ELSEVIER

Journal of Chromatography B, 782 (2002) 393–404

JOURNAL OF CHROMATOGRAPHY B

www.elsevier.com/locate/chromb

Structural characterization of monomeric folding intermediates of recombinant human macrophage-colony stimulating factor β (rhM-CSFβ) by chemical trapping, chromatographic separation and mass spectrometric peptide mapping

H. Peter Happersberger[a,1], Cynthia Cowgill[b], Michael O. Glocker[a,*]

[a]*Proteome Center Rostock, Medical Faculty, University of Rostock, Joachim-Jungius Strasse 9, 18059 Rostock, Germany*
[b]*Chiron Corporation, 4560 Horton Street, Emeryville, CA 94608, USA*

Dedicated to Professor Dr. Max L. Deinzer on the occasion of his 65th birthday

Abstract

We have developed a strategy for the characterization of protein folding intermediates that combines selective modification of bis-cysteinyl thiol groups with melarsen oxide (MEL), chromatographic separation and mass spectrometric characterization of the resulting protein derivatives. In the unfolding reaction of recombinant human macrophage-colony stimulating-factor β (rhM-CSFβ) we observed monomeric M·4MEL and dimeric D·2MEL intermediates. The major locations of the MEL groups in D·2MEL were at C157 and C159. In M·4MEL, MEL groups were predominantly located at C31 and C102. These results indicate the presence of highly structured dimeric and monomeric intermediates. In the completely reduced R·4MEL derivative, MEL groups were distributed such that the smallest ring structures resulted.

Keywords: Peptide mapping; Recombinant human macrophage-colony stimulating factor β; Monomeric folding intermediates

1. Introduction

Understanding of protein folding pathways requires the determination of thermodynamic and kinetic properties of folding reactions and the characterization of intermediate structures. Thus, novel chemical and analytical methods for the structural characterization of folding intermediates are of great interest. Knowledge of folding pathways is a prerequisite for the successful manipulation of complex folding processes, e.g. for biotechnological purposes. To date mass spectrometry has found only little application for the study of protein folding reactions despite its beneficial potential to analyze complex mixtures of proteins that may be present in only femtomolar amounts [1,2]. Structural characterization of the protein derivatives is performed using mass spectrometric peptide mapping analyses [3,4].

Studying unfolding reactions is particularly useful

*Corresponding author. Tel.: +49-381-405-9770; fax: +49-381-405-9686.
E-mail address: michael.glocker@med.uni-rostock.de (M.O. Glocker).
[1]Current address: SCIL Biomedicals GmbH, Fraunhoferstr. 15, 82152 Martinsried, Germany.

1570-0232/02/$ – see front matter
PII: S1570-0232(02)00555-X

for the study of ''late'' folding steps [5], as unfolding is thought to be initiated by local events in the otherwise intact tertiary structure. Reductive unfolding, thus, should start by reduction of the highest accessible disulfide bond by which the protein structure will become more flexible and open up. Subsequent reduction of further, now accessible disulfide bonds, results in complete denaturation of the protein. Alkylation of the resultant thiol groups renders reductive unfolding irreversible. However, distinguishing unfolding events by chemical trapping caused by reductive unfolding from those that are due to denaturing unfolding (e.g. in the presence of chaotropic reagents such as guanidine hydrochloride is not possible with the hitherto applied mono-thiol derivatization and trapping experiments. Additionally, the hitherto applied monothiol derivatizing agents such as iodoacetic acid simultaneously modify protein thiols and thiols from the redox buffer system (e.g. GSH/GSSG; cystine/cysteine), hereby shifting the redox potential during trapping [6,7].

We have developed a strategy for the study of disulfide bond mediated protein (un)folding reactions that enables trapping of otherwise transient intermediates by selective modification of bis-cysteine thiol groups of proteins with melarsen oxide (MEL; *p*-(4,6-diamino-1,3,5-triazin-2-yl)aminophenylarsonous acid). Bis-thiol selective derivatization of intermediates is principally different from mono-thiol trapping strategies [8,9], as bis-thiol modification is a way to crosslink closely spaced cysteine residues pairwise [4,10]. This unique feature of MEL opens interesting possibilities for the study of unfolding reactions, as now reduction will not result in unfolding of the protein since the MEL-bridged protein is maintained as a ''compact'' structure [4]. In addition, MEL does not form stable products with monothiols but selectively traps bis-thiols of the protein folding intermediates without interfering with the redox potential [4].

In this study, denaturing unfolding reactions and structural characterization of the MEL-trapped intermediates were performed with recombinant human macrophage-colony stimulating factor β (rhM-CSFβ). M-CSF is a cytokine that stimulates the survival, proliferation, differentiation, and function of mononuclear phagocytes [11–13]. RhM-CSFβ was cloned and expressed in *E. coli* as a truncated

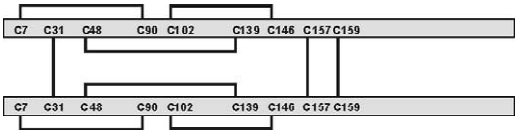

Fig. 1. Schematic representation of the rhM-CSFβ dimer. Cysteinyl residues are depicted and numbered according to their position in the amino acid sequence. Bold lines show disulfide bond connectivities in the mature molecule [15].

form (aa4–221) of the β-cDNA clone [14]. Mass spectrometric analyses have revealed that homodimeric rhM-CSFβ contains nine disulfide bonds, three of which participate in intramolecular disulfide bridges in each monomer (C7–S–S–C90, C48–S–S–C139, C102–S–S–C146) and three form intermolecular bridges (C31–S–S–C31 and pairs of C157/C159–S–S–C157/C159; Fig. 1) [15]. X-ray analysis (of rhM-CSFα, aa4–158) revealed that the dimer is formed by end-to-end disulfide linkage of two four-helical bundles (helices A, B, C and D), which run up–up–down–down [16]. The folding reaction of this protein has been well studied [17,18].

2. Experimental

2.1. Denaturing unfolding of rhM-CSFβ and trapping of intermediates

Highly purified mature rhM-CSFβ dimer (courtesy of Chiron, Emeryville, CA, USA) was stored frozen as lyophilized powder. A thawed aliquot of rhM-CSFβ was redissolved in 1 ml of 2.5 m*M* EDTA, 25 m*M* NaH_2PO_4, pH 8. Protein concentration ($C_{M\text{-}CSF}$) was 4 μg/μl. Denaturing unfolding reactions were carried out as recently described [19]. Briefly, rhM-CSFβ aliquots (200 μg), dissolved in 25 m*M* NaH_2PO_4, 2.5 m*M* EDTA, pH 8, containing different concentrations of guanidine hydrochloride (1–5 *M* GuaCl) were incubated with 44 μ*M* MEL (MEL was synthesized as described recently [4]), dissolved in the respective buffers, to yield a molar ratio of 1:11 (rhM-CSFβ:MEL) and a final volume of 1 ml. The reaction was carried out in a UV cuvette in order to monitor the time course of the modification reaction (for detailed description see [4]). After incubation for 2 min at room temperature, a solution

of 125 m*M* TCEP (Pierce, Rockford, IL, USA; pH was adjusted to 8 with solid NH_4HCO_3) was added (resulting molar ratios of rhM-CSFβ:TCEP were 1:180). After 60 min of reaction at room temperature, a solution of 125 m*M* 4-vinylpyridine (VP; Sigma–Aldrich, Steinheim, Germany) was added resulting in molar ratios of 1:400 (rhM-CSFβ:VP) and the reaction mixtures were incubated for 30 min at room temperature. After acidification by addition of 1.5 ml of a solution consisting of CH_3CN–4.2% TFA (2:1, v/v) to result a final CH_3CN concentration of 40% (v/v), excess of reagents was removed by gel-permeation chromatography (PD-10 columns, Pharamacia and Upjohn, Freiburg, Germany). Samples were eluted with 0.1% TFA dissolved in CH_3CN–H_2O (4:6, v/v), pH 2. Protein containing fractions were collected (in a single fraction), either lyophilized and analyzed by mass spectrometry or immediately subjected to HPLC analysis.

2.2. SDS–PAGE analyses of folding intermediates

SDS–polyacrylamide gels [20] were used directly after polymerization. Protein samples (5 μg) were dissolved in 20-μl sample buffer consisting of 4% SDS, 12% glycerol, 50 m*M* Tris-(hydroxymethyl)-aminomethane (pH 6.8) and 0.02% Coomassie Brilliant Blue G-250 and deposited in one gel-slot. After electrophoresis the gels were fixed in methanol–AcOH–water (5:1:4, v/v) for 0.5–1 h, stained with 0.05% Coomassie Brilliant Blue G-250 in 10% AcOH and destained in 10% AcOH.

2.3. Removal of the MEL modification and carboxamidomethylation of the monomeric *R/M·4MEL/1VP* derivatives of rhM-CSFβ

Solutions of MEL- and VP-modified monomeric rhM-CSFβ derivatives were prepared by dissolving lyophylized HPLC/size-exclusion chromatography (SEC) fractions (100 μg) in 50 m*M* NaH_2PO_4, 2.5 m*M* EDTA, pH 8, containing 6 *M* GuaCl to result a final protein concentration of approximately 0.5–1.0 μg/μl. The MEL-groups were removed by adding 1,4-dithiothreitol (DTT, 100 m*M*; Sigma–Aldrich) in a molar ratio MEL to DTT of 1:100. After incubation at room temperature for 10 min a solution of iodoacetamide (IAA, 125 m*M*; Fluka, Buchs, Switzerland) was added resulting a molar ratio of DTT to IAA of 1:2.5 and rhM-CSFβ–cysteine thiols to IAA of 1:25. After incubation at room temperature in the dark for 30 min the alkylation was stopped by acidification with a mixture of acetonitrile, buffer without GuaCl and glacial acetic acid as described above. These reaction mixtures were subjected to HPLC analyses.

2.4. Proteolytic digestions of *R/M-8CAM/1VP* derivatives of rhM-CSFβ

The lyophilized fractions, containing ~10 μg of M-8CAM-1VP rhM-CSFβ derivatives were completely dissolved in 10 μl of a solvent mixture consisting of 4 volumes of 0.1% TFA in acetonitrile and 6 volumes of water (pH 2). Aliquots (3 μl) were lyophilized again and dissolved in 5 μl of freshly prepared 50 m*M* NH_4HCO_3 buffer, pH 7.8. Proteolytic digestions were carried out by addition of LysC (0.75 μl, 0.1 μg/μl) and AspN (0.75 μl, 0.04 μg/μl; Roche Diagnostics, Switzerland), respectively, resulting in enzyme to substrate ratios (E:S) of 1:40 and 1:100. Digestions were performed for 4 h at 37 °C. The reaction mixtures were lyophilized before MALDI-MS analyses.

2.5. HPLC separation and purification

All RP-HPLC runs were performed using a Waters Millipore HPLC system, consisting of two HPLC pumps (Waters M510 and Waters M45) using solvent A, 0.1% (v/v) TFA in water, and solvent B, 0.1% (v/v) TFA in CH_3CN. For the purification of the reaction mixtures containing the MEL- and VP-modified monomeric and dimeric rhM-CSFβ derivatives, a semipreparative 250×8 mm Chromo-Sil reversed-phase C_4 column (300 Å, 10 μm) equipped with a Chromo-Sil precolumn was used. The flow-rate was set to 2.4 ml/min. The solvent mixture was kept constant at 40% B for 15 min and was then raised to 90% B over a time period of 25 min. The purification of the CAM- and VP-modified monomeric and dimeric rhM-CSFβ derivatives after SEC were carried out in the same manner on an analytical 250×4.6 mm Vydac reversed-phase C_4 column (300 Å, 10 μm) equipped with a Vydac precolumn. The

same column was used for the purification of the M-8CAM-1VP derivatives applying the following solvent gradient: the solvent mixture was kept constant at 40% B for 5 min and was then raised to 90% B over a time period of 25 min. In all cases, protein-containing fractions were collected. Cross-contamination of fractions collected from closely neighboured peaks is inevitable, particularly, when small signals or shoulders are in close proximity to large peptide signals. In our case, cross-contamination was determined by repetitive LC runs and subsequent mass spectrometric analyses and, thus, was minimized. For MALDI-MS analyses aliquots (1 μl) of collected fractions were used without further work-up. For nano-ESI-MS analyses HPLC-fractions were lyophilized and redissolved in 10% (v/v) acetic acid–methanol (9:1, v/v), pH 2, resulting in a final protein concentration of approximately 0.1 μg/μl. For peptide mapping analyses lyophilized fractions of the M-8CAM-1VP derivatives were used.

2.6. Size-exclusion chromatography (SEC)

A binary solvent delivery system 2700 (Bio-Rad, Germany) equipped with a UV–Vis wavelength detector (UV–Vis monitor 1706) that was set at 280 nm and a HPLC interface was used. The SEC was performed using a 300×8 mm YMC-Pack Diol-120 column (S-5 μm, 120 Å) equipped with a YMC-Guardpack-Diol precolumn. Elution was performed applying a solvent mixture consisting of 6 volumes of aqueous 200 m*M* NaCl in <0.05% (v/v) TFA (pH 3) and 4 volumes of acetonitrile. The flow-rate was adjusted to 0.5 ml/min. Under these conditions the molecular-mass cut-off of the system was determined to ~50 000. A molecular excess volume of approximately 7 ml (t_R=14 min) was determined using the proteins: β-amylase (M_r 200 000), BSA (M_r 66 000), native MCSF-dimer (M_r 49 000) and ovalbumin (M_r 47 000). Migration retardation was observed for the following proteins: carbonic anhydrase (M_r 29 000; t_R=14.9 min), myoglobin (M_r 17 000; t_R=15.8 min), cytochrome c (M_r 12 300; t_R=17.3 min), insulin (M_r 5700; t_R=21.6 min) and a mixture of GuaCl and EDTA (M_r <1000, each; t_R=27.5 min). After calibration of the SEC system, the HPLC fractions containing MEL- and VP-modified monomeric and dimeric rhM-CSFβ derivatives (2 R/2 M/D) were analyzed. Two fractions were collected at retention times t_R=13.6 min (2 M/D) and t_R=15.4 min (2 R). These fractions were used for the removal of the MEL modification (see above).

2.7. Mass spectrometry

Nanospray-electrospray ionization mass spectrometry (nano-ESI-MS) was performed using a Vestec-201A quadrupole mass spectrometer (Vestec, Houston, TX, USA). The ion-spray interface temperature was approximately 40–50 °C for all measurements. The mass analyzer with a nominal m/z range of 2000 was operated at 1 unit resolution. A self-developed nano-ESI source [21] was used for nano-ESI-MS analyses. Borosilicate glass capillaries (GC120F-10, Clark Electromedical Instruments, Pangbourne, UK) were used for manufacturing the microcapillaries using a capillary puller (P-97, Sutter Instruments, Novato, CA, USA). The capillaries were coated with a thin layer of gold by a sputter process (International Scientific Instrument). Capillaries were filled by dipping the capillary tip into the sample solution (0.5–2.5 μl). An electrospray voltage at the capillary tip of 1.2–1.5 kV and a repeller voltage of 3–40 V were employed, respectively. Spectra were recorded with a scan rate of 7 s/scan with a mass window of m/z 200–2000. Mass calibration was performed with the 8^+ to 12^+ charged ions of hen egg white lysozyme (M_r 14 305.1) and raw data were analyzed using a Vector-2 data system (Teknivent, Houston, USA). Samples were diluted to a final concentration of 0.1 μg/μl with 10% (v/v) acetic acid–methanol (9:1, v/v), pH 2.

MALDI-MS peptide mapping analyses were carried out using a Bruker Biflex linear time-of-flight mass spectrometer (Bruker Daltonik, Bremen, Germany), equipped with a UV–nitrogen laser (337 nm) and a dual microchannel plate detector. The acceleration voltage was set to 20 kV and spectra were calibrated with insulin. Aliquots of 1.0 μl of the peptide mixtures were mixed with 0.7 μl of matrix solution (10 μg/μl; 4-hydroxy-α-cyanocinnamic acid) dissolved in CH_3CN–0.1% TFA (2:1, v/v) pH 2, directly on the target. Spectra were recorded after

evaporation of the solvent and processed using the X-MASS data system.

3. Results

3.1. Denaturing unfolding of rhM-CSFβ; mixture analysis

Denaturing unfolding of rhM-CSFβ was performed by preincubation of the protein in GuaCl buffer with increasing concentrations. Reduction of the accessible disulfide bonds in the (partially) unfolded intermediates was performed in the presence of MEL, thus, chemically trapping bis-thiol groups pairwise. In order to minimize potential disulfide bond scrambling and exchange of arsonous acid-bridged structures, the trapped intermediates were treated with 4-vinylpyridine (VP). As monomeric rhM-CSFβ contains nine cysteinyl residues (cf. Fig. 1), this step was necessary to irreversibly trap the remaining cysteinyl thiol group that should be present in the monomeric intermediates.

SDS–PAGE analysis (Fig. 2A) showed that the content of monomeric rhM-CSFβ (band migrating at 25 000) increased with increasing GuaCl concentration and the content of dimeric rhM-CSFβ (band migrating at 50 000) disappeared completely when the experiment was performed in a buffer containing 4 *M* GuaCl and higher concentrations thereof. This result shows, that the number of accessible disulfide bonds in rhM-CSFβ increases with increasing concentration of the chaotrop. For the structural characterization of folding intermediates, the protein sample that was incubated in 2 *M* GuaCl was chosen, as in this sample the band intensities of monomeric and dimeric intermediates were approximately equal, suggesting that the amounts of monomeric and dimeric unfolding intermediates were sufficient for further analyses. The nano-ESI mass spectrum of this sample (Fig. 2B) showed the presence of a mixture of rhM-CSFβ. Derivatives were assigned as R/M·4MEL/1VP ($M_{r(exp.)}$ 25 735±11; Table 1) for monomeric folding intermediate (M) and/or completely unfolded monomer (R) with four MEL groups and one VP-group, and as D·2MEL ($M_{r(exp.)}$ 49 582±8; Table 1) for dimeric intermediates with two attached MEL groups, respectively. The distinction between

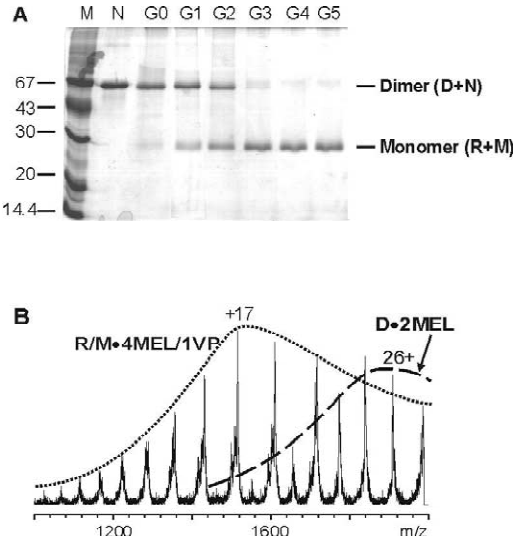

Fig. 2. Analysis of the protein mixture obtained by denaturing unfolding of rhM-CSFβ by increasing GuaCl concentrations in the presence of MEL. (A) SDS–PAGE analysis. Lanes: M = molecular mass markers; N = native rhM-CSFβ dimer; G0–G5 = protein bands observed with the corresponding molar GuaCl concentrations. The positions of monomeric (R and M) and of dimeric (D and N) protein species are marked at the right. Relative molecular masses of the marker proteins (kDa) are given at the left margin. (B) Nano-ESI mass spectrum of the unfolding mixture obtained by incubation of rhM-CSFβ in 2 *M* GuaCl—cf. lane G2 in (A). R/M·4MEL/1VP: monomeric folding intermediate (M) and/or completely unfolded monomer (R) with four MEL groups and one VP-group; D·2MEL: dimeric intermediate with two MEL groups. Charge numbers denote $[M+nH]^{n+}$ ions. The charge state envelopes of the intermediates are indicated by the dashed (dimer) and dotted (monomer) lines around the multiply charged ion series, respectively.

completely reduced monomer (R) and partly unfolded monomer (M) and dimer (D) is derived from the fact, that at 2 *M*/3 *M* guanidinium hydrochloride concentrations a distinct difference in absorption was observed by tryptophan fluorescence studies when compared to completely reduced and native rhM-CSF [19]. The absorption difference is in agreement with the presence of a mixture of completely reduced (R) and partially unfolded monomeric (M) and dimeric (D) protein species. Traces of rhM-CSFβ derivatives termed R/M·3MEL/1VP ($M_{r(exp.)}$ 25 454±6; Table 1) were also present in the ESI-spectra, indicating that MEL was partly removable during sample work-up. Thus, we decided to completely remove MEL from the intermediates after

Table 1
Molecular masses of MEL-trapped rhM-CSFβ folding intermediates

Folding intermediate[a]	M_r[b]	$M_{w\,obsd.}$[b]
R/M·3MEL/1VP	25 449.1	25 454±6
R/M·4MEL/1VP	25 725.2	25 735±11
R/M-8CAM/1VP	25 085.3	25 095±6
D·2MEL	49 581.6[c]	49 582±8

[a] R/M·*n*MEL/1VP, MEL-trapped monomeric intermediates with *n* attached MEL-groups and one VP-modification; R/M-8CAM/1VP, MEL-trapped monomeric intermediate in which the MEL groups were replaced by carboxamidomethyl (CAM) groups.
[b] Average masses.
[c] Calculated values for MEL-modified dimeric derivatives in which unmodified cysteine residues are regarded as being involved in disulfide bonds.

having performed the unfolding and trapping experiment and prior to performing mass spectrometric peptide mapping analyses. MEL groups were removed by incubation with dithiothreitol (DTT) and free thiol groups were carboxamidomethylated (CAM) with iodoacetamide (IAA) which irreversibly blocks free thiols. The MEL vs. CAM exchange reaction went to completion as judged from the molecular mass determinations; the molecular mass of the R/M-8CAM/1VP derivative(s) ($M_{r(exp.)}$ 25 095±6; Table 1) matched that of the calculated mass. The D·2MEL derivative served as a control for both, the double-labeling and the MEL-exchange experiments. When D·2MEL was treated with VP, no uptake of VP was observed showing that no free thiol groups were available (data not shown). Also, exchange of MEL vs. CAM was possible without interfering with the disulfide bonds in the intermediates [19].

3.2. Analysis of monomeric unfolding intermediates of rhM-CSFβ

The MEL-trapped and VP-blocked rhM-CSFβ intermediates were subjected to high-performance liquid chromatography (HPLC) in order to desalt the samples prior to mass spectrometry. To our surprise, a partial fractionation was observed. The chromatogram (Fig. 3A) shows a strong signal (fraction 2) with a preceeding shoulder (fraction 1). The dominant fraction contained a mixture of monomeric and

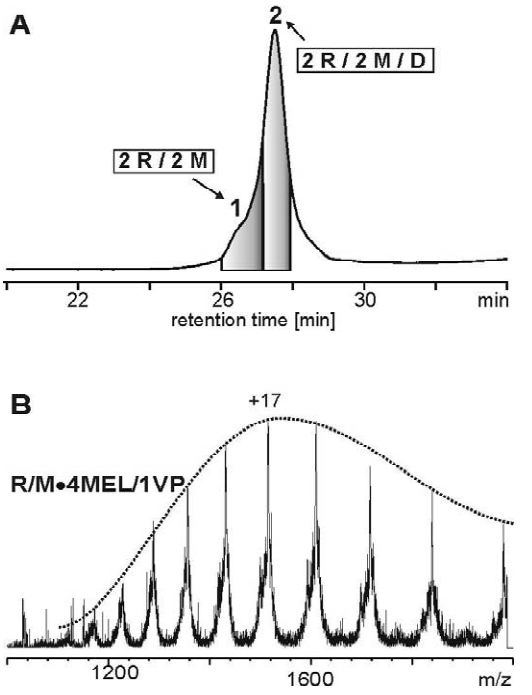

Fig. 3. Analysis of the protein mixture obtained by incubation of rhM-CSFβ in 2 *M* GuaCl. (A) HPLC separation. The chromatogram shows a strong signal at retention time $t_R = 27.5$ min (fraction 2) with a preceeding shoulder at $t_R = 26.5$ min (fraction 1). (B) Nano-ESI mass spectrum of fraction 1. Intense ion signals for the monomeric R/M·4MEL/1VP derivatives are observed. Charge numbers denote $[M+H]^{n+}$ ions. The homogeneic charge state distribution is indicated by the dotted line around the multiply charged ion series.

dimeric protein species (data not shown) whereas the latter contained only monomeric protein species (R/M·4MEL/1VP) as seen by nano-ESI mass spectrometric analyses (Fig. 3B). Intense ion signals for the monomeric derivatives are observed that produced a spectrum with a series of homogeneously distributed multiply charged ion signals with a maximum around the 17+ charged molecular ion signal (Fig. 2B). This fraction was used for detailed mass spectrometric peptide mapping analyses after the MEL groups were replaced by CAM labels. The resulting R/M-8CAM/1VP derivatives were proteolytically cleaved using AspN-protease.

The MALDI spectrum (Fig. 4A) showed abundant signals for the resulting peptides and the cysteinyl residue-containing peptides (Table 2) were in most cases present as doublets (CAM- and VP-modified,

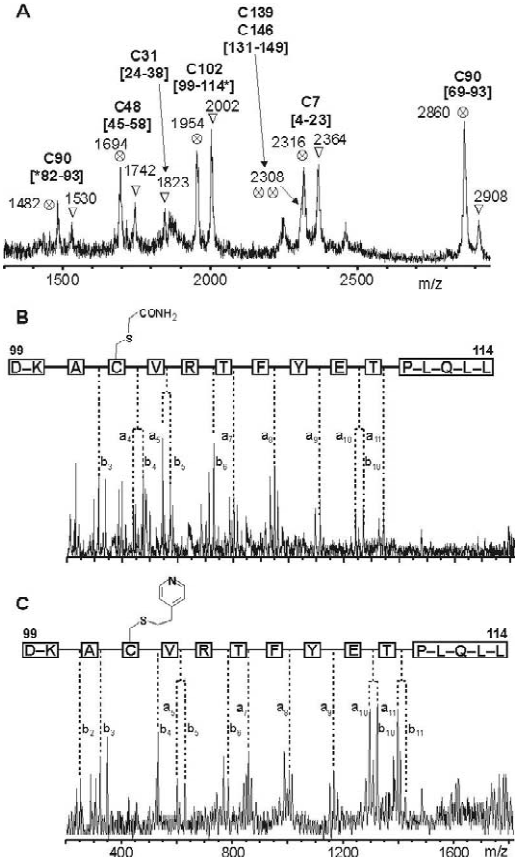

Fig. 4. Mass spectrometric characterization of monomeric unfolding intermediates. (A) MALDI-MS peptide mapping analysis of the AspN digested R/M-8CAM/1VP derivatives of rhM-CSFβ obtained from fraction 1. Only cysteine containing peptides are labeled. Ion signals marked with ⊗ correspond to cysteinyl-CAM-modified peptides, whereas ion signals marked with ∇ indicate cysteinyl-VP-modified peptides. (B) MALDI-postsource decay (PSD) analysis of peptide [99–114*] in which C102 was CAM-modified. (C) MALDI-PSD mass spectrum of peptide [99–114*] in which C102 was VP-modified. The mass spectrometric fragment ions from the a-type and b-type series are indicated [24].

respectively). This showed that a mixture of monomeric intermediates was present in the sample.

With the exception of the peptide that contained C7, the peptides that contained the cysteinyl residues that belong to the four-helix bundle (C7–S–S–C90 and C48–S–S–C139) gave ion signal doublets in which the CAM-labeled signals were predominant (Fig. 4A). This indicates that these residues were mostly MEL-bridged during the trapping experiments. By contrast, the ion signal for peptide [24–38] (m/z 1823) that contained C31 was found VP-modified only. This is consistent with a residue that was not MEL-bridged while performing the trapping experiments. Residue C31 forms an intermolecular disulfide bond to another C31 residue in the homodimer (cf. Fig. 1). Opening this disulfide bond yields in a single thiol with no ''next neighbour'' and, hence, does not form stable products with MEL.

An interesting doublet of signals for which the CAM-modified signal was observed as the dominant signal derived from peptide [99–114*] (m/z, 1954 and 2002) that harbors C102. Interestingly, this peptide is produced by a nonspecific cleavage of AspN. Assignments of these two ion signals were therefore assured by sequence determinations of the respective peptides using postsource decay (PSD)-MALDI-MS experiments (Fig. 4B and C). The PSD-MALDI analysis of peptide [99–114*] (m/z 1954) not only confirmed the assignments of the ion signals but also proved the CAM modification at residue C102 in the peptide (Fig. 4B). Similarly, modification of C102 (PSD of m/z 2002) with VP was proven by the resulting a-type and b-type fragment ion series (Fig. 4C).

3.3. Analysis of completely reduced monomeric protein species of rhM-CSFβ

In order to distinguish the MEL-modified cysteinyl residues derived from the completely reduced monomers (R) from those that were derived from monomeric intermediates (M), we performed SEC in an attempt to further separate the different protein species present in fraction 2 (cf. Fig. 3A). Note that a two step separation procedure (1st low-resolution SEC and 2nd high-resolution HPLC) should be sufficient to separate the studied folding intermediates.

The SEC chromatogram (Fig. 5A) yielded two main fractions. Nano-ESI-MS showed that fraction 3 consisted of a mixture of monomeric and dimeric protein species (data not shown). From the migration behavior it might be speculated that the monomers (M) in this fraction actually were part of noncovalent homodimers (D). By contrast, the dominant fraction (fraction 4) showed series of ion signals for monomeric protein species only (Fig. 5B). Due to the

Table 2
Mass assignments of proteolytically derived peptides after AspN digestion of the folding intermediates M-8CAM/1VP

Cys-residue	Amino acid sequence positions/ cysteine modification	$[M+H]^+_{calcd}$ [a]	$[M+H]^+_{obsd}$ [a]
C7	[4–23]/1×CAM	2317.6	2316
C7	[4–23]/1×VP	2365.7	2364
C48	[45–58]/1×CAM	1695.1	1694
C48	[45–58]/1×VP	1742.2	1742
C90	[*82–93]/1×CAM[b]	1482.8	1482
C90	[*82–93]/1×VP[b]	1530.9	1530
C90	[69–93]/1×CAM	2861.3	2860
C90	[69–93]/1×VP	2908.4	2908
C102	[99–114*]/1×CAM[b]	1956.3	1954
C102	[99–114*]/1×VP[b]	2003.4	2002
C139/C146	[131–149]/2×CAM	2309.5	2308
C139/C146	[131–149]/1×CAM/1×VP	2357.6	–[c]
C139/C146	[131–149]/2×VP	2405.6	–[c]

[a] Average masses.
[b] Asp-N unspecific cleavages are marked with *.
[c] Not observed.

clearly distinct retardation behavior, these were termed R·4MEL/1VP derivatives of rhM-CSFβ. Interestingly, the charge structure around the multiply charged ion series shows two maxima, one at the 17^+ ion signal and the second one at the 22^+ ion signal, indicating the presence of at least two protein species in the sample. Again, small ion signals for traces of a protein species (R·3MEL/1VP) in which MEL was partly dissociated under the ESI conditions were observed. The existence of two maxima in the ESI-spectra provoked to perform additional HPLC experiments with fraction 4 (cf. Fig. 5A). The resulting HPLC chromatogram (Fig. 6A) showed two peaks (fractions 5 and 6). MALDI-MS peptide mapping analysis of the AspN digested rhM-CSFβ protein species R-8CAM/1VP from fraction 5 (Fig. 6B) showed similar spectra than those obtained with the previously described derivatives (cf. Fig. 4A). However, the relative ion signal intensities were different. In particular, a strong ion signal was observed for the C48-VP-modified peptide at m/z 1742 in the presence of a small signal for the C48-CAM-modified peptide (m/z 1694). Further interesting signal intensity distributions were observed for peptide [4–23]. The C7-CAM-modified derivative at m/z 2316 was found in higher abundance than the corresponding C7-VP-modified peptide (m/z 2364). Thus, the corresponding monomeric protein derivative was termed R1. By contrast, the corresponding peptides derived from the R-8CAM/1VP derivative of rhM-CSFβ that eluted in fraction 6 showed the opposite relative intensities in the MALDI spectra (Fig. 6C). Thus, this monomeric derivative of rhM-CSFβ is termed R2. Both R·4MEL derivatives have in common, that MEL groups were found to be distributed such that the smallest ring structures resulted by arsonous acid bridging of cysteinyl residues. The presence of two distinguishable completely reduced monomeric derivatives is consistent with the presence of ''random coil'' structures with no particular impetus on the modification pattern, as no structural predominance was present.

3.4. Synopsis

In the collected samples a large population of the (un)folding intermediates seemed to be present producing a complex mixture that consisted of (i) dimeric (D), (ii) monomeric (M) (un)folding intermediates, and (iii) of completely reduced monomeric protein species (R). Thus, in the peptide mapping experiments of these samples, VP- and CAM-cysteinyl modified peptides were present from all of them. Hence, the comparison of the relative intensities of the VP-modifications of the cysteinyl residues after separation of these species (Table 3), yields in a more detailed picture. Accordingly,

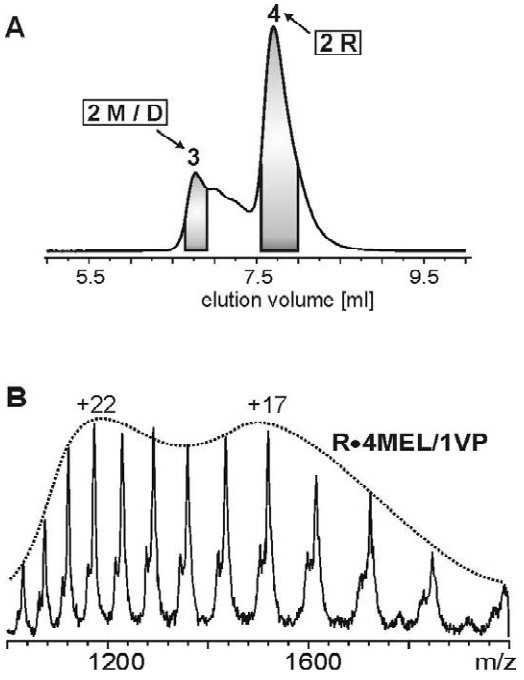

Fig. 5. Analysis of the protein mixture in fraction 2 (cf. Fig. 3A) by size-exclusion chromatography (SEC) and MS-characterization. (A) SEC chromatogram. Two fractions, marked 3 ($V_{el.}=6.8$ ml) and 4 ($V_{el.}=7.7$ ml), were collected. (B) Nano-ESI mass spectrum of fraction 4. R·4MEL/1VP: completely reduced monomer (R) with four MEL groups and one VP-group. Charge numbers denote $[M+nH]^{n+}$ ions. The charge structure around the multiply charged ion series (dotted line) shows two maxima indicating the presence of at least two protein species.

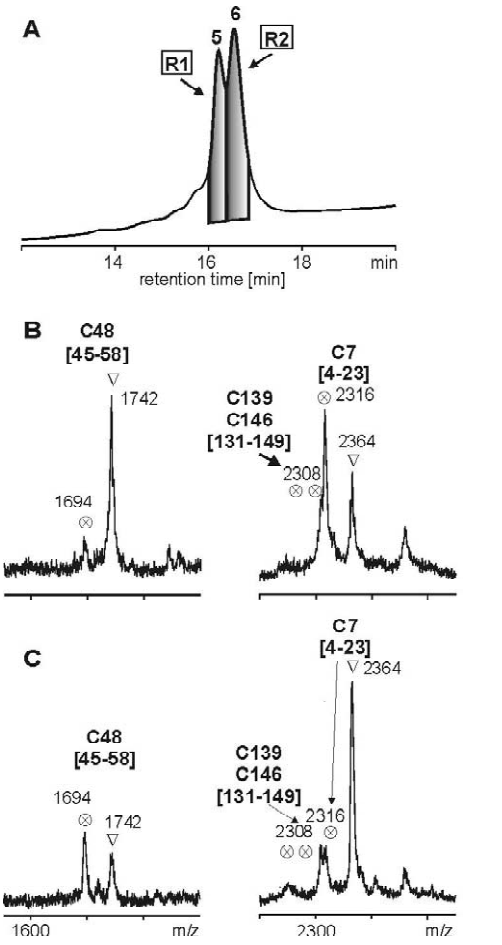

Fig. 6. Characterization of the protein mixture in fraction 4 (cf. Fig. 5A). (A) HPLC separation. The chromatogram shows two signals at $t_R=16.2$ min (fraction 5) and $t_R=16.6$ min (fraction 6). (B) MALDI-MS peptide mapping analysis of the AspN digested rhM-CSFβ protein species R-8CAM/1VP obtained from fraction 5. (C) MALDI-MS peptide mapping analysis of the AspN digested rhM-CSFβ protein species R-8CAM/1VP obtained from fraction 6. Only cysteine containing peptides are labeled. Ion signals marked with ⊗ correspond to cysteinyl-CAM-modified peptides, whereas ion signals marked with ∇ indicate cysteinyl-VP-modified peptides.

subtraction of the VP-modifications that were found to be specific for R1 and R2 leaves predominantly VP-alkylated cysteinyl residues at C31 and C102 of the intermediate (M). Thus, it can be concluded, that similar to the case for C31–S–S–C31, reduction of the disulfide bond C102–S–S–C146 seemed to result in intermediate structures in which the two cysteinyl residues were separated in space and, hence, were not bridged by MEL.

4. Discussion

The trapping strategy and analytical concept applied here enables the characterization of disulfide-bond mediated protein (un)folding pathways by (i) monitoring the average degree of modification reflecting the general unfolding behavior of rhM-CSFβ, and (ii) by pinpointing the modification sites in folding intermediates. Chemical trapping of bis-thiol groups with MEL enables monitoring of conformational changes occurring in different areas of the protein simultaneously on the submolecular level.

From the data described here the following (un)-

Table 3
Relative degree[a] of VP-modification in rhM-CSFβ folding intermediates

Cysteine residue	D[b]	2 M/2 R	R 1	R 2
C7	+	+	+	+++
C31	0	+++	0	0
C48	0	+	+++	+
C90	+	+	0	0
C102	0	+++	0	0
C139	0	N.o.	0	0
C146	0	N.o.	0	0
C157	+++	0	0	0
C159	+++	0	0	0

[a] Relative intensities were observed by MALDI-MS after proteolytic digestion of chemically trapped dimeric and monomeric folding intermediates. Relative signal intensities: 0, no VP-modification observed; +, VP<CAM; ++, VP=CAM; +++, VP>CAM; N.o., no signal observed.
[b] See Ref. [19].

folding intermediates can be deduced. Reduction of the disulfide bonds in the native dimeric protein results in the opening of the intermolecular disulfide bonds in which C159 and C157 are involved, leading to dimeric intermediates (D) [19]. Monomeric intermediates (M) appear after reduction of the intermolecular disulfide bond at C31 and in which cysteinyl residues that are not part of the four-helix bundle (formerly disulfide bond C102–S–S–C146) of rhM-CSFβ are reduced. The latter event opens the monomeric structure such that most of the two resulting thiol groups are placed in large distance to each other (distance must be greater than 3.62 Å, [22]) so that MEL-bridges are not obtained in great abundance. The situation is apparently different for the disulfide bonds (C7–S–S–C90 and C48–S–S–C139) that belong to the four-helix bundle. In this part of the molecule secondary structure elements most likely help to adopt the appropriate conformation in which local specific interactions [23] are keeping the corresponding two neighboring cystinyl-groups in close distance, even in 2 *M* GuaCl buffer. Reduction of these disulfide bonds, hence, produces preferentially closely spaced dithiols that can be bridged by MEL, thereby maintaining a compact structure. This information would have been completely lost when mono-thiol derivatization was performed in the denaturing unfolding experiments.

Evaluation of the X-ray structure of natively

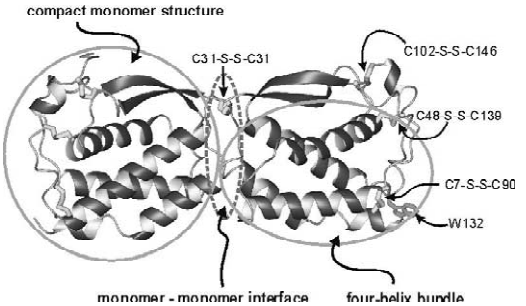

Fig. 7. Ribbon diagram of rhM-CSF. RhM-CSFα is represented including the structurally resolved amino acid residues (4–151) [16]. The monomers (circled) contain the four-helix bundle structure (open oval) with two disulfide bonds (C7–S–S–C90 and C48–S–S–C139) and a flanking β-sheet structure in which disulfide bond C102–S–S–C146 is located. The monomer–monomer interface of the covalent dimer that in the case of rhM-CSFβ contains the disulfide bonds C31–S–S–C31 and C157/159–S–S–C157/159 is indicated (dashed oval). Cysteinyl residues are shown as wireframe models, marking the disulfide bonds. The location of a tryptophan residue (W132) is shown in one monomer using the X-ray co-ordinates of the M-CSFα homodimeric protein.

folded rhM-CSF α (Fig. 7) shows, that three groups of disulfide bonds can be distinguished. The ribbon diagram represents the structure elements of the N-terminal amino acid residues (4–151) [16] of rhM-CSFβ. The monomers (circled) contain the four-helix bundle structure (open oval) with two disulfide bonds (C7–S–S–C90 and C48–S–S–C139) and a flanking β-sheet structure in which disulfide bond C102–S–S–C146 is located. The monomer– monomer interface (dashed oval) of the covalent dimer that in the case of rhM-CSFβ contains the disulfide bonds C31–S–S–C31 and C157/159–S–S–C157/159 is indicated.

It is interesting to note that the differences of the distinct cysteinyl residues in the behavior towards chemical trapping parallels differences as disseminated from the structural data. The function of the disulfide bonds keeping the homodimer together seem obvious. Similarly, the function of the four-helix bundle stabilizing disulfide bonds mainly might serve structure stabilizing purposes. However, one might speculate about the function of the disulfide bond in the flanking region spanning from C102 to C146. One explanation for the presence of this disulfide bond might be that by formation of this

disulfide bond M-CSF receives its remarkable protease resistance [15]. A high resistance towards proteolysis is a prerequisite for a secreted molecule as is the case for M-CSF as a cytokine. Note, that this particular disulfide bond is formed late during folding [17], potentially preventing the formation of protease resistant folding intermediates with incomplete folding and, thus, presumably lacking biological activity.

If this was in fact the major role of this disulfide bond in vivo, this knowledge has great value concerning M-CSF as a biotechnological product. Impacts encompass optimizing production processes in which protein refolding steps are included to potentially tunable pharmacokinetics of designed M-CSF derivatives when administered as a drug.

5. Nomenclature

aa	amino acid
AcOH	acetic acid
CAM	carboxamidomethyl
DTT	1,4-dithiothreitol
D·*n*MEL	*n*-fold MEL-modified rhM-CSFβ dimer
R/M·*n*MEL/*m*VP	*n*-fold MEL- and *m*-fold VP-modified rhM-CSFβ monomer
R/M·*n*CAM/*m*VP	*n*-fold CAM- and *m*-fold VP-modified rhM-CSFβ monomer
GuaCl	guanidine hydrochloride
E:S	enzyme to substrate ratio
ESI-MS	electrospray ionization-mass spectrometry
HCCA	4-hydroxy-α-cyanocinnamic acid
IAA	iodoacetamide
MALDI-MS	matrix-assisted laser desorption/ionization–mass spectrometry
PAGE	polyacrylamide gel electrophoresis
PSD	postsource decay
rhM-CSFβ	recombinant human macrophage-colony stimulating factor β
MEL	*p*-(4,6-diamino-1,3,5-triazine-2-yl)aminophenylarsonous acid (melarsen oxide)
SDS	sodium dodecylsulfate
TCEP	tris-(2-carboxyethyl)phosphine hydrochloride
TFA	2,2,2-trifluoroacetic acid
VP	4-vinylpyridine

Acknowledgements

The authors would like to thank Dr. Detlev Suckau, Bruker-Daltonics, Bremen, and Dr. Martin Hornshaw, Applied Biosystems for performing the MALDI-PSD analyses. We are grateful to Dr. Michael Przybylski in whose laboratories part of this work was carried out. We thank Dr. S.-H. Kim and Dr. K. Koths for providing the X-ray coordinates of rhM-CSFα. This work was supported by grants from the Deutsche Forschungsgemeinschaft (DFG) and by the Fonds der Chemischen Industrie.

References

[1] M. Przybylski, M. Glocker, Angew. Chem. Intl. Ed. Engl. 35/8 (1996) 806.
[2] P. Chaurand, F. Luetzenkirchen, B. Spengler, J. Am. Soc. Mass Spectrom. 10 (1999) 91.
[3] H.P. Happersberger, M.O. Glocker, Eur. Mass Spectrom. 4 (1998) 209.
[4] H.P. Happersberger, M. Przybylski, M.O. Glocker, Anal. Biochem. 264 (1998) 237.
[5] Y.J. Li, D.M. Rothwarf, H.A. Scheraga, Nature Struct. Biol. 2 (1995) 489.
[6] T.E. Creighton, Biol. Chem. 378 (1997) 731.
[7] M. Hoffman, Science 235 (1991) 1357.
[8] T.E. Creighton, Science 256 (1992) 111.
[9] J.S. Weissman, P.S. Kim, Science 256 (1992) 112.
[10] H.P. Happersberger, J. Stapleton, C. Cowgill, M.O. Glocker, in: Proceedings of the 45th Conference on Mass Spectrometry an Allied Topics, 1997, p. 22.
[11] S.L. Aukerman, S. Middleton, A. Sampson-Johannes, P. Ralph, in: M. Symann, P. Quesenberry, G. Morstyn (Eds.), Hematopoietic Growth Factors and Interleukins: Biology and Clinical Applications, Gardiner-Caldwell, Macclesfield, 1992, p. 79.
[12] S.C. Clark, R. Kamen, Science 236 (1987) 1229.
[13] E.R. Stanley, in: N.A. Nicola (Ed.), Guidebook To Cytokines and Their Receptors, 1994, p. 164.
[14] R. Halenbeck, E. Kawasaki, J. Wrin, K. Koths, Bio/Technology 7 (1989) 710.

[15] M.O. Glocker, B. Arbogast, J. Schreurs, M.L. Deinzer, Biochemistry 32 (1993) 482.
[16] J. Pandit, A. Bohm, J. Jancarik, R. Halenbeck, K. Koths, S.-H. Kim, Science 258 (1992) 1358.
[17] M.O. Glocker, B. Arbogast, R. Milley, C. Cowgill, M.L. Deinzer, Proc. Natl. Acad. Sci. USA 91 (1994) 5868.
[18] J.A. Wilkins, J. Cone, Z.I. Randhawa, D. Wood, M.K. Warren, H.E. Witkowska, Protein Sci. 2 (1993) 244.
[19] H.P. Happersberger, J. Stapleton, C. Cowgill, M.O. Glocker, Proteins Suppl. 2 (1998) 50.
[20] H. Schägger, G.v. Jagow, Anal. Biochem. 173 (1988) 201.
[21] T.A. Fligge, M. Przybylski, J.P. Quinn, A.G. Marshall, Eur. Mass Spectrom. 4 (1998) 401.
[22] M. Jauhiainen, K.J. Stevenson, P.J. Dolphin, J. Biol. Chem. 263 (1988) 6525.
[23] K.A. Dill, Biochemistry 29 (1990) 7133.
[24] P. Roepstorff, J. Fohlman, Biomed. Mass Spectrom. 11 (1984) 601.