Protein Science (1994), *3*:1267–1275. Cambridge University Press. Printed in the USA.

Two-step selective formation of three disulfide bridges in the synthesis of the C-terminal epidermal growth factor-like domain in human blood coagulation factor IX

YAN YANG,[1] WILLIAM V. SWEENEY,[1] KLAUS SCHNEIDER,[2] BRIAN T. CHAIT,[2] AND JAMES P. TAM[3]

[1] Department of Chemistry, Hunter College of City University of New York, New York, New York 10021
[2] Laboratory for Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, New York 10021
[3] Department of Microbiology and Immunology, Vanderbilt University, Nashville, Tennessee 37232-2363

(RECEIVED April 4, 1994; ACCEPTED May 20, 1994)

Abstract

The 45-residue C-terminal EGF-like domain in human blood coagulation factor IX has been synthesized by a 2-step method to form selectively 3 disulfide bridges. Four out of 6 cysteines are blocked with either trityl or 4-methylbenzyl, and the remaining 2 cysteines are blocked with acetamidomethyl (Acm). In the first step, 4 free cysteinyl thiols are released concurrently with the removal of all protecting groups except Acm and are oxidized to form 1 of the 3 possible isomers containing 2 pairs of disulfides. In the second step, iodine is used to remove the Acm groups to yield the third disulfide bridge. This approach reduces the number of possible disulfide bridging patterns from 15 to 3. To determine the optimal protecting group strategy, 3 peptides are synthesized, each with Acm blocking 1 of the 3 pairs of cysteines involved in disulfide bridges: Cys^5 to Cys^{16} (Cys 1-3), Cys^{12} to Cys^{26} (Cys 2-4), or Cys^{28} to Cys^{41} (Cys 5-6). Only the peptide having the Cys 2-4 pair blocked with Acm forms the desired disulfide isomer (Cys 1-3/5-6) in high yield after the first step folding, as identified by proteolytic digestion in conjunction with mass spectrometric peptide mapping. Thus, the choice of which pair of cysteines to block with Acm is critically important. In the case of EGF-like peptides, it is better to place the Acm blocking groups on one of the pairs of cysteines involved in the crossing of disulfide bonds.

Keywords: C-terminal EGF-like domain; human blood coagulation factor IX; mass spectrometric peptide mapping; solid phase peptide synthesis; 2-step peptide folding; 3 disulfide bridges

Domains are common building blocks of proteins. They are structurally stable units capable of folding to tertiary structures resembling part of the native proteins. As a result, domains are targets of both biophysical and biochemical studies. The structural stability of domains is often aided by multiple disulfide bridges. One of the most well-studied domains is the epidermal growth factor (Savage et al., 1972). EGF and its family members are potent mitogens for a wide variety of cells in culture. EGF contains 3 disulfide bridges with a 1-3, 2-4, 5-6 pattern for pairing of the cysteinyl residues (Gregory, 1975).

Human blood coagulation factor IX contains 3 distinct classes of domains, including a vitamin K-dependent domain having γ-carboxyglutamic acid (Gla) residues, 2 tandemly linked EGF-like domains, and a serine protease domain (Yoshitake et al., 1985). The Gla domain binds Ca^{2+} and allows the protein to interact with phospholipids on a membrane surface (Astermark et al., 1991). The function of the EGF-like domains is less clear. The N-terminal EGF-like domain binds Ca^{2+} (Huang et al., 1989, 1991; Handford et al., 1990), but little is known about the C-terminal EGF-like domain.

Although the N-terminal domain has been successfully synthesized (Huang et al., 1989), to our knowledge no C-terminal EGF-like domain has been obtained by either biosynthesis or chemical synthesis. Thus, the synthesis of C-terminal EGF-like domain in human blood coagulation factor IX (84–128)

Reprint requests to: James P. Tam, Department of Microbiology and Immunology, Vanderbilt University, A5119 MCN, Nashville, Tennessee 37232-2363.

Abbreviations: Boc, *t*-butyloxycarbonyl; Fmoc, 9-fluorenylmethyloxycarbonyl; Trt, trityl (triphenylmethyl); Meb, 4-methylbenzyl; Acm, acetamidomethyl; TFA, trifluoroacetic acid; HF, hydrofluoric acid; Bzl, benzyl; Cz, benzyloxycarbonyl; Tos, tosyl; Bu, butyl; Pmc, pentamethylchroman; DMSO, dimethyl sulfoxide; TFE, trifluoroethanol; EGF, epidermal growth factor.

(fIX_{EGF-C}) was performed to obtain enough material for biophysical characterizations. For convenience, the amino acids are numbered 1–45, corresponding to 84–128 in the human blood coagulation factor IX sequence (Fig. 1) (Yoshitake et al., 1985).

In general there are 2 approaches for the formation of intramolecular disulfide bridges in synthetic peptides, the single-step approach (Tam et al., 1991) and the sequential approach (Van Rietschoten et al., 1977; Akaji et al., 1992). In the first approach, all disulfide bridges are formed in a single step either by direct oxidation with an oxidant (such as DMSO or air) or disulfide exchange with reduced/oxidized glutathione. This approach depends on the ability of the peptide to fold into the native state. Our initial attempts using this approach to synthesize fIX_{EGF-C} failed to produce a large proportion of the peptide with the desired disulfide pairings (Fig. 2). The amino acid composition and the molecular weight of the linear peptide were confirmed, but the peptide did not yield a single major peak by HPLC under a wide range of folding conditions (e.g., Fig. 2). Thus, the C-terminal EGF-like synthetic peptide exhibited a significant multiple conformation problem that led to misformed disulfides, even though the N-terminal EGF-like synthetic peptide folded in a well-behaved manner to good yield.

The second approach is to form the sulfur–sulfur bridges sequentially. Various schemes using this approach have been employed to guide the formation of 2 disulfide bridges (Barany & Merrifield, 1979; Akaji et al., 1992). Usually the first pair of cysteines is freed and the disulfide bond formed after the acidic cleavage of all non-cysteine protecting groups. The second pair is then liberated by a method that will not interfere with the stability of the previously formed disulfide bridge using an electrophilic agent such as I_2 under acidic conditions. However, sequential formation of a 3-disulfide peptide is a difficult undertaking because of the need for 3 orthogonal sets of thiol protecting groups. We have therefore sought to provide a hybrid 2-step approach that utilizes the selectivity of the sequential approach and the generality of the single-step approach. In the first step, our strategy treats the 3-disulfide-containing peptide as if it has 2 pairs of disulfides, while the remaining disulfide bridge is formed in the second step.

With 6 cysteines, there are 15 possible arrangements of disulfide bridges. All have been found to exist in nature (Warne & Laskowski, 1990). Using our strategy, the possible disulfide patterns can be reduced from 15 to 3 in the first step. With an appropriate choice for which disulfide bridge will be formed in the second step, disulfide isomers that are difficult to synthesize can often be obtained. In this paper, we describe our application of this strategy for the synthesis of fIX_{EGF-C} which contains 3 disulfide bridges (Fig. 1).

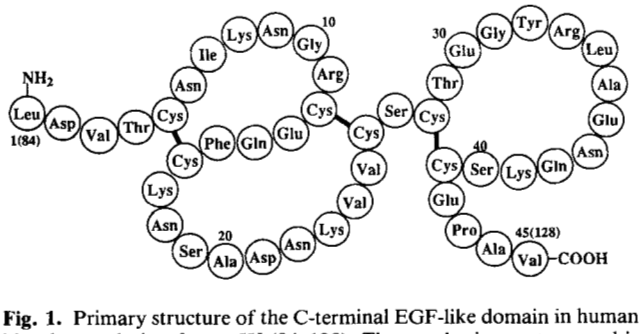

Fig. 1. Primary structure of the C-terminal EGF-like domain in human blood coagulation factor IX (84–128). The numbering system used is based on the synthetic peptide, with the corresponding factor IX sequence position shown in parentheses.

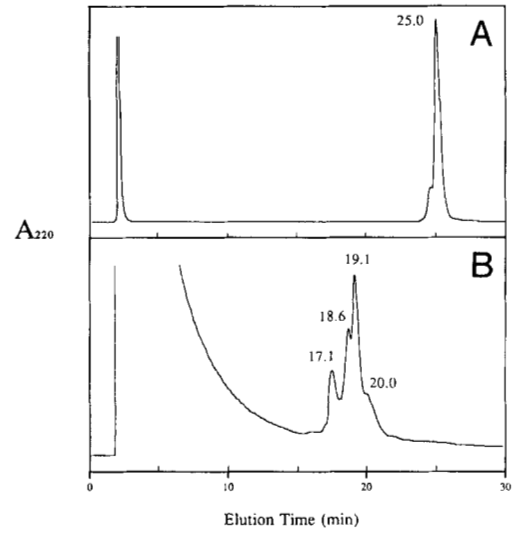

Fig. 2. Reversed-phase HPLC profile showing the folding results of the C-terminal EGF-like domain in human blood coagulation factor IX resulting from using a single-step approach. Peptides were eluted with a 10–40% buffer B linear gradient over 30 min at 1.5 mL/min, monitored at 220 nm. Buffer A contained 5% acetonitrile in 0.037% TFA. Buffer B contained 60% acetonitrile in 0.045% TFA. **A:** Reduced peptide, with all 6 cysteines free, before folding. **B:** Peptide after single-step folding for 29 h at pH 8.0 in the presence of reduced and oxidized glutathione (1 mM, ratio 1:1).

Results

Synthetic approach

The 6 cysteines were selectively protected with 2 different blocking groups: 4 with either trityl for Fmoc chemistry or 4-methylbenzyl for Boc chemistry, and 2 with acetamidomethyl. Disulfide formation was performed in a 2-step procedure (Fig. 3). In the first step, either TFA or HF was used to cleave the peptide from the resin and simultaneously deblock all the side-chain protecting groups except the Acm moiety. Thus, after cleavage, 4 cysteines were free and available to form 2 disulfide bridges in possibly 3 different disulfide isomers. In the second step, iodine was used to remove the Acm groups (Veber et al., 1972; Hiskey, 1981) from the last 2 cysteines to yield the third disulfide bridge. Because this step is conducted in acidic solution (pH 2), disulfide scrambling has been shown not to occur.

The desired product would have disulfide pairing (Cys 1-3/2-4/5-6) analogous to that of EGF (Gregory, 1975; Doolittle et al., 1984; Yoshitake et al., 1985). There are therefore 3 logical choices for blocking the cysteines corresponding to each of the 3 desired disulfide pairs. Acm could be used to block cysteines 5 and 6 (peptide A), cysteines 2 and 4 (peptide B), or cysteines 1 and 3 (peptide C). All 3 peptides A, B, and C were prepared to examine the effect of the position of the Acm groups on the disulfide pairing pattern (Table 1).

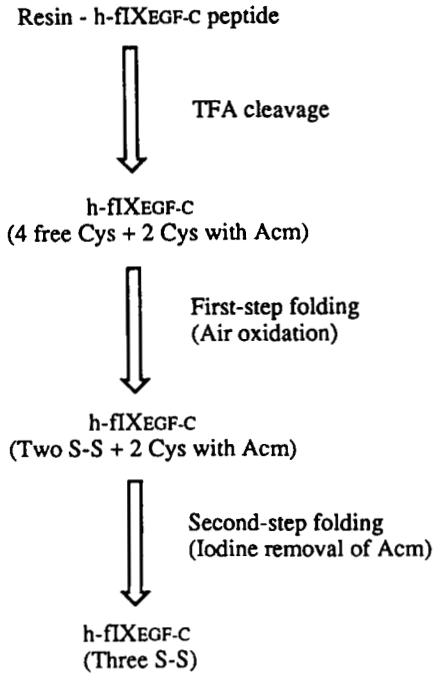

Fig. 3. Scheme for 2-step selective formation of 3 disulfide bridges.

Characterization of linear and folded peptides

The amino acid composition of the linear peptide was examined by amino acid analysis and was found to agree with the sequence (data not shown). The molecular weight of the synthetic peptide with 2 Acm protecting groups was determined by matrix-assisted laser desorption mass spectrometry. The mass found for peptide A with 2 disulfides was 5,108.1 ± 1.0 u (calculated mass 5,107.8 u) (Fig. 4). The masses found for peptides B and C, prior to disulfide formation, were 5,110.9 ± 1.0 u and 5,111.5 ± 1.0 u, respectively (calculated mass 5,111.8 u). The fully folded final products of peptide A and B were also examined. Molecular weights of 4,963.0 ± 1.0 u and 4,963.8 ± 1.0 u were found, consistent with the calculated molecular weight of 4,963.8 u for peptides A and B.

A summary of the pattern of blocking groups used for peptides A, B, and C, together with the final folded products found, is given in Table 2.

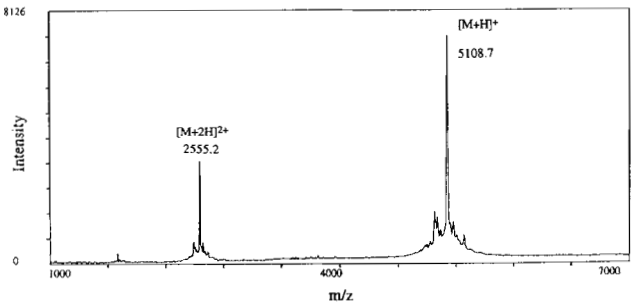

Fig. 4. Matrix-assisted laser desorption mass spectrum for peptide A with 2 disulfide bridges formed after the first step.

Determination of the 2 disulfide bridges in peptide A after first-step oxidation

In peptide A, Cys 5 and Cys 6 were blocked with Acm. The first-step folding and oxidation were performed under a wide range of conditions including pH ranging between 4 and 10, different solvents, with or without oxidized/reduced glutathione in aqueous solution, in 15% DMSO, or in 15% DMSO/60% TFE, in the absence or presence of urea. Only 1 major product, accounting for 80–85% of the starting peptide, was observed by reversed-phase HPLC. The folded peptide eluted at 20.4 min as compared to the unfolded form eluted at 23.0 min in HPLC.

The connectivity of the disulfide bridges formed by air oxidation was determined by proteolytic digestion in conjunction with mass spectrometric peptide mapping. Enzymatic digestion of this product by *Staphylococcus aureus* protease V8, which cleaved the peptide after Glu residues, yielded 4 peaks by HPLC. Each of the 4 peaks examined by matrix-assisted laser desorption mass spectrometry (Table 3) identified peptide fragments ([1–13] and [14–30]) consistent with a disulfide pairing of Cys 1-2/3-4. Thus, by blocking Cys 5 and Cys 6, the major product was a misfolded disulfide isomer (Fig. 5).

Determination of the 2 disulfide bridges in peptide B after first-step oxidation

Peptide B contained Cys 2 and Cys 4 blocked with Acm. To obtain an EGF-like folding pattern, a Cys 1-3/5-6 pairing was required after first-step folding. Under a wide range of folding conditions, similar to those used for peptide A, a single major

Table 1. *Locations of the cysteine residues and their blocking groups in 3 synthetic peptides homologous to the C-terminal EGF-like domain in human blood coagulation factor IX ($fIX_{EGF\text{-}C}$)*[a]

Residue number:	1(84)	5		12		16		26		28		41	45(128)
Sequence:	H-LDVT	C	NIKNGR	C	EQF	C	KNSADNKVV	C	S	C	TEGYRLAENQKS	C	EPAV-OH
Cysteines		1		2		3		4		5		6	
Peptide A[b]		Meb		Meb		Meb		Meb		Acm		Acm	
Peptide B		Trt		Acm		Trt		Acm		Trt		Trt	
Peptide C		Acm		Trt		Acm		Trt		Trt		Trt	

[a] The numbering system used is based on the synthetic peptide, with the corresponding factor IX sequence position shown in parentheses.

[b] In peptide A, a Meb protecting group was used in place of Trt, because Boc chemistry was employed in peptide synthesis.

Table 2. *Summary of the disulfide isomers found after first-step folding of peptides A, B, and C*

Peptide	Blocked cysteines	Folding isomers		Results
		Number	Pattern	
A	5, 6	1	Cys 1-2/3-4	Undesired pairing
B	2, 4	1	Cys 1-3/5-6	Desired pairing
C	1, 3	3	Cys 2-6; Cys 2-5/4-6; Cys 2-4/5-6	Mixture

Table 3. *Identification by matrix-assisted laser desorption mass spectra of 4 HPLC peaks generated in the protease V8 enzymatic digestion of peptide A*

Peak number (retention time)	Determined mass	Calculated mass[a]	Corresponding fragments	Disulfide pairing
1 (3.2 min)	1,045.7	1,046.2	[37-45]	–
	(1,845.8)	–	Unidentified[b]	–
2 (4.4 min)	1,945.5	1,945.2	[14-30]	Cys 3-4
3 (6.8 min)	1,463.3	1,462.7	[1-13][c]	Cys 1-2
	1,735.7	1,736.0	[31-45]	–
	(1,606.8)	–	Unidentified[b]	–
4 (8.8 min)	1,463.5	1,462.7	[1-13][c]	Cys 1-2
	1,928.5[d]	1,945.2	[14-30]	Cys 3-4

[a] Masses are calculated assuming oxidized cysteines.
[b] No disulfide pairing matches this mass.
[c] These fragments have the same mass but elute with different retention times. The reason for this is not known.
[d] The difference between calculated and determined mass, −17, is possibly caused by the loss of NH_3 and the formation of pyrGlu[14] at the N-terminus.

peak was found by HPLC. It accounted for 80–85% of starting peptide and had a retention time of 18.1 min as compared to 23.5 min for the linear peptide. Three different enzyme digestions were performed to determine the disulfide pairing pattern.

Tryptic digestion, which cleaves after Lys and Arg, gave 1 major peak and at least 2 minor peaks in reversed-phase HPLC. The results, including mass spectrum analysis, are shown in Table 4. One of the minor peaks gave a fragment corresponding to either [1-8] + [9-17] or [1-11] + [12-17], indicating a Cys 1-3 bridge. Similarly, the major HPLC peak contained a [1-8] + [12-17] fragment, once again consistent with a Cys 1-3 disulfide bridge. Another of the minor peaks was found to be a misfolded fragment having a Cys 3-5 disulfide link (fragment [12-17] + [24-33]). This misfolded component accounted for less than 15% of the total digestion. To define the other disulfide, protease V8 was used. After V8 digestion, a [1-30] + [37-45] fragment was identified in the digestion mixture, confirming that the Cys 5-6 disulfide was also present (Table 4). In order to further corroborate these results, the peptide was digested by a third enzyme, endoproteinase Lys-C, which cleaves after lysine. Using this enzyme, both disulfide bridges Cys 1-3 and Cys 5-6 were identified in the digestion mixture (fragment [1-8] + [9-17], measured mass 2,058.7 ± 1.0 u, calculated mass 2,058.1 u, and

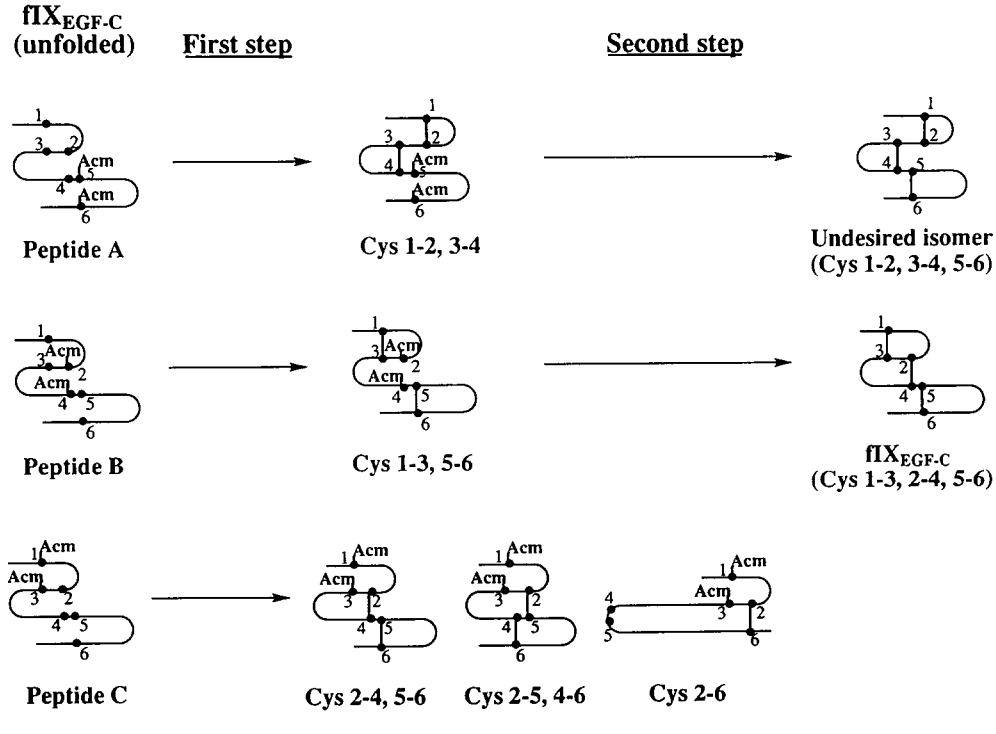

Fig. 5. Results of 2-step selective formation of disulfide bridges in peptides A, B, and C.

Table 4. *Identification by matrix-assisted laser desorption mass spectra of HPLC peaks generated in the tryptic digestion and V8 digestion of peptide B*

Enzyme	Peak number (retention time)	Determined mass	Calculated mass[a]	Corresponding fragments	Disulfide pairing
Trypsin	1 (8.3 min)	2,013.4	2,013.3	[12–17] + [24–33]	Cys 3-5
	2 (12.4 min)	2,059.7	2,058.1	[1–8] + [9–17] (or [1–11] + 12–17])	Cys 1-3
	3 (14.1 min)	1,731.8	1,730.8	[1–8] + [12–17]	Cys 1-3
		2,060.0	2,058.1	[1–8] + [9–17] (or [1–11] + 12–17])	
V8	Digestion mixture	4,435.0	4,436.1	[1–30] + [37–45]	Cys 5-6

[a] Masses are calculated assuming oxidized cysteines.

fragment [24–39] + [40–45], measured mass 2,474.3 ± 1.0 u, calculated mass 2,473.8 u) (Fig. 6). Thus, peptide B gave predominately the properly folded product (Cys 1-3/5-6) (Fig. 5).

Determination of the 2 disulfide bridges in peptide C after first-step folding

In peptide C (Cys 1 and Cys 3 blocked with Acm), a mixture of folding products was observed. There were at least 3 peaks in the HPLC profile (Table 5). Peak 1, accounting for 28% of the total area under the HPLC peaks, was relatively sharp. It contained an isomer, identified by tryptic digestion, with only 1 disulfide bridge formed (Cys 2-6) and 2 free cysteines (Cys 4 and Cys 5) (Table 5). Peaks 2 and 3 (23% and 39%) were partially overlapped. These peaks resulted from 2 isomers with closely matched retention times, differing by less than 1 min. Each isomer contained 2 disulfide bridges. Theoretically, isomers with Cys 2-4/5-6, Cys 2-5/4-6, and Cys 2-6/4-5 were possible, but no fragment consistent with the Cys 2-6/4-5 isomer was found. The Cys 2-4/5-6 and Cys 2-5/4-6 isomers could not be distinguished by tryptic digestion because trypsin does not cleave between Cys 4 and 5 (Cys^{26}-Ser^{27}-Cys^{28}) (Table 5). Thus, the desired Cys 2-4/5-6 isomer was present, although together with the misfolded Cys 2-5/4-6 isomer. Attempts to use partially purified fractions to proceed with the second-step folding failed to yield a significant amount of peptide containing the desired disulfide pairing pattern.

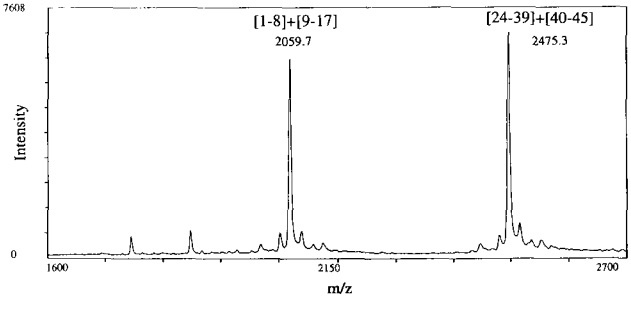

Fig. 6. Portion of the matrix-assisted laser desorption mass spectrum of peptide B after digestion by endoproteinase Lys-C. Two fragments, [1–8] + [9–17] and [24–39] + [40–45], are found that demonstrate the presence of desired Cys^{5}-Cys^{16}/Cys^{28}-Cys^{41} disulfide pairing in peptide B.

Disulfide oxidation in second-step oxidation and folding

The second-step oxidation was performed by iodine in acidic solution to remove the Acm. The oxidation was under kinetic control and was usually completed within an hour. We were unable to determine the location of the 3 disulfide bridges after formation of the third disulfide bond. An attempt was made to locate these bonds using enzyme digestion, but the required cleavage between Cys 4 and Cys 5 (Cys^{26}-Ser^{27}-Cys^{28}) was not accomplished. In any case, however, thiol–disulfide exchange is prevented by the acidic condition (pH 2) employed. Figure 7 shows a chromatogram of the second-step folding process for peptide B. A single major peak accounting for 60% of the starting peptide is observed by reversed-phase HPLC.

Discussion

EGF-like domains are common mosaic blocks in proteins. Our results show that not all such domains are capable of folding to tertiary structures with the desired disulfide pairings. There are a number of reasons to expect that a synthetic peptide may not adopt the native peptide folding pattern to give the desired disulfide isomer. Natural peptides can fold sequentially as they come off the ribosomes, sometimes with the aid of chaperone proteins (Ellis, 1994). Also, the folding of individual domains within a larger protein may be influenced by the presence of the rest of the protein. Both of these factors are absent in the folding of synthetic peptides. We develop here a 2-step strategy to aid the formation of desired disulfide bonds when it is not possible using the conventional chemical approach of simultaneously forming all 3 disulfides.

The number of disulfide bridges in single-chain proteins varies from 0 to more than 12 (Mao, 1989). A molecule containing N disulfide bridges can be crosslinked intramolecularly in $(2N)!/(2^{N}N!)$ different ways. An EGF-like peptide contains 3 disulfide bridges and therefore has 15 different possible disulfide crosslinking patterns.

It is not uncommon to encounter problems in inducing synthetic peptides to fold correctly. Several groups have reported that misfolded isomers are found together with the desired

Table 5. *Identification by matrix-assisted laser desorption mass spectra of 3 HPLC peaks generated after the first-step folding of peptide C*[a]

Peak number (retention time)	Determined mass[b]	Calculated mass[c]	Corresponding fragments	Disulfide pairing
1 (19.8 min)	976.1	975.9	[1–8]	–
28% of total	1,155.7	1,155.3	[9–17]	–
peptide	1,758.2	1,758.0[d]	[9–17] + [40–45]	Cys 2-6
	2,872.1	2,872.2	[9–17] + [24–33] + [40–45]	Cys 2-4/5-6 (or Cys 2-5/4-6)
	3,174.3	3,174.5	[12–13] + [40–45] + H_2O[d]	Same as above
	3,502.2	3,501.8	[9–33] + [40–45] + H_2O[d]	Same as above
2 (21.4 min)	1,155.7	1,155.3	[9–17]	–
23% of total	2,872.7	2,872.2	[9–17] + [24–33] + [40–45]	Cys 2-4/5-6 (or Cys 2-5/4-6)
peptide	3,175.2	3,174.5	[12–13] + [40–45] + H_2O[e]	Same as above
	3,502.2	3,501.8	[9–33] + [40–45] + H_2O[e]	Same as above
3 (22.3 min)	976.1	975.9	[1–8]	–
39% of total	1,155.7	1,155.3	[9–17]	–
peptide	2,872.3	2,872.2	[9–17] + [24–33] + [40–45]	Cys 2-4/5-6 (or Cys 2-5/4-6)
	3,174.8	3,174.5	[12–13] + [40–45] + H_2O[e]	Same as above
	3,501.9	3,501.8	[9–33] + [40–45] + H_2O[e]	Same as above

[a] Enzymatic digestion is by trypsin.
[b] The data indicating incomplete digestion are not listed in this table.
[c] Masses are calculated assuming oxidized cysteines.
[d] Mass 18, named H_2O for convenience, is assembled if the digestion cleavage occurs in a peptide loop. Although new C- and N-termini are produced, the 2 fragments are still linked together by a disulfide bridge.
[e] This mass is calculated assuming only 2 cysteines oxidized.

disulfide-pairing peptide during synthesis of EGF-like domains when a single-step folding approach is used (Violand et al., 1991; Hunter et al., 1993). A coincident problem is the separation and identification of the isomers, always a difficult task. Also, the location of the position of the 3 disulfide bridges in a given isomer can be particularly challenging. Thus, significant advantages are achieved by the present 2-step approach to disulfide bridge formation for synthetic peptides.

In addition to the peptide synthesis reported here, we have used this 2-step approach to synthesize an N-terminal EGF-like domain in human blood coagulation factor X (unpubl. data), and a malaria erythrocyte protein known as a K1 that has an EGF-like disulfide structure (Spetzler et al., 1994). We have also found this strategy to be successful in the synthesis of a neurotoxin (scratcher peptide) that contains a 1-6, 2-4, and 3-5 disulfide structure (C. Rao & J.P. Tam, unpubl. data).

The results of the present study demonstrate that the choice as to which cysteines to block with Acm is important. Only 1 of the 3 possible choices produces a properly folded product in high yield (Table 2; Fig. 5). To rationally design the synthesis, it is necessary to make an educated guess concerning the most probable disulfide pairing patterns to expect for a given placement of the Acm blocking groups.

In general, the 1-2, 3-4 pattern of 2 intrachain disulfide bridges appears to provide much less synthetic difficulty than the other patterns, implying a preference of synthetic peptides to adopt this pairing pattern. A survey of disulfide bonding patterns among native peptides with 2 disulfide bonds (Benham & Jafri, 1993) found that the 1-2, 3-4 cysteine pairing pattern was by far the most common. However, an examination of peptide synthesis papers reporting difficulty with disulfide formation (i.e., those using a sequential approach for disulfide bond formation as opposed to single-step formation) shows that peptides with crossed disulfide bonds (1-3, 2-4 and 1-4, 2-3 pairing patterns) to be reported much more frequently than peptides with 1-2, 3-4 disulfide bonds (Van Rietschoten et al., 1977; Nishiuchi & Sakakibara, 1982; Gray et al., 1984; Kumagaye et al., 1988; Tam et al., 1990; Zhang & Snyder, 1991; Akaji et al., 1992). These findings suggest that an effort to force the peptide into the desired Cys 1-3, 2-4 pattern by blocking the 5-6 pair with Acm will probably fail because of the observed tendency to produce 1-2, 3-4 disulfide pairing. Indeed, for this case we obtained a high yield of the undesired Cys 1-2, 3-4, 5-6 isomer (peptide A, Table 2).

The remaining 2 choices for positioning the Acm blocking groups are the Cys 1-3 pair (peptide C) or Cys 2-4 pair (peptide B). Both peptides were made, but only the peptide B yielded the desired disulfide isomer. An explanation of the difference in behavior between the peptides B and C is found by consideration of the number of residues separating the cysteines. Because there is only 1 residue between the fourth and fifth cysteines (Cys^{26} and Cys^{28}), in peptide C the likelihood of forming an undesired 2-5 cysteine bond might be expected to be almost as large as that for forming the desired 2-4 cysteine linkage. Thus, a peptide with Acm blocking Cys 1-3 (peptide C) would be expected to yield a mixture of peptides, with both 2-5 and 2-4 disulfide bonds present in the isomer mixture. This is precisely what we observed experimentally (Tables 2, 5).

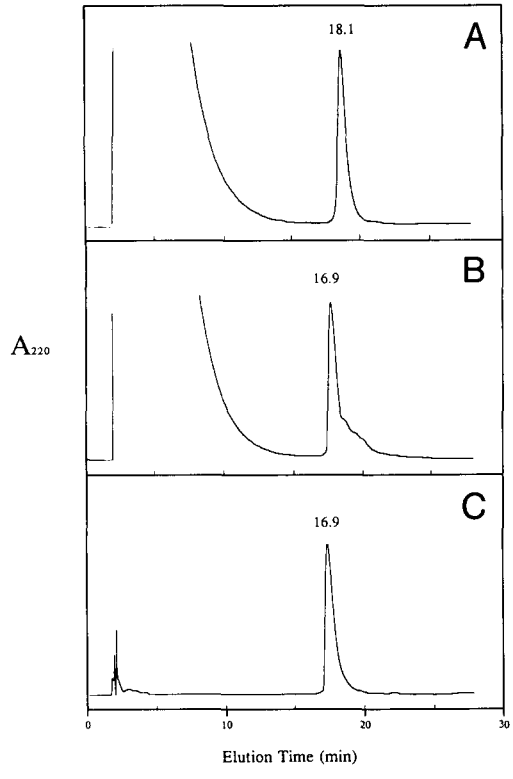

Fig. 7. Reversed-phase HPLC profile showing the formation of the third disulfide bridge for peptide B. Peptides were eluted with a 10–40% buffer B linear gradient over 30 min at 1.5 mL/min, monitored at 220 nm. Buffer A contained 5% acetonitrile in 0.037% TFA. Buffer B contained 60% acetonitrile in 0.045% TFA. **A:** Purified peptide B with 2 disulfide bridges formed after the first step. **B:** Peptide B after the third disulfide bridge was formed by removal of Acm and oxidation in 10% HAc solution with 5 mM iodine. Sodium thiosulfate (50 mM) was added to react with the excess iodine. **C:** The final product with 3 disulfide bridges purified by preparative HPLC.

An additional consideration is that pairs of cysteines with only 1 or 3 intervening residues are very unlikely to form disulfide bonds (Zhang & Snyder, 1989). This is evidenced by the observation that, after the first folding step, peptide C (Acm on Cys 1,3) yielded a peptide, among others, with a 2-6 disulfide bond and 2 free cysteines (Cys 4 and Cys 5). These 2 cysteines are separated by only 1 residue, and the small ring that would be formed by a disulfide bond was so unfavorable as to prevent formation.

The considerations outlined above indicate that an optimum choice for positioning the Acm blocking groups can often be made for peptides containing 3 disulfides with a single crossing of disulfide bonds. It is better to place the Acm blocking groups on one of the pairs of cysteines involved in the crossing of disulfide bonds, preferably on the center loop (if this is a choice). Supplementary considerations regarding the number of intervening residues between cysteines must also be taken into account.

We recommend that a 2-step approach for synthesis of peptides with 3 disulfide bonds be generally used, even for peptides that are not yet known to exhibit difficulty in folding. Our approach is easy to employ and it provides protection against potential unwanted disulfide isomers. If, at a later date, a peptide without Acm blocking groups is desired, it can be easily obtained by treatment of the blocked peptide with iodine.

Materials and methods

Peptide synthesis

Synthetic peptides A, B, and C were prepared using either Boc or Fmoc chemistry with stepwise solid-phase technology (Merrifield, 1963).

Peptide A was synthesized manually using the Boc approach and standard coupling and deprotection techniques. N^{α}-Boc-Val-OCH_2-Pam resin (0.76 mmol/g) was used, where Pam is phenylacetamidomethyl. All amino acids were purchased from Applied Biosystems, Inc. (Foster City, California). Side-chain protecting groups were as follows: Asp and Glu by OBzl, Ser and Thr by Bzl, Tyr by 2-bromo-Cz, Lys by 2-chloro-Cz, and Arg by Tos. Both Meb and Acm were used for blocking Cys. To remove the peptide from the resin, a low/high HF method (Tam et al., 1983) was used. After removal of the HF, the residue was washed with cold ethyl ether/mercaptoethanol 98:2 (v/v) to remove organic contaminants. The crude peptide was then extracted with 8 M urea/0.2 M dithiothreitol/0.1 M Tris-HCl buffer, pH 8.4. The peptide solution was dialyzed (Spectrapor, molecular weight cut-off, 2,000) against deaerated and N_2-purged 8 M, 5 M, 3 M, and 1.5 M urea, all in 0.1 M Tris-HCl (pH 8.4). The dialyzed solution was then diluted with same 0.1 M Tris-HCl buffer to a final urea concentration 0.3 M for the first-step folding (Heath & Merrifield, 1986). The peptide was purified to homogeneity by preparative HPLC. Some purified peptide was then reduced and used to examine various folding conditions.

Linear peptides B and C were synthesized by the RCMI Synthesis and Sequencing Facility at Hunter College. An ABI 430A synthesizer with Fmoc chemistry was used. All amino acids were purchased from Applied Biosystems, Inc. (Foster City, California). Side-chain protecting groups were as follows: Asp and Glu by OtBu, Ser, Thr, and Tyr by *t*-Bu, Asn and Gln by Trt, Lys by Boc, and Arg by Pmc. Both Trt and Acm were used for blocking cysteines. A 4-(hydroxymethyl)phenoxymethyl-Copoly resin (HMP resin/Wang resin) with a substitution of 0.88 mmol/g was used. For peptide B, the initial weight of the HMP resin was 0.284 g, and the final resin containing the 45-residue peptide was 2.53 g (99% of theoretical yield). TFA cleavage simultaneously deblocked all the side-chain protecting groups except Acm. The crude peptide, including the Acm blocking group, was precipitated in cold methyl *t*-butyl ether and collected by centrifugation. From 0.3 g resin B, 0.128 g of crude peptide B was obtained. Similar results were obtained for peptide C. Crude peptides were purified before use for subsequent folding experiments.

Disulfide oxidation

Disulfide oxidation and purification were monitored by analytical C18 reversed-phase HPLC. Peptides were eluted with a 10–40% buffer B linear gradient over 30 min at 1.5 mL/min, monitored at 220 nm. Buffer A contained 5% acetonitrile in 0.037% TFA. Buffer B contained 60% acetonitrile in 0.045% TFA.

Formation of 2 disulfide bridges in the first step

Peptides folding at 10^{-5} M were examined in buffers (10–50 mM) over a range of pH values between 4 and 10 under air oxidation. However, typically folding was carried out at pH 7.5 in 10 mM phosphate buffer at room temperature with slow stirring and monitored by HPLC. The time for completion varied from 24 to 48 h, depending on the temperature and pH. At room temperature and pH 7.5, peptide folding was usually complete within 24–30 h. Because there is always a small portion of misfolded isomers present, the major product was usually purified before carrying on the second step.

Formation of the third disulfide bridge in the second step

Acetic acid was added to the peptide solution to give a 10% acid solution. The solution was bubbled with nitrogen for 10 min before adding 5 mM iodine in methanol dropwise until a brown color remained. The reaction mixture was bubbled continuously with nitrogen in a darkened vessel. The reaction was completed in 1 h, as monitored by HPLC. After the solution was cooled in an ice-bath, 100 mM sodium thiosulfate was added to reduce excess iodine. The solution containing the crude peptide mixture was purified by preparative C18 reversed-phase HPLC (2.5 × 30 cm) in the same manner as described for the analytical HPLC. The product was eluted with aqueous acetonitrile/0.10% TFA at a flow rate of 20 mL/min. The overall synthetic yield was about 10%.

Enzymatic digestion

The identification of the disulfide bridge pairings utilized proteolytic digestion combined with matrix-assisted laser desorption mass spectrometry for the assignment of the disulfide bridge locations. Enzymes were purchased as follows: *S. aureus* protease V8 (Sigma), trypsin (Sigma), and endoproteinase Lys-C (Boehringer Mannheim). In each case, the monitoring and separation of individual fragments of digestion was performed by C18 reversed-phase HPLC.

Protease V8 digestion: A 50 mM ammonium bicarbonate buffer (pH 8.0) was used. The ratio of enzyme to substrate (w/w) was 1:50. The mixture of peptide and protease V8 was incubated at 37 °C for 5 h. *Trypsin digestion:* A 50 mM ammonium bicarbonate buffer (pH 8.0) was used. The ratio of enzyme to substrate (w/w) was 1:25. The mixture of peptide and trypsin was incubated at 37 °C for 4 h. *Endoproteinase Lys-C digestion:* A 50 mM ammonium bicarbonate buffer (pH 8.5) was used. The ratio of enzyme to substrate (w/w) was 1:20. The mixture of peptide and endoproteinase Lys-C was incubated at 35 °C for 4 h.

Mass spectrometry

HPLC-purified peptides, enzymatic digestion mixture of peptides, and enzymatically generated peptide fragments isolated by reversed-phase HPLC were subjected to analysis on a matrix-assisted laser desorption time-of-flight mass spectrometer constructed at The Rockefeller University and described elsewhere (Beavis & Chait, 1989, 1990). The mass spectra were collected by adding individual spectra obtained from 200 laser shots. Samples were prepared for laser desorption mass analysis as follows: The laser desorption matrix material 4-hydroxy-α-cyano-cinnamic acid was dissolved in 0.1% TFA/acetonitrile 2:1 (v/v) to a concentration of 50 mM. A 20 μM aqueous solution of the sample was then added to the matrix solution to give a final sample concentration of approximately 2 μM. A small aliquot (0.5 μL) of this mixture was applied to the metal probe tip and dried at room temperature with forced air. The sample was then inserted into the mass spectrometer and analyzed. Bovine insulin was used to calibrate the mass spectra.

Acknowledgments

This work was supported in part by US PHS grants HL41935 (J.P.T. and W.V.S.), CA36544 (J.P.T.), and RR00862 (B.T.C.), and by US PHS grant RR03037 to the Hunter College Synthesis and Sequence Facility. Some aspects of the peptide synthesis were performed using facilities generously provided by Dr. R. Bruce Merrifield of The Rockefeller University. A preliminary report of part of this work was presented at the 13th American Peptide Symposium (Edmonton, Canada, June 1993).

References

Akaji K, Fujino K, Tatsumi T, Kiso Y. 1992. Regioselective double disulfide formation using silylchloride-sulfoxide system. *Tetrahedron Lett 33*:1073–1076.

Astermark J, Björk I, Öhlin AK, Stenflo J. 1991. Structural requirements for Ca^{++} binding to the γ-carboxyglutamic acid and epidermal growth factor-like regions of factor IX. *J Biol Chem 266*:2430–2437.

Barany G, Merrifield RB. 1979. Solid-phase peptide synthesis. In: Gross E, Meienhofer J, eds. *The peptides: Analysis, synthesis, biology, vol 2. Special methods in peptide synthesis, part A*. New York: Academic Press. pp 233–243.

Beavis RC, Chait BT. 1989. Factors affecting the ultraviolet laser desorption of proteins. *Rapid Commun Mass Spectrom 3*:233–237.

Beavis RC, Chait BT. 1990. High-accuracy molecular mass determination of proteins using matrix-assisted laser desorption mass spectrometry. *Anal Chem 62*:1836–1840.

Benham CJ, Jafri MS. 1993. Disulfide bonding patterns and protein topologies. *Protein Sci 2*:41–54.

Doolittle RF, Feng DF, Johnson MS. 1984. Computer-based characterization of epidermal growth factor precursor. *Nature 307*:558–560.

Ellis RJ. 1994. Roles of molecular chaperones in protein folding. *Curr Opin Struct Biol 4*:117–122.

Gray WR, Luque FA, Galyean R, Atherton E, Sheppard RC, Stone BL, Reyes A, Alford J, McIntosh M, Olivera BM, Cruz LJ, Rivier J. 1984. Conotoxin GI: Disulfide bridges, synthesis, and preparation of iodinated derivatives. *Biochemistry 23*:2796–2802.

Gregory H. 1975. Isolation and structure of urogastrone and its relationship to epidermal growth factor. *Nature 257*:325–327.

Handford PA, Baron M, Mayhew M, Willis A, Beesley T, Brownlee GG, Campbell ID. 1990. The first EGF-like domain from human factor IX contains a high-affinity calcium binding site. *EMBO J 9*:475–480.

Heath WF, Merrifield RB. 1986. A synthetic approach to structure-function relationships in the murine epidermal growth factor molecule. *Proc Natl Acad Sci USA 83*:6367–6371.

Hiskey RG. 1981. Sulfhydryl group protection in peptide synthesis. In: Gross E, Meienhofer J, eds. *The peptides: Analysis, synthesis, biology, vol 3. Protection of functional groups in peptide synthesis*. New York: Academic Press. pp 137–167.

Huang LH, Cheng H, Pardi A, Tam JP, Sweeney WV. 1991. Sequence-specific ^1H NMR assignments, secondary structure, and location of the calcium binding site in the first epidermal growth factor like domain of blood coagulation factor IX. *Biochemistry 30*:7402–7409.

Huang LH, Ke XH, Sweeney W, Tam JP. 1989. Calcium binding and putative activity of the epidermal growth factor domain of blood coagulation factor IX. *Biochem Biophys Res Commun 160*:133–139.

Hunter MJ, Plesniak LA, Komives EA. 1993. Structure/function of the fifth EGF-like domain of thrombomodulin. *Protein Sci 2*(Suppl 1):128.

Kumagaye SI, Kuroda H, Nakajima K, Watanabe TX, Kimura T, Masaki T, Sakakibara S. 1988. Synthesis and disulfide structure determination of porcine endothelin: An endothelium-derived vasoconstricting peptide. *Int J Pept Protein Res 32*:519–526.

Mao B. 1989. Molecular topology of multiple-disulfide polypeptide chains. *J Am Chem Soc 111*:6132–6136.

Merrifield RB. 1963. Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc 85*:2149–2154.

Nishiuchi Y, Sakakibara S. 1982. Primary and secondary structure of conotoxin GI, a neurotoxic tridecapeptide from a marine snail. *FEBS Lett 148*:260–262.

Savage CR Jr, Inagami T, Cohen S. 1972. The primary structure of epidermal growth factor. *J Biol Chem 247*:7612–7621.

Spetzler JC, Rao C, Tam JP. 1994. A novel strategy for the synthesis of the cysteine-rich protective antigen of the malaria merozoite surface protein (MSP-1): Knowledge based strategy for disulfide formation. *Int J Pept Protein Res 43*:351–358.

Tam JP, Heath WF, Merrifield RB. 1983. S_N2 deprotection of synthetic peptides with a low concentration of HF in dimethyl sulfide: Evidence and application in peptide synthesis. *J Am Chem Soc 105*:6442–6455.

Tam JP, Liu W, Zhang JW, Galantino M, De Castiglione R. 1990. D-Amino acid and alanine scans of endothelin: An approach to study refolding intermediates. In: Giralt E, Andreu D, eds. *Peptides*. Leiden: ESCOM Science Publishers B.V. pp 160–163.

Tam JP, Wu CR, Liu W, Zhang JW. 1991. Disulfide bond formation in peptides by dimethyl sulfoxide. Scope and applications. *J Am Chem Soc 113*:6657–6662.

Van Rietschoten J, Pedroso Muller E, Granier C. 1977. Cysteine protection in solid phase synthesis of apamin. In: Goodman M, Meienhofer J, eds. *Peptides. Proceedings of the Fifth American Peptide Symposium*. New York: John Wiley and Sons. pp 522–524.

Veber DF, Milkowski JD, Varga SL, Denkewalter RG, Hirschmann R. 1972. Acetamidomethyl. A novel thiol protecting group for cysteine. *J Am Chem Soc 94*:5456–5461.

Violand BN, Tou JS, Vineyard BD, Siegel NR, Smith CE, Pyla PD, Zobel JF, Toren PC, Kolodziej EW. 1991. Determination of the disulfide bond pairings in bovine transforming growth factor-α. *Int J Pept Protein Res 37*:463–467.

Warne NW, Laskowski M Jr. 1990. All fifteen possible arrangements of three disulfide bridges in proteins are known. *Biochem Biophys Res Commun 172*:1364–1370.

Yoshitake S, Schach BG, Foster DC, Davie EW, Kurachi K. 1985. Nucleotide sequence of the gene for human factor IX (antihemophilic factor B). *Biochemistry 24*:3736–3750.

Zhang R, Snyder GH. 1989. Dependence of formation of small disulfide loops in two-cysteine peptides on the number and types of intervening amino acids. *J Biol Chem 264*:18472–18479.

Zhang R, Snyder GH. 1991. Factors governing selective formation of specific disulfides in synthetic variants of α-conotoxin. *Biochemistry 30*:11343–11348.