# Cysteine Pairing in the Glycoprotein IIbIIIa Antagonist Kistrin Using NMR, Chemical Analysis, and Structure Calculations†

Marc Adler,‡,§ Paul Carter,∥ Robert A. Lazarus,∥ and Gerhard Wagner*,§

*Berlex Laboratories, Inc., 110 East Hanover Avenue, Morristown, New Jersey 07927, Department of Protein Engineering, Genentech, Inc., South San Francisco, California 94080, and Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115*

*Received July 22, 1992; Revised Manuscript Received October 22, 1992*

ABSTRACT: The pairing of the cysteines in disulfide bonds was investigated for the 68-residue RGD-containing protein kistrin, a potent antagonist of the integrin GP IIbIIIa and an inhibitor of platelet aggregation. Kistrin belongs to a family of homologous proteins found in snake venoms termed disintegrins, all of which have a cysteine content. The disulfide pairing of the 12 cysteines was investigated by chemical analysis, NMR spectroscopy, and distance geometry calculations. The data show that the disulfide pairs are 4–19, 6–14, 13–36, 27–33, 32–57, and 45–64. The various means for assigning the disulfide bonds are described, and the results are compared with the cysteine pairings reported for other disintegrin proteins.

Kistrin inhibits platelet aggregation by binding to glycoprotein IIbIIIa (GP IIbIIIa),[1] a calcium-dependent heterodimeric integrin found on the surface of activated platelets (Dennis et al., 1990). Normally, GP IIbIIIa mediates platelet aggregation by binding to a specific site on plasma fibrinogen. Fibrinogen contains two Arg-Gly-Asp (RGD) sequences, a putative adhesion site recognition sequence important for binding to integrins (Ruoslahti & Pierschbacher, 1987). The RGD sequence is also found on several other ligands that bind to GP IIbIIIa (Ruoslahti & Pierschbacher, 1987; Phillips et al., 1988). Kistrin competitively inhibits the binding of GP IIbIIIa to fibrinogen at nanomolar concentrations (Dennis et al., 1990). Mutational analysis of kistrin indicates that the RGD region is critical for GP IIbIIIa antagonist activity (R. A. Lazarus, unpublished results). Such a powerful GP IIbIIIa antagonist is a potential therapeutic for thrombotic diseases (Becker & Gore, 1991); initial results obtained with canine models of coronary artery thrombosis have been promising (Yasuda et al., 1991).

Kistrin is a 68-residue protein purified from the venom of *Agkistrodon rhodostoma* (Dennis et al., 1990) and is highly homologous to several other proteins termed disintegrins obtained from *Viperidae* venoms (Dennis et al., 1990; Gould et al., 1990). These proteins share the RGD sequence, and all have a high cysteine content (ca. 17%). There are 12 cysteines in kistrin; the relative location of the cysteines is well conserved (Figure 1).[2] Published reports, however, indicate that pairing of the disulfides is significantly different in 3 apparently homologous proteins investigated so far: kistrin with 68 residues (Adler et al., 1991), albolabrin with 73 residues (Calvete et al., 1991), and echistatin with 49 residues (Saudek et al., 1991a). The techniques used to established the pattern of cysteine pairing varied from protein to protein. NMR spectroscopy and structure calculations were used for kistrin (Adler et al., 1991) and echistatin (Saudek et al., 1991a,b), and proteolytic digestion followed by mass spectroscopy was used for albolabrin (Calvete et al., 1991). Sequence-specific assignments for echistatin were reported independently by four groups (Chen et al., 1991; Dalvit et al., 1991; Cooke et al., 1991; Saudek et al., 1991b), and a three-dimensional structure of echistatin was published by Saudek et al. (1991a). All data reported for echistatin are consistent with respect to the cysteine pairing, but they are not fully conclusive about the pairing of four of the eight cysteines in echistatin. Echistatin has one cysteine that is not present in kistrin (position 62 in kistrin); thus, differences in the cysteine pairing must exist. On the other hand, kistrin and echistatin share the same global architecture, as judged by NMR spectroscopy (Adler et al., 1991; Saudek et al., 1991a). Furthermore, the diagonal maps of the NOEs from these two proteins are extremely similar. In contrast to echistatin, albolabrin does have all the cysteines that kistrin has, and the apparent differences of the cysteine pairing are somewhat surprising. Here we describe the NMR and proteolytic digestion–mass spectroscopy data used for deriving the cysteine pairing in kistrin and compare the results with cysteine pairings reported for the homologous disintegrin proteins echistatin and albolabrin.

## MATERIALS AND METHODS

*Materials.* Kistrin was purified from the venom of the Malayan pit viper *Agkistrodon rhodostoma* as previously reported (Dennis et al., 1990). Sequencing-grade endoprotease Asp-N from *Pseudomonas fragi*, endoprotease Lys-C from *Lysobacter enzymogenes*, and bovine pancreatic trypsin were purchased from Boehringer Mannheim Biochemicals (Indianapolis, IN).

*Proteolysis of Kistrin.* Several different proteases (trypsin, Lys-C, Asp-N, chymotrypsin, elastase, pepsin, subtilisin, thermolysin, and V8 protease) were tried alone or in various combinations in an attempt to generate fragments of kistrin suitable for disulfide mapping. The combinations of trypsin

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

* To whom correspondence should be addressed.

‡ Berlex Laboratories, Inc.

§ Harvard Medical School.

∥ Genentech, Inc.

[1] Abbreviations: FAB-MS, fast atom bombardment mass spectrometry; GP IIbIIIa, glycoprotein IIbIIIa; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, two-dimensional nuclear Overhauser enhancement spectroscopy; rmsd, root mean square deviation.

[2] The cysteines in all three proteins are numbered by their positions in kistrin. A nomenclature that refers to the relative position of the cysteines cannot be used since echistatin is missing the first five cysteines and has a nonhomologous cysteine inserted at position 62 (Figure 1).

0006-2960/93/0432-282$04.00/0 

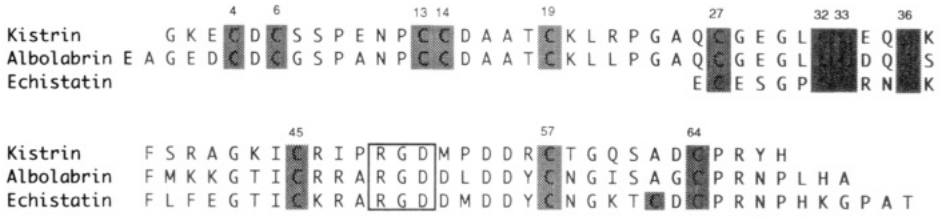


FIGURE 1: Amino acid sequences and cysteine residue numbers of kistrin, albolabrin, and echistatin.

plus Asp-N and of Lys-C plus Asp-N appeared most promising and were performed on a preparative scale. Thus, kistrin (1.75 mg) was digested with trypsin (50 $\mu$g) and Asp-N (10 $\mu$g) for 18 h at 37 °C in the dark in a total volume of 1 mL containing 50 mM Tris-HCl (pH 7.5) and 1 mM iodoacetic acid to minimize disulfide exchange (Marti et al., 1987). Alternatively, kistrin (1.0 mg) was digested with Lys-C (15 $\mu$g) and Asp-N (6 $\mu$g) for 18 h at 37 °C in the dark in a total volume of 0.5 mL containing 50 mM Tris-HCl (pH 7.5), 10 mM $ZnCl_2$, and 1 mM iodoacetic acid. The digestion products were loaded onto a C18 reverse-phase HPLC column (5 $\mu$m, 250 mm × 4.6 mm, Vydac, Hesperia, CA) equilibrated with 0.1% (v/v) trifluoroacetic acid and then eluted with a linear gradient of 0–35% (v/v) acetonitrile in 0.1% (v/v) trifluoroacetic acid over 70 min at 1.0 mL/min. Eluted peptides were monitored at 214 nm, and in addition at 262 and 275 nm to assist in the identification of peptides containing the chromophores F38 or Y67, respectively. Peptides from all major and most minor peaks were then analyzed by fast atom bombardment mass spectrometry (FAB-MS). Isolated peptides were lyophilized, redissolved in 70% (v/v) formic acid, dried onto the probe tip, and resuspended in thioglycerol. FAB-MS data were collected on a JEOL Model HX 110/110 tandem mass spectrometer operated in the 2-sector mode. The identity of cysteine-containing peptides was then confirmed by FAB-MS after reduction on the probe and tandem FAB-MS. As a further check on identity, these peptides were hydrolyzed using 6 N hydrochloric acid for 20 h at 110 °C (Millipore picotag system), and the amino acid composition was derived by analysis using a Beckman Model 6300 analyzer.

*NMR Data.* The sequential proton resonance assignments of kistrin were obtained using standard homonuclear techniques and have been reported previously (Adler & Wagner, 1992). In total, 334 of 374 protons have been assigned. Most of the unassigned resonances belong to methylene protons of residues with long side chains. Stereospecific assignments of several $\beta$-methylene protons were obtained as described previously (Adler & Wagner, 1992). The majority of the NOESY spectra were obtained from a 5 mM sample of kistrin maintained at 20 °C, pH 2.2. However, also 2D spectra at pH 7 were recorded which were very similar to those at low pH, except for larger line widths which made peak identification more difficult. The details of data acquisition and the assignments have been described previously (Adler & Wagner, 1992). The NOE distance constraints used for the structure calculations, which were employed for consolidation of the disulfide mapping, were the same as described previously (Adler et al., 1991). The NOE's were calibrated in the following way. Upper bound distances for intense cross-peaks observed with mixing times of 80 ms or less were set to 3.5 Å; 4.5 Å was used as the limit for weak peaks in these spectra. An upper bound of 5.0 Å was used for all other peaks; 0.5 Å was added to the upper bound to all NOE's involving a methyl group (Kline et al., 1988). Single NOE's to a proton of an unassigned stereopair were referenced back to the nearest heavy atom, and the interatomic distance between the proton and the heavy atom was added to the upper bound.

*Structure Calculations.* Structural calculations were performed using the program Dspace (Hare Research, Inc., Woodinville, WA). Initial calculations were performed without any disulfide bonds. The NOE constraints consist of 549 interresidue NOEs: 260 connect sequential residues, 78 represent medium-range NOEs ($3 \leq i-j \leq 5$), and 216 extend between residues separated by at least 4 residues ($6 \leq i-j$). The structures were first embedded and then minimized. Distance constraints were added to enforce the angular constraints on 25 $\phi$ angles and 15 $\chi_1$. The numerical weight on the constraints that enforce the bond lengths (1–2 interactions), angles (1–3) and planarity (1–4) were increased by 10-, 10-, and 30-fold, respectively, compared to the other constraints. The structures were minimized again and then subjected to a round of simulated annealing. Then selected structures were subjected to 45 rounds of rapid simulated annealing followed by minimization. During this process, the overall penalty function was recorded. The penalty function was the weighted sum of all violations of the distance constraints. If the penalty function rose by more than 8% in a single round or 20% from the best structure obtained so far, the program halted the calculation, read in the best structure, and randomized the coordinates by 0.2 Å, and the process was continued from those coordinates. After the completion of the simulated annealing, the structures were further refined by exhaustive minimization.

The 10 best structures, as judged by the value of the penalty function, were then selected and used to make comparisons between various possible cysteine pairings. The disulfide bonds were added to the 10 structures without altering the atomic coordinates. The structures were first minimized and then subjected to 25 rounds of rapid simulated annealing and minimization as described above. A round of calculations on a single structure took a little more than 2 h on an Iris 4D/25. One set of NMR constraints was used for the different forms of kistrin.

## RESULTS

*Disulfide Mapping of Kistrin by Chemical Techniques.* Cysteine-containing fragments of kistrin obtained by proteolysis with trypsin plus Asp-N (TN peptides) or of Lys-C plus Asp-N (CN peptides) were identified by FAB-MS and amino acid composition analysis (Table I). Peptides TN2 and CN1 each contain four cysteine residues (C27, C32, and C33 plus C57 and C6, C13, and C14 plus C36, respectively) whereas peptide TN1 contains all eight of these cysteines. We were unsuccessful in mapping the disulfides within these three fragments more closely by digestion with V8 protease, which was found to cleave only peptide TN1 on the C-terminal side of E34. However, it is unlikely that adjacent cysteins (C13 plus C14 or C32 Plus C33) are paired because of steric hindrance (Richardson, 1981). The TN1 and TN2 peptides show that there must be one S–S bond within the peptide

Table I: Amino Acid Sequence of Peptides Obtained after Proteolysis of Kistrin with Trypsin plus Asp-N (TN Peptides) or Lys-C plus Asp-N (CN Peptides)

| peptide[a] | $MH^+$ after reduction[b,c] measured | $MH^+$ after reduction[b,c] calcd | $MH^+$ before reduction[b] measured | $MH^+$ before reduction[b] calcd | Cys residues[d] |
|---|---|---|---|---|---|
| TN1 | | | 3409.3 | 3407.8 | |
| 5DCSSPENPCC14 | 1055.3 | 1055.1 | | | 6, 13, 14 |
| 21LRPGAQCGEGLCCEQCK37 | 1796.5 | 1796.1 | | | 27, 32, 33, 36 |
| 57CTGQSA62 | ND | 566.6 | | | 57 |
| TN2 | | | 1869.6 | 1869.1 | |
| 21LRPGAQCGEGLCC[33] | 1307.8 | 1307.5 | | | 27, 32, 33 |
| 57CTGQSA[62] | ND | 566.6 | | | 57 |
| TN3 | | | 856.5 | 857.0 | |
| 3EC4 | ND | 251.3 | | | 4 |
| 15DAATCK20 | 608.6 | 608.7 | | | 19 |
| CN1 | | | 1557.9 | 1557.7 | |
| 5DCSSPENPCC14 | 1055.9 | 1055.1 | | | 6, 13, 14 |
| 34EQCK37 | ND | 507.6 | | | 36 |
| CN2 | | | 1042.2 | 1042.4 | |
| 1GKEC4 | ND | 436.7 | | | 4 |
| 15DAATCK20 | 608.5 | 608.7 | | | 19 |
| CN3 | | | 1396.5 | 1396.5 | |
| 15DAATCK20 | ND | 608.7 | | | 19 |
| 63DCPRYH68 | 790.4 | 790.9 | | | 64 |

[a] The N- and C-terminal residues of each fragment are identified by numbers on the left and right side of the sequence, respectively. [b] The observed mass/charge ratio of the protonated molecular ion ($MH^+$) is shown together with that expected by calculation using natural isotopic abundances both for the intact peptide and after reduction. [c] ND, not detected probably because of the poor signal/noise ratio in this region of the spectrum. [d] Each cysteine in kistrin is identified by the corresponding residue number.

Table II: Analysis of Disulfide Bonds in Kistrin and Homologous Proteins[a]

| kistrin[b] consensus | NOE's[b] | kistrin chem[c,d] | albolabrin chem[d,e] | echistatin NMR[d,f] |
|---|---|---|---|---|
| 4–19 | 6* [b] | 4–19 ?[c] | 4–13/14 | none[g] |
| 6–14 | 0 | 6–13/14 | 6–13/14 | none |
| 13–36 | 5* | 13–6/36 | 13–4/6 | none |
| 27–33 | 2 | 27–32/33 | 27–57 | 27–36 |
| 32–57 | 1* | 32–27/57 | 32–19/36 | 32–57/62 |
| 45–64 | 3* | | 45–64 | 45–64 |
| 19–4 | 6* | 19–4, 19–64 ? | 19–32/33 | none |
| 14–6 | 0 | 14–6/36 | 14–4/6 | none |
| 36–13 | 5* | 36–13/14 | 36–32/33 | 36–27 |
| 33–27 | 2 | 33–27/57 | 33–19/36 | 33–57/62 |
| 57–32 | 1* | 57–32/33 | 57–27 | 57–32/33 |
| 64–45 | 3* | 64–19 ? | 64–45 | 64–45 |
| | | | | 62–32/33 |
| other NOE's in kistrin | | | | |
| 57–64[h] | 1 | | | |
| 14–19[h] | 1 | | | |

[a] Residue numbers always refer back to the homologous positions in kistrin (Figure 1). [b] Pairing for kistrin is based on analysis of the aliphatic NOE's between cysteines as described in the text. The column headed by "NOE's" lists the total number of NOE's, and the asterisk means that at least one of the NOE's is between two $C^\beta$Hs. [c] Results from chemical analysis of kistrin as listed in Table I. The question marks denotes the case where two separate peptides give conflicting results. [d] Analysis assumes that no cysteine is paired with its sequential neighbor. For the remaining ambiguities, both possible cysteines are listed and separated by a slant. [e] Results from the chemical analysis of albolabrin were published by Calvete et al. (1991). [f] NMR results from echistatin published by Saudek et al. (1991). [g] No homologous cysteines in this position. [h] NOE observed between these residues in kistrin. These cysteines are not cross-linked.

D5–C14. Thus, C6 must be paired with either C13 or C14, but not with C36. The same is true for the result of the TN2 peptide. There must be a disulfide within the peptide L21–C33. Thus, C27 must pair with C32 or C33, but not with C57. The resulting possible pairings are shown in Table II. Question marks are used to denote the entries for which contradictory results were obtained from different peptides. This concerns the results from the peptides TN3 and CN2 which indicate a pairing C4–C19. This contradicts the results of CN3 which indicate a pairing C19–C64. The latter is, however, entirely inconsistent with the NOE data (see below). Neglecting the results of the peptide CN3, C45 and C64 are not paired to other cysteines in any of the proteolytic peptides. This suggests, by way of exclusion, that they form a disulfide pair.

*Direct NMR Indications of the Cysteine Pairing from NOE's between Cysteines.* The NOESY spectrum of kistrin was examined for NOE's between the protons of the 12 cysteines. Data from the amide protons were not included in this search. Further analysis has proven that these amide protons are too far from the disulfide bond to give reliable information about the cysteine pairing. The two major conformations of disulfide bonds are a right-handed spiral with torsion angles $\chi^2 = +90°$, $\chi^3 = +90°$, and $\chi^4 = +90°$ and a left-handed spiral with $\chi^2 = -90°$, $\chi^3 = -90°$, and $\chi^4 = -90°$ (Richardson, 1981). In the right-handed case, the $H^{\beta 2}$ protons of both cysteines are sufficiently close in space (3.3 Å) to give rise to NOE's; in the left-handed case, the $H^{\beta 3}$ protons of both cysteines are close. Assignment of the $C^\beta H$ to $C^\beta H$ NOE's was complicated by the extensive overlap in the corresponding spectral region in kistrin. Nevertheless, 21 aliphatic NOE's between cysteines have been assigned (Table II). For the pair *4–19*, we see all four possible NOE's between the two pairs of $H^\beta$ resonances, and both $H^\alpha$ resonances show NOE's to both $H^\beta$ resonances of the other cysteine. For both residues of the putative pair *6–14*, we could not identify unambiguous NOE's to other cysteines. The $H^\beta$ resonances of C14 are degenerate (2.90 ppm) and overlap with one $H^\beta$ of C6. C14 shows one NOE to C19; C19, however, seems to form a disulfide pair with C4. Thus, there is no positive evidence for a disulfide bond 6–14. However, since the other disulfices seem to be unambiguously connected, this one can be deduced by exclusion. For the pair *13–36*, we see one $H^\beta$–$H^\beta$ NOE. It connects the $H^{\beta 2}$ of C13 (2.55 ppm) with one not yet stereospecifically assigned $H^\beta$ of C36. This is most likely the $H^{\beta 2}$ of C36, indicating a right-handed spiral. For the pair *27–33*, the positions of the expected cross-peaks between $H^\beta$ resonances are crowded and overlapped with strong

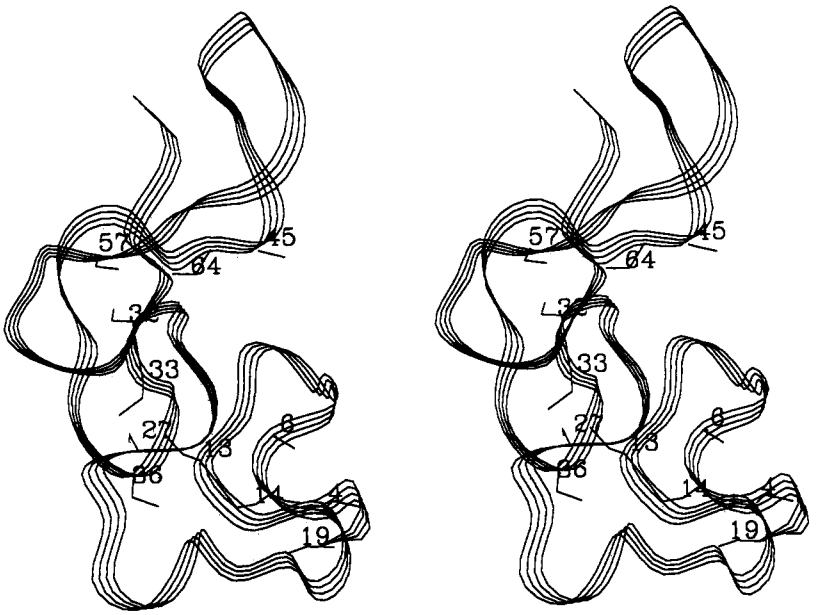


FIGURE 2: Stereodiagram (wall-eye view of a structure of kistrin calculated without disulfides. The backbone is drawn as a ribbon; the 12 cysteine side chains are shown as solid lines. This structure had the lowest value of the residual penalty function.

peaks; however, each $H^{\alpha}$ shows one NOE to one $H^{\beta}$ of the other cysteine. In the pair *32–57*, one $H^{\beta}$ of C32 shows one unambiguous NOE to one $H^{\beta}$ of C57. The pair *45–64* is connected by three NOEs. One of them is between the $H^{\beta}$ of C36 at 3.70 ppm and the $H^{\beta}$ of C64 at 2.79 ppm. Overall, the NOE data are consistent with the chemical analysis of the cysteine pairing (Table II). The chemical analysis also resolves the ambiguities caused by multiple NOE's.

*Cysteine Pairing from Structure Calculations.* A third line of evidence for the cysteine pairing in kistrin was obtained from structure calculations. This was achieved in the following way. First, the structures were calculated from NOE distance constraints and dihedral angle constraints only; no disulfide bonds were included. The 26 structures with the lowest energy were examined by visual inspection for possible cysteine pairing. It was found that cysteines were consistently bunched in three groups containing four cysteines in each (Figure 2). Cysteines-4, -6, -14, and -19, in that order, form the apices of a crude square. Since 4 and 6 faced in opposite directions, the most likely cysteine pairing was 6–14 and 4–19 (however, also pairings 4–6 and 14–19 could not be excluded, although inconsistent with the chemical analysis). Cysteines-36, -13, -33, and -27 formed a trapezoid where Cys-36 and Cys-27 were on the long side while Cys-13 and Cys-33 were on the short one. In most structures, Cys-13 and Cys-36 were the closest pair, but in a significant number of calculated structures, Cys-13 and Cys-33 were the closest pair. On the other hand, Cys-36 and Cys-27 were always distant and usually pointed away from each other. Therefore, the most likely pairings, according to these calculations, are 13–36 and 27–33. The last group of four cysteines, 32, 45, 57, and 64, showed the greatest variation in the calculated structures. Two-thirds of the structures clearly indicated that the disulfides were paired as 32–57 and 45–64. In the other structures, Cys-45 and Cys-64 were separated by the polypeptide strand that contained 54–57. Most of these latter structures had NOE distance violations of greater than 0.5 Å of constraints including residue 56. Two structures, with relatively high penalty functions would indicate an alternate cysteine pairing of 32–45 and 57–64. Summing up the results of the structure calculations, the most likely pairings were 4–19, 6–14, 13–36, 27–33, 32–57, and 45–64.

*Kistrin Structure Calculations with Preassigned Alternative Cysteine Pairings.* In order to test various schemes for cysteine pairing, further structure calculations were performed using different sets of disulfide bonds. Starting from the structures calculated without disulfide assignments, the constraints for the disulfide bonds were added to the DSPACE input data. The resulting conformational strain was minimized using a combination of simulated annealing and minimization (see Materials and Methods). No new embedding was carried out. Since there are 10 395 possible combinations of disulfide bonds for kistrin, a complete search could not be performed. The number $N_m$ of different disulfide topologies, for a protein with $m$ cysteines to be paired, is obtained from the formula:

$$N_m = (m-1)*(m-3)*(m-5)*...*1 \qquad (1)$$

The number $m$ is assumed to be even. The choice of trial sets of disulfide pairs selected for the structure calculations was based on the NMR data and evidence from chemical analysis described above, and on the results reported for the homologous proteins echistatin (Saudek et al., 1991a) and albolabrin (Calvete et al., 1991). For this purpose, the amino acid sequence of kistrin was maintained, but the length of the polypeptide chain was truncated to the fragment overlapping with the homologous protein. For echistatin, in particular, this means the calculations were done for a polypeptide of length 43 residues, which corresponds to the N-terminal sequence of echistatin, excluding the 6 C-terminal residues which are not present in kistrin. Echistatin has one nonhomologous cysteine at the position 62 (alanine in kistrin). For this residue, the names of the three protons of the $\beta$-methyl group of A62 were changed to match the $\gamma$-sulfur and $C^{\beta}H$'s of cysteine. The initial minimization quickly restored the appropriate bond lengths and angles of the modified atoms. All other residues were not changed.

The results from 10 structures, selected for their initial low value of the penalty function, are presented in Table III. In the first three columns and the footnotes of the table, the types of calculations are described. Table III includes the average value of the penalty function as well as the rmsd (root mean square deviation) of the coordinates of the backbone atoms from one structure compared to the same structure

Table III: Comparison of Structure Calculations with Different Sets of Disulfide Bonds

| name[a] | cysteine pairings used as constraints in calcn | av penalty function[b] | av viol/res[c] | rmsd (Å) to starting structure[d] | no. of converged structures (of 10)[e] |
|---|---|---|---|---|---|
| K | | 6.9 | 0.101 | | 7 |
| Kn | 4–19, 6–14, 13–36, 27–33, 32–57, 45–64 | 7.8 ± 0.2[f] | 0.115 ± 0.003 | 1.31 ± 0.06 | 7 ± 1 |
| $Ke_1$ | 4–19, 6–14, 13–33, 27–36, 32–57, 45–64 | 9.5 | 0.140 | 1.23 | 5 |
| $Ke_2$ | 4–19, 6–14, 13–36, 27–33, 32–45, 57–64 | 11.1 | 0.164 | 1.49 | 1 |
| $Ka_1$ | 4–13, 6–14, 19–32, 27–57, 33–36, 45–64 | 38.5 | 0.567 | 2.36 | 0 |
| $Ka_2$ | 4–13, 6–14, 19–33, 27–57, 32–36, 45–64 | 37.9 | 0.557 | 2.19 | 0 |
| $E_1$ | 27–36, 32–57, 33–62, 45–64 | 5.3 | 0.124 | 1.37 | 8 |
| $E_2$ | 27–36, 32–62, 33–57, 45–64 | 11.3 | 0.263 | 1.65 | 0 |
| Ek | 27–33, 32–57, 36–62, 45–64 | 8.3 | 0.193 | 1.43 | 5 |

[a] Identifier for the nine sets of structure calculations. Ten structures were calculated for each set. K, distance geometry calculations for kistrin without disulfide constraints, using only NOE and dihedral angle constraints. Kn, distance geometry calculations for kistrin with disulfide constraints as inferred from the chemical analysis and the NOE's between the cysteines (Table II); $Ke_1$, distance geometry calculations for kistrin with disulfide pairings that are more similar to the echistatin pairing (Saudek et al., 1991a); this means that one pair of disulfides, 13–36 and 27–33, is swapped. $Ke_2$, distance geometry calculations for kistrin with another disulfide pairing more similar to the echistatin pairing, i.e., the pair of disulfides, 32–57 and 45–64, is swapped. $Ka_1$ and $Ka_2$, two sets of distance geometry calculations with pairings analogous to those proposed for albolabrin [see Calvette et al. (1991)]; $E_1$ and $E_2$, distance geometry calculations for a polypeptide truncated at the N-terminus to the length of echistatin and at the C-terminus to the length of kistrin, with two alternative S–S pairings suggested by Saudek et al. (1991) but with NOE's from kistrin. Ek, same as $E_1$ and $E_2$, but with disulfide pairing as in Kn. [b] Average sum of residual violations. Distance constraints that regulate the 1–2, 1–3, and 1–4 covalent interactions are weighted 10, 10, and 30 times more than other violations, respectively. [c] Average sum of residual violations divided by the number of residues used in the calculations, 68 for kistrin and 43 for the echistatin-like compound. The values are arbitrary numbers since constraints of different types are included with different weights. [d] Average root mean square deviations in backbone coordinates to the initial structure which had no disulfides. Only residues within the core of the protein, 4–46 and 56–64, were used in this comparison. [e] Number of acceptable structures generated from 10 starting structures. Acceptable structures have no NOE violations over 0.5 Å, no covalent structure violations over 0.1 Å, and no more than one violation of van der Waals contact over 0.3 Å. [f] "±" gives the root mean square error from three independent calculations performed on the same initial structures.

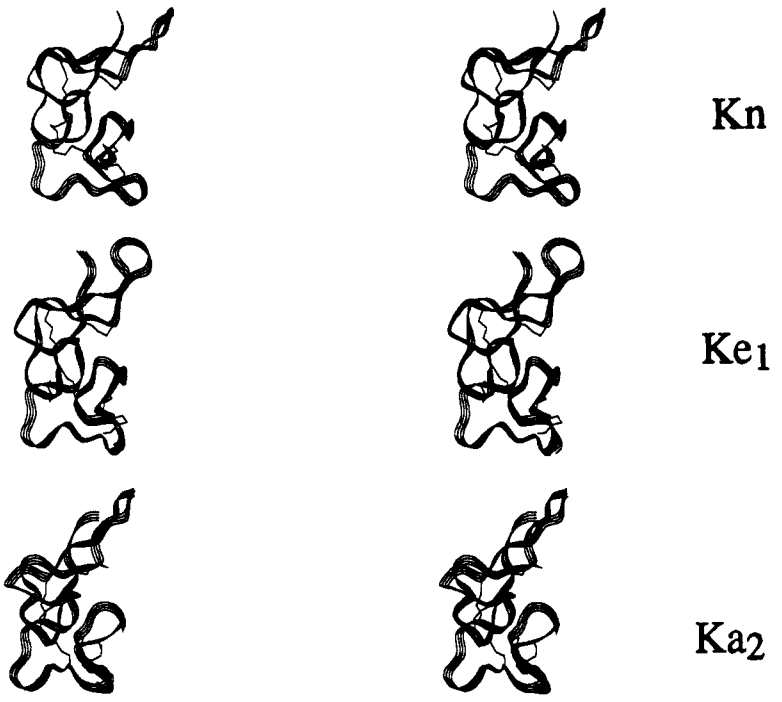


FIGURE 3: Stereodrawings of the structures of kistrin with three different sets of disulfide bonds. Kn represents the best set of bonds for kistrin (Table III). $Ke_1$ is based on homology to echistatin. $Ka_2$ is based on homology to albolabrin. The structure shown in Figure 2 was used for the initial starting coordinates, and the color scheme is also based on Figure 2.

prior to incorporating disulfide constraints. Only residues 4–46 and 56–64 were used for calculation of the average rmsd in this comparison since there are only a few long-range NOE's involving the other residues. The results indicate that the pairing suggested by the combination of NOE's between cysteines and results of the digestion experiments gave the best fit to the data. Cysteine pairings suggested by reports on echistatin (Saudek et al., 1991) and albolabrin (Calvete et al., 1991) are significantly less consistent with the NMR data.

Three sets of cysteine pairings (Kn, $Ke_1$, and $Ke_2$) were tried first. The set suggested by the NOE's of kistrin became the reference set, denoted as Kn in Figure 3 and Table III. In the set $Ke_2$, cysteines-32 and -45 and Cys-57 and -64 were paired while the others remained the same as in Kn. This pairing seemed possible after inspection of some of the initial structures. Finally, we also tried a set that was homologous as possible to echistatin while keeping the full length of the kistrin sequence ($Ke_1$). This contained all the normal pairs of kistrin except the pairs 13 and 33 and 27 and 36. As is shown in Table III, the best results were obtained from the pairings used in the reference set Kn. The residual penalty function rose by 11% compared to the same calculation performed without disulfide bonds. Comparison of the rmsd's

shown in Table III indicates that the introduction of disulfide did not significantly disturb the structure.

Next, the cysteine pairing for albolabrin was investigated by adding the homologous bonds to the initial structures of kistrin that has no disulfides. There are nine possible pairings of the disulfides that are consistent with the published data for albolabrin (Calvete et al., 1991). The ambiguity in the disulfide mapping reported for albolabrin was due to problems with incomplete digestion during chemical analysis. For our calculations, it was assumed that sequentially neighboring cysteines were not paired because of steric reasons (Richardson, 1981), and, wherever possible, the pairings suggested by the kistrin data were retained. This left two possible pairings, denoted as $Ka_1$ and $Ka_2$, that each retained two disulfide bonds from kistrin. As shown in Table III, both these pairings increased the residual penalty function by nearly 5-fold. Therefore, these pairings are not consistent with the NOE data available for kistrin. It is interesting to note that the NOE data maintained the protein in a fold that was very similar to kistrin (Figure 3 and Table III). Other possible pairings suggested by the albolabrin data (Calvete et al., 1991) were not tried because they were even less homologous to the putative kistrin pairings.

The next set of trial pairings was based on homology to echistatin ($Ke_1$). There is a small but significant rise in the residual penalty function (38%) compared to the structures without disulfides. However, 5 out of the 10 structures had no violations of NOE constraints over 0.5 Å. The integrity of the covalent structure (distortions of bond lengths and bond angles) was also within acceptable limits for these structures. Therefore, in the absence of data provided by chemical analysis, it would be difficult to rule out this pairing. This result was unexpected since our initial visual inspection of the structures did not indicate that this was a reasonable bond pairing for any of our structures. For the third set of disulfide bonds, $Ke_2$, the calculation showed a large increase in the penalty function which indicates that this pairing is not consistent with the data.

*Structure Calculations for the Kistrin–Echistatin Overlap.* It is important to note that there are significant differences in the sequences of kistrin and echistatin and the relative locations of the cysteine residues are not completely homologous. When the 2 sequences are aligned (Figure 1), echistatin is missing the first 25 residues which includes 5 cysteines; it also has a nonhomologous cysteine at the position corresponding to Ala-62 in kistrin. Therefore, the results derived using the primary sequence of kistrin may not be relevant to echistatin. The 10 kistrin structures used in the previous calculations were truncated to match the structure of echistatin. The identity of all residues remained the same except that Ala-62 of kistrin was changed to a cysteine (see above). The NMR data from kistrin, including NOE's to Ala-62, were retained wherever possible. Comparisons were made between three different sets of disulfides; two suggested by Saudek et al. (1991a), $E_1$ and $E_2$, and a third set based on the closest possible homology to kistrin, Ek (Table III). Best structures were obtained using $E_1$ (Table III). Of course, this does not mean that the $E_1$ pairing is most likely the one in echistatin. It just means that the NMR constraints from within the truncated kistrin polypeptide strand would be consistent with the $E_1$ pairings. However, the need to pair one of the five cysteines from the N-terminal strand of kistrin with a cysteine from the sequence overlapping with echistatin makes the $E_1$ pairing impossible for kistrin. When the residual penalty value was normalized to the number of amino acids, the results for $E_1$ were nearly identical to those obtained for the best structure of kistrin, Kn. The penalty function was considerably higher for the pairing based on the kistrin data, Ek. Specifically, there was conformational strain introduced by forming a disulfide between 36 and 62. Residue 36 is at the end of a loop which points away from residue 62, and a disulfide bond cannot be introduced without distorting the structure (Figure 4). Also, the location of Cys-27 in kistrin is fixed by NOE's to the first 25 residues. In echistatin, this cysteine becomes the second residue, and the whole N-terminus is free to shift over and form a cross-link to Cys-36 (residue 11 in echistatin). The calculation indicated that the second pairing suggested by Saudek et al. (1991a), $E_2$, is not consistent with the NMR data for kistrin.

## DISCUSSION

The results of the different approaches to characterize the disulfide bonds in kistrin are all consistent and leave only one possible disulfide topology for kistrin. Chemical analysis alone could not provide a complete identification of the disulfide bonds, and also NMR analysis alone would not have been sufficient for a complete and unambiguous disulfide mapping. This statement is valid for kistrin only and cannot be generalized for other disintegrin proteins.

An important outcome of this work is that, in the disintegrin family of cysteine-rich proteins discussed here, all the disulfide bonds are generally not conserved. This is in contrast to most other homologous proteins, where disulfide bonds are highly conserved and seem to be often important for the correct folding of a protein.

For kistrin, the chemical analysis was necessary to complement the NMR data for establishing the cysteine pairings. Without these data, we would be dependent on comparison between structures generated by the distance geometry program. This comparison would be conclusive if the calculations gave a complete and unbiased search of conformational space. Although we have not systematically tested this hypothesis, we have observed that the final structures tend to reflect a subtle dependence on how and when certain constraints are introduced. Also, certain regions of conformational space were not properly explored by the program. Cysteines-13 and -33 and Cys-27 and -36 ($Ke_1$) did not appear as the most likely choice in any of the structures generated without disulfide bonds. However, calculations showed that this pairing gave a good fit to all the data and, therefore, our initial search of conformational space appears to have been incomplete. It is safe to assume that there is a bias in the calculations and that this bias is reflected in the final structures. It is worth noting that most of the sets of disulfide bonds produced at least one structure that had reasonable covalent geometry and no NOE violation over 0.5 Å (Table III). Therefore, most of the pairings give an acceptable fit of the data. Fortunately, the ambiguities raised by the NMR data are resolved by comparison to the results from chemical analysis and vice versa.

The difficulty in assigning the disulfide bonds probably stems from their close packing in these proteins. Kistrin is a small protein in which 17% of the residues are cysteines. Since disulfide bonds are hydrophobic, the cysteines become densely packed in whatever little interior volume there is. Therefore, several cysteines had NOE's to more than one other cysteine. Furthermore, we may assume that during the process of folding, the protein may shuffle the disulfides until the best pairing is obtained from the many possible pairs. On the other hand, it appears to be impossible to totally prevent

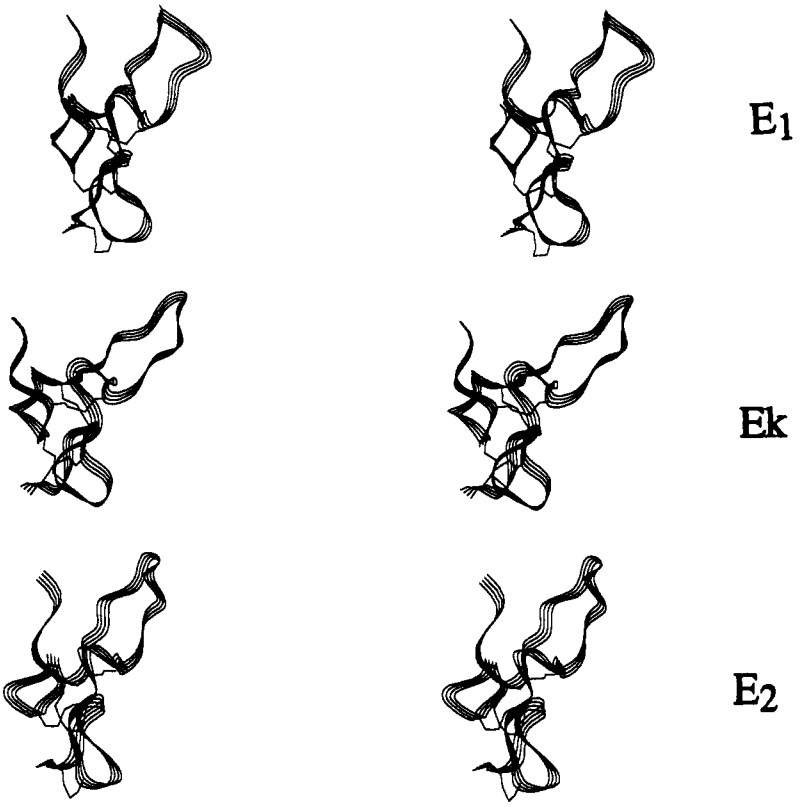


FIGURE 4: Stereodrawings of computer-generated structures of the kistrin–echistatin overlap based on the sequence and NOE's from kistrin (see text) with three different sets of disulfide bonds. The disulfide bonds in $E_1$ and $E_2$ represent the two possibilities suggested by Saudek et al. (1991). The pairing in Ek is based on homology to kistrin. The structure shown in figure 2 was used for the initial starting coordinates.

this reshuffling during proteolytic digestion used in chemical analysis despite the precaution taken by using iodoacetic acid to trap any free thiol generated. According to our chemical analysis, Cys-19 appeared to be paired to both Cys-4 and Cys-64. However, since cysteines-19 and -64 lie on opposites ends of the protein (Figure 1), this disulfide bond must have formed after the protein became denatured.

In the case of kistrin, the disulfide bonds appear to be rather poor constraints for structure determination. The rmsd's between backbone atoms were the same, within experimental error, whether the disulfide constraints are included or not (1.6-Å rmsd between structures for the core region of protein). In fact, the rmsd between structures was the same for both correct and incorrect pairings. When the disulfide bonds were included in the first stages of calculations, they biased the results toward what appears to be a partially incorrect structure. In calculations performed with only two out of six correct disulfide bonds, the protein maintains the correct global fold (Figure 3). We also failed to obtain any structure that had the correct fold in 40 attempts when the disulfide bonds were used as constraints without using any of the NMR data.

Finally, even though kistrin and echistatin are highly homologous, only two of the disulfide bonds are identical. The overall three-dimensional structure of echistatin is similar to the C-terminal region of kistrin. The structure calculations described above suggest that the best disulfides for echistatin are the ones proposed by Saudek et al. (1991a), denoted $E_1$ in Table II, not the ones most homologous to kistrin, $E_k$ (Table III). We are, however, at a loss to explain the difference in cysteine pairing in albolabrin and kistrin, and the calculations shed no new light on this. It is noteworthy that the regions of kistrin important for direct binding to GP IIbIIIa are specifically localized to the RGD sequence; there is no evidence for other residues having any significant role (R.A. Lazarus, unpublished results). The fact that the conformation of the RGD sequence is what is most important for GP IIbIIIa activity in kistrin is supported by data from nonhomologous proteins found in leeches, decorsin (Seymour et al., 1990), and ornatin (Mazur et al., 1991) and in an elapid snake, mambin (McDowell et al., 1992). These proteins are equipotent to kistrin as GP IIbIIIa antagonists yet structurally likely to be unrelated except for the RGD region. It is therefore possible, though surprising, that the homologous disintegrins may have different tertiary structures outside the RGD region as well. Clearly, more experimental data are required to answer this question.

## ACKNOWLEDGMENT

We thank Dr. V. Thanabal for providing considerable assistance in using the spectrometer, Dr. Pat Griffin for mass spectrometry data, and Byron Nevins for performing amino acid analyses. We thank Hare Research, Inc. (Woodinville, WA), for providing the programs FTNMR, FELIX, and DSPACE and Dr. T. Kossiakoff, M. Dennis, and Dr. R. McDowell for helpful discussions. The coordinates of the structure with the native disulfide bonds have been submitted to the Brookhaven Protein Data Bank (P1KST).

## REFERENCES

Adler, M., & Wagner, G. (1992) *Biochemistry 31*, 1031–1039.

Adler, M., Lazarus, R. A., Dennis, M. S., & Wagner, G. (1991) *Science 253*, 445–448.

Becker, R. C., & Gore, J. M. (1991) *Circulation 83*, 1115–1117.

Calvete, J. J., Schäfer, W., Soszka, T., Lu, W., Cook, J. J., Jameson, B. A., & Niewiarowski, S. (1991) *Biochemistry 30*, 5225–5229.

Chen, Y., Pitzenberger, S. M., Garsky, V. M., Lumma, P. K., Sanyal, G., & Baum, J. (1991) *Biochemistry 30*, 11625–11636.
Cooke, R. M., Carter, B. G., Martin, D. M. A., Murray-Rust, P., & Weir, M. P. (1991) *Eur. J. Biochem. 202*, 323–328.
Dahlvit, C., Widmer, H., Bovermann, G., Breckenridg, R., & Metternich, R. (1991) *Eur. J. Biochem. 202*, 315–321.
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*, 2471–2475.
Dubs, A., Wagner, G., & Wüthrich, K. (1979) *Biochim. Biophys. Acta 577*, 177–194.
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*, 168–171.
Kline, A. D., Braun, W., & Wüthrich, K. (1988) *J. Mol. Biol. 204*, 675–724.
Marti, T., Rosselet, S. J., Titani, K., & Walsh, K. A. (1987) *Biochemistry 26*, 8099–8109.
Mazur, P., Henzel, W. J., Seymour, J. L., & Lazarus, R. A. (1991) *Eur. J. Biochem. 202*, 1073–1082.
McDowell, R. S., Dennis, M. S., Louie, A., Shuster, M., Mulkerrin, M. G., & Lazarus, R. A. (1992) *Biochemistry 31*, 4766–4772.
Nagayama, K., & Wüthrich, K. (1981) *Eur. J. Biochem. 115*, 653–657.
Phillips, D. R., Charo, I. F., Parise, L. V., & Fitzgerald, L. A. (1988) *Blood 71*, 831–843.
Richardson, J. (1981) *Adv. Protein Chem. 34*, 167–339.
Ruoslahti, E., & Pierschbacher, M. D. (1987) *Science 238*, 491–497.
Saudek, V., Atkinson, R. A., & Pelton, J. T. (1991a) *Biochemistry 30*, 7369–7372.
Saudek, V., Atkinson, R. A., Lepage, P., & Pelton, J. T. (1991b) *Eur. J. Biochem. 202*, 329–338.
Seymour, J. L., Henzel, W. J., Nevins, B., Stults, J. T., & Lazarus, R. A. (1990) *J. Biol. Chem. 265*, 10143–10147.
Wagner, G., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 347–366.
Wagner, G., Kumar, A., & Wüthrich, K. (1981) *Eur. J. Biochem. 114*, 375–384.
Wider, G., Lee, K. H., & Wüthrich, K. (1982) *J. Mol. Biol. 155*, 367–388.
Wüthrich, K., Wider, G., Wagner, G., & Braun, W. (1982) *J. Mol. Biol. 155*, 311–319.
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*, 1038–1047.