

Accurate Disulfide Formation in *Escherichia coli*: Overexpression and Characterization of the First Domain (HF6478) of the Multiple Kazal-Type Inhibitor LEKTI

Thomas Lauber,* Ute C. Marx,*¹ Axel Schulz,† Peter Kreutzmann,† Paul Rösch,* and Silke Hoffmann*

*Lehrstuhl für Biopolymere, Universität Bayreuth, Universitätsstrasse 30, 95447 Bayreuth, Germany; and †IPF PharmaCeuticals GmbH, Feodor-Lynen-Strasse 31, 30625 Hannover, Germany

Received December 13, 2000, and in revised form February 6, 2001; published online May 10, 2001

The human hemofiltrate peptide HF6478, a putative serine proteinase inhibitor, which is part of the precursor protein LEKTI, was cloned, overexpressed, and purified. HF6478 contains two disulfide bridges with 1-4, 2-3 connectivity, sharing partial homology to Kazal-type domains and other serine proteinase inhibitors. It was expressed as thioredoxin (Trx) fusion protein, and disulfide formation occurred in the oxidative cytoplasm of *Escherichia coli* Origami (DE3) strain which carries a *trxB*⁻/*gor522*⁻ double mutation. The soluble fusion protein was purified using metal-chelating affinity chromatography. Cleavage of the Trx fusion protein with factor Xa and subsequent purification yielded the final product in amounts sufficient for structural studies. Characterization of recombinant HF6478 was done by amino acid sequencing, mass spectrometry, capillary zone electrophoresis, and CD spectroscopy. Taking the blood filtrate peptide HF6478 as example, we present a strategy which should facilitate the expression of different extracellular proteins in the *E. coli* cytoplasm. © 2001 Academic Press

Proteinase inhibitors are essential tools for *in vivo* regulation of endogenous proteinases. Many pathological effects originate from misregulated endogenous proteinases or proteinases encoded by bacteria, viruses, or parasites (1), showing the important function of their

inhibitors. Thus, proteinase inhibitors become more and more important as drugs against currently untreatable diseases. The most prominent example for this kind of therapy is probably the rationally designed HIV-1² proteinase inhibitors (2), although the observed numbers of resistances against several of these drugs are increasing (3).

HF6478 (hemofiltrate peptide with molecular weight 6478 Da) seems to be a serine proteinase inhibitor according to sequence homologies with other inhibitors. It is isolated from human blood and is part of the precursor protein LEKTI that exhibits 15 potential inhibitory domains (4). The amounts of HF6478 (55 amino acids) isolated from human blood were sufficient for amino acid sequencing and mass spectrometry (4). Further studies like inhibition assays and the determination of the three-dimensional protein structure require higher amounts of HF6478, making expression of the recombinant protein (rHF6478) necessary.

Here we report cloning, overexpression, and purification of rHF6478 as thioredoxin fusion protein, using an *Escherichia coli* *trxB*⁻/*gor522*⁻ double mutant that allows disulfide formation in its oxidative cytoplasm (5, 6). After cleavage of the fusion protein, subsequent

² Abbreviations used: BCA, bicinchoninic acid; CD, circular dichroism; CZE, capillary zone electrophoresis; EK, enterokinase; Trx, *E. coli* thioredoxin; HIV, human immunodeficiency virus; HF6478, hemofiltrate peptide with molecular weight 6478 Da; rHF6478, recombinant HF6478; LB, Luria-Bertani; IPTG, isopropyl β -D-thiogalactoside; PMSF, phenylmethylsulfonyl fluoride; SEC, size-exclusion chromatography.

¹ To whom correspondence should be addressed. Fax: ++49 921 553544. E-mail: Ute.Marx@uni-bayreuth.de.

purification yields rHF6478 in amounts sufficient for NMR spectroscopy. The recombinant protein has been analyzed by mass spectrometry, CZE, and circular dichroism (CD) spectroscopy. The disulfide connectivity of rHF6478 was determined by CZE, using a chemically synthesized version of the peptide with known disulfide pattern as reference.

MATERIALS AND METHODS

Construction of the Expression Vector pET-32a-Xa-hf6478

A fragment of the LEKTI cDNA encoding the peptide HF6478 was amplified by PCR (7) using a forward primer which introduces a *KpnI* site and a sequence encoding a factor Xa cleavage site at the 5' end of the gene (GGA GGA GGA TCC GGT ACC ATC GAA GGT CGT AAG AAT GAA GAT CAG GAA ATG). The reverse primer contains a *XhoI* site at its 5' end (GGA GGA CTC GAG TTA CTG TGA TTT TGC TTC TTT TTC CAG). The fragment was inserted into the T7 RNA polymerase-based expression vector pET-32a (Novagen). The resulting pET-32a-Xa-hf6478 is able to produce the peptide HF6478 with an NH₂-terminal *E. coli* thioredoxin fusion (Trx-tag, Novagen). The two proteins are connected by a 43-amino-acid linker containing six histidine residues and a factor Xa cleavage site resulting in the fusion protein Trx-Xa-HF6478.

Preparation of Soluble Extracts

A starter culture of 50 ml LB containing 200 µg/ml ampicillin, 15 µg/ml kanamycin, and 12.5 µg/ml tetracycline was inoculated with the *E. coli* strain Origami (DE3)/pET-32a-Xa-hf6478 and grown overnight at 37°C. The starter culture was used to inoculate 1 liter of LB medium (containing 200 µg/ml ampicillin, 15 µg/ml kanamycin, and 12.5 µg/ml tetracycline) at an OD₆₀₀ of 0.1. Cultures were grown to an OD₆₀₀ of 0.8, and the expression of the recombinant protein was induced by addition of IPTG to 1 mM. After 3 h, cells were harvested and resuspended in 20 ml of buffer A (50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole) with 0.5 mM PMSF, 0.2 mg/ml lysozyme, 0.2 mg/ml DNaseI, and 2 mg/ml benzamidine. The cells were lysed by three 30-s sonication steps (*T* = 4°C; Labsonic U sonicator at 150 W intensity) followed by overnight freezing at -80°C. After thawing, the extract was clarified by centrifugation for 60 min at 40,000 rpm, at 4°C in a TFT 70.38 rotor.

Purification of Recombinant HF6478

The soluble extract (20 ml) was loaded on a Ni²⁺-NTA column (3-ml bed vol, Qiagen, batch column) which was preequilibrated in buffer A. The column was washed

with buffer A and 50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 30 mM imidazole, and the absorbed proteins were eluted with 50 mM Tris/HCl, pH 8.0, 300 mM NaCl, 200 mM imidazole. The fractions containing the fusion protein (as verified by SDS-PAGE) were dialyzed against 50 mM Tris/HCl, pH 8.0, 100 mM NaCl to reduce the high imidazole concentrations. The protein concentration was estimated using the BCA assay (Pierce) (8).

To remove the Trx-tag, the dialyzed protein solution was adjusted to a final concentration of 5 mM CaCl₂ and incubated at room temperature with factor Xa (Novagen; 1 unit/mg of Trx-Xa-HF6478) for 16 h. After stopping digestion by addition of 1 mM PMSF the sample was loaded on a Ni²⁺-NTA column (3-ml bed vol), preequilibrated in buffer A without imidazole. The column was washed with buffer A (without imidazole) and the rHF6478 containing fractions were pooled and dialyzed against buffer B (20 mM ammonium acetate, pH 6.0). After lyophilization the protein was solubilized in 1–2 ml distilled water and the concentrated protein solution (4–5 mg/ml) was loaded on a Superdex 75 HR 30/10 column (Pharmacia Biotech; at a flow rate of 0.15 ml/min), preequilibrated in buffer B. The rHF6478 containing fractions were pooled and dialyzed against distilled water, and the protein concentration was estimated. The lyophilized protein was used for further characterization by amino acid sequencing, mass spectrometry, CZE, and CD spectroscopy.

Characterization of Recombinant HF6478

Electrospray mass spectra were recorded on a biomolecular mass analyzer (Perkin Elmer, Sciex API 100), using a protein sample solubilized in 100 µl of 50% CH₃CN and 0.2% HCOOH.

Capillary zone electrophoresis was carried out on a Bio-Rad CZE CE3000 system (Bio-Rad, Munich, Germany) in 0.1 M phosphate buffer with polymer modifier (pH 2.5) (capillary, 30 cm × 50 µm (uncoated); temperature, 292K; absorbance was detected at 200 nm; run time, 20 min).

RESULTS AND DISCUSSION

Purification of Recombinant HF6478

The gene fragment coding for the peptide HF6478 was combined with a sequence encoding a factor Xa cleavage site at its 5' end, using standard PCR techniques (7) and was cloned into pET-32a. As host strain for the expression of recombinant HF6478, *E. coli* Origami (DE3) (Novagen), a *trxB*⁻/*gor522*⁻ double mutant was used, allowing disulfide formation in the oxidative cytoplasm of this strain (5, 6). Thus, recombinant HF6478 with an NH₂-terminal *E. coli* thioredoxin fusion was expressed by the strain *E. coli* Origami (DE3)/

pET-32a-Xa-hf6478. Expression of the Trx fusion protein (Trx-Xa-HF6478; 212 amino acids; 23.4 kDa) was induced by 1 mM IPTG with an expression level of 40% of total cellular proteins (Fig. 1a, lane 2) as deduced by densitometric analysis of Coomassie-stained SDS-PAGE (data not shown). To avoid reduction of disulfide bridges and renaturation of denatured fusion protein from inclusion bodies, only the soluble fraction of cell lysate was used for further purification, containing fusion protein at 28% of the soluble fraction (Fig. 1a, lane 3). Trx-Xa-HF6478 was purified by one-step Ni²⁺-chelating affinity chromatography (9) with a purity of 66% (Fig. 1a, lane 4). A total of 27 mg of recombinant Trx fusion protein was obtained from 2.4 g wet wt cells (1 liter of culture).

Enterokinase (EK) and factor Xa are, according to their specificity (EK, DDDDK↓; factor Xa, IEGR↓), the only commercially available proteinases that cleave the fusion protein to rHF6478 without additional amino acids at its NH₂-terminus. Since we were unable to cleave the thioredoxin domain from a fusion protein containing the EK cleavage site, as in the original pET-32a vector construct, a modified Trx fusion protein containing a factor Xa cleavage site was used. Compared to the EK cleavage site, sequences similar to that recognized by factor Xa are more prevalent, leading to a higher probability of secondary cleavage (10, 11). The fusion protein (Trx-Xa-HF6478) contains Arg 157 of the

factor Xa cleavage site and Arg 74 as part of the thioredoxin sequence, as well as arginines 129 and 143, located between His-tag and factor Xa cleavage site, as potential secondary cleavage sites. As the sequence of HF6478 does not contain any arginine residues, the overall yield of the peptide would not have been influenced by an unspecific activity of factor Xa. After incubation with the enzyme, a strong decrease in intensity of the fusion protein band (Fig. 1a, lane 5) was observed, showing that cleavage of Trx-Xa-HF6478 was possible. In addition to the expected bands of apparently 17 kDa (Trx-tag) and 7 kDa (rHF6478) (Fig. 1a, lane 5, marked by arrow), two bands with apparent molecular masses of 15 and 14 kDa (Fig. 1a, lane 5) were obtained, indicating that secondary cleavage has indeed occurred after Arg 129 and Arg 143. Size-exclusion chromatography (SEC) to remove secondary cleavage products resulted in substantial loss of rHF6478 (data not shown), necessitating a different purification strategy. As part of all larger peptide fragments (comprising amino acids 1–129, 1–143, and 1–157), the His-tag allowed a second Ni²⁺-chelating affinity chromatography step to purify rHF6478 from secondary cleavage products (Fig. 1, lane 6). After SEC to remove small cleavage fragments (comprising amino acids 129–143, 129–157, and 143–157; Fig. 1b), rHF6478 with apparent homogeneity was obtained (Fig. 2b). Starting with 27 mg of Trx fusion protein containing 7.4 mg of rHF6478, the total of 3 to 3.6

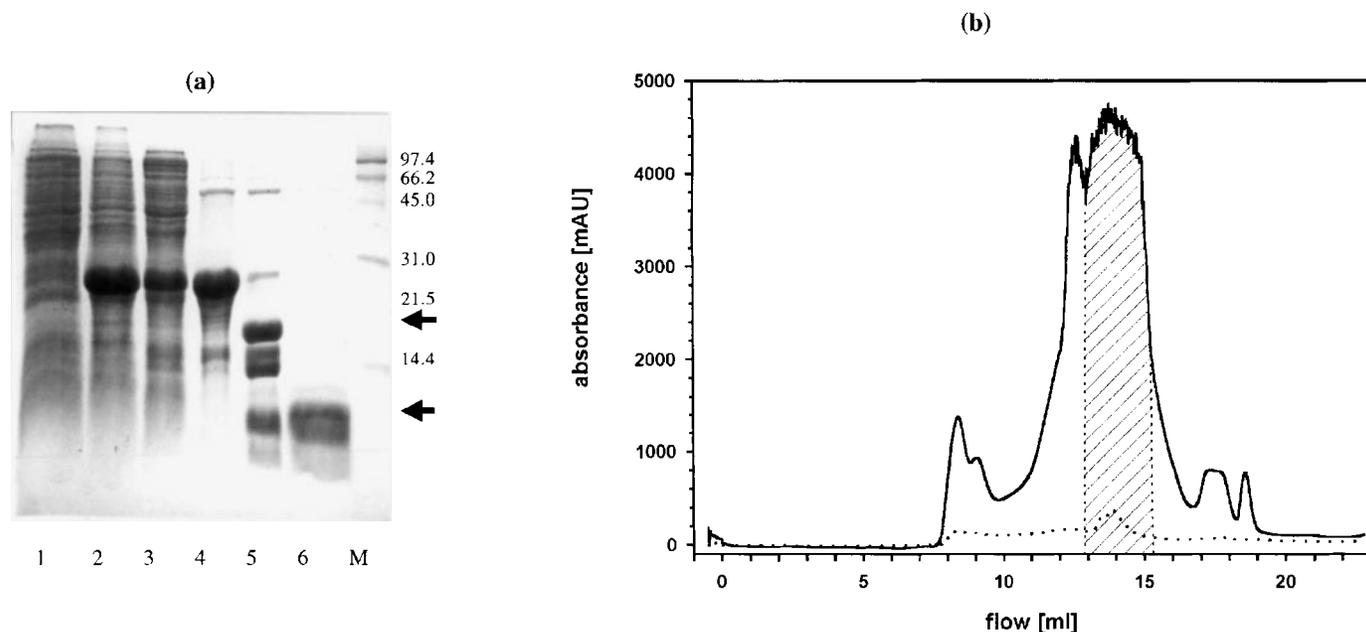


FIG. 1. Purification of recombinant HF6478 (rHF6478). (a) Expression of Trx-Xa-HF6478 in *E. coli* Origami (DE3): lanes 1 and 2, before and after induction with 1 mM IPTG, respectively; lane 3, soluble fraction of cell lysate; lane 4, pooled Trx-Xa-HF6478-containing fractions after Ni²⁺-chelating affinity chromatography; lane 5, after fusion protein cleavage (Trx-tag and rHF6478 are marked by arrows); lane 6, second Ni²⁺-chelating affinity chromatography to remove most of the secondary cleavage fragments; M, molecular weight markers (in kDa). (b) Elution profile of the size-exclusion chromatography to remove small cleavage fragments (as described under Materials and Methods; solid curve, A₂₂₀; dotted curve, A₂₈₀; AU, arbitrary units).

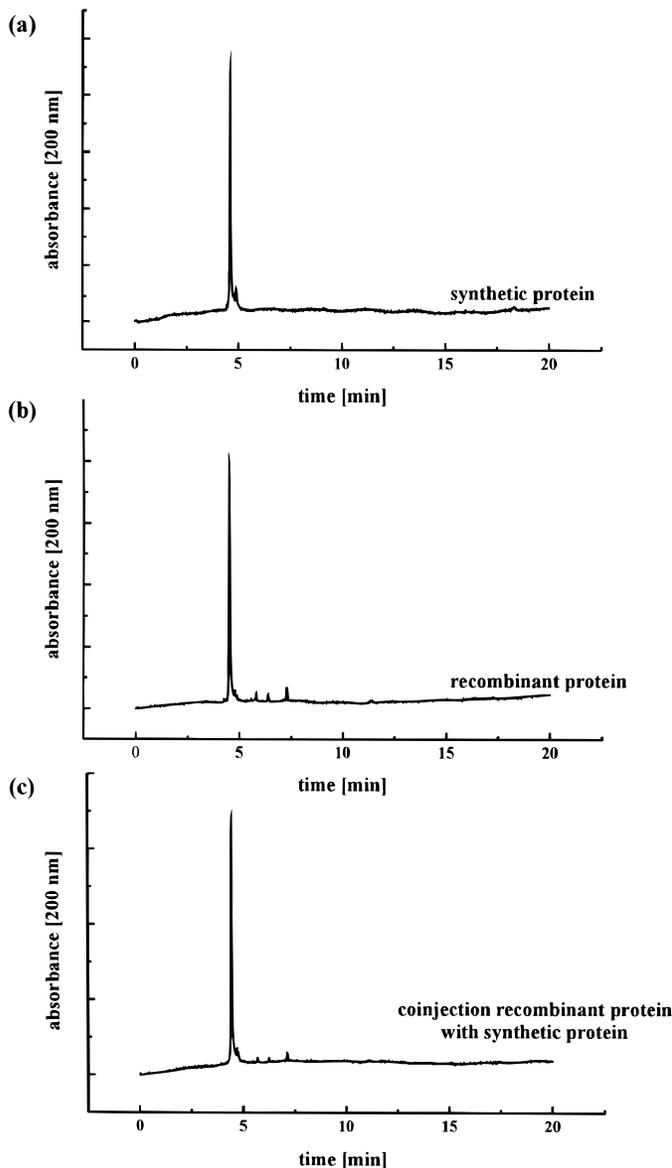


FIG. 2. CZE analysis of rHF6478 (as described under Materials and Methods). (a) Synthetic HF6478 with selectively synthesized disulfide bridges (Cys 8–Cys 44 and Cys 22–Cys 41). (b) Recombinant HF6478 (SEC-pooled fractions corresponding to the hatched area in Fig. 1b). (c) Coinjection of recombinant with synthetic HF6478.

mg of recombinant HF6478 obtained after cleavage and subsequent purification corresponds to an overall recovery yield of 40 to 48%. Isotopic labeling, as required for heteronuclear NMR experiments, can be performed in minimal medium, leading to similar overall recovery yields.

Characterization of Recombinant HF6478

The NH_2 -terminal sequence of rHF6478 (KNEDQEMCH–) was verified by sequencing. Using SDS–PAGE, the apparent molecular weight of

rHF6478 was estimated to be 7 kDa, in agreement with the molecular weight obtained by electrospray mass spectrometry (6477.69 ± 0.68 Da) as well as with the calculated value of 6477.59 Da.

When expression of His-tagged HF6478 was carried out in *E. coli* BL21 (DE3), all of the three possible disulfide isomers were formed, making isolation of the correct isoform necessary. In contrast, expression of rHF6478 as Trx fusion protein in *E. coli* Origami (DE3) resulted in a uniform product. In order to determine its disulfide pattern we took advantage of the significantly different retention times of the three possible disulfide isomers in RP–HPLC (data not shown) and CZE. Coinjection (Fig. 2c) of rHF6478 with a synthetic version of the peptide with selectively synthesized disulfides (Fig. 2a) resulted in a strong single peak, indicating a consistent 1–4, 2–3 (Cys 8–Cys 44; Cys 22–Cys 41) connectivity for both peptides suggested to be the native form in human blood filtrate.

The CD spectrum of rHF6478 from 185 to 250 nm (Fig. 3) is typical for a protein containing both α -helix and β -sheet with a broad minimum between 208 and 222 nm as well as an intense positive band at 195 nm. Evaluation of the spectrum resulted in an estimate of 19.8% α -helix and 33.2% β -sheet (12). As HF6478 contains two disulfide bridges, as well as the side chains of aromatic amino acid residues that influence CD spectra (13), the exact contents of regular secondary structural elements of the peptide will be known only after determination of the detailed 3D structure.

CONCLUSIONS

When extracellular proteins are expressed in conventional *E. coli* strains, missing or wrong disulfide bridges

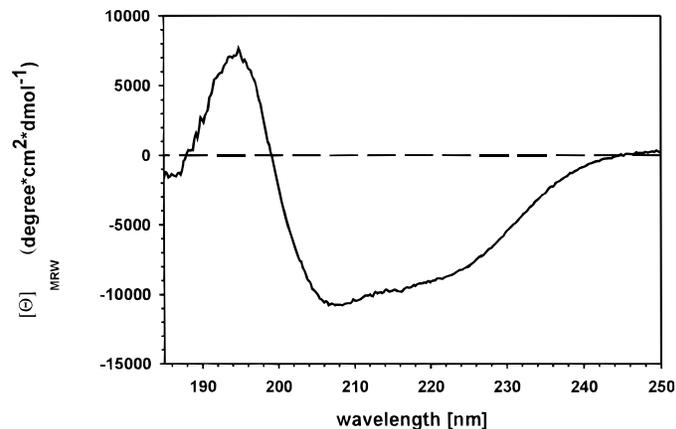


FIG. 3. The CD spectrum of rHF6478 is typical for a protein containing both α -helix and β -sheet. The spectrum was recorded on a Jasco J-810 spectrometer at room temperature with 30 μM rHF6478 in H_2O , pH 4.5.

often cause formation of insoluble aggregates (14). Expression of the putative serine proteinase inhibitor HF6478 in the oxidative cytoplasm of *E. coli* Origami (DE3), however, resulted in soluble, properly folded protein with accurately formed disulfides. Thus, the presented strategy is most likely generally useful for the cytoplasmic expression of different extracellular proteins in *E. coli*.

REFERENCES

1. Seife, C. (1997) Blunting nature's Swiss army knife. *Science* **277**, 1602–1603.
2. Roberts, N. A., Martin, J. A., Kinchington, D., Broadhurst, A. V., Craig, J. C., Duncan, I. B., Galpin, S. A., Handa, B. K., Kay, J., Kröhn, A., Lambert, R. W., Merrett, J. H., Mills, J. S., Parkes, K. E. B., Redshaw, S., Ritchie, A. J., Taylor, D. L., Thomas, G. J., and Machin, P. J. (1990) Rational design of peptide-based HIV proteinase inhibitors. *Science* **248**, 358–361.
3. Boden, D., and Markowitz, M. (1998) Resistance to human immunodeficiency virus type 1 protease inhibitors. *Antimicrob. Agents Chemother.* **42**, 2775–2783.
4. Mägert, H.-J., Ständker, L., Kreutzmann, P., Zucht, H.-D., Sommerhoff, C. P., Fritz, H., and Forssmann, W.-G. (1999) LEKTI, a novel 15-domain type of human serine proteinase inhibitor. *J. Biol. Chem.* **274**, 21499–21502.
5. Derman, A. I., Prinz, W. A., Belin, D., and Beckwith, J. (1993) Mutations that allow disulfide bond formation in the cytoplasm of *Escherichia coli*. *Science* **262**, 744–747.
6. Prinz, W. A., Aslund, F., Holmgren, A., and Beckwith, J. (1997) The role of the thioredoxin and glutaredoxin pathways in reducing protein disulfide bonds in the *Escherichia coli* cytoplasm. *J. Biol. Chem.* **272**, 15661–15667.
7. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1999) *Current Protocols in Molecular Biology*, Unit 15.1, Vol. 3, Wiley, New York.
8. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. (1985) Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76–85.
9. Porath, J. (1992) Immobilized metal ion affinity chromatography. *Protein Express. Purif.* **3**, 263–281.
10. Nagai, K., and Thøgerson, H. C. (1987) Synthesis and sequence-specific proteolysis of hybrid proteins produced in *Escherichia coli*. *Methods Enzymol.* **153**, 461–481.
11. Lottenberg, R., Christensen, U., Jackson, C., and Coleman, P. L. (1981) Assay of coagulation proteases using peptide chromogenic and fluorogenic substrates. *Methods Enzymol.* **80**, 341–361.
12. Sreerama, N., and Woody, R. W. (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. *Anal. Biochem.* **209**, 32–44.
13. van Holde, K. E., Johnson, W. C., and Ho, P. S. (1998) *Principles of Physical Biochemistry*, pp. 431–447, Prentice Hall, London.
14. Mitraki, A., and King, J. (1989) Protein folding intermediates and inclusion body formation. *Biotechnology* **7**, 690–697.