Structural Analysis of a Biologically Active Echistatin Analogue des(46-49)-[Ala\(^8\),37]-Echistatin \(\gamma\) with Three Disulfide Bonds by 2D-NMR and Computer Graphics

Li-Chin Chuang,* Pei-Yeh Chen,* Chinpan Chen,† Tai-Huang Huang,‡ Kung-Tsung Wang,* ‡ Shyh-Horng Chiou,* ‡ and Shih-Hsiung Wu* ‡ 1

*Institute of Biological Chemistry, and †Institute of Biomedical Sciences, Academia Sinica, and ‡Institute of Biochemical Sciences, National Taiwan University, P.O.Box 23-106, Taipei, Taiwan

Received January 16, 1996

An echistatin analogue, designated as des(46-49)-[Ala\(^8\),37]-echistatin \(\gamma\), was synthesized chemically by solid-phase peptide synthesis. The analogue was made by replacing Cys8 and Cys37 residues with two alanines and the deletion of C-terminal peptide 46-49 of echistatin \(\gamma\), resulting in an artificial polypeptide of 45 amino acids with three disulfide bonds. In the platelet aggregation assay, the analogue exhibits almost the same activity as echistatin \(\gamma\), indicating that the linear sequence of des(46-49)-[Ala\(^8\),37]-echistatin \(\gamma\) contains all of the primary-structure information that is required for proper folding of this synthetic polypeptide. The tertiary structure of the analogue, as determined from high-resolution nuclear magnetic resonance (NMR) coupled with dynamic simulated annealing, is very similar to that of echistatin \(\alpha1\) which differs from echistatin \(\gamma\) by 8 residues. In particular the two important sites of the Arg-Gly-Asp (RGD) loop and the C-terminal Lys45, both of which show some degree of disorder, are maintained in similar spatial orientation and proximity as those in echistatin \(\alpha1\) even without the constraint provided by the disulfide bond of the (Cys8-Cys37) pair. These results provide new insights in further defining distinct structural features of echistatin \(\gamma\), which are involved in supporting the active polypeptide conformation to achieve biological activity in the absence of one pair of disulfide bonds.

Platelet aggregation associated with the blood coagulation cascade is now believed to be mediated by the interaction of platelet membrane glycoprotein IIb-IIIa complex with plasma fibrinogen, which eventually leads to the formation of platelet-rich clots [1]. Several polypeptides isolated from snake venoms such as kistrin [2], albolabrin [3], triflavin (or flavoridin) [4–6], trigramin [7,8], barbourin [9] and echistatin [2,10,11] have been shown to possess inhibitory effects on platelet aggregation [12]. These polypeptides can block aggregation through interference with fibrinogen binding to their specific receptors on the platelet surface membrane in a competitive manner and are designated as the “disintegrin” family of venom peptides.

Among disintegrins, echistatins are the smallest members of this family of polypeptides with regard to their molecular sizes. Four different echistatins, designated as echistatin \(\alpha1\), \(\alpha2\), \(\beta\) and \(\gamma\) [2,10,11], were isolated and their amino acid sequences were determined. The exact spatial linkage of disulfide bridges (Cys2-Cys11, Cys7-Cys32, Cys8-Cys37 and Cys20-Cys39) in echistatin \(\alpha1\), which was first predicted by a statistical analysis of the NMR-deduced structures [13] and by sequencing of proteolytic fragments [14], has recently been confirmed by selective chemical-reduction method [15]. The three-dimensional structure of echistatin \(\alpha1\) in solution, which was

1 Corresponding address: S.-H. Wu, Institute of Biological Chemistry, Academia, P.O. Box 23-106, Taipei, Taiwan, Fax: (886)-2-3635038. E-mail: shwu@gate.sinica.edu.TW.

Abbreviations: NMR, nuclear magnetic resonance; DQF-COSY, double-quantum filtered correlation spectroscopy; NOESY, nuclear Overhauser and exchange spectroscopy; NOE, nuclear Overhauser effect; DG, distance geometry; SA, simulated annealing; RMSD, root mean square deviation; HMP resin, \(p\)-hydroxymethylphenoxymethylpolystyrene resin; Fmoc-amino acid, N-fluorenylmethoxycarbonyl-amino acid; IC\(_{50}\), concentration at 50% inhibition.
Synthesis, Characterization and Bioassay of des(46-49)-[Ala₈,₃₇]-echistatin

The structural basis for the presence of multiple disulfide bridges in these small snake toxins. The exact role of each suspected disulfide bond in various disintegrin molecules in order to elucidate the role of the disulfide bond Cys₈-Cys₃₇ in relation to the structure-function aspects of echistatin γ, whose solution structure has not been determined to date. The synthetic analogue coupled with NMR analysis of its structure described here should provide an especially useful synthetic approach for assessing and dissecting the structural basis for the presence of multiple disulfide bridges in these small snake toxins.

MATERIALS AND METHODS

Synthesis, Characterization and Bioassay of des(46-49)-[Ala₈,₃₇]-echistatin γ

Starting with 0.25 mmol (0.295 g) of HMP resin (0.85 mmol/g), synthesis was carried out in a stepwise Fast-Moc protocol using an ABI automated peptide synthesizer [17]. Analytical HPLC was used to monitor the refolding progress by sulphydryl oxidation of synthesized peptide. The oxidized product was lyophilized. The product was dissolved in 0.4% acetic acid solution and filtered to remove the insoluble material. The filtrate was purified by HPLC using a C18 column and were pumped at a flow rate of 1 ml/min using a Carlo Erba HPLC system. Scans were taken in the positive ion mode over the m/z range of 400 to 1600. Mass calibration was carried out with an external horse heart myoglobin reference by using the VG Bio-Q software provided by the instrument supplier.

Amino acid analysis. The synthesized polypeptide was hydrolyzed in 0.6 ml of acid solution which contains 4 parts of 6 M HCl and 1 part of TFA at 140 °C for 3 h and amino acid analysis was carried out in a Beckman 6300 High Performance Analyzer.

Platelet aggregation assay. Platelet aggregation assay was performed in rabbit platelet-rich plasma (PRP). 9 ml of rabbit whole blood was mixed with 1 ml anticoagulant which contains 3.8% sodium citrate and 0.9% sodium chloride. The blood mixture was centrifuged at 1200 rpm for 10 min at 25 °C on Kubota 5700, RS-4/6 swing bucket rotor without brake. PRP was decanted and apyrase was added at a final concentration of 20 kU per ml PRP. This assay was performed on a JASCO CAF-100 Ca²⁺ analyzer. 0.5 ml PRP was incubated with stirring on the analyzer at 37 °C for 30 sec. The inhibitor was added and incubated for 1 min. Then 10 μM ADP was added to initiate the aggregation. The reaction was allowed to proceed for at least 2 min. The extent of inhibition of aggregation was expressed as the percentage of control aggregation observed in the absence of inhibitor.

NMR Measurements

NMR spectra were obtained with samples containing 2 mM peptide in 90% H₂O/10% D₂O. For the determination of the amide proton exchange rates, lyophilized peptides were dissolved in 99.996 atom % D₂O and the NMR spectra were recorded at different time intervals immediately after the dissolution. NMR spectra were recorded at two different temperatures, 298 °K and 302 °K, using a Bruker AMX 600 spectrometer equipped with an X32 computer. A set of standard two-dimensional spectra, including DQF-COSY [18], NOESY [19,20] and TOCSY [21,22] experiment, were accumulated in the phase-sensitive mode [23,24]. The mixing times (τ_m) of the NOESY experiments were 80, 150, 200 and 250 ms. NOESY data set consisted of 2048 data points in the t₁ dimension and 512 (TPPI method) t₂ increments. The TOCSY experiment was achieved by an MLEV-17 spin-lock pulse sequence [22]. For DQF-COSY and TOCSY experiments, 2048 data points were acquired in the t₂ dimension with 960 and 930 t₁ increments.
increments, respectively. FIDs were multiplied by skewed phase-shifted sine bell window functions prior to complex Fourier transformation. Data were zero filled to 2048 real points in both dimensions. Vicinal $J_{\text{Cys-NH}}$ coupling constants were measured from DQF-COSY spectra. Cross-sections taken along the $\omega_2$-axis were subjected to an inverse Fourier transformation zero-filled to 8092 data points and then Fourier-transformed to give a digital resolution of 0.8 Hz/point. Fourier transformation was carried out on an INDIGO workstation using Felix software (version 2.3.0).

Hydropathy Profile and Secondary Structure Prediction

A program analysis using Macintosh computers of the surface hydrophobicity of the native echistatin $\gamma$ and des(46-49)-[Ala$^{8,37}$]-echistatin $\gamma$ along their amino-acid sequences based on the Kyte-Doolittle hydropathy scale [25] is carried out on the MacVector sequence analysis software (International Biotechnologies, Inc., New Haven, CT). The signs of the values have been reversed in order to plot the hydrophilicity instead of hydrophobicity profile. A window of size $N = 7$ was run along the length of toxin segments; for each window, the hydropathy values of the 7 amino acids were summed and divided by 7 to obtain the average hydrophilicity per residue for the window. Values above the axis denote hydrophilic regions which may be exposed on the outside of the protein molecule whereas those values below the axis indicate hydrophobic regions which tend to be buried inside the protein.

Tertiary Structure Computations by Computer Modeling

Three-dimensional structures of des(46-49)-[Ala$^{8,37}$]-echistatin $\gamma$ were computed by using the hybrid distance geometry-dynamic simulated annealing procedure [26] employing the program X-PLOR [27]. During the calculation, the S-S covalent bonds were deleted first and they were reintroduced as pseudo-NOE distances with the S-S distances constrained to the upper limit of 2.1 Å. Thus, by this technique, the folding problem inherent in all real-space calculation methods is avoided. A family of 40 structures was generated using distance geometry algorithm [28]. After carrying out 200 cycles of Powell minimization on these structures, a total of 9.5 ps dynamic simulation was carried out following the procedure described by Brünger [28]. During the first phase of the dynamic simulation at 1000 K, the values of the force constant of the NOE term were kept constant at 50 kcal/(mol · Å$^2$), and those of the repulsion and torsion angle terms were gradually increased from 0.05 to 4.0 kcal/(mol · Å$^2$), and from 5.0 to 200.0 kcal/(mol · rad$^2$), respectively. The resulting structures were finally subjected to 500 cycles of restrained (Powell) minimization with the full CHARMM energy function [29]. In the next step, simulated annealing refinements were carried out on all these structures using the refinement protocol of Nilges et al. [26]. This consists of a 9.0 ps cooling dynamics, followed by 1000 cycles of Powell minimization. From the analysis of the resultant structures, 15 were chosen on the basis of an NOE energy cutoff value of 89.3 kcal/mol with no distance-constraint violation greater than 0.5 Å. All computations were carried out on an INDIGO workstation. The molecular graphics software QUANTA (version 4.0; Polygen/Molecular Simulations INC.) was used for the generation, display, analysis and plotting of molecular structures.

RESULTS AND DISCUSSION

The salient feature of snake venom toxins including the disintegrin family of polypeptides is the high number of disulfide cross-linkages present in various small venom toxins [29,30]. It is well known that disulfide bond formation between cysteine residues is a common occurrence in proteins and disulfide bridges thus formed are intimately linked with three-dimensional folding of the polypeptide chains [31]. The primary sequences of disintegrins are composed of 47-84 amino acid residues and high disulfide contents of four to seven disulfide bonds. They all contain the sequence Arg-Gly-Asp (RGD), a sequence segment that appears in a variety of adhesive proteins and contributes to the inhibition of platelet aggregation [32]. In this study we use solution NMR spectroscopy and molecular modeling methods to determine the tertiary structure of one analogue of echistatin $\gamma$ with three disulfide bonds in order to elucidate the role of the disulfide bond Cys8-Cys37 in relation to the structure-function aspects of echistatin $\gamma$, whose solution structure has not been determined to date.

Synthesis and Characterization

The synthetic peptides, des(46-49)-echistatin $\gamma$ and des(46-49)-[Ala$^{8,37}$]-echistatin $\gamma$, were generated by solid phase peptide synthesis method as described previously [17]. The purity and structure of these peptides were characterized by amino acid analysis, mass spectrometry and CD spectropolarimetry. Amino acid analysis after acid hydrolysis (70 h) showed residue ratios within 5% of those expected values. Sequence analysis of the product carried out for 47 cycles gave the
expected sequence. Further evidence for the correct molecular weight of 4919 was obtained by ion-spray mass spectroscopy. In the previous study [17] the CD spectra of native echistatin \( \gamma \) and des(46-49)-echistatin \( \gamma \) in the far UV-region (200–250 nm) are very similar. On the other hand, the CD spectrum of des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) (data not shown) was found to be slightly different from that of echistatin \( \gamma \), which may reflect some structural differences between des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) and the native echistatin \( \gamma \). The difference is probably due to the absence of a \( \beta \)-turn at residues 42-45 in des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) as determined by solution NMR described below.

Platelet aggregation assay. The data for platelet aggregation assay of echistatin \( \gamma \), des(46-49)-echistatin \( \gamma \) and des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) are shown in Table 1. Both des(46-49)-echistatin \( \gamma \) and des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) have almost similar IC\(_{50}\) of about 700 nM which is higher than that of 430 nM for echistatin \( \gamma \).

\( ^1 \)H NMR Study

In order to determine the structure of des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \), we have obtained a complete set of proton 2D NMR spectra for this synthetic analogue, including DQF-COSY, TOCSY and NOESY at different mixing times. Analysis of the amino acid spin systems began with the amide to aliphatic region of TOCSY spectra of des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) in H\(_2\)O (Fig.1). The \( \mathrm{C}_a \) proton resonances were identified by comparison with the fingerprint region of DQF-COSY spectra (H\(_2\)O). Five groups of spin systems were initially identified from the TOCSY experiment by inspection of the NH to \( \mathrm{C}_a \), \( \mathrm{C}_b \), \( \mathrm{C}_g \) and \( \mathrm{C}_d \) connections. The five groups were (i) Gly with two \( \mathrm{C}_a \) proton resonances, (ii) Thr and Ala spin systems, (iii) Ile and Leu spin systems, (iv) AMX spin systems with \( \mathrm{C}_b \) proton resonance between 2.5 and 4.0 ppm, and (v) Long side chain spin systems (Lys, Arg and Glu) where at least three cross peaks from the side chain to the amide proton were observed.

Des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \) contains three Pro, three Arg, three Lys, two Met, two Glu, three Asn, one Phe, one Tyr, one His, one Ser, six Cys and six Asp residues. Of these spin systems, the three Pro were identified by observing the complete connectivity pattern between the protons. The spin systems of Arg and Lys were distinguishable and were assigned by the sequential assignment procedure. The aromatic ring protons of Phe, Tyr and His were readily determined by inspection of the spectra in the aromatic region.

Using the sequential assignment procedure, specific assignments for the amino acids were obtained (data not shown). The sequential NOE connectivities \( d_{NN} \), \( d_{ON} \) and \( d_{IN} \), together with other NMR parameters are summarized in Fig. 2. The presence of \( \mathrm{CaH}_{i-1}/\mathrm{CaH}_i \) NOE peaks and the absence of \( \mathrm{CaH}_{i-1}/\mathrm{CaH}_i \) cross peaks suggests that all prolines are in the “trans” conformation.

Tertiary Structure Analysis by Computer Graphic Modeling

We have constructed the superposition of 15 structures of the backbone atoms of the des(46-49)-[Ala\(^{8,37}\)]-echistatin \( \gamma \), selected from 40 structures determined from NOE constraints. The structures are well defined and in good superposition among themselves with proper tertiary

<table>
<thead>
<tr>
<th>IC(_{50}) (nM)(^a)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>( \gamma )-Echistatin</td>
<td>430</td>
</tr>
<tr>
<td>des(46-49)-( \gamma )-Echistatin</td>
<td>720</td>
</tr>
<tr>
<td>des(46-49)-( \gamma )-Echistatin-[8-37]Ala</td>
<td>700</td>
</tr>
</tbody>
</table>

\(^a\) The bioassay data were expressed as the concentration of 50% inhibition (IC\(_{50}\)) and were the average value (± 10% S.D.) of triplicate analyses.
folding. The superposition of the three segments of this synthetic analogue regarding the backbone atoms in the two turn regions and α-carbon presentation in the RGD loop region is also excellent with the average RMS deviation from the mean structure is 0.68 Å for the backbone atoms and 1.15 Å for all atoms.

It is clear from the structural statistics (Table 2) that the calculated structures are in good agreement with the experimental data. None of the structures have violations of the NOE constraints greater than 0.5 Å and the deviations from the idealized covalent geometry are small. The nonbonded contacts for these structures are also characterized by a large negative Lennard-Jones van der Waals energy, showing the conversion of different structures into an low energy state of an acceptable conformation by energy minimization.

FIG. 1. TOCSY spectrum for the NH-aliphatic region of des(46-49)-[Ala^{8.37}]-echistatin γ (2 mM) dissolved in 90% H$_2$O/10% D$_2$O at 29°C. Relayed connectivities are indicated by a continuous line.
Comparison of surface hydrophilicity. We have also compared the general distribution of surface-charge groups in two synthetic derivatives of echistatin and the native toxin along their linear primary sequences (Fig. 3) using the program analysis of surface hydrophilicity based on the Kyte-Doolittle hydropathy scale [25]. It is of interest to note that the overall profiles of these three toxin analogues share a great similarity for the distribution of hydrophilic amino acids along the polypeptide chains. The main differences between the synthetic analogues and the native toxin lie in the C-terminal part due to the deletion of residues #46-49. The replacement of two Cys residues at #8 and #37 with alanine does not appear to change much on the surface hydrophobicity of the toxin molecules. Therefore the disulfide pair of Cys^8-Cys^37 may not play an essential role for the structural organization of the toxin polypeptide, which is strengthened by activity assay and NMR

**FIG. 2.** Summary of the observed sequential NOEs, coupling constants, amide proton exchange rate and hydrogen bonds used in the structural assignments. The relative intensities of NOE, classified into strong, medium and weak kinds, are indicated by the thickness of the lines. The filled circles under the amino acid residues indicate slow amide proton exchange. The values (in Hz) of coupling constants (J_{a,NN}) are also shown. Hydrogen bonds are indicated by arrows.

**TABLE 2**
Statistics of the Structure Calculations

<table>
<thead>
<tr>
<th></th>
<th>Average of 15 structures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMS deviations from idealized geometry used within X-PLOR</strong></td>
<td></td>
</tr>
<tr>
<td>Bond (Å)</td>
<td>0.0044 ± 0.0007</td>
</tr>
<tr>
<td>Angle (°)</td>
<td>0.8032 ± 0.0659</td>
</tr>
<tr>
<td>Improper (°)</td>
<td>0.5266 ± 0.0873</td>
</tr>
<tr>
<td>NOE</td>
<td>0.1216 ± 0.0034</td>
</tr>
<tr>
<td><strong>X-PLOR energy terms (kcal/mol)</strong></td>
<td></td>
</tr>
<tr>
<td>E_{NOE}</td>
<td>89.39 ± 42.39</td>
</tr>
<tr>
<td>E_{CDIH}</td>
<td>16.45 ± 11.82</td>
</tr>
<tr>
<td>E_{L-J}</td>
<td>-283.11 ± 11.10</td>
</tr>
</tbody>
</table>

*Note.* E_{L-J} is the Lennard-Jones van der waals energy calculated with CHARMM empirical energy function.
Moreover the C-terminal four residues (#46-49) were shown to have little effect on the biological activity and structure as revealed in our previous study [17]. The present study further demonstrates that the linear sequence of des(46-49)-[Ala8,37]-echistatin with three disulfide bonds (Cys2-Cys11, Cys7-Cys32 and Cys20-Cys39) is sufficient for specifying the biological function and folding pattern.

**CONCLUSION**

In comparison, the tertiary structure of des(46-49)-[Ala8,37]-echistatin is very similar to that of echistatin α1 [16] despite the lack of one pair of disulfide bond. The dynamic nature of the tertiary structure as revealed by the NMR study and computer graphics is shown in Fig. 4. The loop connecting residues #2-11 in the analogue is further away from the central loop due to the absence of the disulfide bond formed by Cys8 and Cys37. Similarly, in the analogue the C-terminal chain (residues #35-39) appears also away from the N-terminal loop (residues #2-11) as compared to that of the corresponding regions in echistatin α1. Interestingly the distance between the RGD loop and the Lys45 residue appears to be similar in the two structures. As we have demonstrated in this report, the RGD loop and the C-terminus are located at the most disordered region, presumably more flexible in terms of their solution conformation, which may facilitate binding of echistatin γ to its receptor.
ACKNOWLEDGMENTS

This work was supported by Academia Sinica and the National Science Council, Taipei, Taiwan. We are grateful to Dr. Ming-Fai Tam at the Institute of Molecular Biology, Academia Sinica, for performing the mass spectrometry.

REFERENCES


FIG. 4. Superposition of the backbone (N, Ca, C') atoms of the mean NMR structure of des(46-49)-[Ala\textsuperscript{8,37}]-echistatin \( \gamma \) (blue), with three disulfide bonds shown by yellow bars and marked with residue numbers, as compared with the mean NMR structure of echistatin \( \alpha 1 \) (red).