RELATIONSHIP BETWEEN MUTABILITY AND LOCATION OF AMINO ACID RESIDUES IN THE THREE-DIMENSIONAL STRUCTURE OF PROTEINS

Mitiko GŌ and Sanzo MIYAZAWA

pp. 33-43 in

Evolution of Protein Molecules

Proceedings of the International Symposium
on Evolution of Protein Molecules
Held at Osaka-Kobe, April 11-13, 1977

Edited by Hiroshi Matsubara and Tateo Yamanaka

Japan Scientific Societies Press
Center for Academic Publications Japan
Tokyo, 1978

RELATIONSHIP BETWEEN MUTABILITY AND LOCATION OF AMINO ACID RESIDUES IN THE THREE-DIMENSIONAL STRUCTURE OF PROTEINS

Mitiko GŌ AND Sanzo MIYAZAWA

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

Owing to the accumulation of results of X-ray analysis, atomic coordinates of over 60 proteins are now available for detailed structural analysis. One fact which these X-ray analyses have revealed is that the tertiary structure of homologous proteins is conserved in spite of substitutions at various amino acid sites (*7*, *13*, *15*). Since the selection pressure on evolving proteins operates on the functions of proteins, this fact implies, among other things, that the whole tertiary structure of a globular protein is essential for its function. Every amino acid substitution in evolving proteins may be assumed to occur only when they conserve tertiary structure. The purpose of this paper is to explore some empirical rules of amino acid substitutions in relation to the three-dimensional structure of proteins.

Amino acid substitutions are anticipated more easily on the surface than in the interior of protein molecules, because in the interior an amino acid residue generally is interacting with more nearby residues than on the surface. The expected result is a greater chance of affecting the tertiary structure by the amino acid replacement in the interior than by that on the surface. In fact less frequent occurrences of invariant amino acid residues on the surface were noticed by X-ray crystallographers for some particular pairs of homologous proteins (*3*, *12*, *17*, *19*). For comparative

studies they utilised three-dimensional models of proteins. In this paper we conduct a more systematic study by defining a quantity which characterizes the location of each amino acid residue in the three-dimensional structure of proteins. The result is to confirm the anticipation stated above, namely, the frequency of residual substitution is higher as a residue is more exposed to the surface of a protein molecule.

It is then natural to ask the following question: What type of atomic interactions are primarily responsible for making the substitutions of interior amino acid residues more difficult? Energy terms, such as those associated with van der Waals forces, hydrophobic interaction, electrostatic interaction and hydrogen bond formation, are responsible for determining tertiary structures of proteins. Among these terms, we examine two simple quantities; a) polarity of each residue (which reflects, for instance, hydrophobic interaction, electrostatic interaction, and hydrogen bond formation) and b) volume of each residue (as a measure primarily of van der Waals forces). In the following sections, we study whether polarity and volume changes associated with substitution are correlated with mutability.

1. Classification of Location, Counting Mutations, and Evolution of Volume Change

1) Location of residues

Classification of amino acid residues according to their locations is done by two different methods. i) The first method is a very simple one. For each amino acid residue we define the number n_r of C^α atoms of other amino acid residues existing within r Å form the C^α of the amino acid residue in question. The distances between C^α atoms are calculated by using the coordinates of C^α atoms determined by X-ray analysis (*1*, *2*, *4*, *18*, *20*). The larger the value of n_r, the deeper from the surface of the protein an amino acid residue is expected to be located, because the number n_r roughly measures the extent to which a particular amino acid residue is interacting with other amino acid residues. An amino acid residue is classified as an interior or exterior one depending on whether or not its n_r is larger than the mean ($\bar{n}_r$) calculated from the maximum and minimum values of n_r. This method of classification of amino acid residues is admittedly very crude, especially for residues having n_r values close to the mean. However, this method has the merit of being very

simple and is an adequate preliminary trial. Four different values of r are employed, *i.e.*, 7 Å, 10 Å, 15 Å, and 20 Å. ii) The second method is a more quantitative one. For this purpose we compute the division of the space occupied by a protein molecule into Voronoi polyhedra. The method of Voronoi polyhedron was introduced by Richards (*16*) in the investigation of the packing problem of atoms in a protein. Each non-hydrogen atom in a protein is located within the smallest polyhedron (Voronoi polyhedron) formed by a set of planes of bisection between the atom under consideration and each of its neighbors. For atoms on the surface (*i.e.*, for atoms whose Voronoi polyhedron can not be closed) water molecules are suitably supplied (*14*) in order to close the polyhedron. The exposed area of an atom is defined as the area of polygons which separate this atom and supplied water molecules. Volume, total surface area and exposed area of each residue in a protein are obtained by summing the contributions from atoms in the residue. Location of residues in the three-dimensional structure are classified by the fraction of the exposed area to the total surface area of the residues. *The surface of a* protein is defined as a set of polygons which separate an atom in the protein and supplied water molecules.

2) Mutations at each residual sites

Dayhoff *et al.* (*6*) constructed probable phylogenetic trees of several protein families from protein sequences now available. Ancestral sequences were determined at the same time. They counted the number of mutations that occurred at each residual site along the phylogenteic trees. An amino acid residue is defined as *invariant* if no mutation is counted at the site.

3) Volume changes accompanied by substitutions

Volume changes accompanied by substitutions at each residual site are calculated. In order to count substitutions by one-step mutations only short branches of the phylogenetic trees are taken into account. Volumes of 20 types of amino acid residues evaluated by Grantham (*11*) are used to calculate their differences caused by residual substitutions.

2. Mutability and Location of Residual Sites

Classification of amino acid residues into interior and exterior ones is

TABLE I

Ratio of the Probability of Being an Invariant Amino Acid Residue of the Interior Amino Acid Residues to That of the Exterior Amino Acid Residues

Protein	Radius r (Å)			
	7	10	15	20
Cytochrome *c*	1.7	1.7	1.9	2.2
Lysozyme[a]	2.0	2.4	2.1	2.4
Ribonuclease	1.4	1.9	1.9	1.6
Myoglobin	1.2	1.1	1.4	1.2
Subtilisin	1.5	1.6	1.7	1.6
Serine protease	5.5[b]	5.5	4.7	4.4
Hemoglobin[c]	1.0[b]	1.3	1.1	1.4

[a] Lactalbumin from three species are included with lysozyme from five species. [b] 7.5Å is used instead of 7Å. [c] A tetramer consisting of two α chains and two β chains.

TABLE II

Ratio of Average Number of Mutations in the Exterior to That in the Interior

Protein	Radius (Å)			
	7.5	10	15	20
Cytochrome *c*	1.7	1.8	2.0	1.6
Lysozyme[a]	1.4	1.5	1.3	1.5
Hemoglobin α[b]	1.2	1.4	1.4	1.4
Hemoglobin β[b]	1.4	1.7	1.7	1.8

[a] Lactalbumin from three species are included with lysozyme from five species. [b] A tetramer consisting of two α chains and two β chains is used to classify the residues into the exterior or interior.

made for seven homologous proteins, cytochrome *c*, lysozyme, ribonuclease, myoglobin, subtilisin, serine protease, and hemoglobin. In all these proteins, the invariant amino acid residues exist more frequently in the interior than the exterior of the proteins. This is shown in Table I. The ratios of the frequency of the invariant amino acid residues do not depend on the value of r used to define the number n_r of neighboring C^α atoms.

The numbers of mutations at each site are averaged over residues in the interior and exterior, respectively. The ratios of these averaged mutations in the exterior to those in the interior are shown in Table II for

cytochrome *c*, hemoglobin α chain, hemoglobin β chain, and lysozyme. For these homologous proteins, the ratio is larger than unity, *i.e.*, the averaged number of counted mutations is larger on the exterior than in the interior. The ratio does not depend on the value of r used to define the number n_r of neighboring C^α atoms.

These results confirm the anticipation stated in the Introduction, namely that substitutions of amino acid residues are more difficult in the interior than on the exterior of three-dimensional structures of proteins. This rule appeares to hold for globular proteins in general.

The above rule is observed for an admittedly rough classification of amino acid residues into the interior and exterior. In order to see if the rule can be refined any further, we employ a more refined classification of residues based on the extent of exposure of each residue to the surface of a protein. The fraction of exposed area in each residue of lysozyme is calculated by computing the Voronoi polyhedron. The fractions range from 0 to 77%. Residues are classified into six groups according to these fractions; *i.e.*, group I: 0~10%, group II: 10~20%, group III: 20~30%, group IV: 30~40%, group V: 40~50%, and group VI: more than 50%. The averaged number of mutations are 1.1, 1.3, 1.4, 1.3, 2.1, and 2.2 for groups I through VI, respectively. The distribution of the number of counted mutations in each group is shown in Fig. 1. It is clear that the number of invariant sites (sites with zero mutation) decreases and the peak of the distribution shifts to the larger number of mutations as the fraction of exposure incresaes.

3. *Mutability and Polarity of Residues*

It is well established that nonpolar residues exist more frequently in the interior than on the exterior and *vice versa* for polar residues. It is therefore necessary to study the correlation between mutability and polarity of residues. (If polar residues are more mutable than nonpolar residues, it explains the correlation confirmed in the last section between mutability and location of residues.) The mutabilities* of each type of amino acid residue in the interior and on the exterior of protein molecules are calculated (*10*) and are tabulated in Table III. No clear correlation between mutability and polarity is observed.

* For Precise definition, see Ref. *10*.

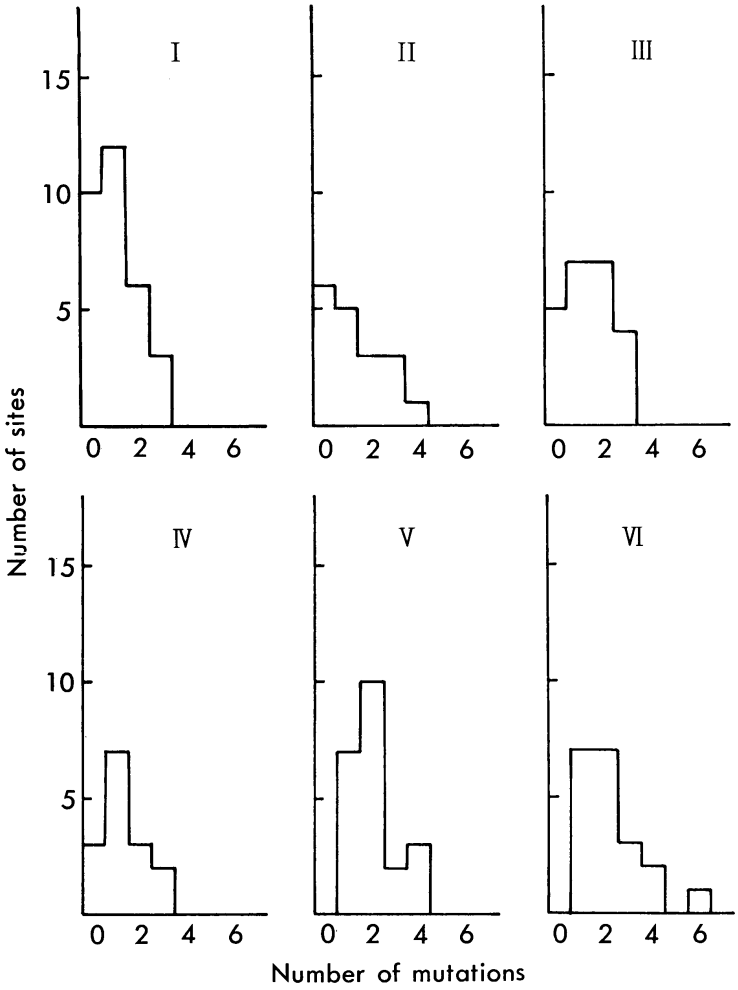

Fig. 1. Distribution of number of residual sites of lysozyme over the number of the mutations occurring on the sites, for each of the residual groups I through VI. Groups I through VI are defined by the fraction of the exposed area of the residue, 0–10%, 10–20%, 20–30%, 30–40%, 40–50%, and more than 50%, respectively.

4. *Volume Changes Accompanied by Substitutions*

The atomic packing densities of various proteins were investigated using the Voronoi polyhedron by several authors (*5*, *8*, *16*). According to these authors the packing in the interior of protein molecules is as good as in crystals of small organic molecules. The differences between the averaged volume changes in the internal and external residues are given in Table III. The volumes of side chains as given by Grantham (*11*) and used in

TABLE III

Volume, Polarity, and Mutabilities in the Exterior and Interior and Difference between Averaged Volume Change on the Exterior and That in the Interior for Each Amino Acid Residue

Amino acid	Volume (Å^3)	Polarity	Mutability (%)		Difference in volume change (Å^3) (ext. − int.)
			ext.	int.	
Arg	124.0	10.5	0.87	1.00	−14.8
Lys	119.0	11.3	0.46	0.51	21.3
His	96.0	10.4	0.66	0.69	0.8
Gln	85.0	10.5	2.04	0.90	−3.1
Asn	56.0	11.6	2.67	1.13	5.4
Asp	54.0	13.0	1.78	1.37	2.3
Glu	83.0	12.3	1.60	2.10	5.8
Ala	31.0	8.1	2.17	1.06	−6.4
Pro	32.5	8.0	0.59	0.38	20.5
Gly	3.0	9.0	1.01	0.46	6.7
Thr	61.0	8.6	1.61	1.14	−2.0
Ser	32.0	9.2	2.60	1.64	−1.4
Cys	55.0	5.5	0.22	1.01	0.0
Val	84.0	5.9	1.15	1.49	−0.4
Met	105.0	5.7	1.05	1.67	−2.3
Ile	111.0	5.2	2.95	2.87	5.9
Leu	111.0	4.9	0.34	0.53	5.2
Phe	132.0	5.2	0.77	0.58	−10.5
Tyr	136.0	6.2	1.34	0.38	5.1
Trp	170.0	5.4	0.00	0.00	
Average					3.8

the present work are also tabulated. The differences in the averaged volume changes vary depending on the amino acids. Lysine and proline show particularly high volume changes on the exterior than in the interior among 20 amino acids. However, the difference in the volume changes is only 3.8 Å^3, when averaged over the total amino acids (counting 272 and 189 amino acids on the exterior and in the interior, respectively). This is a very small quantity when compared with the values of the volumes themselves. Therefore, we are forced to conclude that the volume changes accompanied by amino acid substitutions are almost the same

no matter whether the site is located in the interior or exterior of a protein molecule.

To confirm this result, a different analysis has been carried out which is independent of the classification of the amino acid residues by their locations. The volume changes accompanied by substitutions are plotted (with decreases and increases in volume indicated) in Fig. 2 against the number of mutations counted at the site where the substitution occurred. The substitutions counted in cytochrome *c* are plotted in this figure. Extremely long branches in phylogenetic trees are omitted from this counting. It is seen that the volume changes range from -40 Å^3 to $+30$ Å^3 and they are not dependent on the number of mutations counted at the site where the substitution occurred. This fact, when combined with the correlation confirmed in a previous section between mutability and

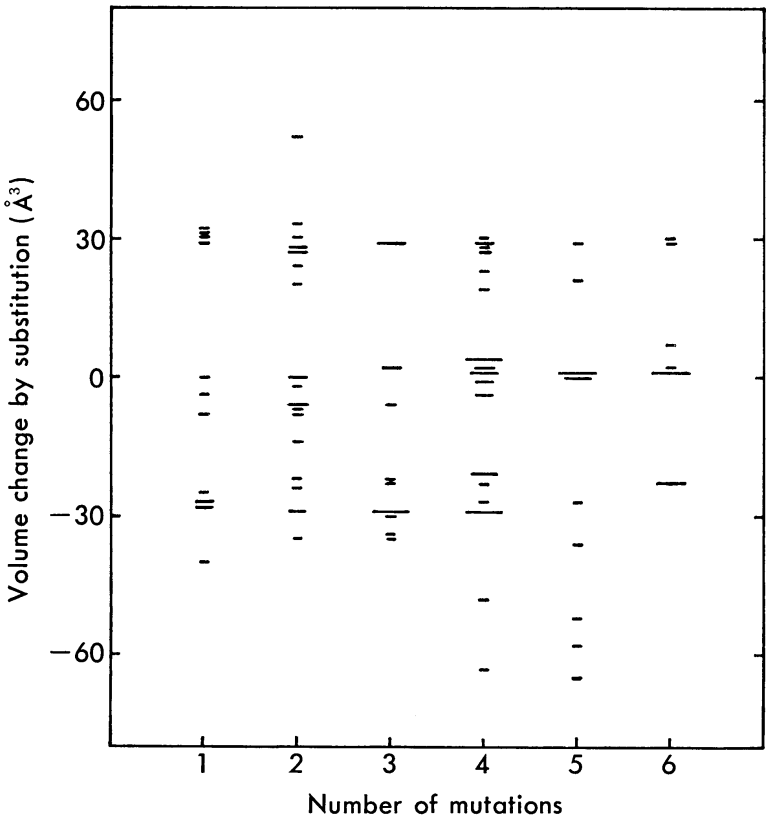

Fig. 2. Plotting of the volume changes accompanied by the residual substitutions in cytochrome *c* against the number of mutations counted at the site where the substitution occurred. The length of the horizontal bars is proportional to the number of the substitutions occurring, ranging from 1 to 4.

location, indicates that the volume changes accompanied by amino acid substitutions are almost independent of the location of the site where the substitution occurred.

DISCUSSION

In the last section we arrived at an unexpected conclusion. Here we discuss the consequences of that conclusion.

First of all, it is suspected that some other type of interaction, which is not related to the volume change, may explain the correlation between mutability and location of residues. As one such possibility, we also studied changes in polarity of a residue accompanied by substitutions (*10*). No significant differences in changes in polarity were observed. The possibility can not be excluded that a new quantity, which is not related either to volume or polarity changes, may exist. The geometrical shape of residues is yet to be examined with this in mind. It is possible to consider two typical mechanisms by which the correlation between mutability and location of residues can be explained. i) All residual sites in a protein have a fairly constant probability of substitution during molecular evolution. However this probability varies from site to site, with exterior sites having, on average, larger probabilities. ii) Probabilities of substitution of residual sites at any time during protein evolution are all or none, meaning that the probability has either vanishing or some finite value. This finite value is roughly independent of the sites. Residual sites with non-vanishing probabilities of substitutions are not fixed, but shift around over most residual sites (with more chances over exterior sites) in a protein, during molecular evolution. Residual sites with non-vanishing probabilities of substitution at any stage of an evolving protein were called covarions by Fitch and Markowitz (*9*).

If no significant difference (between interior and exterior sites) in changes in any quantity exists, then mechanism (i) can not be taken into consideration. We will be forced to employ mechanism (ii).

SUMMARY

Empirical rules of amino acid substitutions in homologous proteins are explored in relation to the three-dimensional structure of proteins. It is shown that the frequency of residual substitution becomes higher as

a residue is more exposed to the surface of a protein molecule. In order to elucidate the mechanism for this rule, polarity and volume of each amino acid residue are investigated. The following results are obtained: i) polarity of amino acid residues does not have clear correlation with mutability; and ii) volume and polarity changes accompanied by residual substitutions do not show significant dependence on the location of residues, *i.e.*, whether or not they are more exposed to surface of a protein molecule than the average. This finding appears to exclude a possibility that all residual sites in a protein have probabilities of substitution which are site-dependent (larger for exterior sites) but fairly constant during molecular evolution. Instead, the following is more probable: probabilities of substitutions at residual sites are all or none at any time of protein evolution. Residual sites with non-vanishing probabilities are not fixed but shift around over most residual sites (with more chances over external sites) in a protein.

Acknowledgments

The authors wish to express thanks to the Medical Research Council, England, for allowing them to use the oxyhemoglobin coordinates which were elaborated by Prof. M. F. Perutz. They are also indebted to the Protein Data Bank in Cambridge and Brookhaven National Laboratory for sending them the coordinates of tuna cytochrome *c*, bovine ribonuclease S, hen egg-white lysozyme, sperm whale myoglobin, tosyl-α-chymotrypsin, and subtilisin BPN'. They acknowledge Prof. H. Matsuda and Dr. N. Go for their encouragement during the work and the critical reading of the manuscript.

REFERENCES

1 Alden, R. A., Birktoft, J. J., Kraut, J., Roberts, J. D., and Wright, C. S. (1971). *Biochem. Biophys. Res. Commun.* **45**, 337–344.
2 Birktoft, J. J. and Blow, D. M. (1972). *J. Mol. Biol.* **68**, 187–240.
3 Blake, C. C. F. and Swan, I. D. A. (1971). *Nature New Biol.* **232**, 12–15.
4 Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C., and Sarma, V. R. (1967). *Proc. Roy. Soc.* **B167**, 365–377.
5 Chothia, C. (1975). *Nature* **254**, 304–308.
6 Dayhoff, M. O., Park, C. M., and McLaughlin, P. J. (1972). *Atlas of Protein Sequence and Structure.* ed. Dayhoff, M. O., pp. 7–16. Washington, D. C.: National Biomedical Research Foundation.

7 Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A., and Margoliash, E. (1971). *J. Biol. Chem.* **246**, 1151–1535.
8 Finney, J. L. (1975). *J. Mol. Biol.* **96**, 721–732.
9 Fitch, W. M. and Markowitz, E. (1970). *Biochem. Genet.* **4**, 579–593.
10 Gō, M. and Miyazawa, S. Unpublished.
11 Grantham, R. (1974). *Science* **185**, 862–864.
12 Hendrickson, W. A. and Love, W. E. (1971). *Nature New Biol.* **232**, 197–203.
13 Love, W. E., Klock, P. A., Lattman, E. E., Padlan, E. A., and Ward, K. B., Jr. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 349–357.
14 Miyazawa, S. and Gō, M. Unpublished.
15 Perutz, M. F. and Kendrew, J. C., and Watson, H. C. (1965). *J. Mol. Biol.* **13**, 669–678.
16 Richards, F. M. (1974). *J. Mol. Biol.* **82**, 1–14.
17 Shotton, D. M. and Watson, H. C. (1970). *Nature* **225**, 811–816.
18 Watson, H. C. (1969). *Prog. Stereochem.* **4**, 299–333.
19 Wright, C. S., Alden, R. A., and Kraut, J. (1969). *Nature* **221**, 235–242.
20 Wyckoff, H. W., Tsernoglou, D., Hanson, A. W., Knox, J. R., Lee, B., and Richards, F. M. (1970). *J. Biol. Chem.* **245**, 305–328.