A COMPARISON OF THE AMINO-TERMINAL SEQUENCES OF SEVERAL CARBOHYDRATE BINDING PROTEINS FROM *ESCHERICHIA COLI* AND *SALMONELLA TYPHIMURIUM*

Robert W. HOGG, Hikaru ISIHARA, Mark A. HERMODSON*, Daniel KOSHLAND Jr., John W. JACOBS and Ralph A. BRADSHAW

Department of Microbiology, Case Western Reserve University, Cleveland, Ohio 44106; Division of Medical Genetics, University of Washington, Seattle, Washington 98195; Department of Biochemistry, University of California, Berkeley California 94720; and Department of Biochemistry, Washington University, School of Medicine, St. Louis, Missouri 63110, USA

Received 23 June 1977

1. Introduction

The identification of periplasmic binding proteins as components of bacterial transport systems has provided a means to study the entry of small molecules into cells at the molecular level. A number of binding proteins have been identified with specificities for amino acids, carbohydrates, inorganic ions, nucleic acids and vitamins [1–9]. The exact manner in which such molecules function in the transport process is unknown, however they are assumed to serve as recognition units which impart specificity and, possibly, alter the kinetic parameters of membrane bound units which translocate the substrate from the environment to the cytoplasm [10].

A number of properties appear to be common to periplasmic binding proteins as a group. Notably, they are all released by osmiotic shock, have molecular weights of about 30 000 and recognized their respective ligands with affinities of approximately 1 μm molar [7,11]. Two binding proteins, the ribose and galactose receptor of *Salmonella typhimurium* have been shown to interact with the same signalling component in the chemotactic system of that bacterium [12]. Thus it would appear that a binding region in the three-dimensional structure of the two proteins is highly similar. Nevertheless, it has been found that there is little cross reaction between the proteins [13]. Antibody directed against the galactose receptor of *Salmonella typhimurium* cross reacts with the galactose receptor of *E. coli*, but does not cross react with the ribose receptor of *Salmonella*. On the other hand, antibody directed against the arabinose binding protein of *E. coli* cross-reacts with galactose binding protein from the same source [13].

The fact that these proteins carry out similar functions in similar environments and yet have clearly different specificities makes their primary amino acid sequence of particular interest. We have, therefore, initiated a study on the amino acid sequences of various carbohydrate binding proteins to determine the degree of similarity and the degree of difference for these proteins carrying out similar functions. The comparison of the arabinose and galactose binding protein of *E. coli* and the arabinose and galactose and ribose binding proteins of *S. typhimurium* is described.

2. Materials and methods

The arabinose and galactose binding proteins (ABP and GBP) from *E. coli* and *S. typhimurium* were purified as described by Parsons and Hogg [14]. The ribose binding protein (RBP) was purified as described by Aksamit and Koshland [15]. Amino terminal sequence analyses of the arabinose and galactose binding proteins were determined on reduced and

*Present address: Department of Biochemistry, Purdue University, West Lafayette, Indiana

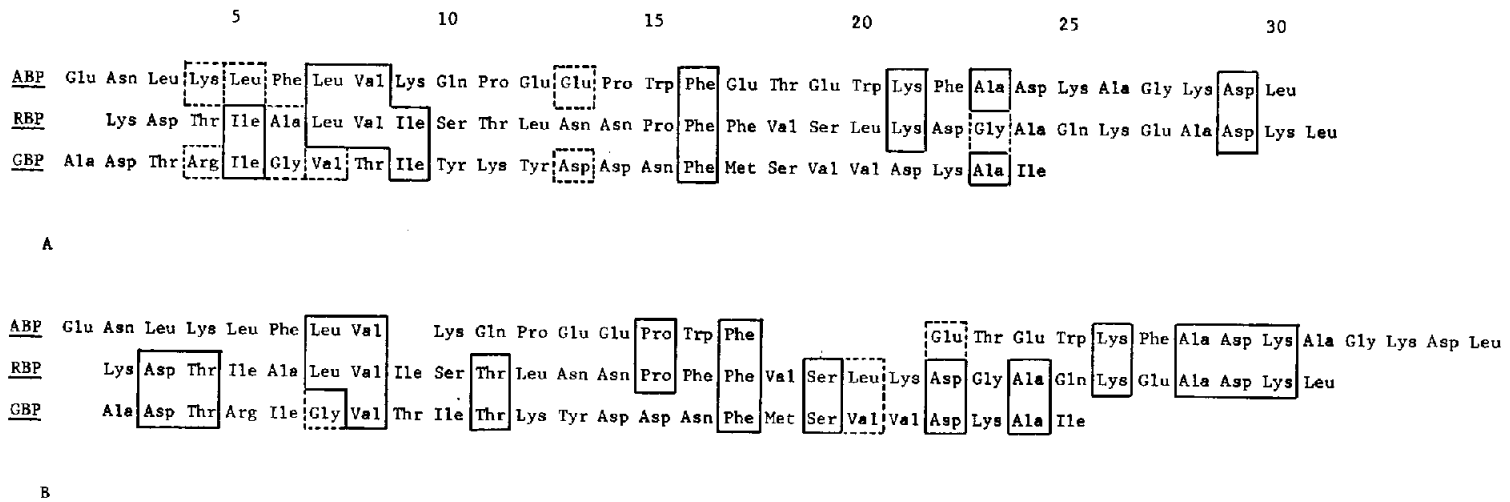

Fig.1. Amino-terminal sequences of the arabinose, galactose and ribose binding proteins.

S-pyridylethylated protein by the procedure of Edman and Begg [17] as modified by Hermodson et al. [18]. The PTH amino acids were identified by gas or thin layer chromatography with the exceptions of PTH-His, PTH-Arg and PTH-Cys which were identified by spot tests on paper or thin-layer chromatography.

3. Results and discussion

The amino terminal sequences of three carbohydrate binding proteins are presented in fig.1A and B in two different alignments of the primary sequences. As can be seen in fig.1A, there are five identities in 30 residues when comparing the ribose binding protein with the arabinose binding protein and three identities in 24 residues comparing the ribose binding protein with the galactose binding protein. These are well above statistical error although only small portions of the sequence have been examined. If one now allows deletions in the sequence, the number of identities between the ribose binding protein and the arabinose binding protein as shown in fig.1B is eight identities in 31 residues and eight identities in 25 residues for the comparison of the ribose binding protein with the galactose binding protein. Again this is above statistical chance.

Other than these identities the sequences bear little similarity to each other, thus suggesting a degree of difference and a degree of similarity which may be intriguing. It might be expected that those portions of the proteins molecule which carry out similar functions, such as interacting with the transport machinery may be more similar than the portions at the amino-terminal end of the molecule.

The arabinose binding protein, isolated from *S. typhimurium* and *E. coli* are identical with respect to the amino terminal eighteen residues. The complete amino acid sequence of the arabinose binding protein of *E. coli* has been determined [19]. and when compared to a preliminary amino acid analysis of Salmonella arabinose binding protein a significant difference is apparent. It can therefore be concluded that differences in the amino acid sequence do occur elsewhere in the molecule. The galactose binding protein isolated from Salmonella and *E. coli* are identical within the first twenty amino-terminal residues.

Only ABP coli has been completely sequenced and further comparisons must await the completion of the sequence analysis of the ribose and galactose binding proteins. Completion of the sequence of other binding proteins (e.g., leucine, histidine or an ion-binding protein) would allow a broader interpretation of any homologous areas.

Acknowledgments

RWH was supported by United States Public Health Service Grant 1-R01-AM-13791 from the National Institute of Arthritis and Metabolic Diseases. MAH was supported by United States Public Health Service Grant GM 15253 from the National Institute of General Medical Sciences. DEK was supported by National Institutes of Health Grant AM9765. RAB was supported by National Institutes of Health Grant NS 10229.

References

[1] Pardee, A. B. (1966) J. Biol. Chem. 241, 5886.
[2] Pardee, A. B. (1968) Science 162, 632.
[3] Medveczky, N. and Rosenberg, H. (1970) Biochim. Biophys. Acta 211, 158.
[4] Anraku, Y. (1968) J. Biol. Chem. 243, 3116.
[5] Anraku, Y. (1968) J. Biol. Chem. 243, 3123.
[6] Anraku, Y. (1968) J. Biol. Chem. 3128.
[7] Boos, W. (1974) Annu. Rev. Biochem. 43, 123.
[8] Penrose, W. R., Nichoalds, G. E., Piperno, J. R. and Oxender, D. L. (1968) J. Biol. Chem. 243, 5921.
[9] Ames, G. F. and Lever, J. E. (1972) J. Biol. Chem. 247, 4309.
[10] Robbins, A. R., Guzman, R. and Rotman, B. (1976) J. Biol. Chem. 251, 3112–3116.
[11] Oxender, D. L. (1972) Ann. Rev. Biochem. 41, 777–814.
[12] Strange, P. G. and Koshland, Jr., D. E. (1975) Proc. Natl. Acad. Sci. USA 73, 762.
[13] Zukin, R. S., Strange, P. G., Heavey, L. R. and Koshland, Jr., D. E. (1977) 16, 381.
[14] Parsons, R. G. and Hogg, R. W. (1974) J. Biol. Chem. 249, 3602–3607.
[15] Aksamit, R. and Koshland, Jr., D. E. (1972) Biochem. Biophys. Res. Comm. 48, 1348–1353.
[16] Jacobs, J. W. and Niall, H. D. (1975) J. Biol. Chem. 250, 3629–3636.
[17] Edman, P. and Begg, (1967) Eur. J. Biochem. 1, 80–90.
[18] Hermodson, M. A., Ericsson, L. H., Titani, K., Neurath, H. and Walsh, K. A. (1972) Biochemistry 11, 4493–4502.
[19] Hogg, R. W. and Hermodson, M. A. (1976) Fed. Proc. 35, 1171.