Rubredoxin from *Clostridium thermosaccharolyticum*

Amino acid sequence, mass-spectrometric and preliminary crystallographic data

Jacques MEYER,* Jean GAGNON,† Larry C. SIEKER,†† Alain VAN DORSSELAER§ and Jean-Marc MOULIS*

*DBMS/Métalloprotéines and †DBMS/Biologie Structurale (CNRS URA1333), CENG, 85X, 38041 Grenoble, France,

‡Department of Biological Structure, University of Washington, Seattle, WA 98195, U.S.A., and §Centre de Neurochimie, CNRS, 67084 Strasbourg, France

Rubredoxin isolated from the thermophilic bacterium *Clostridium thermosaccharolyticum* has been sequenced and crystallized. The 52-residue sequence is similar to those of rubredoxins occurring in other anaerobic bacteria, but displays some unique features, including a tryptophan residue in position 4, two consecutive proline residues in positions 25 and 26, and an aspartic acid residue in position 41. The molecular mass (5988 Da) of the native rubredoxin has been measured by electrospray-ionization m.s., thus establishing the applicability of the technique to this type of iron-sulphur protein. *C. thermosaccharolyticum* rubredoxin crystallizes as dark-red elongated prisms with a flat diamond cross-section. The X-ray diffraction shows symmetry consistent with space group $P2_12_12_1$. Cell parameters are: $a = 2.73$ nm, $b = 2.98$ nm, $c = 6.49$ nm.

INTRODUCTION

Rubredoxins (Rds) are small (6 kDa) electron-transfer proteins occurring in a number of bacteria. Their active sites consist of one iron atom co-ordinated to four sulphur atoms belonging to cysteine residues. Several Rds have been purified from anaerobes (clostridia, sulphate reducers and green photosynthetic bacteria) and characterized in detail: nine amino acid sequences have been reported to date [1–9], and four crystal structures have been elucidated with resolutions better than 0.2 nm [10–13]. Another type of Rd, which differs from the previous ones by its size (14 kDa) and by its iron content (2 atoms/molecule), has been found in *Pseudomonas oleovorans*, where it mediates electron transfer to alkane hydroxylase [14]. The study of natural variants of Rd has already proved to be helpful in determining the residues which are structurally important and is expected to be instrumental in elucidating the features essential for electron transfer. In this context we report the amino acid sequence, the crystallization conditions and initial characterization of crystals of the Rd from *Clostridium thermosaccharolyticum*.

MATERIALS AND METHODS

C. thermosaccharolyticum (A.T.C.C. 7956) cells were grown as described [15], lysed with lysozyme and treated with DNAase. The centrifuged crude extract was loaded on a DE52 (Whatman) column [equilibrated with 20 mM-Tris/HCl (pH 7.4)/0.05 M-NaCl]. Subsequent to washing with 0.15 M- and 0.2 M-NaCl, Rd was eluted with 0.3 M-NaCl, concentrated on a small DE52 column, and chromatographed on Ultrogel AcA 202 [equilibrated with 20 mM-Tris/HCl (pH 7.4)/0.2 M-NaCl]. Contaminating material was removed by using a hydroxyapatite column [equilibrated with 5 mM-potassium phosphate/20 mM-Tris/HCl (pH 7.4)/0.2 M-NaCl]. H.p.l.c. chromatography, amino acid analysis and sequencing showed the protein to be pure. Its u.v.-visible absorption spectrum was essentially identical with those of other Rds, with an A_{380}/A_{280} ratio of 0.50. Apo-Rd was prepared by treating the holoprotein with 2 M-HCl at 95 °C for 5 min, and was carboxymethylated as described in [16]. Peptides were prepared by digestion with trypsin, Glu-C, and Lys-C proteinases (Boehringer). They were purified by

reverse-phase h.p.l.c. (model 130A; Applied Biosystems) on a Spheri-5 RP-18 column (2.1 mm × 100 mm) equilibrated with 0.1 % trifluoroacetic acid in water, and developed with increasing concentrations of acetonitrile containing 0.1 % trifluoroacetic acid. A model 477A sequencer, equipped with on-line detection model 120A (Applied Biosystems), was used for sequence analysis. Amino acid compositions were determined on a Beckman model 6300 analyser after 24 h hydrolysis in 6 M-HCl.

Electrospray-ionization m.s. [17,18] of native Rd was performed on a BIO-Q (VG instruments) machine. Samples were first dissolved in water and then an equivalent volume of methanol was added. These protein solutions (concentration 30–100 pmol/μl) were injected into the ion source at a flow rate of 3 μl/min. The mass spectrometer was scanned from m/z 200 to m/z 2000 in 20 s at unit resolution.

Crystals of Rd were obtained from a 0.5–1.0 % protein solution, buffered at pH 5.5 with 0.05 M-phosphate, adjusted to 2 M-(NH₄)₂SO₄ at room temperature. Crystals were grown by vapour equilibration (hanging drop) or by the batch technique, and developed as dark-red elongated prisms with a flat diamond cross-section. A lath-shaped crystal form, less useful because of a very thin cross-section, developed from solutions buffered with 0.25 M-Tris/maleate, pH 6.6, set to equilibrate against 2.5 M-(NH₄)₂SO₄.

RESULTS AND DISCUSSION

The evidence for the amino acid sequence is summarized in Fig. 1, which shows only the most useful of the ten peptides that have been sequenced. The complete sequence of *C. thermosaccharolyticum* Rd was determined with less than 200 μg of protein. The *N*-terminal methionine residue is blocked: deformylation [19] removed about 20 % of the blocking group, as judged by the sequence yield. The polypeptide chain consists of 52 residues. The determined sequence is consistent with the amino acid analyses, except for serine, glutamate and glycine, which gave values slightly higher than expected from the sequence.

The calculated molecular mass of the protein, including the iron atom and the formyl group of the *N*-terminal methionine, is

	1	10	20	30	40	50
	MEKWQCTVCGYIYDPEVGDPTQNIPPGTKFEDLPDDWVCPDCGVGKDQFEKI					
Cm-Rd	MEKWQCTVCGYIYDPEVGDPTQN					
CNBr	EKWQCTVCGYIYDPEVGDPTQNIPPGTKFEDLP					
Glu-C	VGDPTQNIPPGTKFEDLPDDWVCP-CGV					
Trypsin-1	FEDLPDDWVCPDCGVGKDQFE					
Trypsin-2	EDLPDDWVCPDCGVGKDQFE					
Lys-C	DQFEKI					

Fig. 1. Summary proof of the sequence of *C. thermosaccharolyticum* Rd

The final sequence is shown in bold characters. Only the most useful of the sequenced peptides are shown here. Hyphens stand for unidentified residues. Cm-, carboxymethylated.

5988.5 Da. Electrospray ionization generates multicharged molecular ions resulting from the multiple attachment of cations, most often protons, to the macromolecule. These molecular ions are detected on mass spectra as $(M + zH)/z$ peaks, where z is the number of bound protons. In the case of Rd, the charge contributed by the $\text{Fe}(\text{SCys})_4$ active site is -1 . Thus the number of protons bound to the molecular ions is expected to be one unit larger than the number of ionizing charges, and therefore the positions of the peaks in the mass spectra will be $[M + (z + 1)H]/z$. Electrospray-ionization mass spectra of native Rd displayed a peak at m/z 1498.41, which, for $z = 4$, yielded $M = 5988.64$. This experimental value is in excellent agreement with that calculated from the amino acid sequence. The m.s. analysis thus confirms that the 52 residues identified by Edman degradation constitute the complete sequence of the protein. A peak occurring at m/z 1485.32 could be attributed to apo-Rd ionized by four protons: M 5937.28 Da is in excellent agreement with the value (M 5936.5 Da) calculated from the amino acid sequence. The intensity of the m/z 1485.32 peak was only about 20% of that of the m/z 1498.41 peak, thus showing that Rd is relatively stable under the conditions of electrospray ionization. Counterparts, for $z = 5$, of the m/z 1498.41 and 1485.32 peaks were observed at m/z 1199.20 and 1188.14 respectively. Electrospray m.s. has previously been used with other metalloproteins: myoglobin was detected mainly as the apoprotein, whereas cytochrome *c* was mostly in the holoprotein state [18,20]. For carbonic anhydrase [20] and isopenicillin N synthase [21], the zinc and iron atom respectively appeared to remain bound to the proteins under the electrospray-ionization conditions.

Comparison of the ten Rd sequences determined to date reveals 15 conserved residues (Fig. 2), including the *N*-terminal methionine residue, the four cysteine ligands of the iron atom, and five of the six aromatic residues. The most-conserved features are the aromatic residues, which have been shown to form a hydrophobic core contributing to the stability of the protein structure [10–13], and the two loops holding the iron atom. By contrast, the 21–27 loop, which is deleted in *Desulfovibrio desulfuricans* Rd, and the *C*-terminus (following residue 49), are the most variable parts of the sequences. *C. thermosaccharolyticum* Rd displays some noteworthy changes in its primary structure as compared with the proteins extracted from other anaerobic bacteria. A tryptophan residue occurs in position 4, whereas either phenylalanine or tyrosine are present in the other Rds. Only in the second half of *P. oleovorans* Rd [14] does a tryptophan residue occur in a position corresponding to the fourth position of *C. thermosaccharolyticum* Rd. The threonine residue in position 21 is unique to *C. thermosaccharolyticum* Rd, since in all other sequences of comparable length this position is occupied by an aspartic acid residue (Fig. 2). The presence of two consecutive proline residues in positions 25 and 26 is also a specific feature of *C. thermosaccharolyticum* Rd. An aspartic acid residue in position 41 is not found for any Rd isolated from anaerobes (Fig. 2), but is present in the corresponding position in *P. oleovorans* Rd, as well as in the putative product of the *alkF* gene [14].

The sequence of *C. thermosaccharolyticum* Rd is the first obtained from a thermophilic organism. There are some indications that this protein may be more stable than its

	1	10	20	30	40	50							
CTh Rd	MEKWQCTVCGYIYDPEVGDPTQNIPPGTKFEDLPDDWVCPDCGVGKDQFEKI												
CP Rd	MKKYTCTVCGYIYDPEDGDDGPNPGTDFKIDIPDDWVCPDCGVGKDEFEEVEE												
Cpe Rd	MKKFICDVCGYIYDPAVGDDPDNGVEPGTEFKIDIPDDWVCPDCGVGKDSQFSETEE												
BM Rd	MQKYVCDICGYVYDPAVGDDPDNGVAPGTAFADLPEDWVCPCEGVSKDEFESPEA												
PA Rd	MQKFECTLCGYIYDPALVGPDPDQDG.AFEDVSENWVCPDCGAGKEDFEVYED												
ME Rd	MDKYECISICGYIYDEAEGD.DGNVAAGTKFADLPADWVCPCTGADKDAFVKMD												
DG Rd	MDIYVCTVCGYIYDPAKGDDPSGIKPGTKFEDLPDDWVCPDCGVGKDAFEKQ												
DV Rd	MKKYVCTVCGYIYDPAEGDDPDNGVKGPTSFDDLPADWVCPVCGAPKSEFEAA												
DD Rd	MQKYVNCVCGYIYDPAEHDNVP.....FDQLPDDWCCPVCVGVSKDQFSPA												
Chl Rd	MQKYVCSVCGYIYDPAEGDDPIDPGTGFEDLPEDWVCPVCGVDKDLFEPES												
	1	10	20	30	40	50							
consensus	M	KY	C	VCGY	YDPA	GDPD	PGT	F	DLP	DWVCP	CGV	KD	FE
	*		*	***	**		*		*	*	*	*	*

Fig. 2. Alignment of the ten Rd sequences known to date

The sequences from *C. thermosaccharolyticum* (CTh), *C. pasteurianum* (CPA [3]), *C. perfringens* (Cpe [8]), *Butyrivibrio methylotrophicum* (BM, [9]), *Peptococcus aerogenes* (PA, [1]), *Megaspheira elsdenii* (ME [2]), *D. gigas* (DG, [4]), *D. vulgaris* (DV [7]), *D. desulfuricans* (DD [5]) and *Chlorobium thiosulfatophilum* (Chl [6]) were aligned using MULTALIN [22] with the unit matrix and a gap penalty of 1. The consensus sequence shows those residues which occur in more than half of the aligned sequences. The starred residues below the consensus are conserved in all of the ten sequences.

mesophilic counterparts: the preparation of the apoprotein of *C. thermosaccharolyticum* Rd (treatment with 2 M-HCl at 95 °C for 5 min) appeared to be less straightforward than for those from mesophilic organisms (J. Meyer & J.-M. Moulis, unpublished work). This observation is consistent with a previous report showing that an Rd from the thermophilic sulphate reducer *Thermodesulfobacterium commune* is more stable than that from the mesophile *Desulfovibrio gigas* [23]. However, Rds are intrinsically very stable [23], and it is therefore not certain that those extracted from thermophilic organisms will be found to be considerably more stable than those from mesophilic organisms. These questions, as well as others discussed above, will be addressed more thoroughly after completion of crystallographic investigations.

X-ray precession photos of the prismatic crystal form (see the Materials and methods section) show symmetry consistent with space group $P2_12_12_1$. The cell parameters, determined from the film, are: $a = 2.73$ nm, $b = 2.98$ nm and $c = 6.49$ nm. The unit cell volume calculated from these parameters is 52.799 nm³. By using the molecular mass of 5988 Da for one molecule per asymmetric unit, the ratio of crystal cell volume to molecular mass, V_M , is 0.0022 nm³/Da. This value is in the normal range for protein crystals [24]. Although the crystals tend to develop inclusion problems as they grow large, it is possible to select single crystal fragments as large as 0.1 mm $\times$ 0.2 mm $\times$ 0.5 mm. These should be sufficient for obtaining a resolution of 0.15 nm.

L.C.S. was supported by funds from the Laboratoire d'Ingénierie des Protéines, Commissariat à l'Energie Atomique. We thank Dr. R. Wade for helpful discussions, encouragement and support.

REFERENCES

1. Bachmayer, H., Benson, A. M., Yasunobu, K. T., Garrard, W. T. & Whiteley, H. R. (1968) *Biochemistry* **7**, 986–996
2. Bachmayer, H., Yasunobu, K. T., Peel, J. L. & Mayhew, S. (1968) *J. Biol. Chem.* **243**, 1022–1030
3. Yasunobu, K. T. & Tanaka, M. (1973) in *Iron-Sulfur Proteins* (Lovenberg, W., ed.), vol. 2, pp. 27–130, Academic Press, New York
4. Bruschi, M. (1976) *Biochem. Biophys. Res. Commun.* **70**, 615–621
5. Hornel, S., Walsh, K. A., Prickril, B. C., Titani, K., LeGall, J. & Sieker, L. C. (1986) *FEBS Lett.* **201**, 147–150
6. Woolley, K. J. & Meyer, T. E. (1987) *Eur. J. Biochem.* **163**, 161–166
7. Voordouw, G. (1988) *Gene* **69**, 75–83
8. Seki, Y., Seki, S., Satoh, M., Ikeda, A. & Ishimoto, M. (1989) *J. Biochem. (Tokyo)* **106**, 336–341
9. Saeki, K., Yao, Y., Wakabayashi, S., Shen, G.-J., Zeikus, J. G. & Matsubara, H. (1989) *J. Biochem. (Tokyo)* **106**, 656–662
10. Watenpugh, K. D., Sieker, L. C. & Jensen, L. H. (1979) *J. Mol. Biol.* **131**, 509–522
11. Adman, E. T., Sieker, L. C., Jensen, L. H., Bruschi, M. & LeGall, J. (1977) *J. Mol. Biol.* **112**, 113–120
12. Sieker, L. C., Stenkamp, R. E., Jensen, L. H., Prickril, B. & LeGall, J. (1986) *FEBS Lett.* **208**, 73–76
13. Frey, M., Sieker, L. C., Payan, F., Haser, R., Bruschi, M., Pepe, G. & LeGall, J. (1987) *J. Mol. Biol.* **197**, 525–541
14. Kok, M., Oldenhuis, R., van der Linden, M. P. G., Meulenberg, C. H. C., Kingma, J. & Witholt, B. (1989) *J. Biol. Chem.* **264**, 5442–5451
15. Wilder, M., Valentine, R. C. & Akagi, J. M. (1963) *J. Bacteriol.* **86**, 861–865
16. Crestfield, A. M., Moore, S. & Stein, W. H. (1963) *J. Biol. Chem.* **238**, 622–627
17. Covey, T. R., Bonner, R. F., Shushan, B. I. & Henion, J. D. (1988) *Rapid Commun. Mass Spectrom.* **2**, 249–256
18. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F. & Whitehouse, C. M. (1989) *Science* **246**, 64–71
19. Sarges, R. & Witkop, B. (1965) *J. Am. Chem. Soc.* **87**, 2011–2020
20. Loo, J. A., Udseth, H. R. & Smith, R. D. (1989) *Anal. Biochem.* **179**, 404–412
21. Aplin, R. T., Baldwin, J. E., Fujishima, Y., Schofield, C. J., Green, B. N. & Jarvis, S. A. (1990) *FEBS Lett.* **264**, 215–217
22. Corpet, F. (1988) *Nucleic Acids Res.* **16**, 10881–10890
23. Papavassiliou, P. & Hatchikian, E. C. (1985) *Biochim. Biophys. Acta* **810**, 1–11
24. Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491–497

Received 10 July 1990/20 August 1990; accepted 3 September 1990