N}$ cross peak. A special graphic device was created to accelerate this process. A PLOT macro called "template" created a list of spin systems that was sorted according to the chemical shift of the $C^{\alpha}H$. A second macro called "templatemake" then made a synthetic spectrum where the X-axis represented the rank of a spin system in the list made by "template" and the Y-axis was the chemical shift of all the protons belonging to that spin system. "Cross peaks" in this synthetic spectrum were labeled with the residue number of the spin system. The spectrum was printed on five overhead transparencies using a laserwriter. This graphic device was called the template.

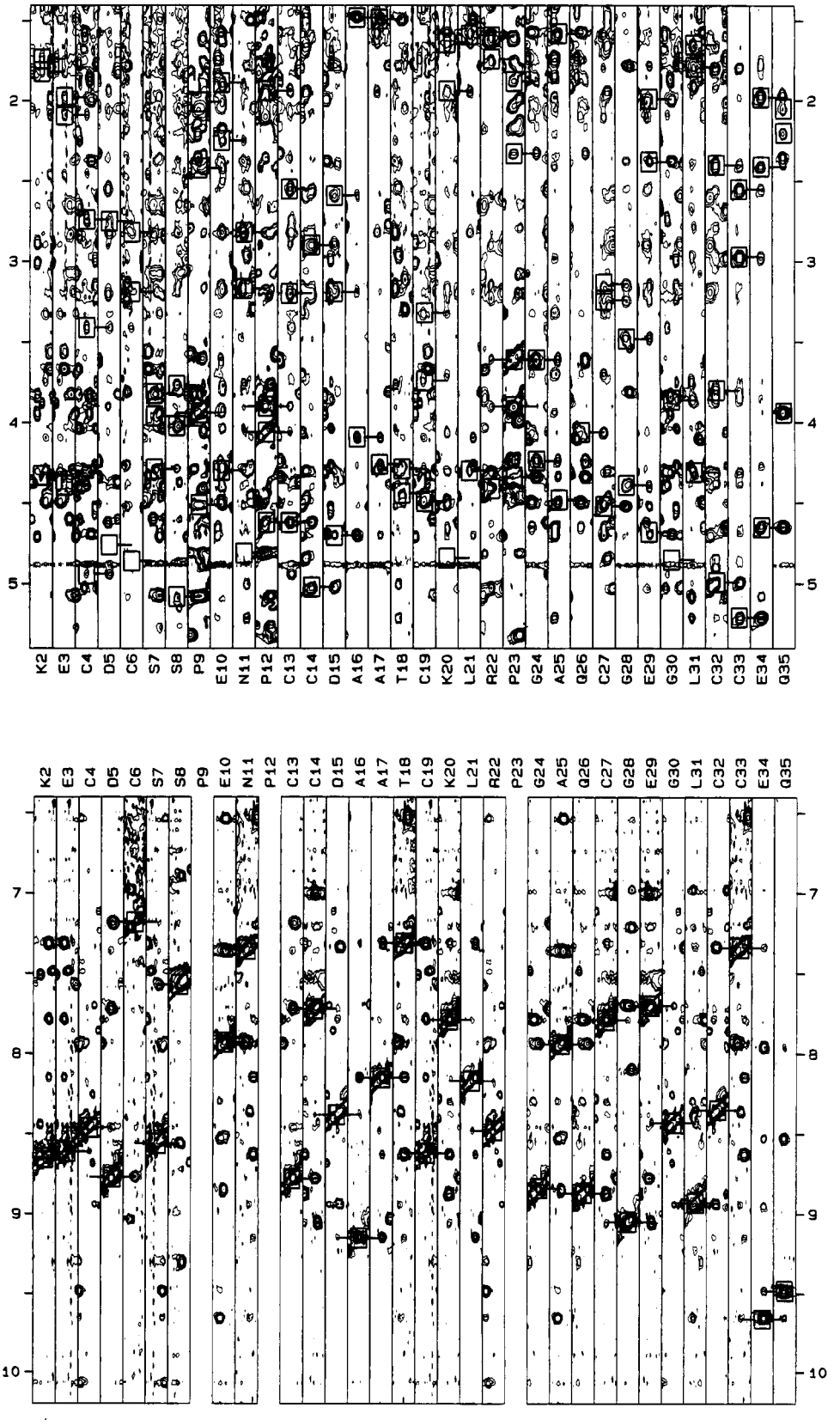
Once a potential sequential $d_{\alpha N}$ was identified, the Y-axis of the template was then lined up with ω_1 of the NOESY spectrum. The X-axis of the template was aligned such that the $C^{\alpha}H$'s of spin systems that potentially matched the $d_{\alpha N}$ were now superimposed on this cross peak. This step also aligned the labels for the HN and $C^{\beta}H$ protons from the sequential residue with the d_{NN} and $d_{\beta N}$ cross peaks. Generally, only one spin system would show the multiple cross peaks that were expected from the sequential residue. The sequential assignments were made using "wallpaper" size plots of the spectrum. Peak identification was done by hand. It was felt that these large paper plots contained more information than the typical computer workstation could conveniently handle. The template greatly simplified the comparison between different spin systems by grouping the information in simple visual form. Thus, it was possible to rapidly identify short stretches of sequentially connected spin systems.

Once the sequential connectivities were established, the appropriate residues were identified by using the "find string" function of a word processor software package together with a text file which contained the one-letter code for the sequence of kistrin. The one-letter code for the amino acids was modified by substituting the letter "X" for the AMX spin systems and the letter "Z" for the long residues. The file was used as follows. Assume a tripeptide had been identified that contained an AMX followed by two long residues. The "find" function of the word processor was then used to search the sequence file for the string "X Z Z". Only one string was found, which corresponded to residues C33–E34–Q35, thus completing the sequential assignments for these residues.

As expected, there was a considerable degree of overlap between amide resonances in certain regions of the spectrum. In most cases, cross peaks to overlapping HN protons could be distinguished by analyzing the peak shape. Almost all the peaks in NOESY spectrum have discernible coupling patterns along the ω_2 axis (Figure 1). This splitting was even more pronounced in spectra obtained at lower sample concentration (data not shown). The remaining ambiguities were resolved by obtaining a complete set of TOCSY and NOESY spectra in H_2O at 20 and 37 °C. No pair of assigned amide protons had degenerate resonances at both temperatures.

The six proline residues were assigned by substituting NOE cross peaks to $C^{\delta}H$'s for the sequential NOE's normally seen from the HN. In all cases, there was an intense sequential NOE to the preceding $C^{\alpha}H$ (Figure 1), which indicates that all six prolines are trans (Wüthrich et al., 1984; Montelione et al., 1986). Assignment of the proline residues was also aided by observation of the intraresidue cross peaks between the Pro $C^{\delta}H$ and $C^{\alpha}H$ in the D_2O TOCSY spectrum (data not shown).

Sequential Connectivities. Figure 1 displays the sequential NOE's for kistrin at 20 °C, pH 2.2. This figure presents a complete documentation of sequential NOE's. The figure is composed from small strips taken along ω_1 from NOESY


K2
E3
C4
D5
C6
S7
S8
P9
E10
N11
P12
C13
C14
D15
A16
A17
T18
C19
K20
L21
R22
P23
G24
A25
Q26
C27
G28
E29
G30
L31
C32
C33
E34
Q35
2
3
4
5
7
8
9
10

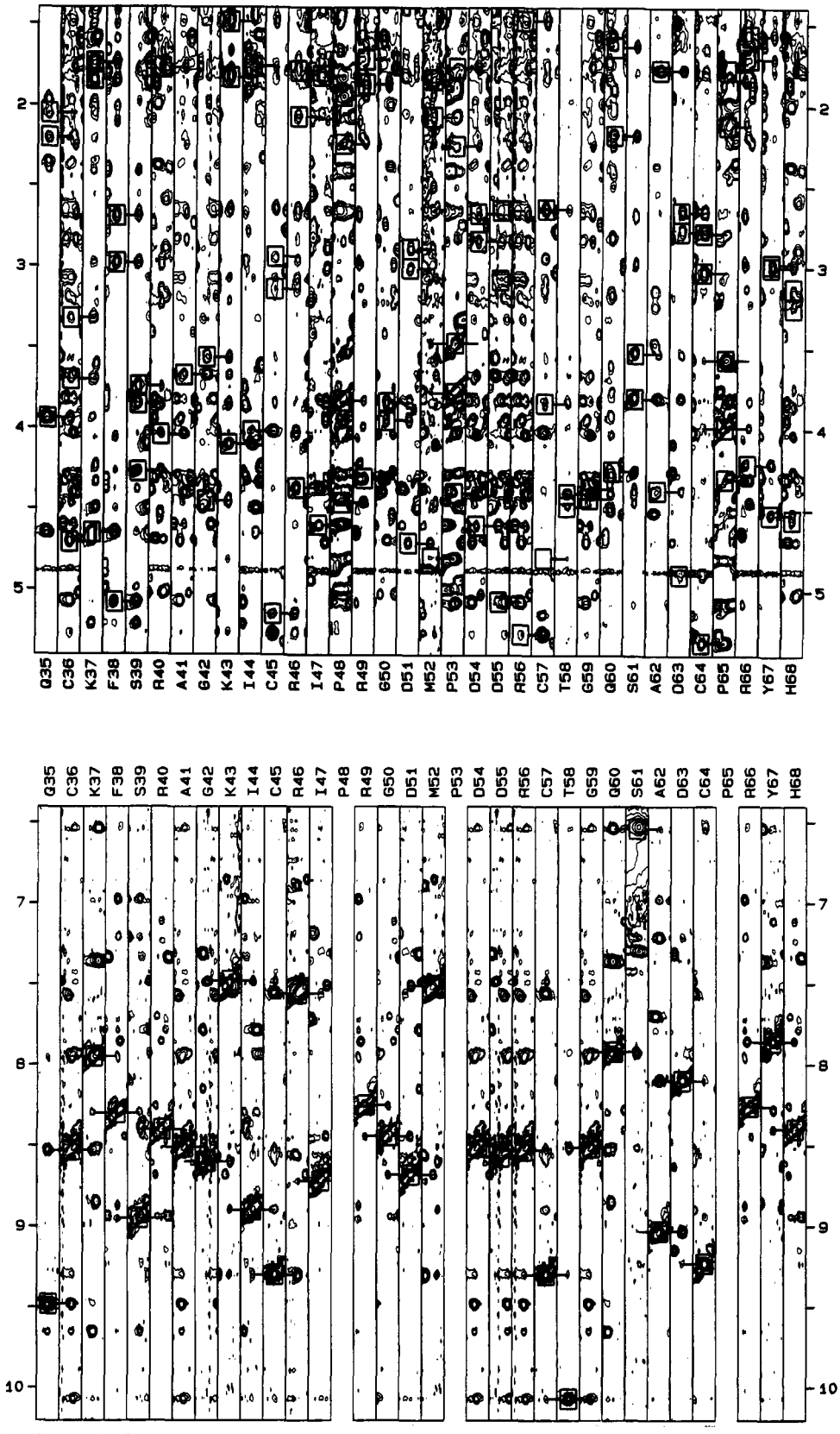


FIGURE 1: Selected slices from NOESY spectra of kistrin taken at 20 °C, pH 2.2. The individual slices from the NH region have been aligned in the sequential order of the residues. Boxes are drawn around the intraresidue cross peaks between the $C^{\alpha}H(s)$ or the $C^{\beta}H(s)$ and the HN. Solid lines are drawn to where the sequential NOE should appear. The level of the minimum contour was selected by hand and varies from slice to slice. A PLOT macro was used to construct this figure using the table of assigned resonances (see text). The lines and boxes were drawn without regard to the actual existence of a cross peak. Several $C^{\alpha}H$ resonances were lost due to the presaturation pulse, and the intraresidue peaks are not visible at this temperature. For the six proline residues, an appropriate slice was selected for a $C^{\beta}H$ instead of the HN. These slices, as well as the aliphatic slices for residues 14, 22, 25, 27, 32–34, 37–39, 43, 46, and 61, were taken from D_2O spectrum acquired with partial exchange of the amide protons. Boxes were placed within 0.01 ppm of the correct chemical shifts. There were minor problems with changes in the chemical shift of some aliphatic protons between the H_2O and D_2O spectra, and the HN resonances had to be calibrated under both conditions. Panels a (previous page, top) and b (previous page, bottom) show the aliphatic and HN region for residues 2–35, respectively. Panel c (this page, top) and d (this page, bottom) show the corresponding regions for residues 35–68.

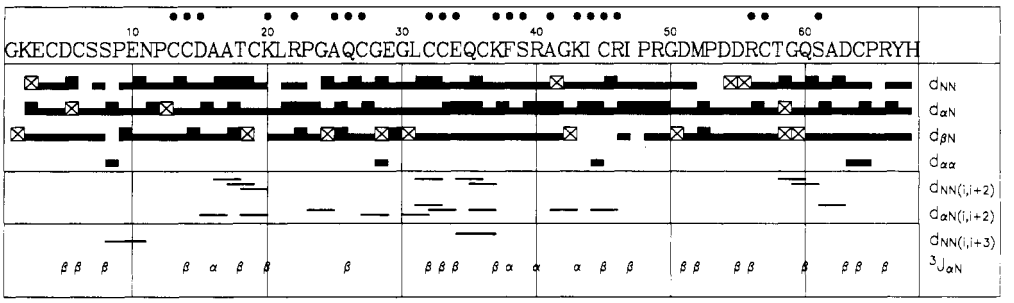


FIGURE 2: Summary of sequential and medium-range NOE's. Solid circles above the letter for a residue represent a slowly exchanging amide proton. The $^3J_{\alpha N}$ are labeled with α's for coupling constants less than 5.5 Hz and β's for coupling constants over 8.0 Hz. These symbols are used for convenience and do not carry any direct information about the secondary structure. For the sequential NOE's, three symbols are used. A thick bar is for NOE's observed in a NOESY with a mixing time of 80 ms or less. The thinner line indicates an observable NOE at longer mixing times (200 ms). The boxes with an "×" in them are used for cross peaks which are obscured by overlaps or are too close to the water line; therefore, the existence of these peaks cannot be verified.

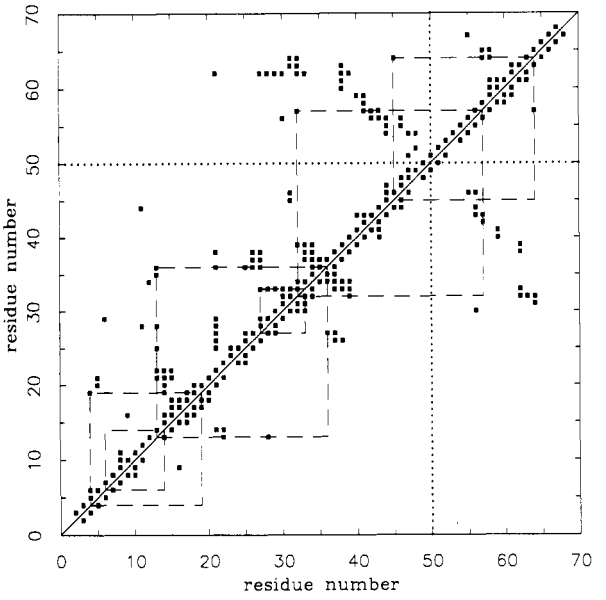


FIGURE 3: Diagonal plot of interresidue NOE's. Points above the diagonal represent an NOE between any two protons in different residues. Below the diagonal, only backbone NOE's are plotted. Boxes formed by dashed lines represent the six disulfide bonds in kistrin. The dotted lines run through center of the RGD adhesion site.

spectra, each covering a small range of the ω_2 frequency centered around the NH resonance of a particular residue. The strips are then arranged in a sequential manner so that the sequential NOE's can readily be documented. Two sets of plots are presented covering $d_{\alpha N}$ and $d_{\beta N}$ connectivities (Figure 1a,c) and d_{NN} connectivities (Figure 1b,d). The connectivities from residues 2 to 35 are given in Figure 1a,b; those for residues 35 to 68 are presented in Figure 1c,d. The positions of intraresidue cross peaks as anticipated from TOCSY spectra and the spin system assignments are identified by rectangles. Sequential connectivities are indicated with horizontal lines. In order to resolve overlap, some of the aliphatic slices have been selected from a NOESY spectrum obtained from the same sample in D_2O taken under conditions where there was limited exchange of amide protons. All the apparent ambiguities caused by overlapping HN were resolved in spectra taken at 37 °C (data not shown). The assignment for 334 proton resonances obtained at 20 °C, pH 2.2, is presented in Table I. A summary of sequential connectivities is displayed in Figure 2. At least two sequential cross peaks were observed for every assigned residue except one (see G59 below). A total of 88% of the potential d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$ cross peaks have been identified, if we count one or two $d_{\beta N}$ cross peaks as a single $d_{\beta N}$ connectivity.

Several residues proved to be difficult to assign. During the initial spin system assignments, S61 was incorrectly classified as glycine on the basis of its appearance in the H_2O TOCSY spectrum. This problem arose because the $C^{\alpha}H$ of this residue is degenerate with one of the $C^{\beta}H$. However, the sequential connectivities for this spin system resulted in a contradiction with the known sequence. The problem was resolved by examining the intraresidue cross peak in the D_2O DQF-COSY spectrum (data not shown). The peak revealed a complex splitting pattern that was incompatible with a two-spin system. Proper identification of the residue became trivial once it was known that it was a three-spin system.

Another problem concerned T58 and G59. The $C^{\alpha}H$ and $C^{\beta}H$ of T58 are not resolved from the $C^{\alpha}H$'s of G59. The amide of G59 is in a very crowded region of the spectrum (8.51 ppm), and intraresidue cross peaks for the amide to the $C^{\alpha}H$'s can only be observed in the DQF-COSY spectrum. All the sequential NOE's of aliphatic protons to the HN of G59 were lost in the overlap. To further complicate matters, there was no observable cross peak between the HN and the $C^{\gamma}H_3$ for T58 in any of the TOCSY spectra, and, therefore, it was not possible to distinguish this residue from a glycine on the basis of the through-bond information. It was obvious from the d_{NN} connectivities that the two-spin system belongs to the 58th and 59th residues, but the exact order was confusing. Final assignment rests on the fact that only T58 shows four sequential NOE's to C57 (Figure 1). The shape of the intraresidue HN–$C^{\alpha}H$ cross peaks from the amide proton in the DQF-COSY spectrum also supports the assignment.

Finally, the G1 has not been assigned. The amide of K2 (8.65 ppm) is not well resolved, and it has not been possible to assign the sequential NOE's.

Stereospecific Assignment of β-Methylene Protons. Stereospecific assignments have been obtained for the $C^{\beta}H$'s of 15 residues (Table I) using methods described previously (Wagner et al., 1987). The assignments of the prochiral protons rest on the measurement of the $C^{\alpha}H$–$C^{\beta}H$ coupling constant and an estimate of the NOE intensity between the HN and $C^{\beta}H$'s. The coupling constants were estimated from a DQF-COSY spectrum of kistrin acquired at 37 °C in D_2O. The information was used to limit the variation of χ^1 to within ±60° of the stable rotor conformation and to assign the observed NOE cross peaks to an individual $C^{\beta}H$'s where appropriate. The prochiral assignments could not be made for most of the residues because overlapping resonances obscured part of the information. For five amino acids with resolved cross peaks, the results were inconclusive and no stereospecific assignments were made.

Table I: Chemical Shifts (ppm) of Assigned Protons of Kistrin at pH 2.2, 20 °C

	H^N	H^α	$H^{\beta a}$	$H\gamma$	others
Lys 2	8.65	4.33	1.82, 1.74	1.39	H^δ, 2.00
Glu 3	8.60	4.39	2.09, 1.98	2.44	
Cys 4	8.46	4.94	*2.74, 3.41*		
Asp 5	8.77	4.76	2.75		
Cys 6	7.18	4.86	3.19, 2.82		
Ser 7	8.56	4.29	3.96, 3.82		
Ser 8	7.57	5.08	*3.77, 4.02*		
Pro 9		4.51	2.42, 2.01	2.06	H^δ, 3.94
Glu 10	7.93	4.30	*1.89, 2.25*	2.46, 2.41	
Asn 11	7.35	4.81	*2.82, 3.17*		H^δ, 7.93, 7.36
Pro 12		4.62	1.94	1.95, 2.10	H^δ, 4.06, 3.90
Cys 13	8.78	4.62	*2.55, 3.20*		
Cys 14	7.74	5.02	2.90		
Asp 15	8.38	4.70	*2.59, 3.19*		
Ala 16	9.15	4.09	1.48		
Ala 17	8.15	4.26	1.49		
Thr 18	7.31	4.44	4.29	1.12	
Cys 19	8.62	4.49	3.74, 3.32		
Lys 20	7.79	4.84	1.94, 1.64	1.47	
Leu 21	8.17	4.30	1.38, 1.59	1.59	H^δ, 0.59, 0.48
Arg 22	8.48	4.41	1.74, 1.60	1.37	H^δ, 3.21, 3.14; $N^\epsilon H$, 6.83
Pro 23		4.34	2.33, 1.88	2.15, 2.02	H^δ, 3.90, 3.61
Gly 24	8.84	4.24, 3.61			
Ala 25	7.94	4.48	1.58		
Gln 26	8.87	4.06	*1.14, 0.48*	2.08, 2.00	H^ϵ, 7.72, 7.01
Cys 27	7.80	4.52	3.24, 3.14		
Gly 28	9.05	4.39, 3.48			
Glu 29	7.70	4.68	2.38, 1.99		
Gly 30	8.43	4.85, 3.86			
Leu 31	8.93	4.31	1.79, 1.65	1.89	H^δ, 1.07, 1.06
Cys 32	8.36	4.99	*2.40, 3.80*		
Cys 33	7.35	5.21	*2.97, 2.55*		
Glu 34	9.66	4.65	2.41, 1.98		
Gln 35	9.48	3.94	2.20, 2.05	2.33	
Cys 36	8.53	4.71	3.70, 3.32		
Lys 37	7.96	4.65	1.84, 1.74	1.73	H^δ, 1.48, 1.40
Phe 38	8.30	5.08	*2.69, 2.98*		H^δ, 6.96; H^ϵ, 7.20; H^ζ, 7.20
Ser 39	8.95	4.27	3.85, 3.74		
Arg 40	8.40	4.04	1.77	1.68	H^δ, 3.22; $N^\epsilon H$, 7.18
Ala 41	8.51	3.67	1.23		
Gly 42	8.60	4.45, 3.56			
Lys 43	7.49	4.10	1.83, 1.49	1.47, 1.36	H^δ, 1.69; H^ϵ, 3.06
Ile 44	8.90	4.02	1.76	1.75, 1.86	$H^{\gamma 2}$, 0.93; H^δ, 0.89
Cys 45	9.30	5.15	3.14, 2.94		
Arg 46	7.56	4.37	*1.79, 2.07*	1.68, 1.32	H^δ, 3.15, 3.27; $N^\epsilon H$, 7.29
Ile 47	8.72	4.60	1.82,	1.42, 0.99	$H^{\gamma 2}$, 0.94; H^δ, 0.87
Pro 48		4.46	2.24, 1.94	1.83, 1.94	H^δ, 3.83
Arg 49	8.25	4.31	1.87, 1.72	1.64	H^δ, 3.22; $N^\epsilon H$, 7.15
Gly 50	8.44	3.95, 3.84			
Asp 51	8.68	4.71	*2.89, 3.01*		
Met 52	7.51	4.80	*2.07, 1.84*	2.63, 2.50	H^ϵ, 2.10
Pro 53		4.40	2.25, 1.77	1.86	H^δ, 3.78, 3.47
Asp 54	8.50	4.60	2.84, 2.67		
Asp 55	8.57	5.07	3.08, 2.65		
Arg 56	8.53	5.27	1.48	1.34	H^δ, 3.17, 3.07; $N^\epsilon H$, 7.42
Cys 57	9.30	4.80	*2.64, 3.84*		
Thr 58	10.07	4.40	4.48	1.35	
Gly 59	8.51	4.43, 4.37			
Gln 60	7.92	4.26	*2.18, 1.64*		
Ser 61	6.54	3.81	3.81, 3.53		
Ala 62	9.03	4.38	1.78		
Asp 63	8.10	4.90	2.78, 2.66		
Cys 64	9.23	5.32	3.03, 2.79		
Pro 65		4.31	2.12, 1.78	2.08, 1.98	H^δ, 3.99, 3.57
Arg 66	8.26	4.22	1.71, 1.57	1.50, 1.45	H^δ, 3.13; $N^\epsilon H$, 7.16
Tyr 67	7.86	4.53	3.00, 2.98		H^δ, 7.10; H^ϵ, 6.80
His 68	8.40	4.56	3.25, 3.13		H^δ, 7.21; H^ϵ, 8.60

[a] Stereospecific assignments were obtained for the H^β's shown in italics. The resonances are then listed in the order $H^{\beta 2}$, $H^{\beta 3}$.

Analysis of the Secondary Structure. To date, 550 inter-residue NOE's have been assigned; 260 NOE's connect sequential residues. The NOE information is summarized in Figures 2 and 3. These figures indicate that there is no extensive formation of regular secondary structure. There appears to be a single residue β-sheet between residues E34 and K37. However, other potential sites for such structures can be eliminated by close examination of the data. For instance, residues 13–20 display a continuous stretch of i to i+2 NOE's which is often found in an α-helix. However, there are unambiguous NOE's between residues at both ends of this region, and these NOE's are incompatible with a helical structure.

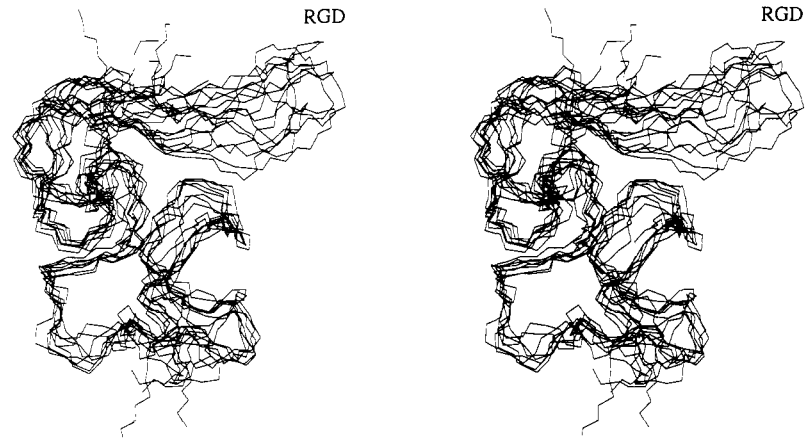


FIGURE 4: Stereoview of the eight superimposed kistrin structures computed by distance geometry methods. The letters "RGD" appear over the approximate location of the RGD adhesion site. For the alignment of the structures, only the backbone atoms of residues 4–46 and 56–65 were used. The average rms deviation to the average structure is 1.1 Å for these atoms.

Figure 3 also indicates that there may be an antiparallel β-sheet and turn formed by residues 41–59. The adhesion site, R49–G50–D51, would be located at the apex of the turn. However, several bits of information tend to rule out this description. There are several i to i+2 and i to i+3 NOE's for residues 41–49. Furthermore, the HN–C$^{\alpha}$H coupling constants for five residues in this region, A41, K43, I44, R46, and C57, are less than 7 Hz. Both the medium-range NOE's and the coupling constants are inconsistent with a fully extended structure found in β-sheets. The results do indicate that there is chain reversal centered at G50, although the exact nature of this structure is still uncertain. The lack of regular secondary structure is surprising given that one-third of the amide protons in kistrin exchange slowly with the solvent (Figure 1). These amide protons were stable enough to be observed in a NOESY experiment recorded over 72 h (Figure 1 and data not shown).

The NOE map (Figure 3) does indicate that kistrin is formed from a number of tight bends. By combining the information from the coupling constants, medium-range NOE's, and slowly exchanging amide protons, five potential β-turns have been identified: R22–P23–G24–A25, G30–L31–C32–C33, E34–Q35–C36–K37, K43–I44–C45–46, and T58–G59–Q60–S61. The amide proton in the fourth position is slowly exchanging in each of these sequences. This is consistent with a 1–4 hydrogen bond often found in β-turns (Richardson, 1981). Also, there is a least one i to i+2 NOE in each of these sequences. The proper identification of the other turns remains uncertain and awaits further refinements in the structure. It is worth noting that the data indicate that the turns at 30–33, 43–46, and 58–61 cannot be type II β-turns (Wagner, 1990). Each sequence has a large value for the HN–C$^{\alpha}$H coupling constant for the third residue, thus ruling out a positive value of ϕ for this residue.

Discussion

Kistrin is a member of an important class of small proteins containing an RGD sequence as recognition site. Thus the assignments of kistrin and the determination of its solution structure (Adler et al., 1991) may provide new insights into the function of these proteins. Due to its highly irregular structure, the assignments were more difficult than expected considering the low molecular weight of the protein. The assignments were also complicated due to the lack of aromatic residues in strategic positions of the protein that could have spread the resonances via ring current shifts. The only aromatic residues are Phe 38, Tyr 67, and His 68. The latter two are at the C-terminal end, which seems to be unstructured. Nevertheless, nearly complete assignments for the backbone and the side chains were obtained.

Preliminary Structure Calculations. Distance geometry calculations have been performed using the information obtained from the 550 NOE's (Figure 3) and the angular constraints from 25 ϕ and 15 χ^1 angles (Adler et al., 1991). These calculations have confirmed the existence of bends at residues G30–L31–C32–C33 and E34–Q35–C36–K37. The existence of the three other bends has not yet been firmly established. The data also indicate that there may be additional bends at S8–P9–E10–N11 and N11–P12–C13–C14. The root-mean-squared deviations for the backbone atoms to the average structure is 1.1 Å for the backbone atoms of residues 4–46 and 56–65. Less is known about the other portions of the molecule since there are few long-range NOE's to these residues. Further improvement in the modeling of the structure is in progress using quantified NOE's. A superposition of eight computed structures is shown in Figure 4.

Conformation of the RGD Site. The adhesion site, residues R49–G50–D51, is believed to play a dominant role in the interaction of GP IIb–IIIa with kistrin (R. A. Lazarus, personal communication). Unfortunately, there are no long-range (i to i+>5) NOE's observed for any of these residues, and only two long-range NOE's for residues 47–53 could be identified. Part of the problem stems from a lack of spectral resolution (none of the protons of G50 are well resolved). However, the methyl groups of both I47 and M52 are resolved, and the lack of long-range NOE's to these protons indicates that these residues are located on the protein surface. Therefore, the detailed conformation of the RGD adhesion site remains somewhat uncertain (Figure 4). However, some of the data indicate that there is some local order. There are 10 medium-range NOE's for residues 47–53, and several of these NOE's are consistent with a partial bend involving *trans*-proline 48. The HN–C$^{\alpha}$H coupling constants for both D51 and M52 are approximately 8 Hz. This implies that these two residues have a fairly extended conformation. The structure calculations consistently displayed the adhesion site as a relatively extended region that is stretched like a bow between two partial bends formed around the *trans*-prolines in positions 48 and 53. Hopefully, further investigations will provide more information about details of the structure of kistrin and, thereby, lead to new insights into the design of antagonist for the platelet surface receptor GP IIb–IIIa.

GP IIb–IIIa is itself a member of the large class of proteins called integrins (Albelda & Buck, 1990). These proteins have

been found on the surface of virtually every eukaryotic cell which has been tested. Integrins play an important role in controlling many cell–cell and cell–matrix interactions. Many of these integrins recognize RGD sequences on their target proteins. Therefore, knowledge gained from the structure of kistrin may shed light on a broad range of biological phenomena. No crystal structure of any RGD protein has been published so far. Recently, we have become aware of a NMR structure analysis for the smaller RGD protein echistatin (Saudek et al., 1991). The conformation of echistatin has high similarities with the conformation of the C-terminal half of kistrin as determined in our laboratory.

ACKNOWLEDGMENTS

We especially thank Dr. R. A. Lazarus and M. S. Dennis for the generous donation of the protein and the support and advice they gave during the course of the project. Dr. V. Thanabal provided considerable assistance in using the spectrometer. We thank Hare Research, Inc. (Woodinville, WA) for the donation of the programs FTNMR and DSPACE and T. Kossiakoff, P. Carter, and R. McDowell for helpful discussions. The coordinates of the structure have been submitted to the Brookhaven Protein Data Bank.

REFERENCES

Adler, M., & Wagner, G. (1991) *J. Magn. Reson. 91*, 450.
Adler, M., Lazarus, R. A., Dennis, M. S., & Wagner, G. (1991) *Science 253*, 445.
Albelda, S. M., & Buck, C. A. (1990) *FASEB J. 4*, 2868.
Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355.
Bax, A., Freeman, R., Frenkiel, T. A., & Levitt, M. H. (1981) *J. Magn. Reson. 43*, 478.
Braunschweiler, L., & Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521.
Billeter, M., Basus, V. J., & Kuntz, I. D. (1988) *J. Magn. Reson. 76*, 150.
Cieslar, C., Clore, G. M., & Gronenborn, A. M. (1988) *J. Magn. Reson. 80*, 119.
Bundi, A., & Wüthrich, K. (1979) *Biopolymers 18*, 285.
Dennis, M. S., Henzel, W. J., Pitti, R. M., Lipari, T. L., Napier, M. A., Deisher, T. A., Bunting, S., & Lazarus, R. A. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 2471.
Dubs, A., Wagner, G., & Wüthrich, K. (1979) *Biochim. Biophys. Acta 577*, 177.
Gould, R. J., Polokoff, M. A., Friedman, P. A., Huang, T.-F., Holt, J. C., Cook, J. J., & Niewiarowski, S. (1990) *Proc. Soc. Exp. Biol. Med. 195*, 168.
Griesinger, C., Otting, G., Wüthrich, K., & Ernst, R. R. (1988) *J. Am. Chem. Soc. 110*, 7870.
Jeener, J., Meier, B. H., Bachmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546.
Kleyweg, G. J., Boelens, R., & Kaptein, R. J. (1990) *J. Magn. Reson. 88*, 601.
Kline, A. D., Braun, W., & Wüthrich, K. (1988) *J. Mol. Biol. 204*, 675.
Kumar, R. M., Ernst, R. R., & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun. 95*, 1.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967.
Montelione, G. T., Hughes, P., Clardy, J., & Scheraga, H. A. (1986) *J. Am. Chem. Soc. 108*, 6765.
Nagayama, K., & Wüthrich, K. (1981) *Eur. J. Biochem. 115*, 653.
Nelson, S. J., Schneider, D. M., & Wand, A. J. (1991) *Biophys. J. 59*, 1113.
Pardi, A., Billeter, M., & Wüthrich, K. (1984) *J. Mol. Biol. 180*, 741.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800.
Phillips, D. R., Charo, I. F., Parise, L. V., & Fitzgerald, L. A. (1988) *Blood 71*, 831.
Rance, M., Sørensen, O. W., Bodenhausen, G., Wagner, G., Ernst, R. R., & Wüthrich, K. (1984) *Biochem. Biophys. Res. Commun. 117*, 479.
Richardson, J. (1981) *Adv. Protein Chem. 34*, 167.
Ruggeri, Z. M. (1989) *Circulation 80*, 1920.
Ruoslahti, E., & Pierschbacher, M. D. (1987) *Science 238*, 491.
Saudek, V., Atkinson, R. A., & Pelton, J. T. (1991) *Biochemistry 30*, 7369.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169.
Stein, B., Fuster, V., Israel, D. H., Cohen, M., Badimon, L., Badimon, J. J., & Chesebro, J. H. (1989) *J. Am. Coll. Cardiol. 14*, 813.
Wagner, G. (1990) *Prog. Nucl. Magn. Reson. Spectrosc. 22*, 101.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347.
Wagner, G., Kumar, A., & Wüthrich, K. (1981) *Eur. J. Biochem. 114*, 375.
Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Go, N., & Wüthrich, K. (1987) *J. Mol. Biol. 196*, 611.
Wand, A. J., & Nelson, S. J. (1991) *Biophys. J. 59*, 1101.
Wider, G., Lee, K. H., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 367.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311.
Wüthrich, K., Billeter, M., & Braun, W. (1984) *J. Mol. Biol. 180*, 715.
Yasuda, T., Gold, H. K., Leinbach, R. C., Yaoita, H., Fallon, J. T., Guerrero, L., Napier, M. A., Bunting, S., & Collen, D. (1991) *Circulation 83*, 1038.