J. Biosci., Vol. 14, Number 2, June 1989, pp. 133–142. © Printed in India.

Hydrophilicity and antigenicity of proteins—A case study of myoglobin and hemoglobin

A. SHARADA DEVI, TASNEEM REHANA*, A. S. KOLASKAR** and M. W. PANDIT†

Centre for Cellular and Molecular Biology, Hyderabad 500 007, India

*Present Address: Department of Zoology, Osmania University, Hyderabad 500 007, India

**Department of Zoology, University of Poona, Pune 411 007, India

MS received 31 January 1989; revised 12 April 1989

Abstract. Hydrophilicity index is used to locate antigenic determinants on two related groups of proteins—myoglobin and hemoglobin. The data on 41 species (including 34 mammals) of myoglobin show that average hydrophilicity for the complete myoglobin molecules as well as the average hydrophilicity for all hydrophilic regions put together seem to remain constant; the variation in the size and location of the antigenic determinants in these species is very small indicating that the antigenic sites are not shifted during evolution. In the case of both the proteins there is a good agreement between the antigenic sites picked up by using hydrophilicity index and the experimentally determined antigenic sites. The data on 56 species of hemoglobin α-chains and 44 species of hemoglobin β-chains showed that although there are few sites on hemoglobin which have remained invariant during evolution, there is a significant variation in other sites in terms of either a splitting of a site, or a drastic change in the hydrophilicity values and/or a length of the site. Comparison of the hydrophilicity data on these two groups of proteins suggests that hemoglobins which perform a variety of functions as compared to myoglobins are evolving faster than myoglobins supporting the contention of earlier workers.

Keywords. Hydrophilicity; antigenic sites; myoglobin; hemoglobin; evolution.

Introduction

Immunological properties of proteins have been used widely to study their structure. However, determination of complete immunogenic structure of a single protein from a given species is not very easy and some times takes a long time. In fact, there are only a few proteins such as sperm whale myoglobin (Atassi, 1975), hemoglobin (Kazim and Atassi, 1980, 1982) and lysozyme (Atassi, 1978) for which complete immunogenic structure is available. The experimental studies on proteins, particularly on lysozyme, have also pointed out that antigenic sites on proteins may be formed either by sequential continuous regions or by bringing together several antigenic determinants to form antigenic sites. The elaborate experimental procedures used for these studies and the time spent on them have necessitated the development of a theoretical approach for the prediction of immunogenic structure of a protein. One of the approaches which seems to have high potential to delineate antigenic determinants of the protein molecule is by Hopp and Woods (1981) and by Fraga (1982). This approach was recently used to predict and confirm the antigenic determinants of proteins such as RNAase A and seminalplasmin (Pandit, 1985). The success of this approach prompted us to apply the method to the

†To whom all correspondence should be addressed.

proteins for which complete sequences are known for large number of species and at least in one case where both the 3-dimensional structure as well as complete immunogenic structure is known. The natural choice for our case study was therefore myoglobin and hemoglobin. The purpose of choosing these proteins was to examine the antigenic sites on functionally related groups of proteins and to see if it is possible to throw some light on the differences between myoglobins and hemoglobins as hemoglobin is involved in much more diversified functions than myoglobin. These studies showed that indeed the level of confidence in the prediction of antigenic sites can be increased substantially if, instead of applying the approach of Hopp and Woods (1981) to a single protein, it is applied to a group of related proteins of known sequences. The results obtained from the analysis of such related protein sequences are discussed along with their significance in the succeeding sections of this paper.

Materials and methods

Amino acid sequences of myoglobin from 41 species (34 mammals, 5 reptiles and 2 birds) were taken from protein sequence data bank of NBRF. The species for which myoglobin sequences were analysed are given in figure 1. All the myoglobin sequences were arranged to get maximum homology among their sequences as has been done earlier by Hunt *et al.* (1978). The system of numbering the sequence after alignment was the same as followed by Hunt *et al.* (1978). This sequence data was further used to determine the hydrophilic regions on proteins. The algorithm which is quite similar to the one used by Hopp and Woods (1981, 1983) and Pandit (1985) to determine the hydrophilic regions is briefly discussed below. For each overlapping hexapeptide a profile of hydrophilic values as a function of the position in the sequence of the first amino acid of the hexapeptide is constructed using a computer program. Hydrophilicity values used are those given by Levitt (1976) with the adjustments suggested by Hopp and Woods (1981). Each hexapeptide was addressed by the position of the first amino acid in a hexapeptide sequence. Whenever the hydrophilicity values of at least 4 consecutive overlapping hexapeptides as well as the average hydrophilicity of all the amino acids consisting the region composed of such hexapeptides was greater than or equal to zero, such a region was picked up as the most probable antigenic site. The search was continued for delineating antigenic determinants for the entire sequence. This algorithm was applied to the sequences of myoglobin and hemoglobin of the species mentioned in figures 1 and 3 respectively, and antigenic determinants were determined.

Results and discussion

Probable antigenic sites along with their average hydrophilicity values for myoglobin sequences for all the species are given in figure 1. It can be seen from figure 1 that in case of myoglobin, 7 probable antigenic sites were obtained. These sites along with the experimentally observed antigenic sites for human myoglobin as reported by Atassi (1975) and Westhoff *et al.* (1984) are given in table 1. It can be seen from figure 1 and table 1 that all experimentally observed sites are picked up by our method except the site 166–172. The only theoretically predicted additional antigenic determinant is 155–166. In all other cases, however, one finds a very good

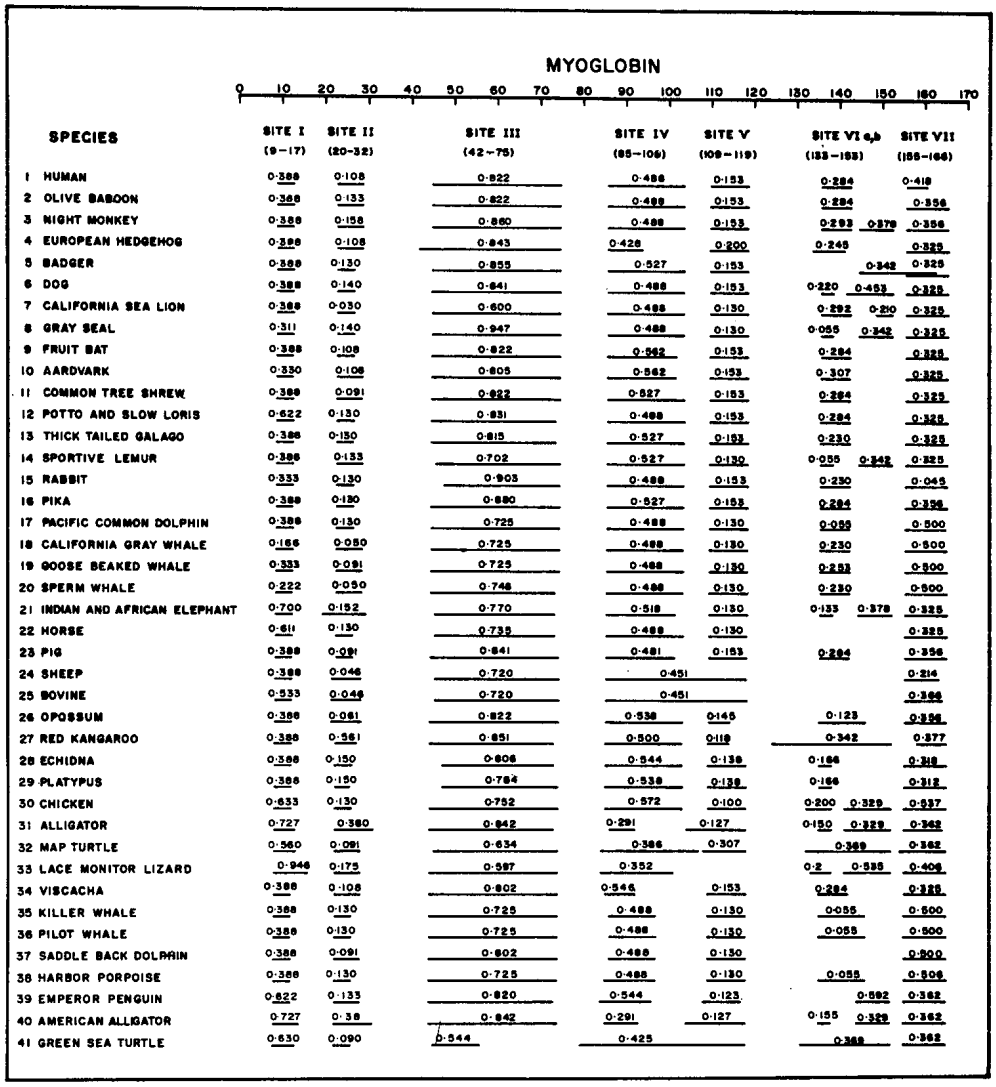

Figure 1. Hydrophilic sites on myoglobin from various species (the average hydrophilicity values of individual sites are given along with the sites).

agreement between experimentally observed and theoretically predicted antigenic sites. It can also be seen that many of the sites such as sites I to IV and VII are present in all the species without exception and there is a very little variation in the length of these sites. A close scrutiny of the hydrophilicity values indicates that the values for sites II and V are not very high when compared with the hydrophilicity of the other sites. In fact, according to the rationale established by Hopp and Woods (1981, 1983) a good correlation between hydrophilicity value and antigenic site exists only for those two or three regions which have highest hydrophilicity values. It means that sites II and V in myoglobin species, being relatively weak in their hydrophilic character, may not coincide with the antigenic sites if the choice is based only on the hydrophilicity data. However, experimental data on human myoglobin given in table 1 indicate that the sites II and V are antigenic. Therefore,

Table 1. Comparison of sites obtained on myoglobin by using average hydrophilicity values with the sites observed experimentally by earlier workers.

Site No.	Site based on hydrophilicity[a]	Experimentally observed sites	
		Atassi (1975)	Westhoff *et al.* (1984)
I	9– 17	—	9– 15
II	20– 32	24– 31	24– 31
III[b]	42– 68	—	31– 68
	69– 75	69– 74	69– 74
IV	85–108	—	85–108
V	109–119[c]	113–121	113–120
VI	133–147[d]	134–141	134–140
	142–153	—	142–148
VII	155–166	—	—
VIII	—	166–172	166–172

[a]Sites are depicted by the range covered by various species given in figure 1.
[b]Picked up as single site but shown separately for comparison.
[c]Except in case of alligator where the site was 105–119.
[d]Note a small overlap between the subsites of the site VI.

this study points out that the antigenic sites with relatively low hydrophilicity values (such as sites II and V) can be picked up if one carries out analysis similar to the one mentioned above on a similar protein from different species. It is interesting to note from the myoglobin data that although the order of the sequence-homology among proteins from various animals which are evolutionarily distant is less than 50%, the variation in the size and the location of antigenic determinants in these species is very small indicating that the antigenic sites are not shifted during evolution.

The quantitative variation in the average hydrophilicity values of different sites on myoglobin for all species studied is given in figure 2. A closer look at these values in few cases indicates that for a given species the change in the hydrophilicity value of one of the antigenic sites is associated with a compensatory change in the hydrophilicity value of the other site (ex. compare the hydrophilicity values of sites III and VII for species No. 15). Although one cannot generalize in regard to such a behaviour, this suggests that the delicate balance of the total hydrophilicity value for antigenic sites is maintained from one species to another species. This is further supported by the finding that the average hydrophilicity for complete myoglobin molecules as well as the average hydrophilicity values for all hydrophilic regions put together (see two lowermost profiles in figure 2) seem to remain constant. The maintenance of the above mentioned delicate balance of hydrophilicity of the protein molecule may be of a general nature and is probably necessary for an intrinsic functional requirement of the protein.

Hydrophilicity data obtained for hemoglobin species are given in figures 3 and 4. Kazim and Atassi (1980, 1982) have reported the antigenic sites on the α-chains of hemoglobins and few other species (table 2) by using synthetic approach in their confirmation of antigenic sites on these proteins. Comparison of the sites obtained by them with the sites picked up by using hydrophilicity values is given in table 2. It can be seen from the results that there is an appreciable amount of overlap between the experimentally confirmed antigenic sites in the species mentioned and

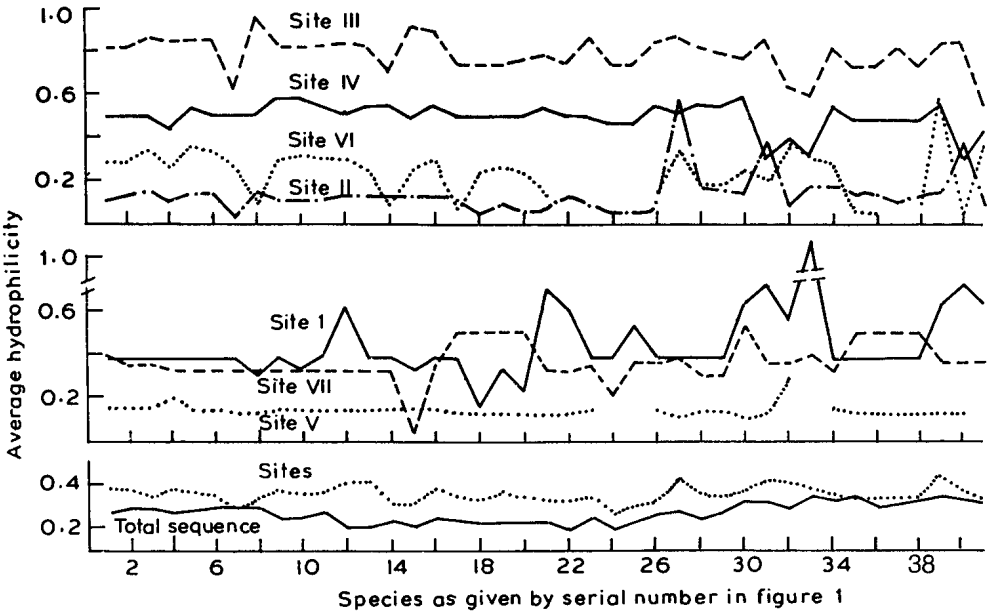

Figure 2. Average hydrophilicity values for different sites on the myoglobins from various species.

the range indicated by the sites picked up by using hydrophilicity values. Although sites I and IV picked up by the later approach are not found antigenic experimentally, reasonably good correlation was found in case of other sites. Several hemoglobins from various species included in the study are evolutionarily distant and have very little sequence homology. However, it can be seen from figures 3 and 4 that sites I, III and V in the case of α-chains of hemoglobins and sites I and V in the case of β-chains of hemoglobins have changed very little during the evolution. It is interesting to note that in the case of sites II and IV of α-chains of hemoglobins there is a significant variation in terms of either a splitting of a site, or a drastic change in the hydrophilicity values, or even a complete absence of any of these sites. In addition, site VI is missing in most of the α-chains of hemoglobins with very few exceptions. In the case of β-chains of hemoglobins there is a significant variation in sites II, III and VI, and a complete absence of site II in most of the birds and the reptiles analysed. These specific observations may have a bearing on the evolution of these species. Comparison of hydrophilicity data of hemoglobins with myoglobins from all the species shows that site VII present on myoglobin is completely deleted in case of α- and β-chains of hemoglobins.

These observations open up many questions in relation to their significance and suggest the need for more scrutiny of such data. It was thought that it would be possible to meaningfully interpret these results by comparing the data on myoglobins, α- and β-chains of hemoglobin from the same species. We could pick up 17 species where the sequences for all the 3 proteins were available from the data bank. These data are summarised in figure 5. In the case of myoglobin, there is a tendency of hydrophilic sites to remain constant in terms of length as well as hydrophilicity values except in case of site VI which sometimes splits into two with

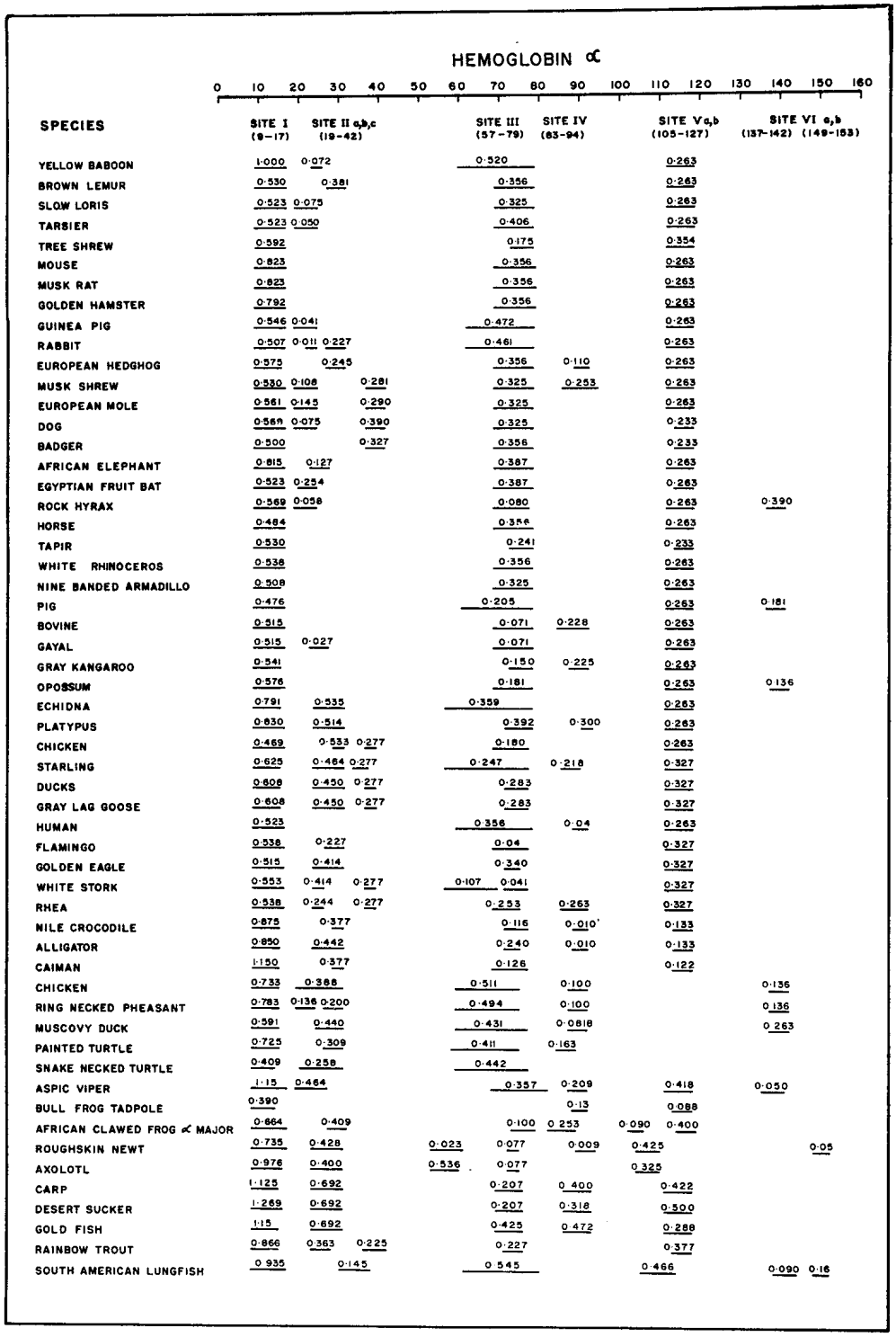

Figure 3. Hydrophilic sites on α-chains of hemoglobins from various species (the average hydrophilicity values of individual sites are given along with the sites).

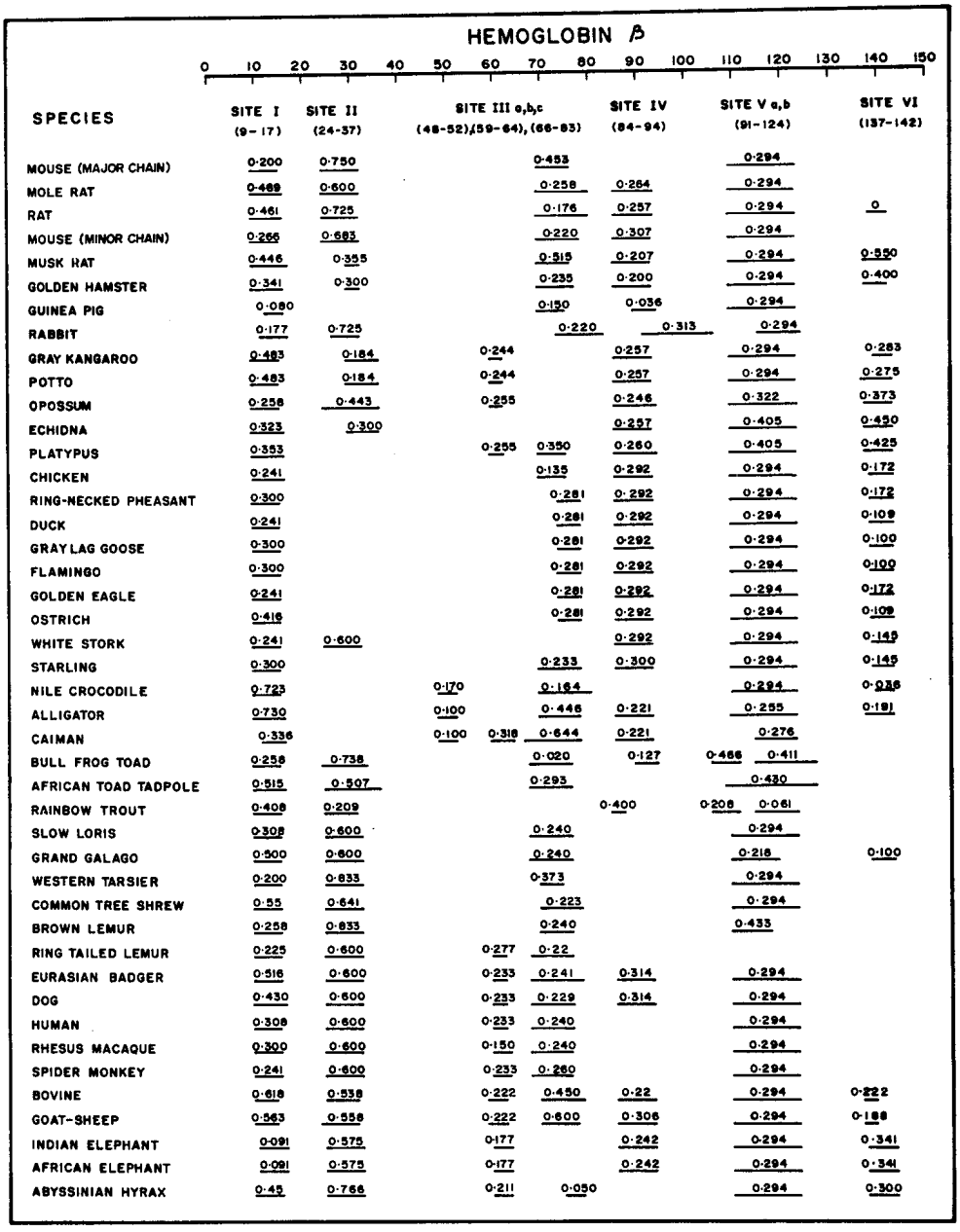

Figure 4. Hydrophilic sites on β-chains of hemoglobin from various species (the average hydrophilicity values of individual sites are given along with the sites).

an appreciable variation in hydrophilicity value. There was also a merging of sites IV and V in the case of myoglobins from two species. In the case of α- and β-chains of hemoglobins the invariance in length of sites and their hydrophilicity values is maintained in only few sites (sites I and V). The variations, wherever they occurred, are reflected in terms of either splitting or shifting of the hydrophilic site, or change in the length of the site. In their studies on mammalian hemoglobins and

Table 2. Comparison of the sites on α-hemoglobin obtained by using average hydrophilicity values with those observed experimentally by earlier workers[b,c].

Site No.	Site based on hydrophilicity[a]	Experimentally observed sites			
		Human[b]	Rabbit[c]	Mouse[c]	Goat[c]
I	9– 17	—	—	—	—
IIa,b,c	19– 42	20– 37	21– 36	21– 36	24– 32
III	57– 79	53– 84	70– 83	70– 83	62– 75
IV	83– 94	—	—	—	—
Va,b	105–127	106–122	107–121	107–121	107–121
VIa	137–142	128–142	126–134	126–134	129–141
VIb	149–153	138–162	149–161	149–161	146–154

[a]Sites are depicted by the range covered by various species shown in figure 3.
[b]Kazim and Atassi (1980).
[c]Kazim and Atassi (1982).

myoglobins based on the nucleotide replacements, Barnabas *et al.* (1978) have suggested that primate myoglobins evolve at a slower rate than primate hemoglobins. Our observations are also suggestive of the fact that hemoglobins are evolving faster than myoglobins. Tetrameric hemoglobin is known to carry out variety of functions not observed with myoglobin. Relatively faster evolution of hemoglobins may be the manifestation of the much diversified functional demands placed on them.

Comparison of hydrophilicity values averaged over all the species for the 3 proteins are given in figure 6. This figure allows one to see the differences in the hydrophilic character of various sites on different proteins irrespective of species being studied. It can be seen from figure 6 that hydrophilic characteristics of sites I to IV on myoglobins are drastically different from the other proteins while they are more or less same in case of sites V and VI. Site III on myoglobins has highest hydrophilic character amongst all the sites, while site II is appreciably weaker in this respect. In keeping with the single crystal structures obtained by X-ray diffraction method for deoxymyoglobin and hemoglobin (Dickerson and Geis, 1983), almost all the residues which we have predicted as antigenic are found to be on the surface of the molecules. It may be mentioned here that we have determined antigenic determinants on α- and β-chains independently in their monomeric forms. However, when hemoglobin molecule is formed, a few residues which are on the surface of the individual α- and β-chains do not remain on the surface but fall in the interior of the molecule. In the case of hemoglobins site I is more hydrophilic in α-chains as compared with β-chains while reverse is the case with site II, thus explaining the maintenance of delicate balance in hydrophilic character of the protein as mentioned earlier. In general it can be said that in all species hydrophilic regions show preference for the N-terminal half of the protein chain making these regions of the chain more exposed to the environment. The data further suggest that in case of hemoglobins the hydrophilic regions are much more flexible and have evolved due to the strong interactions with the environment. It is known that hemoglobins have evolved from myoglobins, leading to α- and β-forms most probably through the mechanism of gene duplication.

The above observations indicate that in case of related proteins the changes in

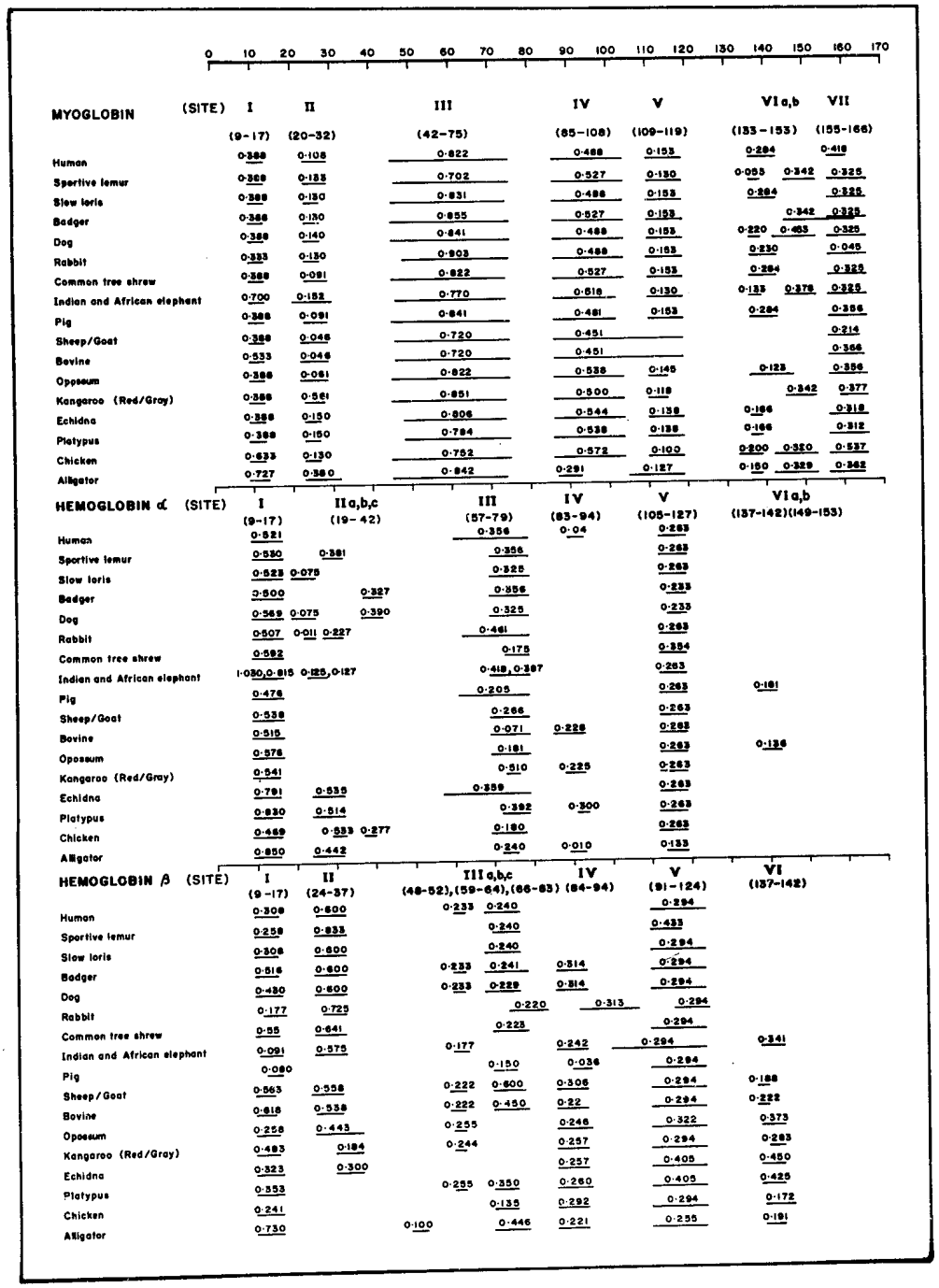

Figure 5. Comparative data on hydrophilic sites on myoglobins, α- and β-chains of hemoglobins from various species.

the characteristics of antigenic sites are maintained at the minimal level and, therefore, the hydrophilic sites on a group of related proteins could be used to tune the data on the antigenic determinants obtained experimentally for a single protein.

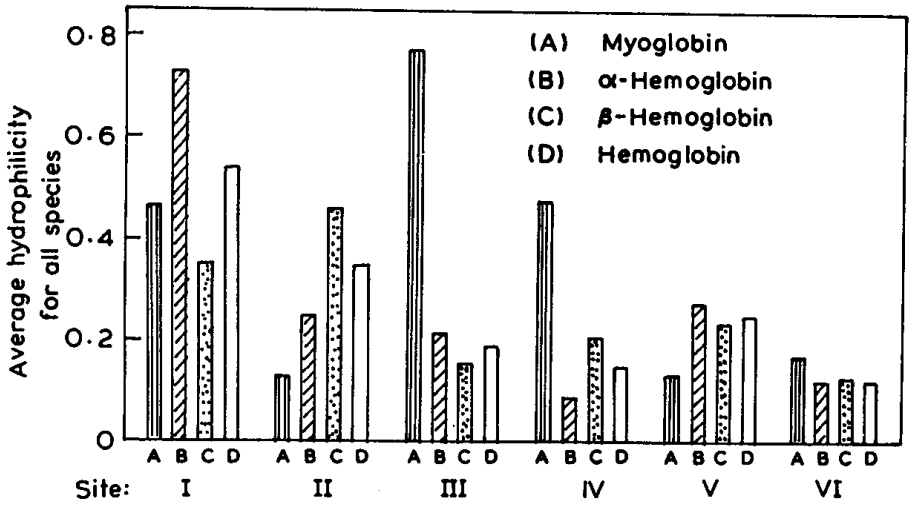

Figure 6. Hydrophilicity for various sites for 3 proteins averaged over all the species analysed.

The results also support the view that hemoglobins which perform a variety of functions are evolving faster than myoglobins.

References

Atassi, M. Z. (1975) *Immunochemistry*, **12,** 423.

Atassi, M. Z. (1978) *Immunochemistry*, **15,** 909.

Barnabas, J., Mathew, P. A., Ratnaparkhi, M. V. and Barnabas, S. (1978) *Indian J. Biochem. Biophys.*, **15,** 388.

Dickerson, R. E. and Geis, I. (1983) in *Hemoglobin: Structure, function, evolution and pathology* (ed. Philip Hagopien) (California: The Benjamin/Cummings Publishing Co.) p. 19.

Fraga, S. (1982) *Can. J. Chem.*, **60,** 2606.

Hopp, T. P. and Woods, K. R. (1981) *Proc. Natl. Acad. Sci. USA*, **78,** 3824.

Hopp, T. P. and Woods, K. R. (1983) *Mol. Immunol.*, **20,** 483.

Hunt, L. T., Hurst–Calderone, S. and Dayhoff, M. O. (1978) in *Atlas of protein sequence and structure*, (ed. M. O. Dayhoff) (Maryland: The National Biomedical Research Foundation) Vol. 5, Suppl. 3, p. 229.

Kazim, A. L. and Atassi, M. Z. (1980) *Biochem. J.*, **191,** 261.

Kazim, A. L. and Atassi, M. Z. (1982) *Biochem. J.*, **203,** 201.

Levitt, M. (1976) *J. Mol. Biol.*, **104,** 59.

Pandit, M. W. (1985) *Int. J. Peptide Protein Res.*, **25,** 197.

Westhoff, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A. and Van Regenmortel, M. H. V. (1984) *Nature (London)*, **311,** 123.