Journal of Chromatography, 507 (1990) 33–44
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

CHROMSYMP. 1710

High-performance liquid chromatography of amino acids, peptides and proteins

CI[a]. Identification and characterisation of coulombic interactive regions on sperm whale myoglobin by high-performance anion-exchange chromatography and computer-graphic analysis

A. N. HODDER
Department of Biochemistry and Centre for Bioprocess Technology, Monash University, Clayton, Victoria 3168 (Australia)
K. J. MACHIN
St. Vincent's Institute of Medical Research, Princes St., Fitzroy, Victoria 3065 (Australia)
and
M. I. AGUILAR and M. T. W. HEARN*
Department of Biochemistry and Centre for Bioprocess Technology, Monash University, Clayton, Victoria 3168 (Australia)

SUMMARY

A combination of high-performance liquid chromatographic and computer-graphic analysis of the electrostatic interactive surface of sperm whale myoglobin has allowed the putative anion-exchange binding domains to be identified and characterised. These studies have established the existence of four regions of high electrostatic potential in terms of negatively charged amino acid side chain groups. These binding sites are comprised of seven glutamic acid residues, two aspartic acid residues and two propionic acid groups located on the heme moiety which are arranged in a continuous topographic band around one section of the myoglobin surface. The number of charged groups in these interactive patches correlated with the range of experimentally determined Z_c values under different elution conditions of varying gradient time, flow-rate and displacer salt. The changes in Z_c values observed under some chromatographic conditions have been interpreted in terms of the conformational reorientation of the myoglobin molecule as it interacts with the charged stationary phase surface.

[a] For Part C, see ref. 8.

0021-9673/90/$03.50

INTRODUCTION

Recent advances in the development of retention models to describe the interactive properties of proteins separated by high-performance liquid chromatography have demonstrated a complex dependence of chromatographic behaviour on protein three-dimensional structure[1,2]. In particular, the ability of the stationary phase to act as a topographic probe for the surface of a protein solute has provided significant insight into the factors which control the mechanistic basis of protein–surface interactions. For example, protein retention in high-performance ion-exchange chromatography (HPIEC) arises from complex electrostatic interactions between the protein solute and the coulombic sorbent and is dependent on both the number and distribution of charged sites on the protein surface. As demonstrated in various recent studies[3–7], over a limited range of displacing salt concentration, c, the dependence of protein retention of log k' (capacity factor) on log $(1/c)$ can be approximated by the following linear relationship

$$\log k' = \log K + Z_c \log (1/c) \quad (1)$$

where K is an ion-exchange distribution coefficient obtained by extrapolation of plots of log k' versus log $(1/c)$. According to the stoichiometric displacement model[3], the Z_c term, which is the slope of these plots, represents the number of charges involved in the binding of the protein to the stationary phase surface. In anion-exchange chromatography, the magnitude of Z_c is dependent on the buffer pH[3,4], the salt or buffer concentration[5], the type of displacer salt[6,7] and the mode of elution[6]. This variation in Z_c is therefore associated with changes in the interactive surface structure, or the ionotope of the protein. The ability to predict the number of surface regions of high electrostatic potential and to determine the relative interactive or ionotopic dominance of these sites under a particular set of chromatographic conditions would clearly provide significant insight into protein retention behaviour. In an associated study[8], we have characterised the anion-exchange binding site on hen egg white lysozyme (HEWL) in terms of the relative distribution of charged regions on the protein surface through correlation of chromatographic properties and X-ray crystallographic data. Furthermore, variations in the range of Z_c values were interpreted in terms of conformational flexibility of the HEWL molecule during the retention process. The present paper reports the results of a similar analysis of the anion-exchange chromatographic properties and crystallographic data of sperm whale myoglobin (SWM). Regions of negative charge arising from spatially or sequentially juxtaposed glutamic acid and aspartic acid residues were located and analysed in terms of their charge density and surface accessibility, and the data utilised in the development of a model for the interaction of SWM with the positively-charged anion-exchange support surface.

MATERIALS AND METHODS

High-performance anion-exchange chromatography

All chromatographic experiments were performed with a Pharmacia (Uppsala, Sweden) fast-protein liquid chromatography (FPLC) system as previously described[5].

Computer-assisted molecular modelling

Protein-molecular modelling studies were carried out with a Silicon Graphics (Mountain View, CA, U.S.A.) Iris 3120 computer system.

RESULTS AND DISCUSSION

Myoglobin is a single polypeptide chain comprised of 153 amino acid residues, the primary structure of which is shown in Fig. 1. Compared to many globular proteins the myoglobin molecule is unique in terms of its high α-helix content. X-ray crystallographic[9] studies indicate that *ca.* 75% of the amino acid residues are arranged in segments of right-handed α-helices. These helical arrangments lead to a compact and roughly spheroidal molecule with dimensions 45 × 35 × 25 Å. In terms of the anatomy and taxonomy of proteins, myoglobin is representative of the Greek key helix bundle sub-group of proteins characterised by an antiparallel arrangement of α-helical domains[10]. Most polar amino acid residues are located on the surface of the protein while the interior is largely comprised of non-polar residues.

Myoglobin functions as an oxygen-storing protein in skeletal muscle cells. This is achieved by the formation of a complex between oxygen and the prosthetic heme group of the protein. The iron-containing heme group is planar and located in a small pocket between two α-helices, termed the E- and F-helices[9]. The iron atom, located at the centre of the heme, is bound at the fifth coordination site to the imidazole nitrogen of histidine 93 while oxygen complexes at the sixth coordination site. While two polar propionic acid groups of the heme are orientated at the surface of the protein, the heme moiety is predominantly buried within the myoglobin structure.

Myoglobin has an isoelectric point (pI_{SWM}) of 7.68. In anion exchange chromatographic systems where the buffer pH is greater than pI_{SWM}, SWM will carry an overall net negative charge, due to the ionisation of glutamic acid (Glu) and aspartic acid (Asp) residues. Myoglobin contains 21 acidic residues which contribute to the surface charge of the protein which could all potentially interact with a positively charged stationary phase surface. While there will be several areas of surface negative charges, the interaction of SWM-related proteins with charged surfaces will, however, be mediated through regions of highest electrostatic potential in terms of negative charge density. In chromatographic systems, the charge stoichiometry of this interaction, which depends on the degree of protein ionisation and the extent of salt–charge interactions in the dominant ionotope, is quantitated by the parameter Z_c. Table I shows the experimentally determined Z_c values representing the number of charged

```
        *         *         *         *         *         *

VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASED

        *         *         *         *         *         *

LKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHP

        *         *         *

GDFGADAQGAMNKALELFRKDIAAKYKELGYQG
```

Fig. 1. Primary structure of sperm whale myoglobin. The asterisks denote every tenth residue.

TABLE I

Z_c VALUES FOR MYOGLOBIN AT pH 9.60

Salt	Z_c		
	Elution mode		
	Varied gradient time, constant flow-rate	*Varied flow-rate, constant gradient time*	*Constant flow-rate, isocratic elution*
LiCl	17.75 ± 3.23	6.72 ± 3.11	0.82 ± 0.13
LiBr	1.96 ± 0.66	4.11 ± 0.44	–[a]
NaF	5.49 ± 1.71	1.46 ± 0.14	2.36 ± 0.65
NaCl	4.61 ± 1.57	5.09 ± 0.80	2.76 ± 0.31
NaBr	1.43 ± 0.02	3.48 ± 0.59	2.36 ± 0.69
KF	1.98 ± 0.27	1.32 ± 0.12	–[a]
KCl	4.54 ± 0.75	1.47 ± 0.15	1.43 ± 0.14
KBr	4.84 ± 0.75	6.03 ± 2.26	–[a]

[a] Not eluted with this salt.

interactions between SWM and the positively charged Mono-Q column. These data were derived under elution conditions of pH 9.6 with a variety of alkali metal halide salts as the displacer salt. The Z_c values obtained under these elution conditions, where all acidic amino acid side chain groups will be ionised, were generally in the range between 1 and 6. This result suggests that only a portion of the negatively charged amino acid side chain groups are involved in the interactive process.

Computer simulation of the three-dimensional structure for SWM was therefore carried out to search for clusters of surface accessible glutamic acid (Glu) and aspartic acid (Asp) residues. Initial inspection of the X-ray crystallographic data revealed a complex surface distribution of negative and positive charges as shown in Figs. 2 and 3, respectively. Aspartic acid and glutamic acid residues are shown in red and yellow while pink and blue areas represent the positively charged lysine (Lys) and arginine (Arg) residues, respectively. Because of spatial constraints associated with the protein–ligand interaction four areas on the SWM surface were considered to have the potential to act as anion-exchange binding sites. Generally, it is anticipated that surface interactive patches occur as a result of either sequentially-linked amino acids or via topographic arrangements involving through-space alignment of acidic residues. The patches of negative charge identified on the SWM surface all arise from non-sequential juxtaposition of the Glu and Asp residues.

The first interactive site is shown in Fig. 4, where Glu 105 and Glu 109 located in the G-helix, and Glu 136 in the H-helix, were found to be in close spatial proximity. Another potential anion-exchange binding site located nearby and at the top of the molecule in Fig. 4 incorporates Glu 36, Glu 41 which are both located in the C-helix and Asp 44 in the non-helical CD region. The remaining two sites which are potentially involved in the anion-exchange binding process both incorporate the heme prosthetic group. Fig. 5 shows a view into the cleft containing the heme group where the depth of the picture has been reduced by Z-clipping (a function of the graphics software) to allow clear observation of the potential interactive sites. The surface-

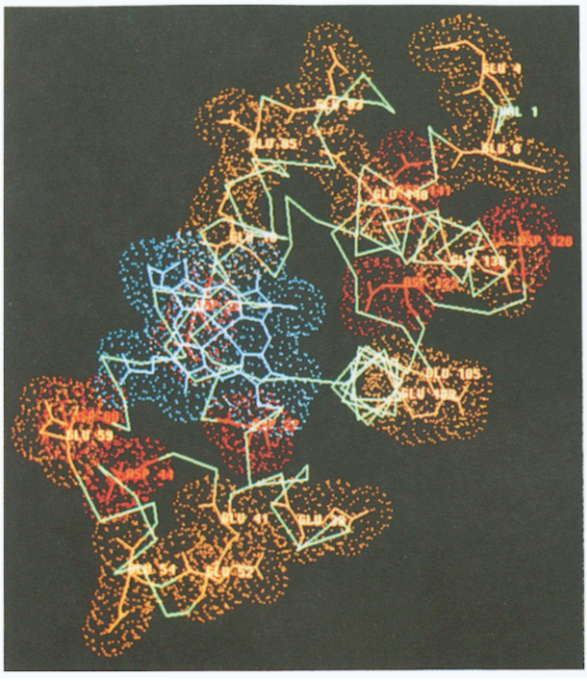

Fig. 2. Computer-graphic simulation of the X-ray crystal structure of SWM showing the location of aspartic acid (red) and glutamic acid (yellow) residues. The prosthetic heme group is shown in blue. Picture produced by a computer program written by A. M. Lesk and K. D. Hardman[22].

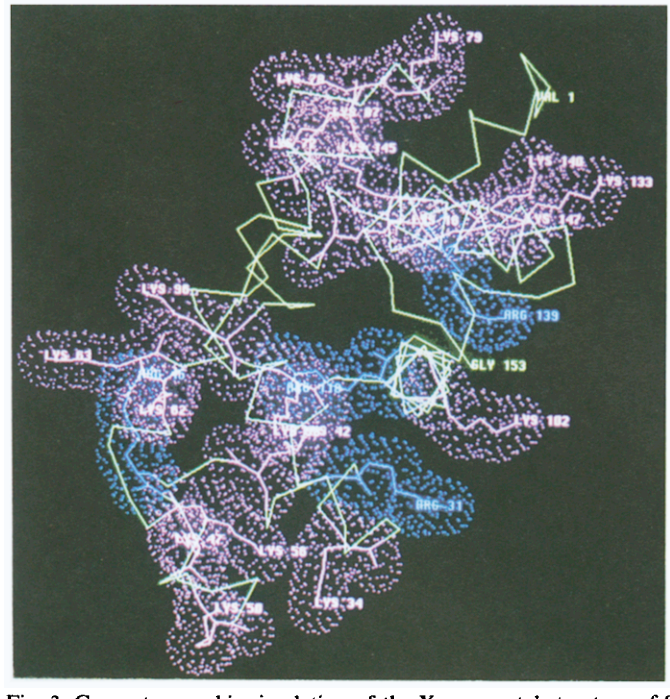

Fig. 3. Computer-graphic simulation of the X-ray crystal structure of SWM indicating the location of lysine (pink) and arginine (blue) residues. Picture produced by a computer program written by Lesk and Hardman[22].

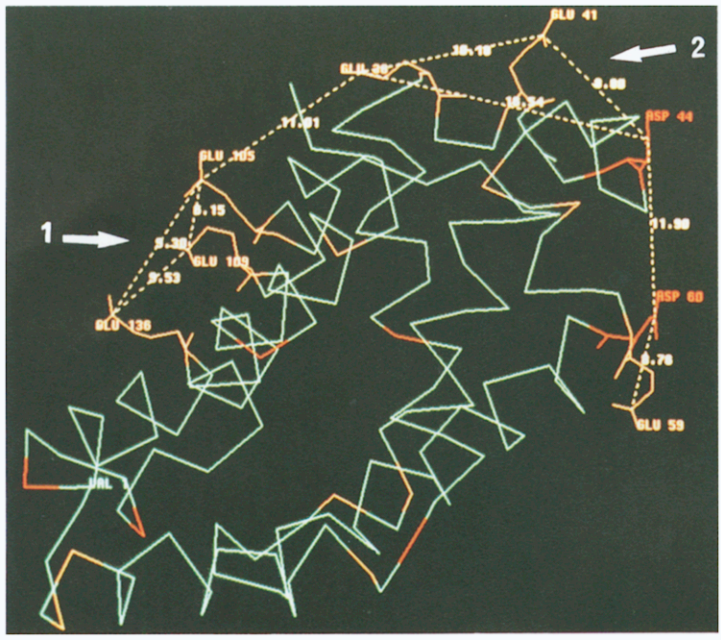

Fig. 4. Computer-graphic representation of the orientation of negatively charged groups located around the heme-derived propionic acid side chains. Contour lines were drawn to obtain the distances between the Glu 105, 109 and 136 in site 1 and Glu 41, 36 and Asp 44 in site 2. The charge densities of these triangles are listed in Table II. Picture produced by a computer program written by Lesk and Hardman[22].

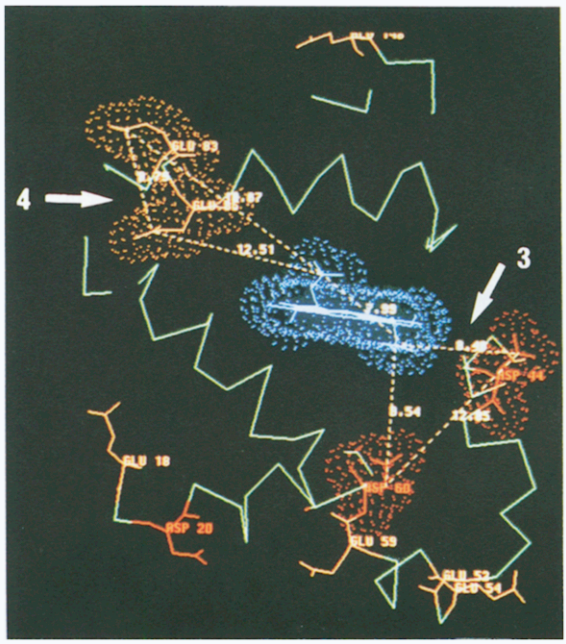

Fig. 5. Computer-graphic analysis of the distances between Asp 44, 60 and one heme propionic acid group (site 3) and Glu 83, 85 and a second heme propionic acid group (site 4). The resulting charge density values are listed in Table II. The spatial depth of the picture has been reduced by Z-clipping to facilitate visual analysis. Picture produced by a computer program written by Lesk and Hardman[22].

accessible propionic acid side-chains of the porphyrin ring have a *trans*-configuration. At high pH both propionic acid groups will be negatively charged and have the potential to participate in electrostatic interactions. The close spatial proximity of Asp 60, 44 and Glu 83, 85 to the propionic acid groups generates one if not two areas that have the potential to participate in the anion-exchange binding process.

Protein–surface interactions in anion-exchange chromatography will be mediated through the negatively charged regions of highest electrostatic potential. The ranking of these areas in terms of the magnitude and direction of the charge vector will provide further information on the relative significance of these four regions in terms of their role in the approach and orientation of the protein solute to the stationary phase in the intitial stages of adsorption. The charge density of the four potential interactive sites was therefore determined to assess the relative influence of each site in both steering the molecule toward the support surface and its position in the interactive hierarchy. In Figs. 4 and 5 the distances between proximal negatively charged groups are illustrated. Sites 1 and 2 are shown in Fig. 4 while sites 3 and 4 incorporating the heme group are shown in Fig. 5. Estimates of the charge density (charge per unit area) were calculated using the areas obtained for the dotted contours of each binding site and the resultant values are shown in Table II. Site 1, which encompasses Glu 105, 109 and 136 has the highest charge density, namely 0.37 charge/10 $Å^2$. This information indicates that these charged residues located in the G- and H-helices will under conditions of near equilibrium be important in steering the myoglobin molecule from the bulk solvent toward the stationary phase via attractive, electrostatic forces.

While the regions of high electrostatic potential will dominate the interaction of the protein solute with the anion-exchange sorbent, the relative spatial disposition of neighbouring positive and even neutral groups will also play an important role in the approach and orientation of the protein solute at the stationary phase surface. The alignment of SWM shown in Fig. 2 was rotated to allow the influence of neighbouring positively charged groups to be assessed. This indicated that the site 1 residues comprising Glu 105, 109 and 136 project from the protein surface to form an unobstructed region of negative charge. However, in site 2, comprised of Glu 36, 41 and Asp 44, the positively charged Lys 42 was found to project out of the plane directly between residues 38 and 44. As a consequence, this residue will sterically interfere with the binding process and significantly inhibit the approach and orientation of this potential binding site to the anion-exchange support surface.

TABLE II

CHARGE DENSITY FOR POTENTIAL BINDING SITES ON SPERM WHALE MYOGLOBIN

Location: Site 1 = Glu 105, 109 and 136; site 2 = Glu 41, 36 and Asp 44; site 3 = Asp 44, 60 and heme propionic acid; site 4 = Glu 83, 85 and heme propionic acid.

Site	*Charge (per 10 $Å^2$)*
1	0.37
2	0.26
3	0.25
4	0.21

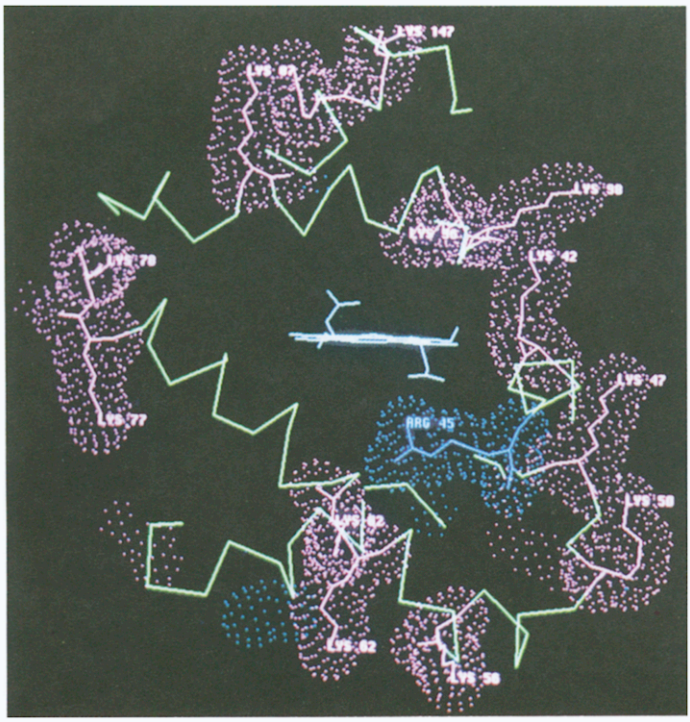

Fig. 6. Computer-graphic representation of the distribution of positively charged lysine (pink) and arginine (blue) residues in the anion-exchange binding sites 3 and 4. The spatial depth of the picture has been reduced by Z-clipping to facilitate visual analysis. Picture produced by a computer program written by Lesk and Hardman[22].

In Fig. 6, the distribution of the positively charged Lys and Arg residues around sites 3 and 4 is shown from the same viewpoint as depicted in Fig. 5. The general distribution of positive charge in this area does not directly interfere with the potential binding sites shown in Fig. 5. However, Arg 45 is located between Asp 44 and 60 and is orientated toward the lower propionic acid groups as shown in Fig. 6. In the X-ray crystal structure, Arg 45 is reported[11] to be salt-linked to this particular acid group. During binding to the anion-exchanger it is possible that electrostatic repulsion will reorientate Arg 45 away from the propionic acid side-chain to permit increased electrostatic interaction between site 3 and the stationary phase.

Close inspection of Fig. 4 also reveals that the four areas identified as anion-exchange binding sites are arranged in a continuous topographic band around a specific region on the myoglobin molecule. If the native structure is maintained, then steric restrictions will prevent all of these residues interacting with the support surface at the same time. However, a wide range of Z_c values was obtained for myoglobin eluted by various neutral salts and elution modes at buffer pH 9.6, which suggests that the molecule is undergoing some degree of conformational reorientation or unfolding at the surface of the sorbent. Inspection of the Z_c values for each of the three different elution modes (Table I) indicates that neutral salts significantly influence the size of

the anion-exchange binding site on myoglobin. Neutral salts can influence the structure of macromolecules via complicated changes to the layer of hydration and electrostatic potential at the protein surface. The small fluctuations in Z_c values observed for different salts (*i.e.*, when Z_c values remain between 1 and 3) may indicate small movements in the structural subunits or domains, which influence the degree of electrostatic interaction between residues of a single binding site and the stationary phase. Studies on the structural dynamics of myoglobin[12] indicate that this conformational equilibrium is possible as the α-helices in this protein do not only move as rigid units but may also experience rippling or breathing modes. Larger changes in the magnitude of Z_c values will indicate the occurrence of more extensive structural reorientations resulting in an increased number of residues interacting with the stationary phase surface. For example, the change from NaF to NaCl, with varied flow elution, resulted in the Z_c value increasing from *ca.* 1.5 to 5.1, respectively. If binding sites 1 to 4 are involved in the retention process then some unfolding will be necessary to permit five negatively charged residues to interact simultaneously with the support surface.

In Table I the magnitude of the Z_c value is also observed to change significantly with variation in the mode of elution. For example, the elution of SWM by NaF resulted in Z_c values of *ca.* 5.5, 1.5 and 2.4 for the varied gradient time, varied flow-rate and isocratic modes of elution. Consequently, the rate of change of the displacer salt concentration not only determines the column residence time for the protein but also affects the dynamic geometry of the protein molecule[13]. The fluctuations in Z_c observed between each elution mode in Table I could therefore reflect the preferential solute–stationary phase interaction through different binding sites on the surface of SWM.

Inspection of Fig. 1 reveals that there are 21 Glu and Asp residues in the primary sequence of SWM, while there are 11 residues implicated in anion-exchange binding sites 1–4. Although the Z_c values shown in Table I vary from *ca.* 1 to 18, the majority of values fall between 1 and 6. If the three-dimensional structure of SWM as it interacts with the stationary phase surface, is approximated by the X-ray crystal structure, then allowing for some steric restrictions, the range of Z_c values of 1–6 in Table I indicates that only a proportion of these sites will be involved in the binding process. However, when myoglobin is eluted with LiCl, under varied gradient time conditions, the Z_c value (*i.e.*, $Z_c = 17.75$) approaches the total number of Glu and Asp residues present in this protein. This result almost certainly reflects the situation where the majority of Glu and Asp groups have electrostatically bound to the stationary phase during elution. Such an event would require the total unfolding of the myoglobin molecule. One mechanism through which this apparent surface-mediated phenomenon may occur is via the initial binding of the "steering" residues in site 1 to the stationary phase. The molecule could then unfold in a rolling manner as adjacent sites (*i.e.*, 1 to 4) progressively bind as a preferred sequence of events. This phenomenon, shown schematically in Fig. 7, is analogous to the unrolling of a carpet, and would permit an increased number of electrostatic interactions to arise from the geometrical flattening of the myoglobin molecule.

The physicochemical processes which underly the interaction of proteins with chromatographic surfaces are analogous to the environmental and structural factors which control cellular biorecognition processes such as hormone–receptor and antibody–antigen reactions. The surface structures of a number of proteins have been

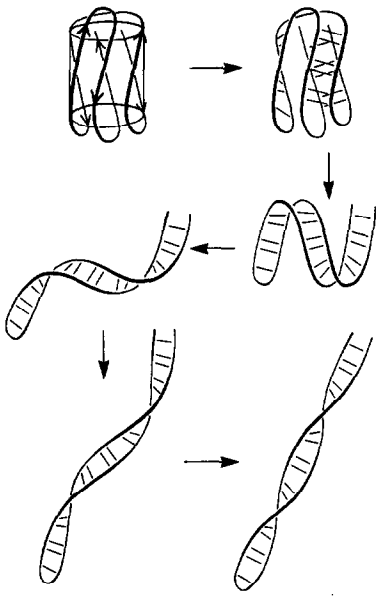

Fig. 7. Schematic representation of the conformational unravelling of the helical domains of SWM.

probed in considerable detail by immunochemical techniques. In particular, a variety of antigenic determinants (epitopes) have been mapped on SWM using anti-myoglobin antibodies and T-cell clones. Up to nine continuous epitopes, comprising sequentially linked amino acid residues, have been identified immunologically by recognition of a variety of peptides derived from the digestion of the native protein[14–18]. Furthermore, monoclonal antibodies to SWM have also been shown to recognise determinants that differ from the continuous epitopes. These differences were attributed to the variation in conformation associated with topographically —rather than sequentially— linked residues on the protein surface involved in antibody binding. Similarly, two further antigenic determinants on the G- and H-amphipathic helices in myoglobin have been defined by reactivity with T-cell clones[19–21]. The factors which predispose particular amino acid residues to be antigenic or immunogenic are not fully known or understood. While it is generally considered that accessibility of surface amino acid residues is a major contributing factor, the precise physicochemical requirements of the epitope structure, such as the number of hydrophobic, hydrophilic and neutral residues, have not yet been determined. In Table III the amino acid residues identified in each antigenic determinant of SWM are compared to the residues predicted to constitute electrostatic interactive ionotopes at the surface of SWM. Most significantly, all the residues that define the anion-exchange binding sites are also located within the known B-cell and T-cell epitopes. Moreover, the fourteen epitopes listed in Table III all contain charged residues (*i.e.*, Glu, Asp, Lys or Arg) and eleven of these epitopes contain Glu or Asp groups.

It is apparent from these observations that the distribution and surface accessibility of Glu and Asp residues play an important role in the overall antigenicity of myoglobin. The results of the present study therefore demonstrate that the correla-

TABLE III

COMPARISON OF THE POTENTIAL ANION-EXCHANGE BINDING SITES TO THE KNOWN ANTIGENIC DETERMINANTS ON SPERM WHALE MYOGLOBIN

Residues	
Anion-exchange sites	*Antigenic sites*
	B-cell continuous epitopes[14–18]
1 = 105, 109 and 136	1–6, 15–22, 22–55, 56–62,
2 = 36, 41 and 44	72–89, 94–99, 113–119,
3 = 44, 60 and heme propionic acid	121–127 and 145–151
4 = 83, 85 and heme propionic acid	
	B-cell topographic epitopes[19]
	Mab 1 = 4 and 79
	Mab 2 = 83, 144 and 145
	Mab 3 = 140, ?
	T-cell epitopes[20–21]
	clone 1.2, 9.27 = 102–118
	clone 14.5 = 132–153

tion between anion-exchange chromatographic data and X-ray structural data can be used to partially define and characterise immunologically active amino acid residues on SWM. These studies also document further an experimental approach pertinent to the characterisation of protein–surface reactivity in biological systems via detailed topographic mapping of protein-interactive surfaces through a combined analysis of chromatographic data derived from electrostatic, hydrophobic or related modes of protein–ligand interactions.

ACKNOWLEDGEMENTS

Research grants to M.T.W.H. from the National Health and Medical Research Council, the Australian Research Council and the Monash University Research Fund Committee are gratefully acknowledged.

REFERENCES

1 M. T. W. Hearn and M. I. Aguilar, in Z. Deyl, (Editor), *Separation Methods*, Elsevier, Amsterdam, 1988, p. 107.
2 F. E. Regnier, *Science (Washington, D.C.)* 238 (1987) 319.
3 W. Kopaciewicz, M. A. Rounds, J. Fausnaugh and F. E. Regnier, *J. Chromatogr.*, 266 (1983) 3.
4 M. T. W. Hearn, A. N. Hodder, P. G. Stanton and M. I. Aguilar, *Chromatographia*, 24 (1987) 769.
5 M. T. W. Hearn, A. N. Hodder and M. I. Aguilar, *J. Chromatogr.*, 458 (1988) 27.
6 W. Kopaciewicz and F. E. Regnier, *Anal. Biochem.*, 133 (1983) 251.
5 M. T. W. Hearn, A. N. Hodder and M. I. Aguilar, *J. Chromatogr.*, 443 (1988) 97.
8 M. T. W. Hearn, A. N. Hodder and M. I. Aguilar, *J. Chromatogr.*, submitted for publication.
9 J. C. Kendrew, R. E. Dickerson, B. E. Stradberg, R. G. Hart, D. R. Davies, D. C. Phillips and V. C. Shore, *Nature (London)*, 185 (1960) 422.
10 J. S. Richardson, *Adv. Protein Chem.*, 24 (1981) 167.
11 J. C. Kendrew, *Brookhaven Symp. Biol.*, 15 (1962) 216.
12 H. Frauenfelder, G. A. Petsko and D. Tsernoglou, *Nature (London)*, 280 (1979) 558.
13 A. N. Hodder, M. I. Aguilar and M. T. W. Hearn, *J. Chromatogr.*, 476 (1989) 391.

14 M. Z. Atassi, *Eur. J. Biochem.*, 145 (1984) 1.
15 H. E. Schmitz, H. Atassi and M. Z. Atassi, *Mol. Immunol.*, 19 (1982) 1699.
16 C. R. Young, H. E. Schmitz and M. Z. Atassi, *Mol. Immunol.*, 20 (1983) 567.
17 H. E. Schmitz, H. Atassi and M. Z. Atassi, *Immun. Commun.*, 12 (1983) 161.
18 S. J. Leach, *Biopolymers*, 22 (1983) 425.
19 J. A. Berzofsky, G. K. Buckenmeyer, G. Hicks, F. R. N. Gurd, R. J. Feldmann and J. Minna, *J. Biol. Chem.*, 257 (1982) 3189.
20 H. Z. Streicher, I. J. Berkower, M. Busch, F. R. N. Gurd and J. A. Berzofsky, *Proc. Natl. Acad. Sci., U.S.A.*, 81 (1984) 6831.
21 K. B. Cease, I. Berkower, J. York-Jolley and J. A. Berzofsky, *J. Exp. Med.*, 164 (1986) 1779.
22 A. M. Lesk and K. D. Hardman, *Science (Washington, D.C.)*, 216 (1982) 539.