J. Mol. Evol. 10, 155–160 (1977)

Journal of Molecular Evolution

On the Cysteine and Cystine Content of Proteins

Differences between Intracellular and Extracellular Proteins

Robert C. Fahey, John. S. Hunt and Gayle C. Windham

Department of Chemistry, University of California, San Diego, La Jolla, CA 92093, USA

Summary. Analysis of published data on the cysteine and half-cystine content of proteins indicates that most intracellular proteins may be classified as sulfhydryl proteins (those containing cysteine but little or no half-cystine) and that such sulfhydryl proteins have a low cysteine content. The mean cysteine content found for 32 intracellular mammalian proteins was 1.6 % and intracellular proteins of many bacteria have similar or lower values. Extracellular mammalian proteins are primarily disulfide proteins (i containing half-cystine but little or no cysteine) and have a high half-cystine co , the mean value found for some 34 extracellular mammalian proteins being 4.1 %. This is contrasted with many of the extracellular proteins from facultative bacteria which are cyst(e)ine-free proteins, being lacking in both cysteine and half-cystine. These and related observations are interpreted in terms of the evolution of life in a reducing atmosphere and the subsequent transition to an oxidizing environment. It is suggested that disulfide proteins evolved primarily after the accumulation of oxygen in the atmosphere.

Key words: Cysteine – Cystine – Protein – Evolution.

Over the course of the past decade considerable interest has been focussed upon the statistical and theoretical analysis of the amino acid composition of proteins with special attention being given to the relationship between the observed content and that predicted on the basis of the genetic code (Smith, 1966; Kimura, 1968; King and Jukes 1969; Ohta and Kimura, 1971; Reeck and Fisher, 1973; Gatlin, 1974; Jukes et al., 1975; Holmquist, 1975; Gatlin, 1976). Since 2 out of the 61 triplets which code for amino acids specify cysteine, random occurrence of the nucleic acid bases would predict that 3.3 % of the amino acids residues in proteins should be cysteine or its oxidized counterpart, half-cystine. The average observed values for cyst(e)ine (cysteine plus half-cystine) are similar to or lower than this number depending upon the specific proteins surveyed (Smith, 1966; Kimura, 1968; King and Jukes, 1969; Ohta and Kimura, 1971; Reeck and Fisher 1973; Jukes et al., 1975). However, such surveys have generally not differentiated between the sulfhydryl form of this amino acid, cysteine, and the disulfide form, cystine. The purpose of this com-

munication is to point out some important differences in the distribution pattern and in the average content of cysteine and half-cystine residues as found in intracellular and extracellular proteins.

Since they nearly always contain cysteine rather than half-cystine residues, it is possible to classify most proteins which function in an intracellular environment as sulfhydryl proteins. These sulfhydryl proteins are considered to be maintained in the reduced state by glutathione (Barron and Singer, 1943) which is itself reduced by NADPH through the action of glutathione reductase (Williams, 1976). Not included here are proteins which are stored in lysosomes or other subcellular organelles not known to contain glutathione, or which function extracellularly after release from the cell. With this exclusion, exceptions to the generalization are few and usually involve the occurrence of cystine only as a component of the disulfide-dithiol redox cycle which functions in the catalytic mechanism of enzymes such as glutathione reductase (Williams, 1976). To analyze the occurrence of cysteine in proteins we have examined some 32 mammalian intracellular proteins taken from the compilations of Jocelyn (1972) and Reeck (1970). Identical proteins from different species and very closely related proteins have been used only once, and only proteins having more than 50 amino acid residues have been included. In each case the sulfhydryl assignment and numerical value were checked in the original literature. It can be seen in the accompanying Figure that none of the intracellular proteins have a cysteine content greater than the 3.3 % predicted from the code and that the mean value of 1.6 % is only about half the predicted value. Thus, intracellular proteins have a low, but generally nonzero, cysteine content.

This low cysteine content is not a special property of intracellular proteins of mammalian origin. In fact, the intracellular proteins of many bacteria appear to have even less cysteine. Pollock and Richmond (1962) found the cyst(e)ine content of the total cell protein from some nine different species of bacteria to range from 0.7 – 1.6 %, the mean value being 1.1 ± 0.3 %. This accords well with the mean value of 1.07 ± 0.66 % obtained by Reeck and Fisher (1973) for 57 purified prokaryotic enzymes, although this sample includes some extracellular enzymes. Proteins from bacteria of the genus *Bacillus* appear to have an especially low cyst (e) ine content (Pollock and Richmond 1962; Bieber and Zimmermann, 1971). In some instances there is sufficient data to see these trends in a specific enzyme from different organisms. Thus, the cysteine content of malate dehydrogenase has been found to be 0 % when obtained from *Bacillus subtilis*, 0.8 % from *E. coli*, 1.25 % from *Neurospora crassa*, 1.24 % from chicken heart supernatant, and 2.1 % from chicken heart mitochondria (Murphey et al., 1967), the latter two values being generally typical of the animal enzymes (Banaszak and Bradshaw, 1975).

Why do intracellular proteins tend to have low but finite levels of cysteine? A plausible answer to this question can be derived from a consideration of the special reactivity properties of the thiol group. Thus, the thiol group can enter into more facile oxidation, substitution, addition, and metal-binding reactions under physiological conditions than can any other group found in the common amino acids (Friedman, 1973). Such reactivity makes thiol groups useful in catalytic processes of enzymes, e.g. glyceraldehyde-3-phosphate dehydrogenase (Harris and Waters, 1976), in binding of coenzymes, prosthetic groups, or activators, e.g. cytochrome c (Dickerson and

Timkovitch, 1975) and ferrodoxin (Malkin, 1973), and in regulatory reactions. Since it is likely that thiols were stable, relative to disulfides, under the reducing conditions of the primitive earth, the incorporation of this valuable group into proteins is understandable. However, the high reactivity of the thiol group also makes it a site for adverse reactions with heavy metal ions and with other thiol-reactive agents present in the environment. With the accumulation of oxygen in the atmosphere, molecular oxygen and compounds derived from it logically came to represent the major threat to protein thiol groups. The toxic effects of oxygen upon cells and the involvement of thiol groups in the mechanisms of oxygen toxicity have been thoroughly reviewed and discussed by Haugaard (1968). The level of occurrence of cysteine in proteins is then reasonably expected to represent a balance between the beneficial and the detrimental reactions of protein thiol groups, moderated by whatever mechanisms are present within the cell to protect against the adverse reactions.

The observation cited above that the proteins of bacilli have an unusually low cysteine content supports the latter conclusion. It was reported long ago by Miller and Stone (1938) that bacilli and cocci have very low levels of soluble thiol, considered to represent glutathione. Thus, in the absence of appreciable levels of glutathione or other protective thiol, the cysteine content of intracellular proteins appears to be restricted to a minimal level.

In contrast to intracellular sulfhydryl proteins, extracellular proteins of higher organisms can be classified as disulfide proteins. Since these proteins function in an environment unprotected from the effects of atmospheric oxygen and since thiols are

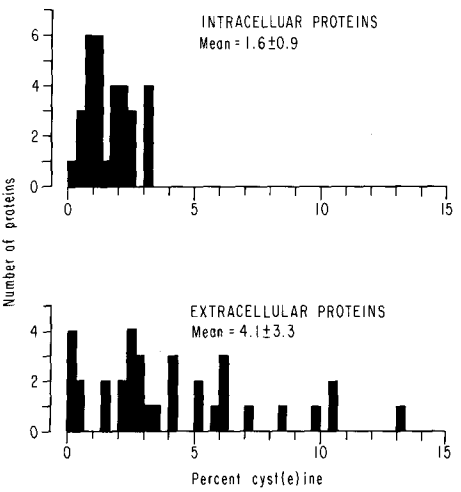

Fig. 1. Cyst(e)ine content of mammalian proteins, as % of total residues, with mean value and standard deviation. Proteins are identified in order of increasing cyst(e)ine content by the table number and reference letter from the tabulation by Jocelyn (1972) or by the reference number of the tabulation by Reeck (1970). Intracellular proteins: 96, 6-h, 8-h, 70, 71, 34, 6-q, 3k, 83, 9-1, 114, 151, 50, 77, 142, 5-h, 6-f, 8, 69, 76, 191, 150, 124, 148, 190, 5-i, 5-g, 78, 5-a, 202, 10, 82. Extracellular proteins: 40, 122, 123, 143, 32, 121, 122, 146, 86, 7-l, 11-d, 102, 7-b, 17, 74, 84, 10-j, 115, 64, 110, 7-q, 87, 62, 196, 189, 11-i, 67, 173, 29, 37, 7-g, 10-f, 195, 97

unstable relative to disulfides in the presence of oxygen, it is expected that half-cystine should be favored relative to cysteine. The latter should then occur only when essential to the protein function or when specifically protected against oxidation. The occurrence of half-cystine residues in proteins has been examined with some 34 extracellular mammalian proteins taken from the tabulations of Jocelyn (1972) and Reeck (1970) with the same restrictions as described above for intracellular proteins. The results, presented in the accompanying Figure, show that the values for half-cystine content exhibit a wide range (0 % – 13 %) and that the mean value (4.1 %) is greater than that expected on the basis of the genetic code (3.3 %). Both the pattern and the mean content differ substantially from that found for intracellular sulfhydryl proteins. The marked variation in half-cystine contents found for disulfide proteins presumably reflects widely varying requirements for rigidity in the tertiary structure and it is noteworthy that some of the highest half-cystine contents occur in protease inhibitors, as high as 20 % for those from pineapple stem (Reddy et al., 1975).

The occurrence of cysteine and cystine in the extracellular proteins of bacteria is more complex, but appears to be generally understandable in terms of the oxidation state of the environment in which the protein functions. Strictly anaerobic bacteria produce extracellular proteins which must function in a reducing environment where disulfides would not be expected to be stable. Such proteins are expected to be either sulfhydryl proteins, as in the case of clostripain (Mitchell and Harrington, 1968), or cyst(e)ine-free proteins (lacking both cysteine and half-cystine), as in the case of collagenase from *Clostridium histolyticum* (Harper and Seifter, 1974), but not disulfide proteins. The extracellular proteins of facultative bacteria must potentially face both reducing anaerobic environments and oxidizing aerobic conditions with the consequence that neither cystine nor cysteine would have universal stability. The great majority of the extracellular proteins produced by species of *Bacillus* and other bacteria having some capacity for both aerobic and anaerobic growth are cyst(e)ine-free proteins (Pollock and Richmond, 1962; Pollock, 1962; Matsubara and Feder, 1971; Markland and Smith, 1971; Saheb, 1976). An exception is streptoccocal protease which is a sulfhydryl protease having a single cysteine at the active site (Liu and Elliott, 1971). Finally, strictly aerobic bacteria would be expected to produce cyst(e)ine-free or disulfide extracellular proteins but not sulfhydryl proteins. The serine proteases obtained from the aerobes *Streptomyces griseus* (Gertler and Trop, 1971; Olafson et al., 1975) and *Sorangium* sp. (Olson, et al., 1970) are indeed disulfide proteins which show substantial homology with the proteases of mammalian origin. These observations suggest that extracellular disfulfide proteins are produced primarily by strictly aerobic organisms and have evolved largely after the transition to an oxidizing atmosphere.

The present analysis in terms of strictly intracellular and extracellular proteins is clearly oversimplified. Different subcellular organelles and sites may well differ significantly in their thiol-disulfide redox state and therefore also in the cysteine and cystine content of their proteins. In fact, this could be an important underlying factor in the separation of various metabolic processes, especially those involving oxygen, in compartments isolated from the highly reduced environment of the cytoplasm. Also interesting is the status of the plasma membrane proteins, which at the intracellular surface are exposed to a highly reduced environment and at the extracellular surface to an oxidized environment. The data on the cysteine and cystine content of proteins from different

subcellular locations is at present too limited to permit an adequate analysis, but it can be noted that the cyst(e)ine content of the proteins from human erythrocyte plasma membrane is only 1.1 % (Rosenberg and Guidotti, 1968).

The examples and analysis given here show that the cyst(e)ine content of proteins is highly variable, apparently more so than that of any other amino acid. The reasons for this variability seem to be generally understandable in terms of the redox state of the environment in which the protein functions, taken in conjuction with the different roles played by thiol and disulfide groups in proteins. The profound consequences of the transition from a reducing to an oxidizing atmosphere seem nowhere more clear than in their impact upon the redox state of cyst(e)ine in proteins, and it seems certain that this factor will prove to be of central importance as we gain further insight into the chemical mechanisms which have directed the evolution of aerobic organisms.

We thank Russell Doolittle, Joe Kraut, Jack Kyte, Stanley Miller, and Leslie Orgel for helpful discussions. This work was supported by a grant (GM 22122) from the National Institute of General Medical Sciences.

References

Banaszak, L.J., Bradshaw, R.A.,(1975). The enzymes, P.D. Boyer, ed., 3rd ed., Vol. XI, p. 369. New York: Academic Press

Barron, E.S.G., Singer, T.P. (1943). Science **97**, 356

Bieber, H., Zimmermann, W., (1971). Zbl. Bakt., I. Abt. Orig. **217**, 47

Dickerson, R.E., Timkovitch, R. (1975). The enzymes, P.D. Boyer, ed., 3rd ed., Vol. XI, p. 397. New York: Academic Press

Friedman, M. (1973). The chemistry and biochemistry of the sulfhydryl group in amino acids., peptides and proteins. New York: Pergamon Press

Gatlin, L.L. (1974). J. Mol. Evol. **3**, 189

Gatlin, L.L. (1976). J. Mol. Evol. **7**, 185

Gertler, A., Trop, M. (1971). Eur. J. Biochem. **19**, 90

Harper, E., Seifter, S. (1974). Isr. J. Chem. **12**, 515

Harris, J. I., Waters, M. (1976). The enzymes, P.D. Boyer, ed., 3rd ed., Vol. XIII, p. 1. New York: Academic Press

Haugaard, N. (1968). Physiol. Rev. **48**, 311

Holmquist, R. (1975). J. Mol. Evol. **4**, 277

Jocelyn, P.C. (1972). Biochemistry of the SH Group, Chapt. 1. New York: Academic Press

Jukes, T.H., Holmquist, R., and Moise, H. (1975). Science, **189**, 50

Kimura, M. (1968). Genet. Res., Camb. **11**, 247

King, J.L., Jukes, T.H. (1969). Science **164**, 788

Liu, T.Y., Elliott, S.D. (1971). The enzymes, P.D. Boyer, ed., Vol. III, p. 609. New York: Academic Press

Malkin, R. (1973). Iron-Sulfur Proteins, W. Lovenberg, ed., Vol. II, p. 1. New York: Academic Press

Markland, Jr., F.S., Smith, E.L. (1971). The enzymes, P.D. Boyer, ed., 3rd ed., Vol. III p. 561. New York: Academic Press

Matsubara, H., Feder, J. (1971). The enzymes, P.D. Boyer, ed., 3rd ed., Vol. III, p. 721. New York: Academic Press
Miller, T.E., Stone, R.W. (1938). J. Bacteriol. **36**, 248
Mitchell, W.M., Harrington, W.F. (1968). J. Biol. Chem. **243**, 4683
Murphey, W.H., Barnaby, C., Lin, F.J. and Kaplan, N.O. (1967). J. Biol. Chem. **242**, 1548
Ohta, T., Kimura, M. (1971). Science **174**, 150
Olafson, R.W., Jurasek, L., Carpenter, M.R., and Smillie, L.B. (1975). Biochemistry **14**, 1168
Olson, M.O.J., Nagabhushan, N., Dzwinill, M., Smillie, L.B., and Whitaker, D.R. (1970). Nature **228**, 438
Pollock, M.R., (1962). The bacteria, I.C. Gunsalas, R.Y. Stanier, eds., Vol. IV, p. 121. New York: Academic Press
Pollock, M.R., Richmond, M.H. (1962). Nature **194**, 446
Reddy, M.N., Keim, P.S., Heinrikson, R.L., Kezdy, R.J. (1975). J. Biol. Chem. **250**, 1741
Reeck, G.R., (1970). Handbook of biochemistry, H.A. Sober, ed., 2nd ed., p. C282. Cleveland: Chemical Rubber Company
Reeck, G.R., Fisher, L. (1973). Int. J. Peptide Res. **5**, 109
Rosenberg, S.A., Guidotti, G. (1968). J. Biol. Chem. **243**, 1985
Saheb, S.A. (1976). Biochimie **58**, 793
Smith, M.H. (1966). J. Theoret. Biol. **13**, 261
Williams, Jr., C.H. (1976). The enzymes, P.D. Boyer, ed., 3rd ed., Vol. XIII, p. 90. New York: Academic Press

Received March 24, 1977