N-TERMINAL AMINO ACID SEQUENCES OF *AZOTOBACTER VINELANDII* AND *RHODOSPIRILLUM RUBRUM* FLAVODOXINS†

Martha L. MacKnight*, William R. Gray**, and Gordon Tollin*

*Department of Chemistry, University of Arizona
Tucson, Arizona 85721

**Department of Biology, University of Utah
Salt Lake City, Utah 84112

Received June 4,1974

SUMMARY

The amino-terminal sequences of *A. vinelandii* and *R. rubrum* flavodoxins have been determined. These proteins have larger molecular weights (mw≃23,000) than do other flavodoxins for which sequences have previously been reported (mw≃14,000). A comparison of these results with those obtained in other laboratories for *P. elsdenii*, *C. pasteurianum*, *C. MP* and *D. vulgaris* flavodoxins shows that significant homology exists in the FMN phosphate binding site region in all of these flavoproteins. Furthermore, the sequences reveal relationships among these various proteins which correlate with properties such as riboflavin binding ability and visible circular dichroism spectra.

INTRODUCTION

Flavodoxins are useful for studies of flavoprotein structure/function relationships because of their relative simplicity; e.g., low molecular weights, single polypeptide chains, and one FMN molecule bound per protein molecule. Particular interest has been stimulated by the recent publication of crystal structures for *Desulfovibrio vulgaris* (*D.v.*, 1) and *Clostridium MP* (*C.MP*, 2) flavodoxins. This work indicates that the binding site of the ribityl phosphate of the FMN is located near the amino terminus in these proteins. It is thus significant that the amino acid sequences of the flavodoxins from *D.v.* (3), *C.MP* (Data of K. Yasunobu quoted in (2)), *Peptostreptococcus elsdenii* (*P.e.*, 4), and *Clostridium pasteurianum* (*C.p.*, 5) reveal marked homology in this region. Residues corresponding to Ser-10, Thr-12, Thr-15, and Asn-14 in the *D.v.* flavodoxin are strictly conserved in all of the published flavodoxin sequences. In the *D.v.* and *C.MP* flavodoxins, the first three of these residues

† Supported in part by grants from the National Institutes of Health (1 RO1-AM 15057), from the National Science Foundation (GB 20983); by National Science Foundation Postdoctoral Fellowship stipends to MLM; and by a USPHS Career Development Award to WRG.

serve to hydrogen bond to the FMN phosphate, while the last residue forms a hydrogen bond to the 4'-hydroxyl group of the ribityl side chain (2,6).

On the basis of properties such as ability to bind riboflavin, and visible circular dichroism spectra, the flavodoxins have been separated into two classes, one of which contains the *P.e.*, *C.p.*, and *C.MP* flavodoxins, and the other of which contains the flavodoxins from *D.v.*, *Rhodospirillum rubrum* (*R.r.*) and *Azotobacter vinelandii* (*A.v.*, Shethna flavoprotein)(7-9). The last two proteins are larger than the others, having molecular weights of 23,000 compared with 14,000. The *A.v.* protein, although it has flavodoxin activity (10), is not induced by iron deprivation and probably does not replace a ferrodoxin, although it functions as part of the nitrogen-fixation system (11). In view of these relationships, it is of interest to compare the primary structures of these various flavodoxins. We present here the N-terminal sequences of the *A.v.* and *R.r.* flavodoxins. Work on the complete sequence of the *A.v.* protein is in progress. These studies show that considerable similarities exist in the ribityl phosphate binding site of all of the flavodoxins. This is undoubtedly significant in terms of the function of these flavoproteins, inasmuch as it has been previously shown that the presence of the ribityl phosphate of the flavin cofactor is essential for the stabilization of the half-reduced form of these proteins (8).

MATERIALS AND METHODS

A. vinelandii (strain Wisconsin "O") was grown and the flavodoxin was isolated using previously published methods (12, 13). *R. rubrum* (strain 2.1.1, Van Niel) was grown on an iron-deficient medium and the flavodoxin was isolated as described by Cusanovich and Edmondson (14).

The amino-terminal sequences of the intact flavodoxins were determined using a Beckman model 890 automatic protein sequencer. As an internal standard, BTH*-norleucine was added to each of the tubes in the sequencer. Following conversion of the thiazolinone derivatives to the ethyl acetate soluble PTH-amino acids (15), the more volatile PTH-amino acids were identified by gas-liquid chromatography using a Packard model 7400 Gas Chromatographic System as described previously (16). Confirmation of the PTH-amino acids was accomplished by HI hydrolysis (17) to yield the free amino acids, which were identified by the Beckman 121 automatic amino acid analyzer. All analyses were quantitative, with corrections applied for differential hydrolysis of the various PTH-amino acids.

In addition to the intact protein, the mixture of peptides obtained from

* Abbreviations used: BTH- , 2-benzyl-5-thiohydantoin; PTH- 2,phenyl-5-thiohydantoin.

the cyanogen bromide (CNBr) cleavage of the *A.v.* flavodoxin in 70% formic acid (18) was analyzed on the sequencer.

RESULTS

The amino-terminal sequences of the *A.v.* and *R.r.* flavodoxins are presented in Table I, along with parts of the known sequences of the *D.v.* , *P.e.*, *C.MP*, and *C.p.* flavodoxins. The numbering, based on *D.v.* flavodoxin, and the gaps used to align the sequences, are the same as published by Dubourdieu *et al.* (3). The first 36 amino acid residues from the N-terminal end of the *R.r.* flavodoxin were identified using the sequencer. In the case of the *A.v.* protein, it was possible to extend the sequence determination by the use of CNBr cleavage. The sequencer analysis of the untreated protein revealed a methionyl residue at position 30. Since the amino acid composition of the *A.v.* flavodoxin (19, 20) indicated the presence of only one to two methionines, CNBr cleavage was performed, and the resulting unseparated mixture of peptides was analyzed on the sequencer. On the basis of the sequence determination of the untreated protein, and despite the low(10%) yield of cleavage at the Met-Ser bond, it was possible to distinguish the CNBr peptide sequence from that of the N-terminus of the intact protein, and thus to establish the overlap of the sequence at Met-30.

DISCUSSION

As has been found to be the case for the other flavodoxins (3, 21), considerable homology is evident in the amino-terminal regions of the *A.v.* and *R.r.* flavodoxins. Although it was known from circular dichroic and redox properties (7-9) that *A.v.* flavodoxin was similar to the other flavodoxins, the amino-terminal sequence data are consistent with the recent demonstration of flavodoxin activity by this enzyme (10) and add to the justification that it be properly termed a flavodoxin. The *A.v.* and *R.r.* flavodoxins, because of their molecular weights of about 23,000, have an additional length of peptide chain (approximately 80 residues) as compared with the *C.p.*, *P.e.*, and *D.v.* flavodoxins (molecular weights about 14,000). Because of the homology noted in the sequences of Table I, it can be concluded that these extra residues do not occur at the N-terminus unless the additional length of peptide chain arises from a duplication of the original N-terminal sequence.

Within the highly conserved homologous region (positions 8 through 22), five residues are invariant in all of the flavodoxins in Table I. Two of these residues, Ser in position 10 and Thr in position 15, have been shown to function in both the *D.v.* and *C.MP* flavodoxins to hydrogen bond to a single ribi-

Table I Amino-terminal sequences of flavodoxins.

	1			5						10						15						20					
C. pasteurianum	Met	-	Lys	Val	Asn	Ile	Ile	Tyr	Trp	Ser	Gly	Thr	Gly	Asn	Thr	Glu	Ala	Met	Ala	Asn	Glu	Ile	Glu	Ala			
C. MP	Met	-	Lys	-	-	Ile	Val	Tyr	Trp	Ser	Gly	Thr	Gly	Asn	Thr	Glu	Lys	Met	Ala	Glu	Leu	Ile	Ala	Lys			
P. elsdenii	Met	-	-	Val	Glu	Ile	Val	Tyr	Trp	Ser	Gly	Thr	Gly	Asn	Thr	Glu	Ala	Met	Ala	Lys	Leu	Ile	Ala	Glu			
D. vulgaris	Met	Pro	Lys	Ala	Leu	Ile	Val	Tyr	Gly	Ser	Thr	Thr	Gly	Asn	Thr	Glu	Tyr	Thr	Ala	Glu	Thr	Ile	Ala	Arg			
R. rubrum			Gly	Thr	Thr	Val	Ile	Tyr	Gly	Ser	Asp	Gly	Gly	Thr	Thr	Glu	Gly	Val	Ala	Lys	Arg	Ile	Ala(A/S)					
A. vinelandii		Ala	Lys	Ile	Gly	Leu	Phe	Phe	Gly	Ser	Asn	Thr	Gly	Lys	Thr	Arg	Lys	Val	Ala	Lys	Ser	Ile	Lys	Lys				

	25				30					35					40					45							
C.p.	Ala	Val	Lys	Ala	Ala	Gly	Ala	Asp	Val	Glu	Ser	Val	Arg	Phe	Glu	Asp	Thr	Asn	Val	Asp	Asp	Val					
C.MP	Gly	Ile	Ile	Glu	Ser	Gly	Lys	Asp	Val	Asn	Thr	Ile	Asn	Val	Ser	Asp	Val	Asn	Ile	Asp	Glu	Leu					
P.e.	Gly	Ala	Gln	Glu	Lys	Gly	Ala	Gln	Val	Lys	Leu	Leu	Asn	Val	Ser	Asp	Ala	Lys	Glu	Asp	Asp	Val					
D.v.	Glu	Leu	Ala	Asx	Ala	Gly	Tyr	Glu	Val	Asp	Ser	Arg	Asp	Ala	Ala	Ser	Val	Glu	Ala	Gly	Gly	Leu					
R.r.	Pro	Leu	Glx	Ala	Lys	Val	Val	Asx	Ile	Lys(Val)Ala	Thr(Thr)																
A.v.	Arg	Phe	Asx	Asx(Glx)Thr	Met(Ser)Asp	Ala	Leu	Asx	Val(Asx)Arg	Val	xxx	Ala(Glx)(Asx)Phe	Ala														

tyl phosphate oxygen in the FMN cofactor (2,6). Thus it is not unreasonable to suggest that these residues play an homologous role in the other flavodoxins. A second phosphate oxygen in the *D.v.* and *C.MP* proteins is hydrogen bonded to Thr-12 and to the peptide -NH of Asn-14. It is interesting that Thr-12 is replaced in the *R.r.* flavodoxin by a glycyl residue although it is invariant in the other flavodoxins. Asn-14 in the *D.v.* and *C.MP* flavodoxins also forms a hydrogen bond through its amide to the 4'-hydroxyl of the ribityl side-chain. Though it is replaced by Thr and Lys in the *R.r.* and *A.v.* proteins, respectively, these residues can hydrogen bond to this or another side-chain hydroxyl group. It is also possible that the positively charged Lys of the *A.v.* flavodoxin might interact electrostatically with the phosphate, and perhaps be involved in stabilizing the semiquinone radical.

A third invariant residue, Gly-13, in the conserved region is intimately involved in the phosphate binding site and is probably necessary for the proper alignment of the peptide backbone in this region. In contrast, no clear chemical function (such as hydrogen bonding) is apparent for Ala-19 and Ile-22, the remaining two of the five invariant residues. These residues are spatially removed from the phosphate binding region in the *C.MP* and *D.v.* proteins; however, they could be important for the proper association of the N-terminal chain with the other regions of the protein involved in side-chain binding. For example, Ser-54 and Ser-89 in the *C.MP* protein, and Ser-58 in the *D.v.* protein also form hydrogen bonds to the phosphate.

The *A.v.* flavodoxin is unique in being very basic in the amino-terminal region, with five Lys and two Arg residues between positions 14 and 25 in Table I. This region is removed from the phosphate-binding site in the *D.v.* and *C.MP* flavodoxins, yet the presence of basic groups could affect the flavin environment in the *Azotobacter* protein. It is possibly significant in this context that the latter forms the most stable half-reduced form of all of this group of flavoproteins (8).

Published sequence data on the small flavodoxins, as well as crystallographic structures, leave little doubt that they are homologous proteins derived from a common ancestor. The data from the present work, though less extensive, reveal some interesting possibilities concerning the relationship of the other two proteins (*R.r.* and *A.v.*) to this group. Pairwise comparisons were made among the proteins to assess the degree of similarity. The results are shown in Table II, listing separate figures for the highly conserved region (residues 8-22), and for the less constrained region following (3-7 and 23-38).

The *R.r.* protein appears to be related to the smaller proteins in both areas, and we may tentatively conclude that it is homologous with them. The *A.v.* protein, however, does not show this clear-cut picture. It shows a simi-

Table II Comparisons of flavodoxin sequences.

	C.p.	*C.MP*	*P.e.*	*D.v.*	*R.r.*	*A.v.*
C.p.	-	.40(.95)	.35(1.00)	.70(1.13)	.80(1.00)	.90(1.44)
C.MP	.40(.95)	-	.20(.90)	.53(.80)	.67(1.33)	.67(1.28)
P.e.	.35(1.00)	.20(.90)	-	.65(1.19)	.65(1.06)	.80(1.50)
D.v.	.70(1.13)	.53(.80)	.65(1.19)	-	.90(1.38)	.85(1.38)
R.r.	.80(1.00)	.67(1.33)	.65(1.06)	.90(1.38)	-	.85(1.44)
A.v.	.90(1.44)	.67(1.28)	.80(1.38)	.85(1.38)	.85(1.44)	-

Table II Values tabulated are the minimum mutations per codon for pairwise comparisons of the sequences in Table I. Main entries are for the conserved regions (residues 8-22). Values in parentheses are for the non-conserved regions (3-7 plus 23-38); these values are slightly above those for the comparisons of whole proteins whose complete sequences are known. Proteins showing no significant relationship to each other usually have values of 1.4-1.5 mutations/codon.

Deletions were counted as one mutation; Asx or Glx to any amino acid was counted as the lesser of Asp or Asn to that amino acid, etc.

larity to the other proteins (and to that of *R.r.*) only in the extremely conservative region. Elsewhere the relationship is random. That the *R.r.* flavodoxin is more closely related to the others than is the *A.v.* protein is also indicated by similar trends in the visible flavin ORD and CD bands (7), the redox properties of the semiquinone forms of the proteins (8), the apoprotein fluorescence maxima, riboflavin equilibrium binding constants, and the fluorescence quenching rate constants upon FMN binding (9). In addition, the *A.v.* protein is unique among the flavodoxins in that it does not replace a ferredoxin, and is not induced by iron-deficiency in the bacterial growth medium.

In the conserved region, the *C.p.*, *P.e.*, and *C.MP* flavodoxins are clearly the most closely related among this group of flavodoxins. The *D.v.*, *R.r.*, and

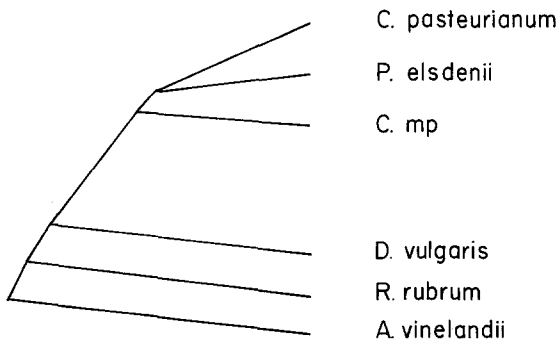

Fig. 1 Tentative scheme for evolution of flavodoxins, based on data of Table I. The large mutation distances leave a good deal of uncertainty, especially in the very early branch points. They also render any time-scale strongly skewed, so the early divergences probably occurred much longer ago than appears at first glance.

A.v. proteins are about equally remote from the *C.p.-C.MP-P.e.* line, and are even more remote from one another. This is interesting in view of the suggested grouping of the flavodoxins into two sub-classes (*P.e.*, *C.MP* and *C.p.* *vs.* *D.v.*, *R.r.* and *A.v.*) based on physicochemical properties (7-9). The sequence data indicate that the *D.v.*, *R.r.* and *A.v.* proteins do not form a closely related group as do the *C.p.*, *C.MP* and *P.e.* flavodoxins. In view of the data of Table II, therefore, we should seriously consider some alternative relationships:

(1) All six flavodoxins share a common ancestor, and the lack of an observed relatedness of the *A.v.* protein is due merely to the long period since divergence. The most conservative scheme for the proteins' evolution is shown in Fig. 1.

(2) The *A.v.* protein is of fully independent origin; the observed similarity in the amino-terminal region is due to convergent evolution because of the stringent requirements of the flavin-binding site.

(3) The *A.v.* protein is largely of independent origin, and had a flavin-binding site "grafted on" by a genetic crossing-over event.

It is premature to try to choose among these alternatives. Sequence analysis is being continued, and the complete sequence, plus comparison of crystal structures, should reveal much about this important group of proteins.

ACKNOWLEDGMENTS

We are grateful to Dr. M. A. Cusanovich for his assistance in the purification of *R. rubrum* flavodoxin, to E. Hostetter for technical assistance and

to M. A. McNulty, R. P. Mecham, and R. Coltrain for their assistance in sequence determinations.

REFERENCES

1. Watenpaugh, K.D., Sieker, L.C., Jensen, L.H., LeGall, J., and Dubourdieu, M., *Proc. N.A.S. 69*, 3185 (1972).
2. Burnett, R.M., Darling, G.D., Kendall, D.S., LeQuesne, M.E., Mayhew, S.G., Smith, W.W., and Ludwig, M.L., *J. Biol. Chem.*, in press.
3. Dubourdieu, M., LeGall, J., and Fox, J.L., *Biochem. Biophys. Res. Commun. 52*, 1418 (1973).
4. Tanaka, M., Haniu, M., Yasunobu, K.T., Mayhew, S., and Massey, V., *Biochem. Biophys. Res. Commun., 44*, 886 (1971).
5. Fox, J.L., Smith, S.S., and Brown, J.R., *Zeit. für Naturf. 27b*, 1096 (1972).
6. Watenpaugh, K.D., Sieker, L.C., Jensen, L.H., *Proc. N.A.S. 70*, 3857 (1973).
7. Edmondson, D.E., and Tollin, G., *Biochemistry 10*, 113 (1971).
8. Edmondson, D.E., and Tollin, G., *Biochemistry 10*, 133 (1971).
9. D'Anna, J.A., and Tollin, G., *Biochemistry 11*, 1073 (1972).
10. VanLin, B., and Bothe, H., *Arch. Mikrobiol. 82*, 155 (1972).
11. Yates, M.G., *FEBS Letters 27*, 63 (1972).
12. Hinkson, J.W., and Bulen, W.A., *J. Biol. Chem. 242*, 3345 (1967).
13. Edmondson, D.E., Ph.D. Dissertation, University of Arizona (1970).
14. Cusanovich, M.A., and Edmondson, D.E., *Biochem. Biophys. Res. Commun. 45*, 327 (1971).
15. Edman, P., and Begg, G., *Eur. J. Biochem. 1*, 80 (1967).
16. Foster, J.A., Bruenger, E., Gray, W.R., and Sandberg, L., *J. Biol. Chem. 248*, 2876 (1973).
17. Smithies, O., Gibson, D., Fanning, E.M., Goodfliesh, R.M., Gilman, J.G., and Ballantyne, D.L., *Biochemistry 10*, 4912 (1971).
18. Gross, E., *Methods in Enzymol. 11*, 238 (1967).
19. Edmondson, D.E., and Tollin, G., *Biochemistry 10*, 124 (1971).
20. MacKnight, M.L., and Gray, W.R., unpublished results.
21. Tanaka, M., Haniu, M., Matsueda, G., Yasunobu, K.T., Mayhew, G., and Massey, V., *Biochemistry 10*, 3041 (1971).