Analysis, by electrospray ionization mass spectrometry, of several forms of *Clostridium pasteurianum* rubredoxin

Yves PETILLLOT,* Eric FOREST,*† Isabelle MATHIEU,† Jacques MEYER† and Jean-Marc MOULIS†

*Laboratoire de Spectrométrie de Masse des Protéines, Institut de Biologie Structurale, 41, avenue des Martyrs, 38027 Grenoble Cedex 1, and †Laboratoire des Métalloprotéines, Département de Biologie Moléculaire et Structurale, CENG, 85X, 38041 Grenoble Cedex, France

Clostridium pasteurianum rubredoxin and its recombinant counterpart purified from *Escherichia coli* have been analysed by electrospray ionization m.s. (e.s.i.m.s.). Whereas the N-terminal methionine of the native protein is formylated, the recombinant one has a free N-terminal methionine. *E. coli* cells also produce a colourless protein from the cloned gene. This protein is absent from *C. pasteurianum* and was shown to be zinc-substituted rubredoxin. The molecular forms of rubredoxin detected by e.s.i.m.s. depended on the experimental conditions used. Signifi-

cant conversion into apo-rubredoxin occurred when the proteins were ionized at acidic pH and detected in the positive-ion mode. This conversion was quantitative in the case of Zn-rubredoxin. In contrast, when the proteins were analysed at neutral pH in the negative-ion mode, only the holoproteins, i.e. the species initially present in the solutions, were detected in the spectra. The e.s.i.m.s. experimental conditions set up here may prove useful for the analysis of other acidic metalloproteins with weakly bound metals.

INTRODUCTION

The availability of an increasing number of recombinant proteins, often synthesized in *Escherichia coli*, widens the scope of the studies on the structure–function relationships in these molecules. A preliminary step in the use of recombinant proteins is their detailed characterization. Indeed, depending in part on the origin of the gene, the overexpressing host may not handle the gene and its product in exactly the same manner as the natural organism. For instance, experimental conditions leading to overproduction of the target protein may alter the capacity of *E. coli* to carry out all of the post-translational processes occurring in the natural host.

We have recently expressed in *E. coli* the *Clostridium pasteurianum* gene encoding rubredoxin (Mathieu et al., 1992), a small protein (54 amino acids) the active site of which consists of an iron atom co-ordinated to four cysteinyl sulphur atoms. *E. coli* cells were shown to efficiently produce a protein having the expected size and active site. However, this protein in its apo form exhibited a molecular mass differing by 28 mass units from that of the natural rubredoxin. In addition, a protein very similar to rubredoxin, but lacking the iron atom, was also isolated from *E. coli* cells transformed with the plasmid containing the rubredoxin gene (Mathieu et al., 1992).

In order to characterize these species further, we have implemented electrospray ionization m.s. (e.s.i.m.s.) which has proved to be a valuable technique for precise mass determinations of biopolymers (Fenn et al., 1989; Chait and Kent, 1992). However, as the interface conditions are selected to provide efficient ionization and sufficient excitation for a good desolvation, labile species are most often dissociated. The iron–sulphur proteins are a typical case of such an occurrence; since the iron–sulphur bonds are weaker than the other bonds involved in proteins, the metal centres are likely to be removed during implementation of the method. In line with this prediction, we have found the

apoprotein contribution to be dominant, and sometimes exclusive, in the e.s.i.m.s. spectra of several iron–sulphur proteins (Y. Petillot, E. Forest, J. Meyer and J.-M. Moulis, unpublished work). We have nevertheless succeeded in obtaining mass spectra with no other contribution than that of the holoprotein in the cases of the rubredoxin from *Clostridium thermosaccharolyticum* (Meyer et al., 1990) and of the high-potential iron protein ('HiPIP') from *Chromatium tepidum* (Moulis et al., 1993). However, since these proteins are unusually stable, in part due to their being produced by thermophilic bacteria, getting similar results with less stable iron–sulphur proteins will probably require the implementation of milder experimental conditions.

Thus we have carried out a comparative analysis of several forms of *C. pasteurianum* rubredoxin with two goals in mind. First, the investigation of an iron–sulphur protein less stable than those analysed previously is a step towards the eventual set up of e.s.i.m.s. conditions applicable to a wide range of iron–sulphur proteins. Secondly, the structural differences between the native and recombinant rubredoxins are reflected in small mass differences for the measurement of which e.s.i.m.s. is an ideal technique. The analysis of these differences may prove important for the understanding of post-translational processing of metalloproteins in *E. coli*.

MATERIALS AND METHODS

Clostridium pasteurianum rubredoxin (hereafter referred to as 'N Rd') was isolated as previously described (Mathieu et al., 1992). For e.s.i.m.s. measurements, a 200 µl sample containing 50 nmol of protein in 10 mM sodium phosphate buffer, pH 7.5, was incubated at room temperature for 2 h under Ar with 200 µl of 1 M-ammonium acetate, pH 7.0. The solution was filtered on a 10 ml column containing Bio-Gel P-2 (Bio-Rad Laboratories) equilibrated with 10 mM ammonium acetate, pH 7.0. The eluted protein was concentrated to ~0.2 mM on a YM5 membrane

Abbreviations used: e.s.i., electrospray ionization; LB, Luria broth; N Rd, native rubredoxin from *Clostridium pasteurianum*; R Rd, recombinant rubredoxin from *Escherichia coli*; R*Rd, colourless recombinant rubredoxin from *E. coli*.

† To whom correspondence should be addressed.

(Amicon) and used as a stock solution in the e.s.i.m.s. experiments.

In order to improve the production of recombinant rubredoxin (hereafter referred to as 'R Rd'), the rubredoxin gene (Mathieu et al., 1992) was put under the dependence of the T7 RNA polymerase promoter $\Phi 10$. Plasmid pCPRD2 was digested with *EcoRI* and *HindIII* and the mixture separated on a 1.5%-agarose gel (Seakem; FMC Bioproducts) in Tris/borate buffer. The 518 bp band containing the rubredoxin gene was sliced from the gel and eluted by low-speed centrifugation through sterile glass wool in a tip for an automatic pipette. The resulting solution was precipitated with ethanol, and the dry DNA pellet was dissolved in water; 0.1 pmol of this DNA fragment was ligated to 20 fmol of *EcoRI*-*HindIII*-cleaved plasmid pT7-5 (Tabor, 1990) and used to transform competent *E. coli* DH5 α . The plasmid content of the transformants was analysed by restriction analysis and sequencing. One purified plasmid displaying the expected sequence was named pTRD1 and used to transform *E. coli* K38 containing plasmid pGP1-2 (Tabor, 1990). One resulting single colony was used to grow a preculture which eventually served to inoculate 1 litre of Luria broth (LB) medium. Production of R Rd essentially followed a standard protocol (Tabor, 1990) with synthesis of the protein for 90 min after the end of the induction period. Rubredoxin purification was carried out as described elsewhere (Mathieu et al., 1992). Holo-rubredoxin was separated from a colourless rubredoxin form (hereafter referred to as 'R*Rd') in the last anion-exchange, purification step. The final yield of pure holo-rubredoxin ($A_{490}/A_{280} = 0.44$) was 5.5 mg/litre of culture in these conditions. R Rd and R*Rd were prepared for e.s.i.m.s. experiments as described above for native rubredoxin.

M.s. analyses of the proteins were carried out on a Perkin-Elmer Sciex API III triple quadrupole mass spectrometer equipped with a nebulizer-assisted electrospray (ionspray) source. The spectra were recorded in the 1000–2100 range of mass-to-charge (m/z) ratios in steps of 0.5 m/z , with a 2 ms dwell time and with a resolution of 1 mass unit. The signal was averaged over seven scans. The instrument was tuned and calibrated in the positive mode (for the detection of cations). For the negative-mode operations (detection of anions), all the instrumental parameters were unaltered, with the exception of reversing the polarities and decreasing the absolute value of ion-spray voltage from 5000 to 4000 V. Samples were infused into the source using a Harvard 22 syringe pump at a flow rate of 3 μ l/min and a Valco C6W injector equipped with a 0.5 μ l internal loop. Proteins were dissolved at a concentration of 1 mg/ml in the solvents used for infusion, which were 0.1% formic acid, pH 2.4, in positive mode or 10 mM ammonium acetate/ammonia, pH 8, in negative mode. About 150 pmol of protein was used for each spectrum. The reconstructed molecular-mass profiles were obtained by using a deconvolution algorithm (Perkin-Elmer Sciex).

RESULTS

The amino acid sequence of the rubredoxin from *C. pasteurianum* has recently been redetermined by Edman degradation and by translation of the corresponding gene (Mathieu et al., 1992). The protein is composed of 54 amino acids, with a large excess of acidic (eight aspartate residues and five glutamate residues) over basic (four lysine residues) residues. Accordingly, its calculated (MacBioSpec software; Perkin-Elmer Sciex) pI is 4.6.

R Rd spectra obtained in negative and positive modes are respectively presented in Figures 1(a) and 1(b). A coherent series

of multiply charged ions, corresponding to molecules bearing from three to six charges is clearly observed in both cases. The presence of several species is better revealed from the reconstructed molecular-mass profiles of the mass spectra (Figures 2a and 2b). In both spectra, the measured molecular mass of the main component (6099 and 6100 ± 1 Da) is in good agreement with the isotope-averaged mass deduced from the amino acid sequence with an iron atom bound (6099.5 Da). A partial loss of the iron ion appears in the positive-mode spectrum with a peak at 6047.5 Da, whereas there is no evidence for apoprotein formation in the negative mode. This indicates that the conditions of ionization in the positive mode are somewhat detrimental to the protein. Other components, corresponding to sodium adducts (from one to three) are also observed in Figures 2(a) and 2(b). These cation adducts are frequently revealed in mass spectrometry and are likely to arise from strong ionic interactions with carboxylic acid side-chains.

The reconstructed spectra of N Rd display, in both the negative and the positive modes (Figures 3a and 3b respectively), a major component at a measured molecular mass of 6128 ± 1 Da. This increase of 28 Da corresponds to the mass of a formyl group, as previously observed with the apo-rubredoxins and in agreement with the blocking of the N-terminal methionine inferred from amino acid sequencing (Mathieu et al., 1992). Both spectra of N Rd also display contributions of the unblocked

Figure 1 E.s.i.m.s. spectra of R Rd

(a) Negative-ion mode; (b) positive-ion mode. The net charge of the ions is indicated above the peaks which correspond to multiprotonated forms (from three to six) of R Rd.

Figure 2 E.s.-reconstructed molecular mass profiles of R Rd

These spectra were obtained by using a deconvolution algorithm; (a) negative-ion mode; (b) positive-ion mode. The molecular mass of the main species is indicated above the corresponding peaks.

Figure 3 E.s.-reconstructed molecular-mass profiles of N Rd

These spectra were obtained by using a deconvolution algorithm; (a) negative-ion mode; (b) positive-ion mode. The molecular mass of the main species is indicated above the corresponding peaks.

protein at 6100 Da (Figure 3). In the positive detection mode, but not in the negative one, apoproteins arising from both the blocked and the unblocked rubredoxin can be observed at 6075 and 6048 Da respectively. It is noteworthy that the ratio of holoprotein to apoprotein is much larger in the case of the

Figure 4 E.s.-reconstructed molecular-mass profiles of R*Rd

These spectra were obtained by using a deconvolution algorithm; (a) negative-ion mode; (b) positive-ion mode. The molecular mass of the main species is indicated above the corresponding peaks.

formylated rubredoxin than in the case of the non-formylated one (Figure 3b). This suggests a greater stability of the formylated rubredoxin, as compared with the non-formylated one. From the spectrum recorded in negative-ion mode (Figure 3a), a crude estimation shows that the formylated rubredoxin represents more than 75% of the protein extracted from *C. pasteurianum*. Peaks corresponding to masses higher than 6128 Da, and probably assignable to sodium and potassium adducts, are observed in the spectra of Figure 3.

Reconstructed mass spectra of the colourless rubredoxin expressed in *E. coli*, R*Rd, are presented in Figures 4(a) and 4(b) (negative and positive modes respectively). The negative-mode spectrum shows only one peak corresponding to a protein with a molecular mass of 6109.5 ± 1 Da. This value corresponds to the mass calculated for rubredoxin with a zinc atom replacing iron (6109.1 Da). In contrast, the major peak in Figure 4(b) (positive mode) at 6046.5 Da corresponds to the non-formylated apo-rubredoxin. Peaks present at higher mass in Figure 4(b) are again due to sodium and potassium attachments on the apoprotein. It is noteworthy that, in the positive-ion mode, no holoprotein is detected and, conversely, in the negative-ion mode, no apoprotein is present.

DISCUSSION

E.s.i. (Chait and Kent, 1992) is an efficient method to produce multiply charged gas-phase ions from protein molecules in solution. The composition of the solvent and its pH have been shown to be important parameters determining the relative abundance of the obtained ions (Loo et al., 1991). In the case of metalloproteins, the metal ions may be separated from the polypeptide chain during the ionization process (Y. Petillot, E. Forest, J. Meyer and J. M. Moulis, unpublished work). We have therefore investigated conditions under which optimal results could be obtained with three molecular forms of *C. pasteurianum* rubredoxin.

The measured mass of the recombinant rubredoxin purified from *E. coli* is identical to the one expected from the amino acid sequence with one iron atom bound (Figure 2). This confirms that the N-terminal methionine residue is not blocked, as inferred from the initial sequencing yield and from preliminary m.s. measurements on apo-rubredoxin (Mathieu et al., 1992). In contrast, a large fraction (> 75%) of the rubredoxin extracted from *C. pasteurianum* is formylated, most likely on the N-terminal methionine (Figure 3), in agreement with the previous report of low (~5%) initial sequencing yields by Edman degradation (Mathieu et al., 1992). Thus the post-translational processing of rubredoxin differs in the two bacteria. The de-formylase activity is presumably larger in *E. coli*, as only a small minority of soluble proteins are found in the formylated state (Adams, 1968). Corresponding data for *C. pasteurianum* are not available. In any case, the recombinant protein is more homogeneous (totally unblocked N-terminus) than the native one (mixture of blocked and unblocked N-terminus). The homogeneity of the former protein might be beneficial to some investigations, but it should nevertheless be borne in mind that the native protein has yielded crystals of outstanding quality, despite its N-terminus heterogeneity (Watenpaugh et al., 1980).

In the last step (anion-exchange h.p.l.c.) of R Rd purification, a colourless protein was isolated, which migrated identically with rubredoxin in SDS/PAGE and which could partially be reconstituted into holo-rubredoxin upon incubation in the presence of Fe³⁺ and thiol. For these reasons it was tentatively identified as aporubredoxin (Mathieu et al., 1992). The presently reported e.s.i.m.s. data (Figure 4a) demonstrate that this protein, R*Rd, is in fact Zn-substituted rubredoxin. The negative-ion detection mode spectrum shows that no apo-rubredoxin is associated with this material. The e.s.i.m.s. data afford no direct evidence that the zinc atom replaces iron at the same site on rubredoxin, but it has independently been shown, by n.m.r. of Zn- and ¹¹³Cd-substituted rubredoxins, that these metals replace iron iso-structurally (Blake et al., 1992). In the positive-ion detection mode, the zinc atom is quantitatively lost by rubredoxin (Figure 4b). This result indicates that the binding of zinc to rubredoxin is more acid-labile than the binding of iron, and explains why zinc-substituted rubredoxin could be partially converted into iron-rubredoxin in the presence of the latter metal (Mathieu et al., 1992).

The synthesis of metalloproteins containing a metal different from the naturally occurring one has previously been observed in two cases of heterologous expression in *E. coli*. The use of a synthetic gene encoding a rubredoxin differing slightly from *C. pasteurianum* rubredoxin led to the production of several species, including one lacking iron and containing one atom of zinc per molecule of protein (Eidsness et al., 1992). Expression of the *azu* gene of *Pseudomonas aeruginosa* afforded a mixture of authentic Cu(II)-azurin and of another azurin-like molecule which had lost the ability to bind Cu (van de Kamp et al., 1990). The latter molecular form was subsequently found, by a combination of atomic-absorption analysis and X-ray crystallography, to have incorporated Zn instead of Cu at the metallic centre of the protein (Nar et al., 1992). Metal substitutions such as those described above do not occur frequently, as the failure to correctly incorporate the naturally present metal ion(s) often results in the production of metal-free proteins (e.g. Karlsson et al., 1989; Varadarajan et al., 1985; Smith et al., 1990).

Zinc tetrahedrally co-ordinated to four cysteinyl sulphur atoms occurs in mammalian alcohol dehydrogenases and in aspartate transcarbamylase, but is not directly involved in the catalytic properties of these enzymes (Vallee and Auld, 1990). Zn-substitution by another metal is usually possible but, generally,

with a lower affinity (Bergman et al., 1992). The same co-ordination pattern occurs in the subset of the Zn-finger proteins (Berg, 1990) sometimes designated as Zn-twists (Vallee et al., 1991). In model compounds of these, copper (Hutchens et al., 1992) or iron (Omichinski et al., 1993) can replace zinc. As iron does not impair the specific DNA-binding properties of the erythroid transcription factor GATA-1, the proposal that Zn may not be the only naturally occurring metal in such systems has even been put forward (Omichinski et al., 1993).

It is worth noting that neither in the case of azurin nor in that of rubredoxin has any evidence been found for the presence of the Zn-substituted species in the natural hosts, *P. aeruginosa* or *C. pasteurianum*, respectively. As discussed for azurin (Nar et al., 1992), the selection for one particular metal may be regulated *in vivo* by metal homeostasis or by selective carriers. The same arguments might be proposed for the synthesis of rubredoxin in *C. pasteurianum*, but a simpler explanation follows from the present e.s.i.m.s. experiments. The formation of Zn-rubredoxin, owing to its low stability in acidic conditions (Figure 4) is probably less favoured in the relatively acidic cytoplasm of fermenting *C. pasteurianum* cells (Grupe and Gottschalk, 1992) than in *E. coli*.

In two previous e.s.i.m.s. studies of iron-sulphur proteins, holoproteins only were detected in the spectra, despite the use of the positive-ion detection mode (Meyer et al., 1990; Moulis et al., 1993). However, in view of the unusually high stability of these proteins, similar experimental conditions could not be expected to produce as good results with a wide range of iron-sulphur proteins. Indeed, the three proteins investigated here lose their metal ion to various extents in the positive-ion mode. The loss is even complete in Zn-rubredoxin (Figure 4b). The negative-ion detection mode could be anticipated to be more adapted to the investigation of small iron-sulphur proteins such as rubredoxins and ferredoxins; most of these proteins are acidic, and are therefore negatively charged at neutral pH, where their stability is highest. Negative-ion detection has indeed proved here to be a very efficient alternative; in no case was any apoprotein detected, even for Zn-rubredoxin, which was quantitatively converted into apoprotein in the positive-ion mode. This opens the way to the detection and analysis, as holoproteins, of less stable iron-sulphur proteins. Along these lines we are currently completing an investigation of [2Fe-2S] and [2Fe-4S] ferredoxins. In the long run, analysis of larger and less stable iron-sulphur proteins might become feasible. One might then be able, by e.s.i.m.s., to gain access to an often missing and most important piece of structural information, namely the precise composition of the metal centres.

REFERENCES

- Adams, J. M. (1968) *J. Mol. Biol.* **33**, 571–589
- Berg, J. M. (1990) *J. Biol. Chem.* **265**, 6513–6516
- Bergman, T., Jörnvall, H., Holmquist, B. and Vallee, B. L. (1992) *Eur. J. Biochem.* **205**, 467–470
- Blake, P. R., Park, J. B., Adams, M. W. W. and Summers, M. F. (1992) *J. Am. Chem. Soc.* **114**, 4931–4933
- Chait, B. T. and Kent, S. B. (1992) *Science (Washington, D.C.)* **257**, 1885–1894
- Eidsness, M. K., O'Dell, S. E., Kurtz, D. M., Jr., Robson, R. L. and Scott, R. A. (1992) *Protein Eng.* **5**, 367–371
- Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. and Whitehouse, C. M. (1989) *Science (Washington, D.C.)* **246**, 64–71
- Grupe, H. and Gottschalk, G. (1992) *Appl. Environ. Microbiol.* **58**, 3896–3902
- Hutchens, T. W., Allen, M. H., Li, C. M. and Yip, T.-T. (1992) *FEBS Lett.* **309**, 170–174
- Karlsson, B. G., Pascher, T., Nordling, M., Arvidsson, R. H. A. and Lundberg, L. G. (1989) *FEBS Lett.* **246**, 211–221
- Loo, J. A., Ogorzalek Loo, R. R., Udseth, H. R., Edmonds, C. G. and Smith, R. D. (1991) *Rapid Commun. Mass Spectrom.* **5**, 101–105

- Mathieu, I., Meyer, J. and Moulis, J.-M. (1992) *Biochem. J.* **285**, 255–262
- Meyer, J., Gagnon, J., Sieker, L. C., Van Dorsselaer, A. and Moulis, J.-M. (1990) *Biochem. J.* **271**, 839–841
- Moulis, J.-M., Scherrer, N., Gagnon, J., Forest, E., Pétilot, Y. and Garcia, D. (1993) *Arch. Biochem. Biophys.* **305**, 186–192
- Nar, H., Huber, R., Messerschmidt, A., Filippou, A. C., Barth, M., Jaquinod, M., van de Kamp, M. and Canters, G. W. (1992) *Eur. J. Biochem.* **205**, 1123–1129
- Omichinski, J. G., Trainor, C., Evans, T., Gronenborn, A. M., Clore, G. M. and Felsenfeld, G. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 1676–1680
- Smith, A. T., Santama, N., Dacey, S., Edwards, M., Bray, R. C., Thorneley, R. N. F. and Burke, J. F. (1990) *J. Biol. Chem.* **265**, 13335–13343
- Tabor, S. (1990) in *Current Protocols in Molecular Biology* (Ausubel, F. A., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K., eds.), pp. 16.2.1–16.2.11. Greene Publishing and Wiley Interscience, New York
- Varadarajan, R., Szabo, A. and Boxer, S. G. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 5681–5684
- Vallee, B. L. and Auld, D. S. (1990) *Biochemistry* **29**, 5647–5659
- Vallee, B. L., Coleman, J. E. and Auld, D. S. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 999–1003
- van de Kamp, M., Hali, F. C., Rosato, N., Finazzi Agro, A. and Canters, G. W. (1990) *Biochim. Biophys. Acta* **1019**, 283–292
- Watenpaugh, K. D., Sieker, L. C. and Jensen, L. H. (1980) *J. Mol. Biol.* **138**, 615–633

Received 28 June 1993/27 July 1993; accepted 2 August 1993