

RESEARCH ARTICLE

Rigid and flexible regions in lysozyme and the invariant features in its hydration shell

Madhusudan and M. Vijayan

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

Water-mediated transformations provide a useful handle for exploring the flexibility in protein molecules and the invariant features in their hydration shells. Low-humidity monoclinic hen egg white lysozyme, resulting from such a transformation, has perhaps the lowest solvent content observed in any protein crystal so far and has a well-ordered structure. A detailed comparison involving this structure, low-humidity tetragonal lysozyme, and the other available refined crystal structures of the enzyme permits the delineation of the relatively rigid, moderately flexible and highly flexible regions of the molecule. The relatively rigid region forms a contiguous structural unit close to the molecular centroid and encompasses parts of the main β -structure and three α -helices. The hydration shell of the protein contains 30 invariant water molecules. Many of them are involved in holding different parts of the molecule together or in stabilizing local structure. Five of the six invariant water molecules attached to the substrate-binding region form part of a water cluster contiguous with the side-chains of the catalytic residues Glu-35 and Asp-52.

FLEXIBILITY OF PROTEIN MOLECULES¹⁻⁷ AND THE INVARIANT

CURRENT SCIENCE, VOL. 60, NO. 3, 10 FEBRUARY 1991

features in their hydration shell⁸⁻¹² are problems of considerable current interest. Water-mediated transformations, first described in haemoglobin in the early days of protein crystallography^{13,14} and recently shown by us to occur in many protein crystals^{8,15,16}, provide a useful handle for exploring these two related problems. In these transformations, the unit cell dimensions, the diffraction pattern and the solvent content of protein crystals change abruptly, typically in the relative humidity range 90–93%, when the environmental humidity is systematically varied¹⁶. In terms of composition, the difference between the native and the low-humidity forms is only in the amount of bulk water in the crystals. The change in the amount of bulk water, however, leads to significant changes in the hydration shell, which in turn cause structural perturbations in the protein molecule⁸. Withdrawal of a small amount of water from the solvent regions in the crystal, as in the water-mediated transformations outlined above, is perhaps the gentlest way to cause a structural transformation. The changes that accompany the transformation are therefore likely to correspond to the

intrinsic structural variability of the protein molecule and its hydration shell.

Monoclinic hen egg white (HEW) lysozyme¹⁷ exhibits the most remarkable water-mediated transformation observed so far¹⁶. The two crystallographically independent molecules in the native crystals become equivalent in the low-humidity form and the volume of the solvent regions decreases by nearly 40% to 22%. The transformation to the low-humidity form is also accompanied by a substantial improvement in the quality of the diffraction pattern. The structure of this form has been solved by the molecular replacement method¹⁸ and refined using the restrained least-squares method¹⁹ to an *R* value of 0.175 for 7684 photographically observed reflections in the 10–1.75-Å resolution range (Madhusudan, R. Kodandapani and M. Vijayan, unpublished results). The atomic positions in the structure are well defined, with average *B* values of 8.1, 13.8, 10.8 and 29.7 Å² for main-chain atoms, side-chain atoms, all protein atoms and solvent atoms respectively. Also, this form has been analysed at a higher resolution than the other crystal forms of HEW lysozyme, for which detailed descriptions are currently available. The extremely low solvent content, the high accuracy of the structure and the high resolution at which the structure has been analysed make low-humidity lysozyme a suitable candidate for use as a reference when comparing structural features in different crystal forms of the enzyme.

HEW lysozyme is the third protein and the first enzyme to be X-ray analysed²⁰. The three-dimensional structure of the enzyme derived from its tetragonal crystals is very well known²¹. The structure has been refined at 2-Å resolution (D. C. Phillips, private communication). Recently we have analysed the low-humidity form of tetragonal lysozyme at 2.1-Å resolution⁸. 2-Å resolution structures of triclinic lysozyme²² and a high-pressure form of tetragonal lysozyme^{23, 24} are also currently available. None of the other crystal forms of HEW lysozyme reported so far has been completely refined at high resolution.

The native, the low-humidity and the high-pressure tetragonal lysozyme, the native triclinic lysozyme, and the low-humidity monoclinic lysozyme represent a variety of environmental conditions and crystal packing. Comparison among the five crystal structures should therefore help in delineating the relatively rigid and flexible regions in the protein molecule and in identifying the invariant features in its hydration shell.

Rigid and flexible regions in the molecule

In order to delineate the rigid and flexible regions, a composite difference distance matrix^{23, 25} was constructed from 10 individual matrices, each corresponding to a

pair of the five crystal structures. In each individual matrix, an element y is the distance between the α -carbons of the i th and j th residues in one structure minus that in the other. In the composite matrix, shown in Figure 1, an element is left blank if the corresponding element in all the ten matrices is less than 0.5 Å. It appeared reasonable to treat a set of α -carbon atoms as representing the relatively rigid region of the molecule if all the distances among them lead to blank elements in the composite matrix. The 43 α -carbon atoms that fulfil this condition have 903 unique non-zero vectors among them. The condition implies that the magnitude of none of these vectors differs by more than 0.5 Å between any pair of the five structures considered. Thus, the criterion employed for delineating the relatively rigid region of the molecule is rigorous although the choice of the cut-off value as 0.5 Å is somewhat arbitrary. A careful examination of error estimates, where available, in the five structures and other refined protein structures analysed at comparable resolutions indicated that differences greater than 0.5 Å could be considered significant.

The 43 α -carbon positions representing the relatively rigid region of the molecule are illustrated in Figure 2. They belong to residues 5 to 8, 10, 28 to 36, 38, 50 to 65, 69, 75 to 77, and 89 to 96. The longest stretch of polypeptide chain in the region is the 16-member fragment 50 to 65, which encompasses a major part of

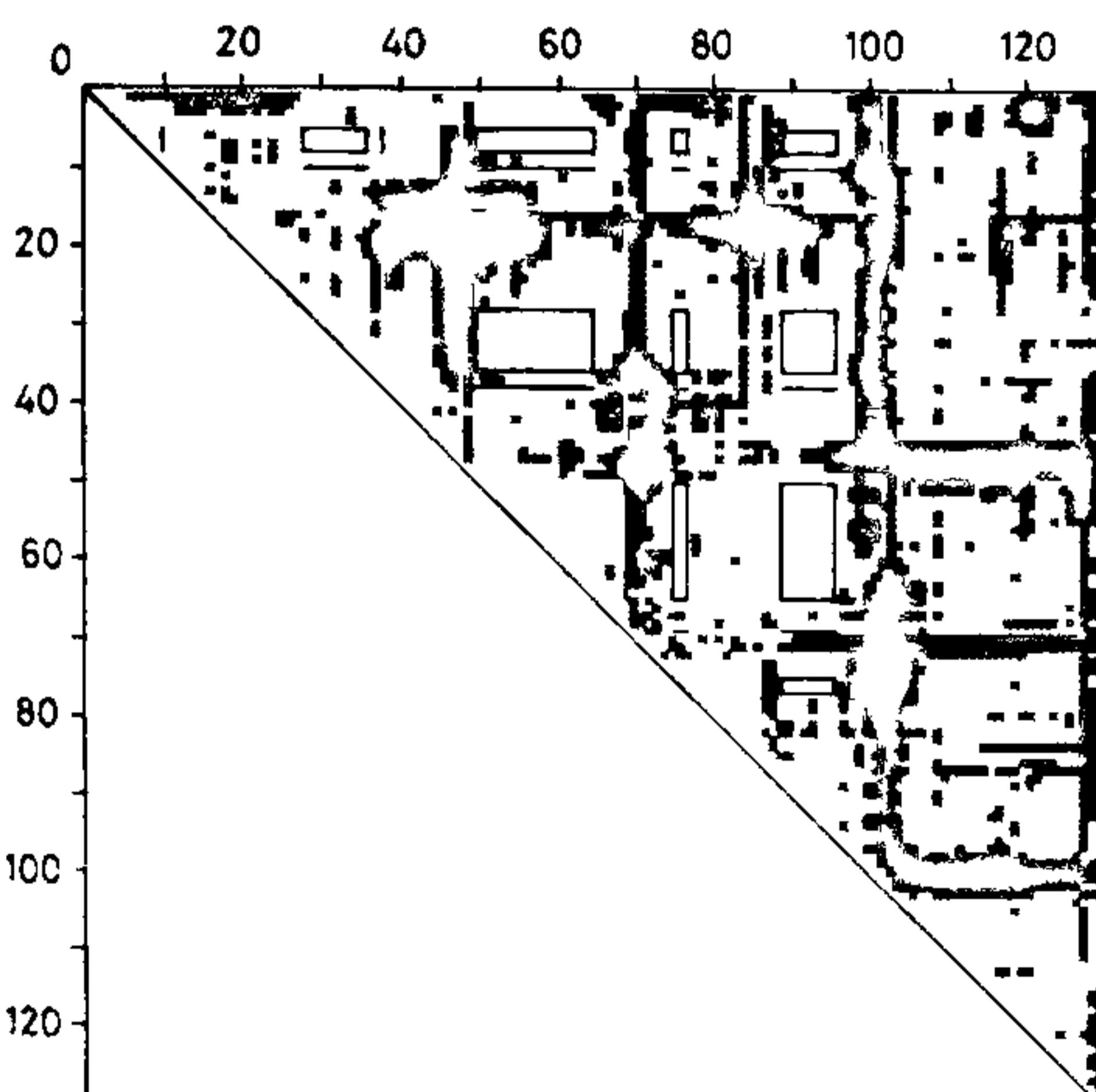


Figure 1. Composite difference distance matrix for α -carbon atoms, involving five crystal structures of HEW lysozyme. An element in it is left blank if the corresponding element in all the ten matrices is less than 0.5 Å. The boxes and lines correspond to the α -carbons that are relatively rigid. No vectors among these α -carbons differ by more than 0.5 Å between any pair of the structures concerned.

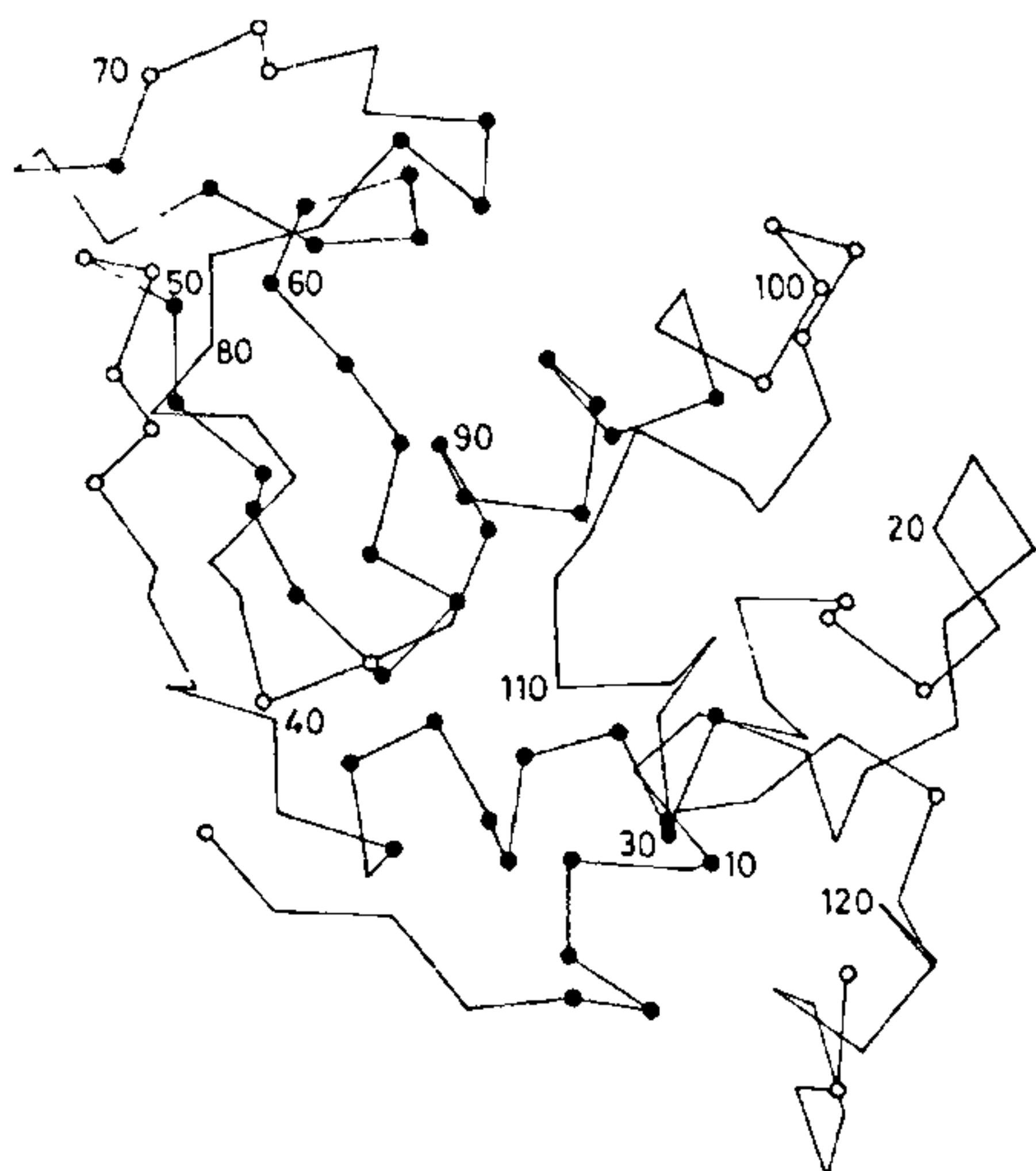


Figure 2. α -Carbon positions in the HEW lysozyme molecule. Atoms in the relatively rigid region and those in the highly flexible regions are denoted by filled and open circles respectively. Residue numbers are indicated.

the main β -structure in lysozyme. Fragments 28 to 36 and 89 to 96 encompass major parts of the second and the third α -helices respectively in the molecule, while residues 5 to 8 and 10 form part of the first α -helix. A careful examination shows that the different peptide stretches and residues in the relatively rigid region of the molecule are connected among themselves through hydrogen bonds, a disulphide bridge and van der Waals interactions, and the region thus forms a contiguous structural unit.

A second composite difference distance matrix was constructed, now with a cut-off value of 1 \AA . The α -carbon atom positions, 107 in number, all the distances among which have blanks in the composite map, belong to residues 2 to 15, 19 to 44, 50 to 69, 73 to 85, 88 to 98, 104 to 116, and 118 to 127. They encompass a substantial part of the molecule, including all the four α -helices, the two 3_{10} helices, a major part of the β -structure, and part of the main loop. Excluding the 43 residues belonging to the relatively rigid region, which obviously form a subset of the 107 residues obtained on the basis of a cut-off value of 1 \AA , the remaining 64 residues may be considered to constitute the moderately flexible regions of the molecule. There are 22 more residues (1, 16 to 18, 45 to 49, 70 to 72, 86, 87, 99 to 103, 117, 128 and 129) in the polypeptide chain, and they obviously form the flexible regions of the molecule (Figure 2).

Atomic temperature factors (B values) derived from X-ray analysis often provide an estimate of molecular flexibility^{1,2,26}. In the low-humidity monoclinic form, the main-chain atoms in the relatively rigid region have an average B value of 4.4 \AA^2 . The corresponding average B values for the moderately flexible and the highly flexible regions are 6.2 and 20.5 \AA^2 respectively. Thus the conclusions regarding the relative flexibility of different regions of the molecule are corroborated, on an average, by the distribution of thermal parameters. They are also broadly consistent with earlier detailed analyses of temperature factors in tetragonal HEW lysozyme and orthorhombic human lysozyme^{1,26}.

The α -carbons in the relatively rigid, moderately flexible and highly flexible regions have average distances of 10.6 , 13.8 and 16.5 \AA respectively from the molecular centroid. Thus, on an average, regions closer to the centroid are more rigid than those farther away from it. No general correlation between flexibility or rigidity and the location of the residue on the surface or in the interior of the molecule^{1,26} was, however, readily discernible.

Of the 16 residues in the molecule (34, 35, 37, 44, 46, 52, 57, 59, 62, 63, 101, 103, 107, 108, 109 and 114) implicated in substrate binding^{21,27-29}, seven belong to the relatively rigid region, six to the moderately flexible regions and three to the highly flexible regions. Thus, as far as the main-chain conformation is concerned, the binding region does not appear to be any more flexible or rigid than the rest of the molecule. However, some of the side-chains in the region, particularly that of Trp-62, exhibit comparatively high conformational flexibility. Residues 70 to 76, although not directly involved in interactions with the substrate, are known to move during substrate binding²⁷⁻²⁹. Three of them belong to a highly flexible region.

Invariant water molecules in the hydration shell

The number of water molecules identified in the five structures varies widely. In each case, the water sites that surround and interact with the protein molecule were identified using appropriate symmetry elements. In these calculations a water molecule was considered as interacting with the protein molecule if the water oxygen is at a distance of 3.6 \AA or less from a nitrogen atom or an oxygen in the protein molecule⁸. The same distance criterion was used for defining interactions between water molecules. The number of interactions between a water molecule and the protein molecule varies from none to six. The distribution of water molecules in different structures as a function of number of interactions with the protein molecule is illustrated in Figure 3. The widest variation in the number of water molecules occurs when they do not

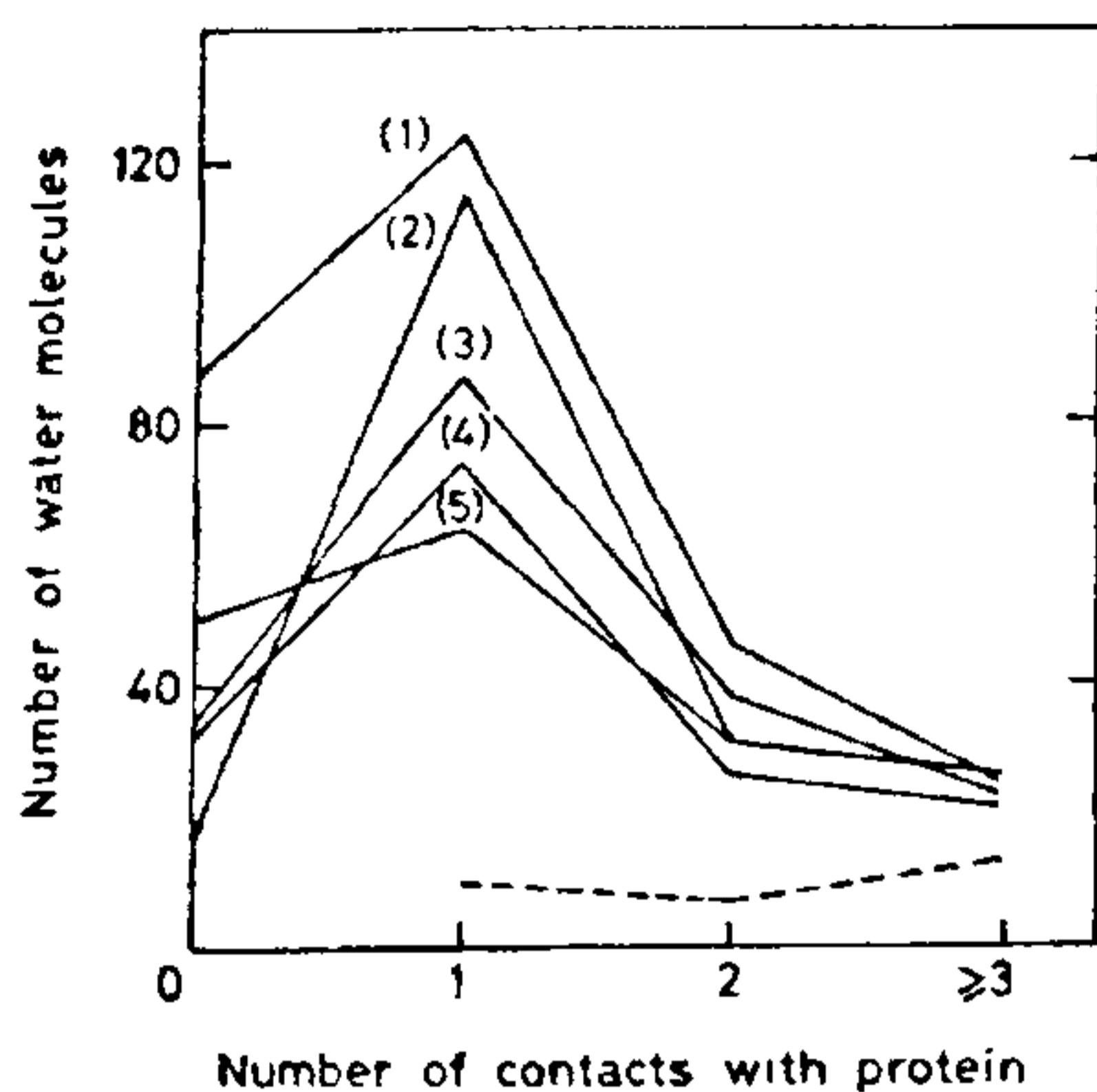


Figure 3. The distribution of water molecules as a function of the number of interactions with the protein in triclinic lysozyme (1), low-humidity monoclinic lysozyme (2), and high-pressure (3), native (4) and low-humidity (5) tetragonal lysozyme. The broken line illustrates the corresponding distribution of invariant water molecules in low-humidity monoclinic lysozyme. The hydration shell of an isolated protein molecule, constructed using appropriate symmetry elements, was considered when arriving at the numbers of water molecules with one or more interactions with the protein.

interact with the protein at all. Among the water molecules in the hydration shell, substantial variation exists when there is only one interaction. The variation is considerably reduced when there are two interactions. The number of water molecules with three or more interactions is nearly the same in all the five structures. The protein molecule, along with the water molecules in its first hydration shell, in each structure was superposed on that in the low-humidity monoclinic form using orientation matrices and translation vectors based on α -carbon positions³⁰. A water molecule in the hydration shell was considered invariant if it interacts with at least one common protein atom in all the five structures and if, in addition, the distance between the corresponding molecules in any pair of structures is less than 1.8 Å (half the maximum distance between two water molecules to be treated as interacting with each other)⁸. Such invariant water molecules, common to all structures and numbering 30, are listed in Table 1. Inasmuch as they occur in all the five structures, despite differences in the amount and the composition of the solvent, crystal packing and hydrostatic pressure, they may be considered to constitute the invariant features of the hydration shell of the enzyme. The fact that 22 of them are found among the strongly bound water molecules in human and tortoise lysozymes⁹ indicates that a substantial part of the invariant water molecules in HEW lysozyme are conserved with respect to species variation also.

Table 1. Invariant water molecules in the hydration shell of HEW lysozyme. The numbering of the water molecules corresponds to that in low-humidity monoclinic lysozyme.

Water	The protein atoms with which the water molecule interacts in all the structures	Additional protein atoms with which the water molecule interacts in at least one structure
107 W3'	93 O	—
107 W4'	90 O	—
109 W1	Asn-27 OD1	115 O, 117 N
109 W4'	Lys-96 NZ	—
110 W1	53 O, 56 N, Ser-91 OG	55 N, 57 N
111 W1	22 O, 24 N, Asn-27 ND2	Ser-24 OG
114 W2	82 O, 87 O, 91 N, Ser-91 OG	83 O, Ser-85 OG, 90 N
114 W4	34 O	—
115 W4	114 O	—
115 W5'	87 O, Asp-87 OD1, 90 N	Ser-85 OG, Asp-87 OD2, 89 N
116 W2	Thr-118 OG1, Trp-123 NE	119 O
117 W1	85 O	83 O, 87 O, 88 N, Ser-91 OG
119 W2'	13 O, 18 N	17 N
120 W1'	84 O	1 N, Thr-40 OG1, Gln-41 OE1, 86 N, Ser-86 OG
123 W2'	1 N, Asn-39 OD1	Gln-41 OE1
131 W3	5 N	—
132 W6'	68 N, 69 N, 69 O	Asn-65 OD1, Asn-65 ND2, 67 N, Ser-72 OG
133 W5'	126 N	—
134 W4'	64 O, 66 N, Asp-66 OD1, Ser-72 OG	60 O, Asn-65 OD1, Asn-65 ND2, Thr-69 OG1
136 W2'	49 O, Ser-51 OG, Asp-66 OD2, Thr-69 OG1	Arg-45 NH2, 51 N, Arg-68 NE, Arg-68 NH1
139 W2'	107 O, 109 N	—
140 W2'	59 N	—
141 W5'	Asp-48 OD2, Ser-50 OG, Asn-59 OD1	Asn-59 ND2, Arg-61 NE
151 W2	57 O	Glu-35 OE2
153 W3	Asp-52 OD1, Asn-57 OE1	Glu-35 OE1, Asn-44 ND2
165 W3	36 O	Asn-39 ND2
186 W1'	Asn-65 OD1 or ND2	—
201 W2	57 O	Asp-52 OD1
211 W1	118 N, 118 O	117 N
222 W1'	86 N	Ser-85 OG, Ser-86 OG

The distribution of the invariant water molecules as a function of the number of interactions with the protein, as found in low-humidity monoclinic lysozyme, is also shown in Figure 3. It clearly shows that a substantial proportion of, but by no means all, water molecules with three or more interactions with the protein are invariant in the hydration shell. The proportion decreases with the decrease in the number of interactions with the protein.

There are eight water molecules that make three or more interactions with the protein in all the structures. These include the three described as internal in native tetragonal lysozyme²¹. In addition to the eight, there are nine more water molecules which make three or more interactions in at least one of the structures. Many of them are involved in holding different regions of the molecule together or in stabilizing local structure. For example, 110 W1 connects the main β -structure

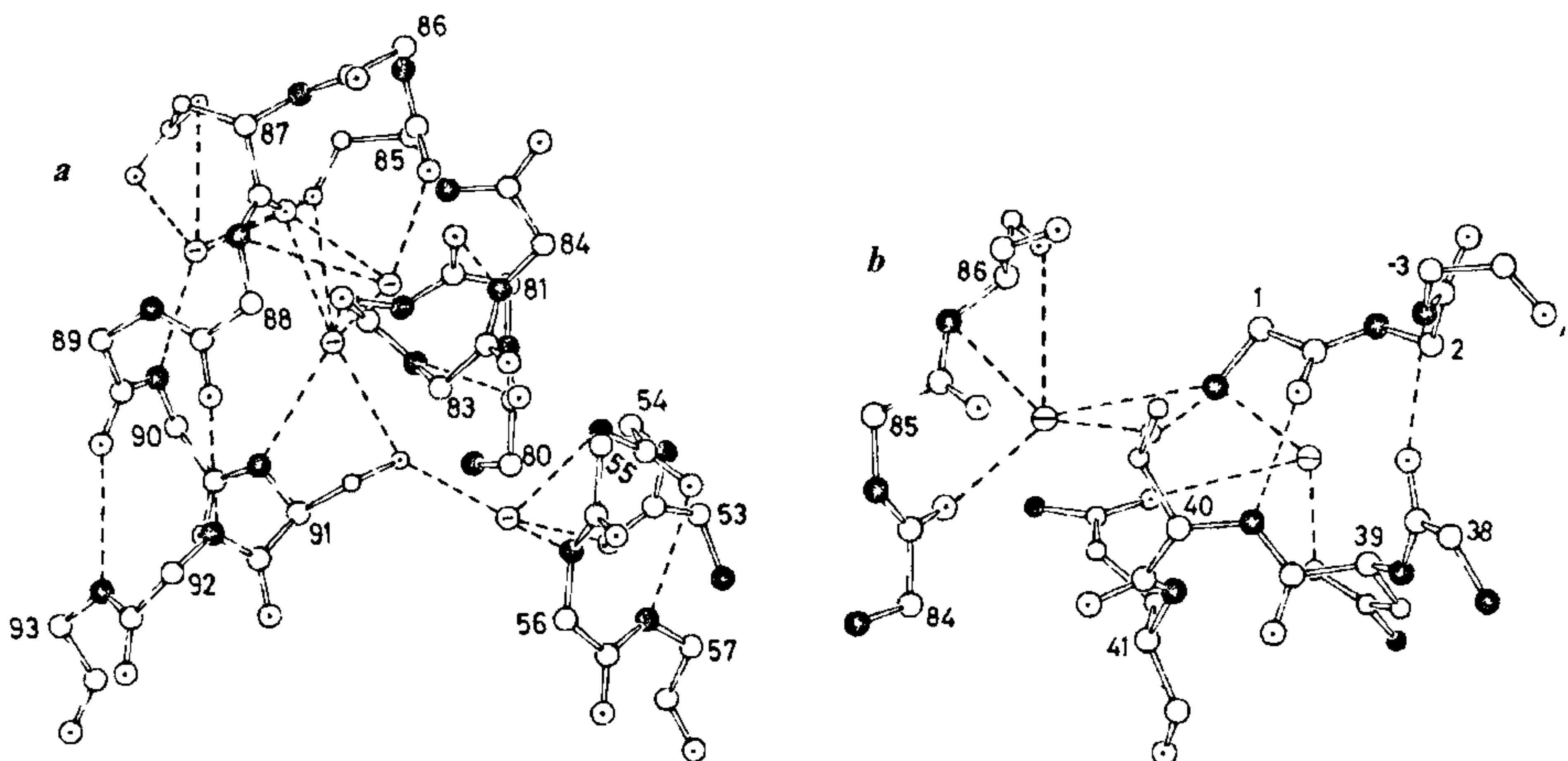


Figure 4. Examples for the role of some invariant water molecules in stabilizing the three-dimensional structure of lysozyme. **a**, Interactions of four invariant water molecules with the 80-84 3_{10} helix, part of the 88-101 α -helix, the loop between them, and part of the major β -structure. **b**, Additional stabilization of the minor β -structure by an invariant water molecule and the role of another in interconnecting the minor β -structure region and a segment of the 85-87 loop. Residue numbers are given against α -carbon positions. In this and the subsequent figure, nitrogens are represented by filled circles, carbonyl oxygens by dotted circles and water oxygens by dashed circles. Only relevant side chains are represented by smaller circles.

and the third α -helix, while 114 W2, 115 W5' and 117 W1, between them, add stability to the region involving the 80-84 3_{10} helix, the third α -helix and the intervening loop (Figure 4, a). Water molecule 120 W1' not only connects this loop to, but also, along with 123 W2', lends additional stability to, the minor β -structure (Figure 4, b). 134 W4' and 136 W2' connect the main β -structure and the main loop. 153 W3 and 141 W5' additionally stabilize the main β -structure. 119 W2' connects the region between the first and the second α -helices to the first helix, while 111 W1 connects the region to the second. Thus many of the invariant water molecules contribute substantially to the stability of the three-dimensional structure of the protein.

There are nine water molecules that interact with only one protein atom each in all the structures. There are three more water molecules that interact with only one protein atom in all but one structure. A majority of the protein atoms involved in such interactions are main-chain carbonyl groups, followed by main-chain amino groups. A water molecule can directly contribute to the stability of the protein structure only when it interacts with two or more protein atoms. Therefore no structural role is immediately obvious for the invariant water molecules that interact with only one protein atom each.

The substrate-binding region of the enzyme along with the water molecules attached to atoms implicated in substrate binding^{21,27-29} are illustrated in Figure 5.

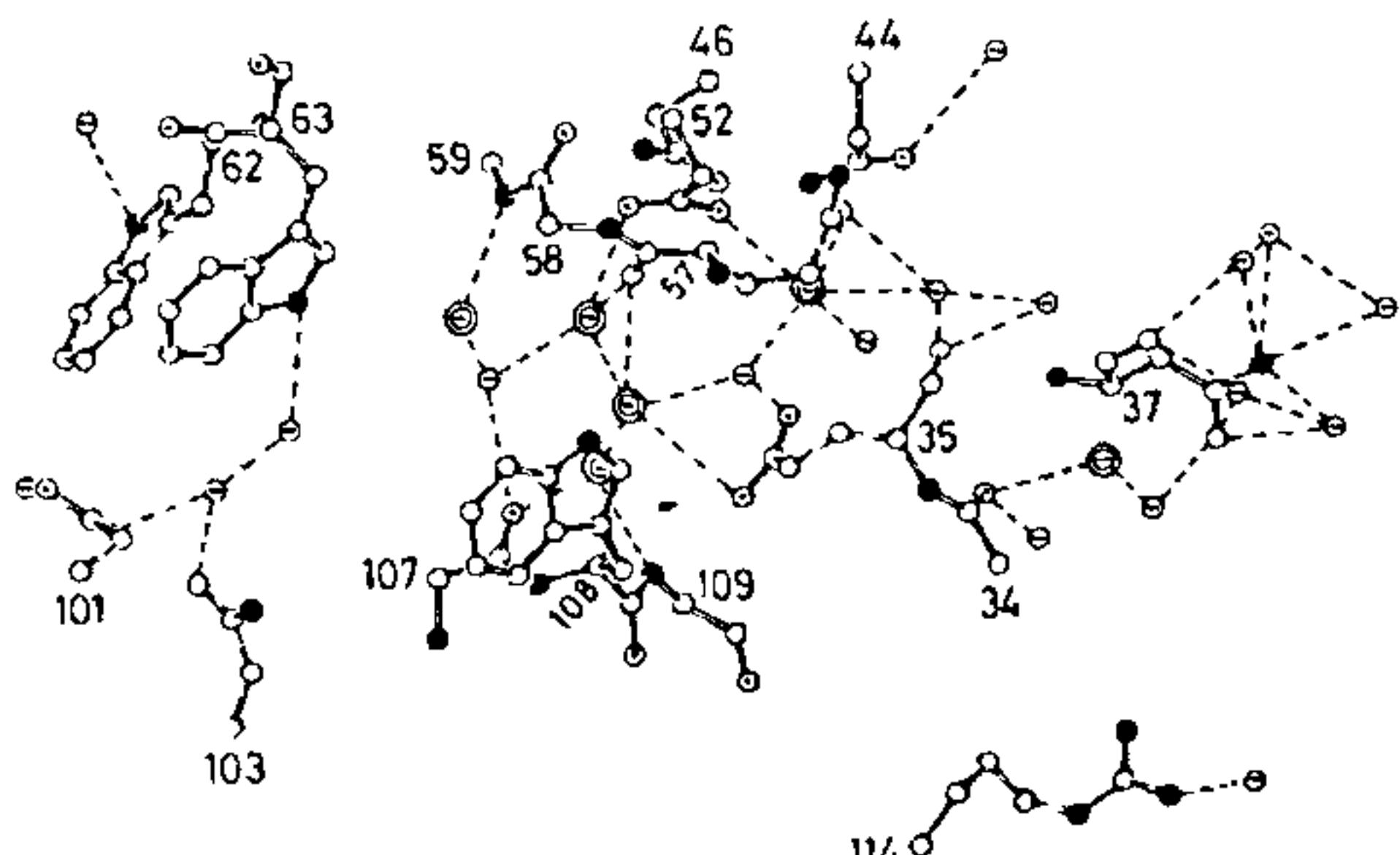


Figure 5. Residues implicated in substrate binding and water molecules attached to them in low-humidity monoclinic lysozyme. The oxygen atoms that belong to invariant water molecules are circled.

Of the 24 such water molecules, six are invariant. Out of these six invariant water molecules, three interact directly with the side-chains of the catalytic residues Glu-35 and Asp-52. Indeed, five of the six invariant water molecules in the active-site region form part of a water cluster contiguous with these side-chains.

1. Artymuk, P. J. et al., *Nature*, 1979, **280**, 563.
2. Ringe, D. and Petsko, G. A., *Prog. Biophys. Molec. Biol.*, 1985, **48**, 197.

3 Huber, R and Bennet, W S, *Biomaterials*, 1983, 22, 261
 4 Brook III, C L and Karplus, M J, *J Molec. Biol.*, 1989, 208, 159
 5 Doucet, J and Benoit, J P, *Nature*, 1987, 325, 643.
 6 Caspar, D L D *et al.*, *Nature*, 1988, 332, 659.
 7 Williams, R J P, *Eur J Biochem*, 1989, 183, 479
 8 Kodandapani, R., Suresh, C. G. and Vijayan, M., *J. Biol. Chem.*, 1990, 265, 16126
 9 Blake, C. C. F., Pulsford, W. C. A and Artymuk, P J, *J. Molec. Biol.*, 1983, 167, 693
 10 Baker, E N and Hubbard, R E, *Prog Biophys. Molec. Biol.*, 1984, 44, 97.
 11. Savage, H and Wlodawer, A., *Methods Enzymol.*, 1986, 127, 162.
 12. Thanki, N., Thornton, J M. and Goodfellow, J. M., *J. Molec. Biol.*, 1988, 202, 637.
 13. Huxley, H E. and Kendrew, J. C., *Acta Crystallogr.*, 1953, 6, 76
 14. Perutz, M F., *Proc R Soc*, 1954, A225, 264.
 15. Salunke, D. M., Veerapandian, B. and Vijayan, M., *Curr. Sci.*, 1984, 53, 231.
 16. Salunke, D M *et al.*, *Acta Crystallogr.*, 1985, B41, 431.
 17. Rao, S T., Hogle, J. and Sunderlingam, M., *Acta Crystallogr.*, 1983, C39, 237.
 18. Rossmann, M G. (ed.), *Molecular Replacement Method*, Gordon and Breach, New York, 1972.
 19. Hendrickson, W. A and Konnert, J H., *Computing in Crystallography* (eds. Diamond, R, Ramaseshan, S and Venkatesan, K.), Indian Academy of Sciences, Bangalore, 1980, pp. 1301-1326.
 20. Blake, C. C. F. *et al.*, *Nature*, 1965, 206, 757
 21. Imoto, T *et al.*, *The Enzymes* (ed. Boyer, P D.), Academic Press, New York, London, 1972, pp. 665-808
 22. Ramanadham, M., Seiker, L. C and Jensen, L H., *Acta Crystallogr.*, 1990, B46, 63.
 23. Kundrot, C. E. and Richards, F M., *J. Molec. Biol.*, 1987, 193, 157
 24. Kundrot, C. E and Richards, F M., *J. Molec. Biol.*, 1988, 200, 401.
 25. Nishikawa, K. *et al.*, *J. Phys. Soc. (Jpn)*, 1972, 32, 1331.
 26. Sternberg, M. J. E., Grace, D. E. P. and Phillips, D C., *J. Molec. Biol.*, 1979, 130, 231.
 27. Ford, L O *et al.*, *J. Molec. Biol.*, 1974, 88, 349.
 28. Perkins, S J. *et al.*, *Biochem. J.*, 1978, 173, 607.
 29. Kelly, J. A *et al.*, *Nature*, 1979, 282, 875
 30. Rossmann, M. G. and Argos, P., *J. Biol. Chem.*, 1975, 250, 7525.
 31. Johnson, C. K., ORTEP. Report ORNL-3794, Oak Ridge National Laboratory, 1970.

ACKNOWLEDGEMENTS. We thank Professor Sir D. C. Phillips for making available the refined atomic coordinates in the native tetragonal lysozyme. The atomic coordinates of high-pressure tetragonal lysozyme and triclinic lysozyme were obtained from Brookhaven Protein Data Bank. Our thanks are due to Dr R. Kodandapani for useful discussions. Two of the figures were prepared using the DEC version of ORTEP³¹, and two others with help from Mr R Sankararamakrishnan and Ms K. P. Viji. The work was supported by the Department of Science and Technology, Government of India. One of us (Madhusudan) has been a CSIR fellow during the course of this work.

Received 22 November 1990; accepted 4 December 1990

RESEARCH COMMUNICATIONS

U-Au occurrence in Nogli Valley, Shimla District, Himachal Pradesh

R. V. Singh, D. B. Sen, R. K. Talra, R. K. Gupta and Rajendra Singh

Atomic Minerals Division, West Block VII, R K Puram, New Delhi 110 066, India

Anomalous gold values in association with uranium have been found in the Middle to Lower Proterozoic formations southwest of Kasha village in Nogli Valley, Himachal Pradesh. This is the first significant occurrence of gold from this part of the Higher Himalaya.

URANIUM occurrences in the Nogli Valley, Shimla District, Himachal Pradesh, are known since the early sixties¹ but gold mineralization, or uranium-gold association in particular, has not been reported so far from the Precambrian formations of this part of the Higher Himalaya. This note presents the first account of gold mineralization in association with uranium in the Rampur Group of rocks in the upper reaches of the Nogli Valley.

The area under reference occupies the southeastern

corner of the Rampur Window, about a kilometre south-southwest of Kasha village on the left bank of Nogli Gad (Figure 1). Stratigraphically the mineralization is confined within the Rampur Group of rocks, which are tectonically overlain by paragneisses and

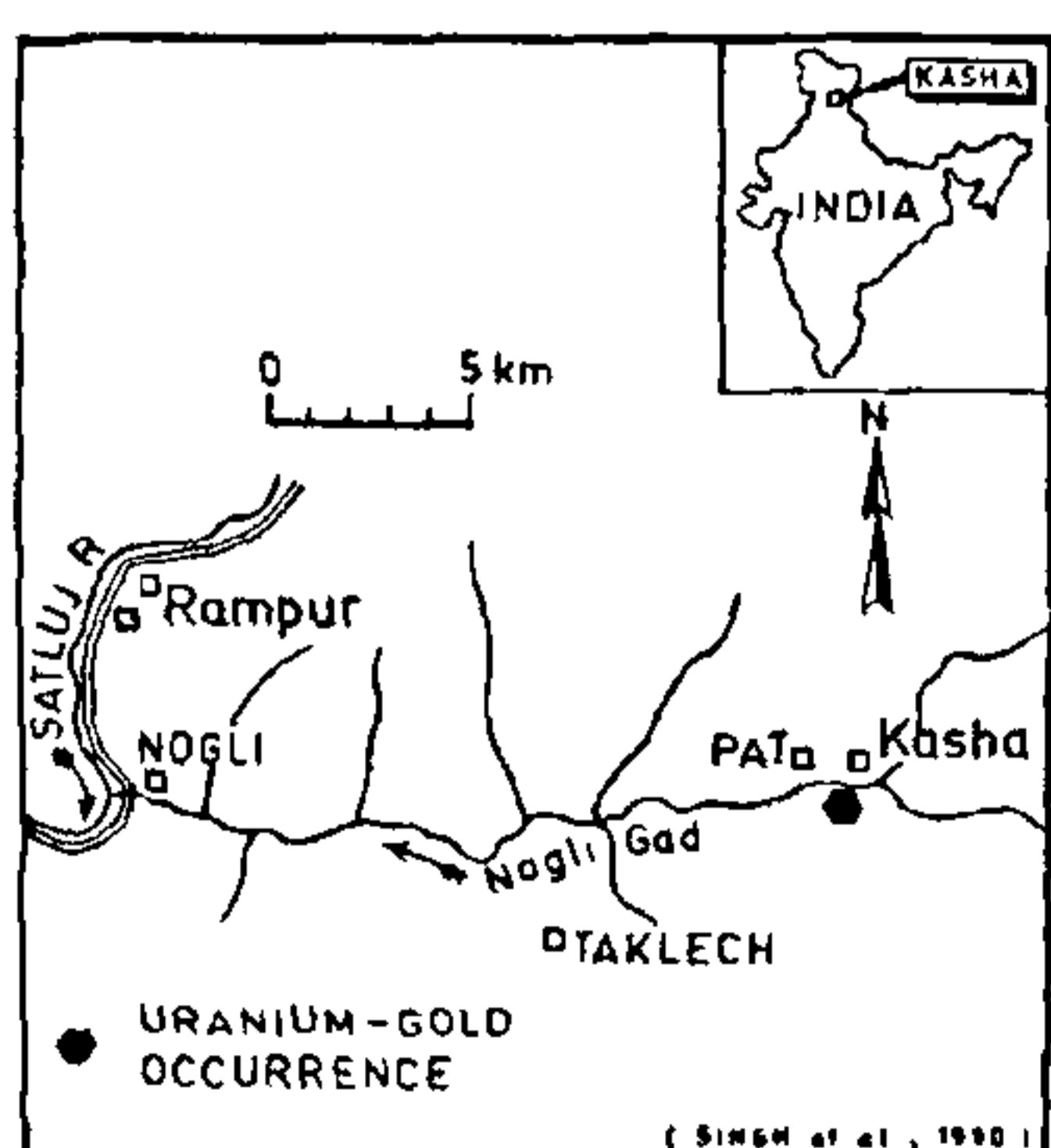


Figure 1. Location map of U-Au occurrence in Nogli Valley