Reports

The Predicted Structure of Immunoglobulin D1.3 and Its Comparison with the Crystal Structure

CYRUS CHOTHIA, ARTHUR M. LESK, MICHAEL LEVITT, ADOLFO G. AMIT, ROY A. MARIUZZA, SIMON E. V. PHILLIPS, ROBERTO J. POLJAK

Predictions of the structures of the antigen-binding domains of an antibody, recorded before its experimental structure determination and tested subsequently, were based on comparative analysis of known antibody structures or on conformational energy calculations. The framework, the relative positions of the hypervariable regions, and the folds of four of the hypervariable loops were predicted correctly. This portion includes all residues in contact with the antigen, in this case hen egg white lysozyme, implying that the main chain conformation of the antibody combining site does not change upon ligation. The conformations of three residues in each of the other two hypervariable loops are different in the predicted models and the experimental structure.

ANTIBODIES ACHIEVE THEIR RANGE of specificity in antigen recognition through the modulation of the conformation of specific loops by changes in the amino acid sequence. An understanding of how the conformations of these loops are determined is essential for an appreciation of the basic mechanism of antibody affinity and specificity, and for the design of modifications of immunoglobulin structures.

We describe here the prediction of the conformation of the antigen-binding domains of immunoglobulin G (IgG) D1.3, and the comparison of the predicted models with the experimental structure determined from the 2.8 Å resolution electron density map (*1*). Two methods of prediction were used. One was based on analysis of the known structures of the variable (V) domains of light (L) and heavy (H) chains (V_L and V_H) (*2*) of the myeloma proteins NEWM, McPC603, KOL, REI, and RHE (Table 1). From these structures, we identified the relatively few residues that, through packing, hydrogen bonds, or ability to assume unusual conformations (Gly, Pro, Asn, Asp), are primarily responsible for the conformations of the regions in these domains involved in antigen binding: the hypervariable loops. In contrast, de la Paz *et al.* have based predictions on the length and overall sequence homology of the hypervariable regions (*3*). Our other predictions were based on conformational energy calculations (*4*).

Comparisons of the atomic structures of variable domains confirm and refine the distinction between a conserved framework consisting of the β sheets of the individual domains and the V_L–V_H packing, and hypervariable loops in the regions that link β sheets (*5–11*). In the five known V_L structures, 69 residues in two β sheets have the same conformation. The root-mean-square differences in position (Δ) of their main chain atoms are between 0.5 and 1.6 Å for different pairs of V_L domains. The β-sheet residues adjacent to the hypervariable regions differ in position by less than 1.5 Å (*2*). Similarly, 79 residues in the β sheets of three V_H domains have the same conformations. The relative geometry of the V_L–V_H packing is also conserved (*12*). This conserved framework structure is produced by conservation of residues buried between the sheets and at the V_L–V_H interface. In the monoclonal antibody IgG D1.3 (*13*), 56 of the 62 buried framework residues (Table 2) are identical to those in known structures; the other six differ in size by no more than a methyl group. We predicted the framework structure of D1.3 to be the same as that of the known structures.

Table 1. Immunoglobulin variable domains known to atomic resolution from x-ray crystal structure analysis. A Fab is a fragment of an immunoglobulin containing variable domains V_L and V_H and constant domains C_L and C_{H1}.

Protein	Chain type		Resolution (Å)	Reference
	L	H		
Fab' NEWM	λI	γII	2.0	(*6*)
Fab McPC603	κ	γI	2.7	(*7*)
Fab KOL	λI	γIII	1.9	(*8*)
V_L REI	κ		2.0	(*9*)
V_L RHE	λI		1.6	(*10*)

Table 2. Residues buried in sheet-sheet packings in immunoglobulin variable domains.

At the interface between the two sheets in V_L domains:
4, 6, 19, 21, 23, 25, 33, 35, 37, 47, 48, 62, 64, 71, 73, 75, 82, 84, 86, 88, 90, 97, 99, 101, 102, 104

At the interface between the two sheets in V_H domains:
4, 6, 18, 20, 22, 24, 34, 36, 38, 48, 49, 51, 69, 78, 80, 82, 86, 88, 90, 92, 104, 106, 107, 109

At the interface between the V_L and V_H domains:
V_L: 36, 38, 44, 87, 98
V_H: 37, 39, 45, 47, 91, 93, 103

The antigen-binding site contains three hypervariable loops from V_L and three from V_H, denoted L1, L2, L3, and H1, H2, H3 (*11*). The residues, numbered according to Kabat *et al.* (*5*), are:

V_L		V_H	
Region	Residues	Region	Residues
L1	26–32	H1	26–32
L2	50–52	H2	53–55
L3	91–96	H3	96–101

These residues are outside the common β-sheet framework of the molecules of known structure. [The complementarity-determining regions (CDR's) defined by Kabat *et al.* (*5*) on the basis of sequence comparisons are more extensive.] The homology of the hypervariable loops of D1.3 with those of known structures of the same class is shown in Table 3. Fewer than 10 of the 32 residues in the hypervariable regions of IgG D1.3 are the same as those in homologous positions in any single known structure.

We analyzed the determinants of the conformation of the hypervariable loops.

L1: The structure of the L1 region is characteristic of the class of the domain (*14*). The V_κ class includes the V_L domains of REI, McPC603, and the target structure D1.3 (Table 1). In McPC603, L1 is 13 residues long and 7 in REI. The common residues have the same main chain conformation: for 26 to 29 and 32, Δ = 0.5 Å.

C. Chothia, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England, and Christopher Ingold Laboratory, University College London, 20 Gordon Street, London WC1H 0AJ, England.
A. M. Lesk, MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England, and Fairleigh Dickinson University, Teaneck-Hackensack Campus, Teaneck, NJ 07666.
M. Levitt, Department of Chemical Physics, Weizmann Institute of Science, 76100 Rehovot, Israel.
A. G. Amit, R. A. Mariuzza, R. J. Poljak, Département d'Immunologie Structurale, Institute Pasteur, 25–28, rue du Docteur Roux, 75724 Paris, France.
S. E. V. Phillips, Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, England.

They adopt an extended conformation, across the top of the V_L domain. The six additional residues in McPC603 form a loop extending away from the domain. The L1 conformation is determined primarily by packing of residue 29 against residues 2, 25, 33, and 71 in the interior of the framework. In D1.3, L1 has the same length as in REI, and the residues packing against the framework are conserved:

Residue:	2	25	29	33	71
D1.3	Ile	Ala	Ile	Leu	Tyr
REI	Ile	Ala	Ile	Leu	Tyr
McPC603	Ile	Ser	Ile	Leu	Tyr

We predicted L1 of D1.3 to have the REI conformation.

L2: In four V_L structures L2 is a three-residue turn (in NEWM, L2 is deleted). Their structures are similar, as conformations of short turns are restricted by the allowed values of the main chain torsion angles (*15*). We predicted that L2 in D1.3 would also have this common structure.

L3: The L3 regions in REI and McPC603 are each six residues long and have the same structure: $\Delta = 0.4$ Å. The conformations usually observed for such hairpin loops would involve main chain hydrogen bonds between residues 92 and 95. The residue Pro 95, present in most V_κ domains including REI, McPC603, and D1.3, precludes this conformation (Fig. 1). In REI and McPC603, Pro 95 has a *cis* peptide, and the side chains of Asn 90 (REI) and Gln 90 (McPC603) form hydrogen bonds to L3 main chain atoms. In D1.3, L3 is also six residues long and includes Pro 95 and His 90. If His 90 could form the same hydrogen bonds as Gln or Asn do in REI and McPC603, we should expect a similar L3 conformation in D1.3.

H1: In KOL and McPC603 the seven residues have similar conformations (Δ = 0.4 Å). In KOL and McPC603, Gly 26 produces a sharp turn. Residues 28 to 32 are a helix with Phe 29 packing in a cavity formed by the side chains of residues 34 and 94 and the main chain of 72 and 77. In D1.3, H1 also has seven residues, and the important side chains are similar but not all identical:

Antibody	Residue				
	26	27	29	34	94
D1.3	Gly	Phe	Leu	Val	Tyr
KOL	Gly	Phe	Phe	Met	Tyr
McPC603	Gly	Phe	Phe	Met	Tyr

In particular, two important residues are both smaller in D1.3. The total volume of Leu 29 and Val 34 in D1.3 is 65 Å^3 less than the total volume of Phe 29 and Met 34 in KOL and McPC603.

H2: In KOL, NEWM, and McPC603, H2 differs in size and conformation. In NEWM, the three residues in H2 are part of a seven-residue hairpin turn with a Gly in the fourth position. Such turns have the same conformation whenever they occur in known structures (*15*). As H2 in D1.3 also has three residues and a Gly at this position, we predicted it to have the NEWM conformation.

H3: The length and conformation of H3 differ in KOL, NEWM, and McPC603. Medium and large H3 loops extend from the surface of the domain. In McPC603 the conformation of the stem of H3 is determined primarily by the packing of Tyr 100b into a cavity between the V_L and V_H domains, its OH group remaining accessible. The homologous residue in D1.3, Arg 99, is also large. We predicted it would pack in the same manner to produce a similar H3 conformation.

We began the construction of the model based on the analysis of known structures with the framework regions of V_L REI and V_H KOL. Of those domains of known structure of the same classes as those of D1.3, these had been determined at the highest resolution. The overall sequence homologies with D1.3 are 66 percent (V_L)

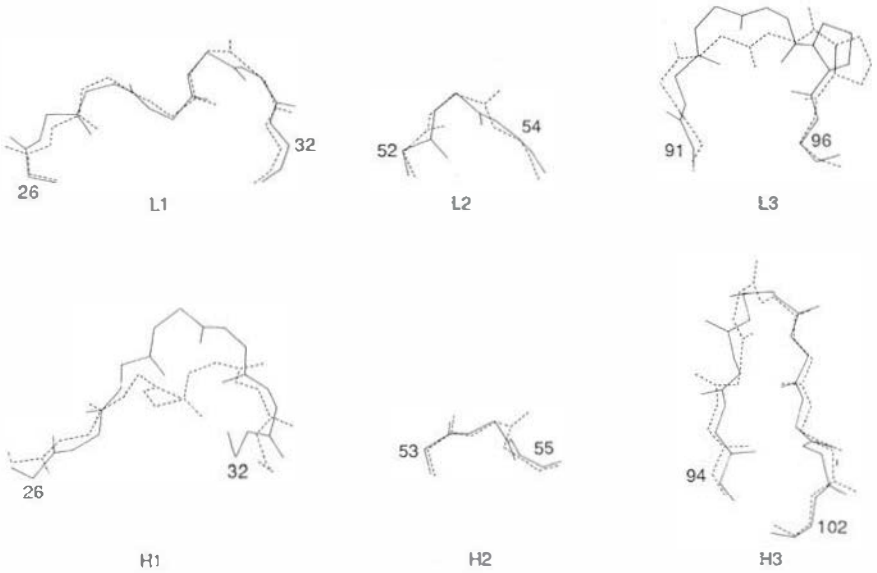

Fig. 1. The main chain atoms of the hypervariable regions in the experimental structure (– – –) and in the predicted model based on analysis of the known structures (——). In a 2.8 Å electron density map, the exact orientations of peptides outside secondary structures may not be well determined. This figure contains independent pictures of these regions, each shown in a viewpoint perpendicular to the mean plane of the loop.

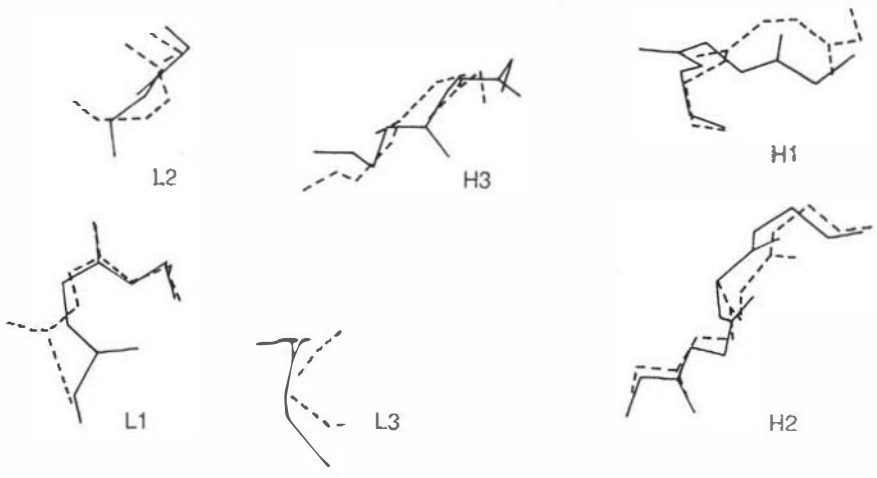

Fig. 2. The main chain atoms of the part of the binding site that makes contact with the antigen (*1*), in the experimental structure (– – – –) and in the predicted model based on analysis of the known structures (——). The actual relative positions of the regions in the structure are shown from a viewpoint roughly perpendicular to the plane of the antigen-antibody interface.

and 49 percent (V_H). To combine the domains we superposed V_L REI onto V_L KOL by fitting the conserved V_L residues of the domain-domain interface. We took L1 and L2 from V_L REI and H2 from V_H NEWM. For L3, noting the conserved Pro, and believing on the basis of model building that His 90 could make the same hydrogen bonds as Gln, we used the loop from REI. In the H1 region of D1.3, two crucial residues are smaller than those in the known structures. Nevertheless, we did not expect this loop to change its conformation radically, and took H1 from V_H KOL.

No known structure has an H3 region of the same length as D1.3. Faced with alternative conformations in known structures, we first chose to build a loop based on NEWM H3 (for reasons we now believe to be incorrect). Informed that this initial prediction was qualitatively incorrect, we then based our prediction on McPC603, which proved to be the proper model. To correct the length we removed three residues from the apex of the loop—these were chosen to preserve the pattern of hydrogen bonding—and closed the chain using the interactive graphics program FRODO (*16*). Side chains of D1.3 were substituted wherever its sequence differed from the parent structures, and bad steric contacts relieved by changing their conformations.

Two other models were constructed independently with the use of conformational energy calculations. One was based exclusively on McPC603, the molecule of known structure having the highest homology with D1.3 in both V regions. The side chains of McPC603 were replaced with their D1.3 homologs, and insertions and deletions made without, initially, preserving the continuity of the backbone. Equivalent atoms were taken from the McPC603 structure whenever possible; 125 missing atoms (out of 1977) were placed at random within 1 Å of the atom of known position to which they are bonded. Energy minimization closed the gaps in the chain; eliminated close contacts; formed strong hydrogen bonds; and brought all bond lengths, bond angles, and torsion angles close to their standard values (*4*). After 600 conjugate gradient cycles, the total energy was −1655 kcal/mol.

A second model merged features from each of the three known Fab structures, by selecting seven stretches of residues locally most homologous in sequence to D1.3: From McPC603 V_L, residues 1 to 17 and 27 to 107; KOL V_L, 18 to 26; NEWM V_H, 1 to 67 and 90 to 101; and KOL V_H, 68 to 89 and 102 to 115. Only 100 atoms were missing. Minimization reduced the total energy to −1681 kcal/mol.

We compared the predicted structures with the experimental structure of D1.3 derived from the 2.8 Å electron density map of a crystal containing the Fab fragment of D1.3 binding the antigen hen egg white lysozyme (*1*). In Fig. 1, the experimental hypervariable regions and those predicted from structural analysis are superposed. The observed antigen-binding site (*1*) and the homologous regions of this predicted structure are shown in Fig. 2. Deviations in atomic positions of observed and predicted structures are shown in Tables 4 to 7. Results are given for the framework structures, the residues adjacent to the hypervariable regions, the hypervariable regions themselves, and a maximal set of well-fitting residues that includes the lysozyme binding site. Atoms included in superposing the individual loops are N, Cα, C, and Cβ. The carbonyl O is omitted because in a 2.8 Å electron density map the orientations of peptides outside secondary structures may not be well determined. The Cβ is included to evaluate the general positioning of side chains.

In this model, the framework and the folds of L1, L2, H2, and H3 were predicted correctly (Table 6). Three residues in the middle of L3 are different. In the current interpretation of the electron density map, the peptide of Pro 95 has a *trans* conformation; in the predicted structure it is *cis* (Fig. 1). Although the *trans* conformation is favored by the current map, a *cis* conformation cannot be ruled out. Three residues at the center of H1 are very different; in particular, in the experimental structure the side chain of Leu 29 does not lie in the cavity occupied

Table 3. The homology of the hypervariable loops of IgG D1.3 and other immunoglobulins of known structure and the same class. References to the structures are given in Table 1. Residue numbers for the hypervariable regions are given in the text.

L1	REI	Ser	Gln	Asp	Ile	Ile	Lys							Tyr
	McPC603	Ser	Gln	Ser	Leu	Leu	Asn	Ser	Gly	ASn	Gln	Lys	Asn	Phe
	D1.3	Ser	Gly	Asn	Ile	His	Asn							Tyr
	REI	Glu	Ala	Ser										
L2	McPC603	Gly	Ala	Ser										
	D1.3	Tyr	Thr	Thr										
	REI	Tyr	Gln	Ser	Leu	Pro	Tyr							
L3	McPC603	Asp	His	Ser	Tyr	Pro	Leu							
	D1.3	Phe	Trp	Ser	Thr	Pro	Arg							
H1	KOL	Gly	Phe	Ile	Phe	Ser	Ser	Tyr						
	NEW	Gly	Thr	Ser	Phe	Asp	Asp	Tyr						
	McPC603	Gly	Phe	Thr	Phe	Ser	Asp	Phe						
	D1.3	Gly	Phe	Ser	Leu	Thr	Gly	Tyr						
H2	KOL	Asp	Asp		Gly	Ser								
	NEW	Tyr	His			Gly								
	McPC603	Asn	Gly	Lys	Asn	Lys								
	D1.3	Gly	Asp			Gly								
H3	KOL	Gly	Gly	His	Gly	Phe	Cys	Ser	Ser	Ala	Ser	Cys	Phe	Gly	Pro	Asp
	NEW	Leu	Ile	Ala									Gly	Cys	Ile	Asp
	McPC603	Tyr	Tyr	Gly	Ser							Thr	Trp	Tyr	Phe	Asp
	D1.3	Arg	Asp										Tyr	Arg	Leu	Asp

by Phe 29 in KOL and McPC603, but is external. Although the experimental structure may have to be modified, as the three residues have main chain conformations usually disallowed because of steric hindrance, the predicted structure is incompatible with the electron density. In retrospect, it appears that the reduced volume of two important side chains (29 and 34) may have changed the packing to produce a different fold.

None of the residues at which the model differs from the observed structure is involved in antigen contacts. The main chain conformation of the binding site itself is predicted accurately (Fig. 2).

Both the models based on conformational energy calculations correctly predicted the framework and the folds of L2, H2, and H3. Loop L1 is very different from the experimental structure because it was built from the extended part of McPC603 L1 rather than from the part common to other V_κ chains. L3 and H1 are similar to those in the model based on structure analysis.

The comparisons show that the prediction of the main chain conformation of D1.3 was largely successful (Tables 4 to 7 and Figs. 1 and 2). They support the premise that the binding site conformation is determined principally by specific interactions of a few residues, and that these residues can be identified and used to formulate rules valid for structure prediction. A structure of D1.3 at higher resolution will permit a more detailed evaluation of the predictions, as well as the refinement and extension of the rules.

The predictions treated an isolated V_L–V_H dimer. The observed D1.3 structure is contained in a Fab-antigen complex. The close similarity of the observed and predicted D1.3 binding site and framework implies that the association with antigen does not significantly alter the main chain conformation of the antibody.

Table 4. Fit of V_L-V_H framework. Comparisons of predicted and observed structures. In Tables 4 to 7, the model SA is based on structure analysis, and model CE on conformational energy calculations starting from a structure derived initially from McPC603. Δ is the root-mean-square deviation in angstroms after superposition.

Residues	Δ (N, Cα, C, and O atoms)	
	SA	CE
V_L: 4–6, 19–25, 33–48, 52–54, 61–76, 84–90, 97–107 V_H 3–12, 17–25, 33–52, 56–60, 68–82, 88–95, 102–113	1.0	1.0

Table 5. Shifts of framework residues adjacent to hypervariable regions. Shift is the difference (angstroms) in position of residue after superposition of the framework.

Residues		Shift			
		SA		CE	
V_L: 25	33	0.6	0.3	2.5	0.8
48	52	0.6	0.8	0.2	0.6
90	97	0.8	0.4	0.9	0.8
V_H: 25	33	0.9	1.8	0.6	1.2
52	56	1.5	1.8	1.7	0.7
95	102	1.2	1.0	1.3	0.6

Table 6. Fits of hypervariable loops.

Residues	Δ (N, Cα, C, and Cβ atoms)	
	SA	CE
L1 26–32	0.85	3.76
L2 49–51	0.63	0.47
L3 91–96	0.97	1.10
H1 26–32	2.07	1.68
H2 53–55	0.50	0.89
H3 99–104	0.86	0.87

Table 7. Fit of maximal well-fitting portion, including framework and all lysozyme contact residues (model SA).

Residues	Δ (N, Cα, and C atoms)
V_L: 2–6, 19–65, 69–92, 96–104 V_H: 3–7, 14–27, 30–59, 67–75, 78–81, 84–113	1.0

REFERENCES AND NOTES

1. A. G. Amit, R. A. Mariuzza, S. E. V. Phillips, R. Poljak, *Science* 233, 747 (1986).
2. C. Chothia and A. M. Lesk, in preparation.
3. P. de la Paz, B. J. Sutton, M. J. Darsley, A. R. Rees, *EMBO J.* 5, 415 (1985).
4. M. Levitt, *J. Mol. Biol.* 170, 723 (1983).
5. E. A. Kabat, T. T. Wu, H. Bilofsky, M. Reid-Miller, H. Perry, Eds., *Sequences of Proteins of Immunological Interest* (U.S. Public Health Service, Washington, DC, 1983).
6. F. Saul, L. Amzel, R. Poljak, *J. Biol. Chem.* 253, 585 (1978).
7. D. Segal *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 71, 4298 (1974).
8. M. Marquart, J. Deisenhofer, R. Huber, W. Palm, *J. Mol. Biol.* 141, 369 (1980).
9. O. Epp, E. Lattman, M. Schiffer, R. Huber, W. Palm, *Biochemistry* 14, 4943 (1975).
10. W. Furey, B. C. Wang, C. S. Yoo, M. Sax, *J. Mol. Biol.* 167, 661 (1983).
11. J. Novotný *et al.*, *J. Biol. Chem.* 258, 14433 (1983).
12. C. Chothia, J. Novotný, R. Bruccoleri, M. Karplus, *J. Mol. Biol.* 186, 651 (1985).
13. M. Verhoeyen, unpublished.
14. A. M. Lesk and C. Chothia, *J. Mol. Biol.* 160, 325 (1982).
15. B. L. Sibanda and J. M. Thornton, *Nature (London)* 316, 170 (1985); C. Chothia and A. M. Lesk, unpublished results.
16. T. A. Jones, in *Computational Crystallography*, D. Sayre, Ed. (Clarendon, Oxford, 1982), pp. 303–317.

17. We thank Dr. M. Verhoeyen for making the amino acid sequence available; Drs. A. Feinstein, C. Milstein, and G. Winter for discussion; and the Royal Society, the U.S. National Science Foundation (PCM83-20171), the National Institute of General Medical Sciences (GM25435), and the European Molecular Biology Organization (ASTF 4475) for support.

16 April 1986; accepted 30 June 1986

Cambrian River Terraces and Ridgetops in Central Australia: Oldest Persisting Landforms?

A. J. STEWART, D. H. BLAKE, C. D. OLLIER*

Fluvial sediments in paleovalleys cut in the Ashburton surface of the Davenport province of central Australia form terrace remnants that appear to retain their original depositional tops and have probably existed as subaerial landforms since their inception. Marine fossils in sediments conformable with the fluvial sediments near the southeast margin of the province give a Cambrian age for the terraces; the Ashburton surface forming the ridgetops between the paleovalleys is Cambrian or older.

INLAND AUSTRALIA IS WELL KNOWN for the flatness of its landscape and the antiquity of its erosion surfaces (*1*, *2*). The highest surface in the Tennant Creek region of central Australia, the Ashburton surface, has long been thought to be Cretaceous or older (*2*). We report evidence from the Davenport province of the Tennant Creek region (Fig. 1) that suggests that, during the Cambrian, fluvial sediments were deposited in valleys between ridges whose tops are remnants of the Ashburton surface. Subsequent dissection of the sediments formed terraces and mesas, but where preserved, the relation of the terraces to the adjoining ridges indicates that the terraces and mesatops have existed as subaerial landforms since the Cambrian sedimentation. The Ashburton surface itself is therefore Cambrian or older.

The Davenport province (*3*) is a broad

Bureau of Mineral Resources, Canberra, 2601, Australia.

*Present address: Department of Geography, University of New England, Armidale, 2350, Australia.