Int. J. Peptide Protein Res. **12**, 1978, 237–241
Published by Munksgaard, Copenhagen, Denmark

VOLUME AND POLARITY CHANGES ACCOMPANIED BY AMINO ACID SUBSTITUTIONS IN PROTEIN EVOLUTION

MITIKO GŌ and SANZO MIYAZAWA

Department of Biology, Faculty of Science, Kyushu University, Fukuoka, Japan

Received 13 February, accepted for publication 25 May 1978

We evaluated the volume and polarity changes accompanied by amino acid substitutions along branches of the phylogenetic trees of cytochrome c, myoglobin and hemoglobin α and β chains. In most cases the volume changes accompanied by the substitutions were found to be much larger than the volume of cavities existing in the interior of X-ray-analysed proteins. This implies that the interior of the proteins is very flexible and the necessary space for a larger amino acid residue substitution can be provided by adjusting nearby structures. Also, the volume and polarity changes are not particularly dependent on whether the substituted site is located in the exterior or interior of the proteins. This result supports the concept of the covarions by Fitch and Markowitz, when combined with the known fact that the exterior sites are more variable than the interior ones during protein evolution.

Key words: covarion; flexibility of protein conformation; interior and exterior of proteins; X-ray coordinates.

During molecular evolution the tertiary structures of homologous proteins are conserved in spite of many amino acid residue replacements (Perutz *et al.*, 1965; Love *et al.*, 1971; Dickerson *et al.*, 1971). These replacements are observed to occur more frequently on the surface of proteins (Wright *et al.*, 1969; Shotton & Watson, 1970; Blake & Swan, 1971; Hendrickson & Love, 1971; Fitch, 1976). Also, the replacements are observed to occur more frequently between amino acid residues with similar physicochemical properties (Sneath, 1966; Epstein, 1967; Clarke, 1970; Dayhoff *et al.*, 1972*c*; Grantham, 1974). Are the changes in physicochemical properties by the replacements larger on the surface sites than in the buried sites? In order to find the answer to this question we will first classify the residual sites into exterior and interior sites. Volume and polarity changes accompanied by the amino acid substitutions will be compared at these two types of sites. Analysis of the results has implications in two very different fields of protein study, a) the flexibility of protein conformations and b) the mechanism of protein evolution.

Methods

The classification of residual sites is done simply by calculating for each residue the number (n_r) of c^α atoms in neighbouring residues within a given distance (r) of the c^α atom in the residue in question. If n_r is smaller than the mean, the residual site is defined as exterior; otherwise it is interior. The results presented in the following are not dependent on the value of r, but to be specific, r = 10 Å is used. The exterior sites thus defined

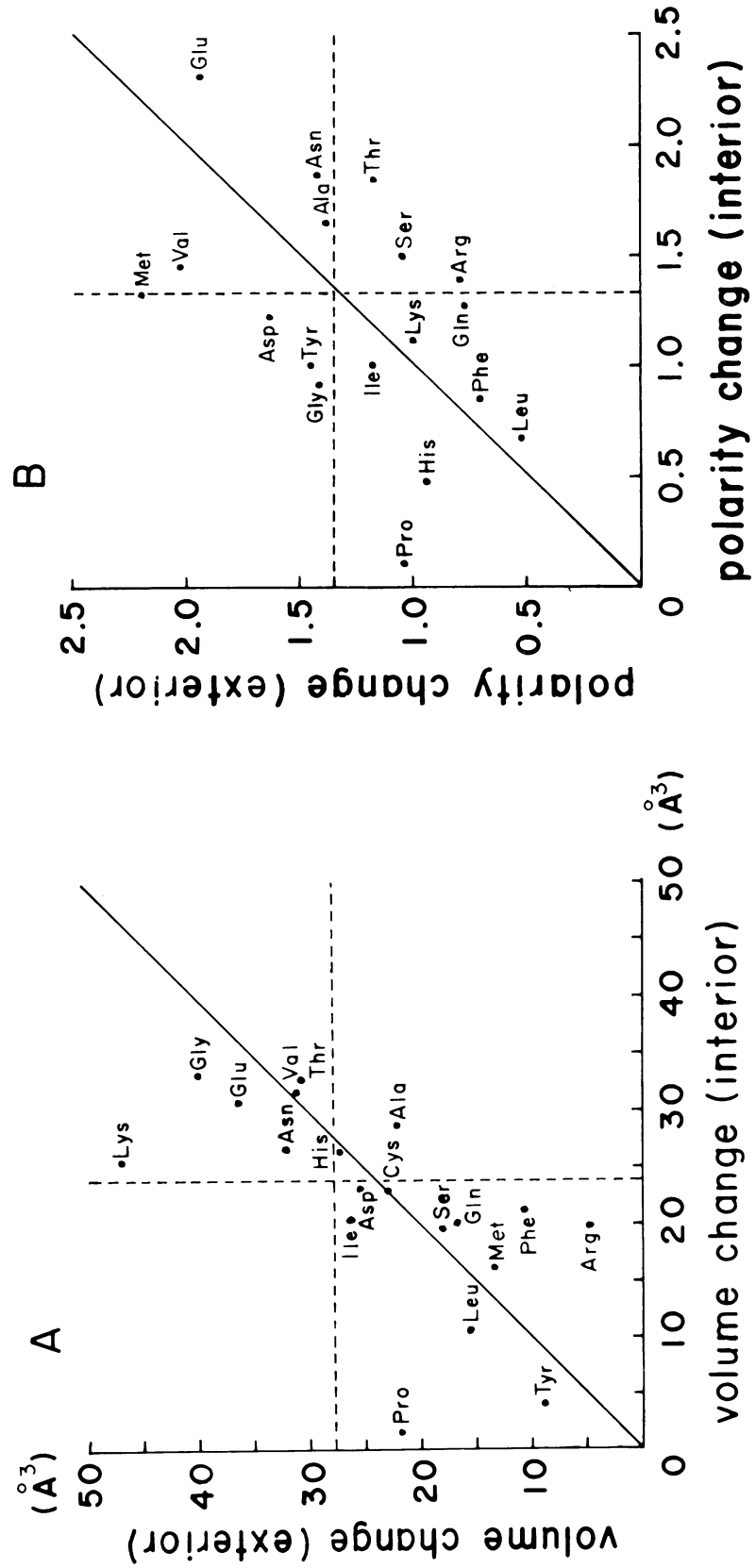

FIGURE 1

(A) The average volume changes (in absolute values) in the interior (abscissa) and exterior (ordinate) are plotted for each type of amino acid residue. The dotted lines show the average volume changes over amino acid residues in the interior and exterior, respectively. The solid line indicates the hypothetical case where the volume change in the interior is the same as that in the exterior. (B) The same as (A) but for polarity changes. The polarity scale of Grantham (1974) is used.

are confirmed to have higher variabilities in seven sets of homologous proteins (Gō & Miyazawa, submitted).

This method of classifying residual sites has the merit of being simple, but is admittedly rather crude. In order to check how well the number (n_r) reflect the true extent of exposure of residues, the accessible surface area to water molecules (Lee & Richards, 1971; Shrake & Rupley, 1973; Chothia, 1975) was evaluated for cytochrome c. A good correlation was observed between n_r and the extent of exposure as determined in terms of the accessible surface area. For instance, the residues are classified into exterior and interior in terms of the extent of exposure as defined by the accessible surface area. About 80% of the exterior residues as defined by the two methods, n_r and the extent of exposure, are found to overlap. Because of this good correlation, the main results obtained in this paper are expected to remain true even when we use the more refined, yet more complex, method of classifying residues by using the accessible surface area.

What type of atomic interaction is primarily responsible for making the substitutions easier in the exterior sites than in the interior sites? Energy terms, such as those associated with van der Waal's forces, hydrophobic interactions, electrostatic interactions and hydrogen bond formation, are responsible for determining the tertiary structure of proteins. From among these terms we examine two simple quantities, (a) volume of each residue (as a measure primarily of van der Waal's forces) and (b) polarity of each residue (which reflects hydrophobic interactions, electrostatic interactions and hydrogen bond formation). Volume change ΔV_{ij} and polarity change ΔP_{ij}, which are associated with substitution of amino acid i by amino acid j, are calculated for each substitution step. We use the volume and polarity scale of the 20 naturally occurring amino acid residues evaluated by Grantham (1974). On first consideration, it might be expected that more variable sites have larger values of ΔV_{ij} and ΔP_{ij}. In other words, a site would be more variable if it can accommodate larger changes of volume and/or polarity. If this is so, the fact that the exterior sites are more variable means that the exterior sites have larger values of volume and/or polarity changes than the interior sites.

RESULTS AND DISCUSSION

Absolute values of volume and polarity changes are averaged in the interior and exterior, respectively, for each type of substituted amino acid residue (Fig. 1). The X-ray coordinates of c^{α} atoms used to define the interior and exterior sites are from tuna cytochrome c (Swanson *et al.*, 1977), sperm whale myoglobin (Watson, 1969) and horse oxyhemoglobin (Perutz, personal communication). For hemoglobin chains α and β, the classification of the residual sites is done for a tetrameric state consisting of two α and β chains. Only single-base changes of short branches (less than 17 PAM's) of the phylogenetic trees of cytochrome c (Dayhoff *et al.*, 1972*a*), hemoglobin α and β chains (Dayhoff *et al.*, 1972*b*) and myoglobin (Romero-Herrera *et al.*, 1973) are used. Lysine and proline show markedly higher volume changes in the interior than in the exterior. Proline and methionine show higher polarity changes in the exterior than in the interior. However, there is no sign of a systematic tendency for the exterior sites to have larger volume and polarity changes.

Figure 2A gives the distributions of the absolute values of volume changes in the exterior and interior, respectively, each being normalized to a total of 100%. Figure 2B gives the same but for the polarity changes. Substitution data are those described for Fig. 1. The large absolute values of volume changes in the interior of globular proteins should be especially noted. Lee & Richards (1971) reported in their paper on the packing of atoms in proteins that in the interior of each of the three proteins studied there are three, five and 13 cavities which are hardly big enough to contain a water molecule. The volume of such a cavity is about 12 $Å^3$. The absolute values of the volume changes shown in Fig. 2A are much larger than this, which means that a protein molecule accommodates a new sub-

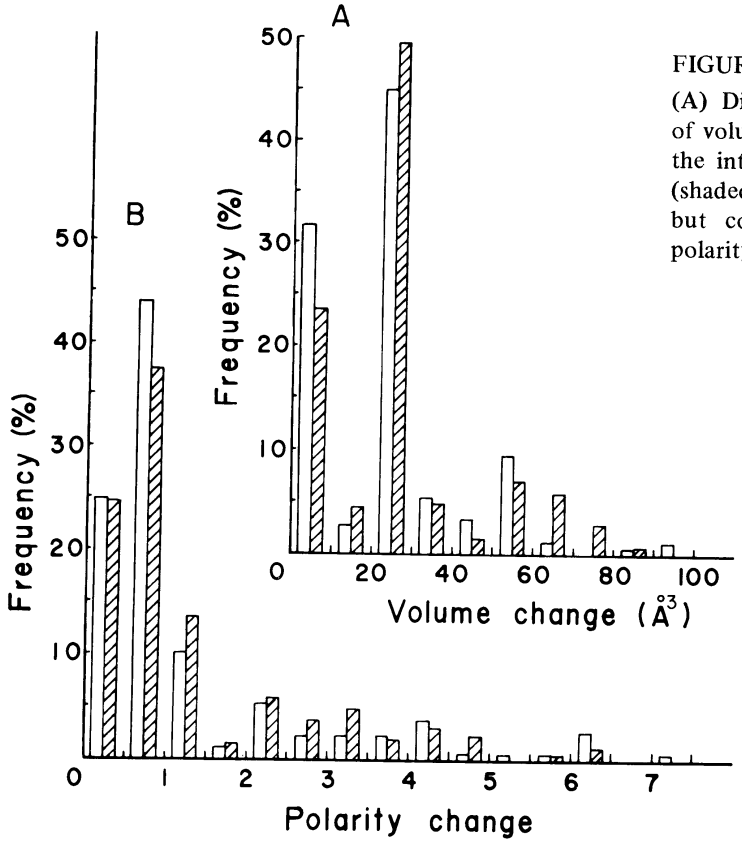

FIGURE 2
(A) Distribution of the absolute values of volume changes in 10 Å^3 intervals for the interior (open column) and exterior (shaded column). (B) The same as (A) but concerns polarity changes in 0.5 polarity scale intervals.

stituting amino acid residue not simply by inserting into a preexisting vacancy, but by providing the necessary space by adjusting nearby structures. This flexibility of the interior of proteins is one of the conclusions to be drawn from Fig. 2A.

There is no clear indication in Figs. 2A and B that the exterior sites have larger volume and polarity changes than the interior sites. When all substitutions in the exterior and interior are averaged respectively, the volume and polarity changes show very similar values, the difference being 3.8 Å^3 and 0.02, respectively. These values are very small compared with the absolute values of volume and polarity changes shown in Fig. 2. Therefore, we are forced to conclude that volume and polarity changes accompanied by the amino acid substitutions have no clear correlation with the location of the sites.

This unexpected conclusion indicates that the substituting amino acid residues in the exterior and interior are not markedly different. How can this result be reconciled with the confirmed fact that the exterior sites are more variable than the interior sites? The most plausible mechanism would be that at any stage of protein evolution residual sites are classified into two types: completely invariable sites and fully variable sites. The variety of substituting amino acid residues at fully variable sites are not markedly dependent on the location of the site. The fact that exterior sites are more variable can be explained by the fact that there are more fully variable sites in the exterior than interior. Different sets of sites may constitute the fully variable sites at different stages of protein evolution. However, the average number of fully variable sites must be larger in the exterior than in the interior. A set of fully variable sites at any stage of protein evolution is conceptually similar to what Fitch & Markowitz called covarions (Fitch & Markowitz, 1970). The reason why there are more fully variable sites in the exterior than interior can not be given from the analysis in this paper.

ACKNOWLEDGMENTS

We thank the Protein Data Bank for sending us all the protein coordinates except oxyhemoglobin, and the Medical Research Council, England for allowing us to use the oxyhemoglobin coordinates.

REFERENCES

Blake, C.C.F. & Swan, I.D.A. (1971) *Nature New Biol.* **232,** 12–15

Chothia, C. (1975) *Nature (Lond.)* **254,** 304–308

Clarke, B. *Nature (Lond.)* (1970) **228,** 159–160

Dayhoff, M.O., Park, C.M. & McLaughlin, P.J. (1972*a*) in *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.), pp. 7–16, The National Biomedical Research Foundation, Maryland

Dayhoff, M.O., Hunt, L.T., McLaughlin, P.J. & Jones, D.D. (1972*b*) in *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.), pp. 17–30, The National Biomedical Research Foundation, Maryland

Dayhoff, M.O., Eck, R.V. & Park, C.M. (1972*c*) in *Atlas of Protein Sequence and Structure* (Dayhoff, M.O., ed.), pp. 89–99, The National Biomedical Research Foundation, Maryland

Dickerson, R.E., Takano, T., Eisenberg, D., Kallai, O.B., Samson, L., Cooper, A. & Margoliash, E. (1971) *J. Biol. Chem.* **246,** 1511–1535

Epstein, C.J. (1967) *Nature (Lond.)* **215,** 355–359

Fitch, W.M. & Markowitz, E. (1970) *Biochem. Genet.* **4,** 579–593

Fitch, W.M. (1976) *J. Mol. Evol.* **8,** 13–40

Grantham, R. (1974) *Science* **185,** 862–864

Hendrickson, W.A. & Love, W.E. (1971) *Nature New Biol.* **232,** 197–203

Lee, B. & Richards, F.M. (1971) *J. Mol. Biol.* **55,** 379–400

Love, W.E., Klock, P.A., Lattman, E.E., Padlan, E.A. & Ward, K.B., Jr (1971) *Cold Spring Harbor Symposia on Quantitative Biol.* **36,** 349–357

Perutz, M.F., Kendrew, J.C. & Watson, H.C. (1965) *J. Mol. Biol.* **13,** 669–678

Romero-Herrera, A.E., Lehmann, H., Joysey, K.A. & Friday, A.E. (1973) *Nature (Lond.)* **246,** 389–395

Shotton, D.M. & Watson, H.C. (1970) *Nature (Lond.)* **225,** 811–816

Shrake, A. & Rupley, J.A. (1973) *J. Mol. Biol.* **79,** 351–371

Sneath, P.H.A. (1966) *J. Theoret. Biol.* **12,** 157–195

Swanson, R., Trus, B.L., Mandel, N., Mandel, G., Kallai, O.B. & Dickerson, R.E. (1977) *J. Biol. Chem.* **252,** 759–775

Watson, H.C. (1969) *Progr. Stereochem.* **4,** 299–333

Wright, C.S., Alden, R.A. & Kraut, J. (1969) *Nature (Lond.)* **221,** 235–242

Address:
Dr. *Mitiko Gō*
Department of Biology
Faculty of Science
Kyushu University
Fukuoka, 812
Japan