25. Ollier, C. D. *Tectonics and Landforms* (Longman, New York, 1981).
26. Fitch, F. J. & Miller, J. A. *Spec. Publs geol. Soc. S. Afr.* **13**, 247–266 (1984).
27. Cox, K. G. *J. Petrol.* **21**, 629–650 (1980).
28. England, P. C. & Molnar, P. *Geology* (in the press).
29. Petri, S. & Fulfaro, V. J. *Geologia do Brasil* (Editora da Universidade de São Paulo, 1983).
30. Tankard, A. J. *et al. Crustal Evolution of Southern Africa* (Springer, New York, 1982).
31. Craddock, C. (ed.) *Antarctic Geoscience* (University of Wisconsin Press, Madison, 1982).
32. King, L. C. *South African Scenery* 2nd edn (Oliver and Boyd, Edinburgh, 1951).
33. King, L. C. *The Morphology of the Earth* (Oliver and Boyd, Edinburgh, 1962).
34. Partridge, T. C. & Maud, R. R. *S. Afr. geol. J.* **90**, 179–208 (1987).
35. Cahen, L., Snelling, N. J., Delhal, J. & Vail, J. R. *The Geochronology and Evolution of Africa* (Clarendon, Oxford, 1984).
36. Brown, R. W. *6th Int. Conf. Fission Track Dating* Abstr. Vol. Université de Franche-Comté, Besançon, 1988)

Conformations of immunoglobulin hypervariable regions

Cyrus Chothia*†, Arthur M. Lesk*‡, Anna Tramontano‡, Michael Levitt*§, Sandra J. Smith-Gill‖, Gillian Air¶, Steven Sheriff#, Eduardo A. Padlan#, David Davies#, William R. Tulip††, Peter M. Colman††, Silvia Spinelli‡‡, Pedro M. Alzari‡‡ & Roberto J. Poljak‡‡**

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK
‡ European Molecular Biology Laboratory, Meyerhofstrasse 1, Postfach 1022.09, D-6900 Heidelberg, FRG
§ Department of Cell Biology, Stanford University Medical School, Stanford, California 94305, USA
† Christopher Ingold Laboratory, University College London, 20 Gordon Street, London WC1H 0AJ, UK
‖ National Institute of Cancer and # National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA
¶ Department of Microbiology, University of Alabama, Birmingham, Alabama 35294, USA
†† CSIRO Division of Biotechnology, 343 Royal Parade, Parkville 3052, Australia
‡‡ Unité d'Immunologie Structurale, Département d'Immunologie, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris, France

On the basis of comparative studies of known antibody structures and sequences it has been argued that there is a small repertoire of main-chain conformations for at least five of the six hypervariable regions of antibodies, and that the particular conformation adopted is determined by a few key conserved residues. These hypotheses are now supported by reasonably successful predictions of the structures of most hypervariable regions of various antibodies, as revealed by comparison with their subsequently determined structures.

THE relationships between the amino-acid sequences of immunoglobulins and the structures of their antigen-binding sites are important for understanding the molecular mechanisms of the generation and maturation of the immune response and for designing engineered antibodies. Antigen-binding sites are formed by six loops of polypeptide, the hypervariable regions; three from the variable domain of the light chain (VL) and three from the variable domain of the heavy chain (VH), denoted L1, L2, L3, and H1, H2, H3, respectively (Fig. 1*a*). Within the domains, the loops are connected to a β-sheet framework whose structure is conserved[1,2]. The specificity and affinity of the binding sites are governed by the structures of the six hypervariable regions[3,4].

Two models can be proposed for the relationship between the amino-acid sequence and structure of the binding-site loops. In one model, different sequences produce different conformations for both the main chain and side chains of the loops. Because hypervariable regions have different sequences in different antibodies, this model implies that each region adopts a different conformation in different antibodies. In the other model, antibodies have only a few main-chain conformations or 'canonical structures' for each hypervariable region. Most sequence variations would only modify the surface provided by the side chains on a canonical main-chain structure. Sequence changes at a few specific sets of positions would switch the main chain to a different canonical conformation.

Canonical structure model

Experimental evidence indicates that the canonical structure model describes the relationship between amino-acid sequence and structure for at least five of the six hypervariable regions[5–9]. Kabat *et al.*[5] found conserved residues at sites within certain sets of hypervariable regions and suggested that they had a structural role. Padlan and Davies[6], and more recently de la Paz *et al.*[7], showed that some of the hypervariable regions in the immunoglobulins of known structure have the same main-chain conformation in spite of several differences in sequence.

Chothia and Lesk[8] identified the residues that through packing, hydrogen bonding, or the ability to assume unusual values of the torsion angles ϕ, ψ or ω, are primarily responsible for the main-chain conformations of the hypervariable regions in the structures then known—the Fab fragments of NEW (ref. 10), McPC603 (ref. 11), KOL (ref. 12) and J539 (ref. 13) and the VL domains of REI (ref. 14) and RHE (ref. 15). The conformations are determined by the interactions of a few residues at specific sites in the hypervariable regions and, for certain loops, in the framework regions. Hypervariable regions that have the same conformations in different immunoglobulins have the same or very similar residues at these sites (Fig. 1 and Table 1). Examination of the amino-acid sequence of the antibody D1.3 showed that its hypervariable regions are the same size as those in known structures and contain the same or similar residues at the sites responsible for known conformations[9]. On the basis of these observations the atomic structure of the VL–VH dimer of D1.3 was predicted before its experimental determination. Comparison of this predicted structure with the preliminary crystal structure showed that the conformations of four of the hypervariable regions had been predicted correctly; the conformation of L3 was significantly different from that predicted, and

** Present address: The Squibb Institute for Medical Research, PO Box 4000, Princeton, New Jersey 08543-4000, USA.

TABLE 1 Sequences and conformations of Vκ and VH hypervariable regions of known structure

L1 Regions†

Canonical Structure	Protein	26	27	28	29	30	31	a	b	c	d	e	f	32	2	25	33	71
					*										*	*	*	*
1	J539	S	S	S	V	S	—	—	—	—	—	—	—	S	I	A	L	Y
	HyHEL-5	S	S	S	V	N	—	—	—	—	—	—	—	Y	I	A	M	Y
	NQ10	S	S	S	V	R	—	—	—	—	—	—	—	Y	I	A	M	Y
2	REI	S	Q	D	I	I	K	—	—	—	—	—	—	V	I	A	L	Y
	D1.3	S	G	N	I	H	N	—	—	—	—	—	—	Y	I	A	L	Y
	HyHEL-10	S	Q	S	I	G	N	—	—	—	—	—	—	N	I	A	L	F
	NC41	S	Q	D	V	S	T	—	—	—	—	—	—	A	I	A	L	Y
3	McPC603	S	E	S	L	L	N	S	G	N	E	K	N	F	I	S	L	F
4	4-4-20	S	Q	S	L	V	H	S	—	N	G	N	T	Y	V	S	L	F

Total no. of sequences known for L1 regions: human, 95; mouse, 299.

Canonical structure	1	2	3	4
Human sequences that fit (%)	—	60	5	5
Mouse sequences that fit (%)	15	25	20	10

L2 Regions

Canonical Structure	Protein	50	51	52	48	64
					*	*
1	REI	E	A	S	I	G
	McPC603	G	A	S	I	G
	J539	E	I	S	I	G
	D1.3	Y	T	T	I	G
	HyHEL-5	D	T	S	I	G
	HyHEL-10	Y	A	S	I	G
	NC41	W	A	S	I	G
	NQ10	D	T	S	I	G
	4-4-20	K	V	S	I	G

Total no. of sequences known for L2 regions: human, 69; mouse, 183.

Canonical structure	1
Human sequences that fit (%)	95
Mouse sequences that fit (%)	95

L3 Regions

Canonical Structure	Protein	91	92	93	94	95	96	90
						*		*
1	REI	Y	Q	S	L	P	Y	Q
	McPC603	D	H	S	Y	P	L	N
	D1.3	F	W	S	T	P	R	H
	HyHEL-10	S	N	S	W	P	Y	Q
	NC41	H	Y	S	P	P	W	Q
	4-4-20	S	T	H	V	P	W	Q
	NQ10	W	S	S	N	P	L	Q
2	J539	W	T	Y	*P	L	I	*Q
3	HyHEL-5	W	G	R	N	*P	—	*Q

Total no. of sequences known for L3 regions: human, 52; mouse, 152.

Canonical structure	1	2	3
Human sequences that fit (%)	90	—	2
Mouse sequences that fit (%)	80	10	1

H1 had a very different fold from that predicted[9]. (We report below that the refined conformation of D1.3 corresponds more closely to the predicted structure.)

An examination of the library of the known immunoglobulin sequences shows that many immunoglobulins have hypervariable regions that are the same size as those in the known structures and contain the same or closely related residues at the sites responsible for the known conformations[8]. These observations indicate that for at least five of the hypervariable regions there is only a small repertoire of canonical main-chain conformations and that the conformation actually present can often be predicted from the sequence by the presence of specific residues.

The accuracy of the canonical structure model for immunoglobulin binding sites depends on (1) the correct determination of the sets of residues responsible for the observed conformations and (2) changes in the identity of residues at other sites not significantly affecting the conformations of the canonical structures. The model can be tested, refined and extended by using it to predict the atomic structures of binding sites in immunoglobulins before their structures have been determined by X-ray crystallography.

We have now tested the canonical structure model by using it to predict the structures of four immunoglobulins before their structures had been experimentally determined. These immunoglobulins are HyHEL-5 (ref. 16), HyHEL-10 (ref. 17), NC41 (ref. 18) and NQ10 (S.S., P.M.A. and R.J.P., manuscript in preparation). The analysis of the amino-acid sequences of these immunoglobulins indicated that 19 of their 24 hypervariable regions should have conformations close to known canoni-

H1 Regions‡

Canonical Structure	Protein	26	27	28	29	30	31	32	34	94
		*	*		*				*	*
1	McPC603	G	F	T	F	S	D	F	M	R
	KOL	G	F	I	F	S	S	Y	M	R
	J539	G	F	D	F	S	K	Y	M	R
	D1.3	G	F	S	L	T	G	Y	V	R
	HyHEL-5	G	Y	T	F	S	D	Y	I	R
	NC41	G	Y	T	F	T	N	Y	M	R
	NQ10	G	F	T	F	S	S	F	M	R
	4-4-20	G	F	T	F	S	D	Y	M	G
1′	NEW	G	S	T	F	S	N	D	Y	R
	HyHEL-10	G	D	S	I	T	D	D	W	N

Total no. of sequences known for H1 regions: human, 50; mouse, 321.

Canonical structure	1
Human sequences that fit (%)	50
Mouse sequences that fit (%)	80

H2 Regions§

Canonical Structure	Protein	52a	b	c	53	54	55	71
							*	
1	NEW	—	—	—	Y	H	G	
	D1.3				G	D	G	
	HyHEL-10	—	—	—	Y	S	G	
		*					(*)	*
2	HyHEL-5	P	—	—	G	S	G	A
	NC41	T	—	—	N	T	G	L
						(*)		*
3	KOL	D	—	—	D	G	S	R
	J539	P	—	—	D	S	G	R
	NQ10	S	—	—	G	S	S	R
						(*)	*	*
4	McPC603	N	K	G	N	K	Y	R
	4-4-20	N	K	P	Y	N	Y	R

Total no. of sequences known for H2 regions: human, 54; mouse, 248.

Canonical structure	1	2	3	4
Human sequences that fit (%)	15	1	40	15
Mouse sequences that fit (%)	15	40	5	20

The residues listed here (single-letter code) are those that form the hypervariable regions and those in the framework regions that are important for the observed conformations of these regions[8]. The hypervariable regions are taken as those outside the framework β-sheet[8]. Except for H2, they are similar to, but not identical with the regions that show high sequence variations and which Kabat *et al.*[26] use to define hypervariable regions. The sequences are grouped so that those that have the same main-chain conformation, or canonical structure, are adjacent. The canonical structure numbers used below refer to the conformations shown in Fig. 1. The residues in the hypervariable and framework regions that are mainly responsible for these conformations[8] are indicated by an asterisk. The classification and sequence requirements of the H2 conformations have been revised in the light of work described here and elsewhere[28]. For each hypervariable region the number of human and mouse sequences listed by Kabat *et al.*[26] are given. We also give the percentage of these sequences that are the same size as the known canonical structures and have the same residues at the positions marked by an asterisk.

† Canonical structure 4 is illustrated in Fig. 4. Although the size of the known L1 structures varies between 6 and 13 residues, they have closely related folds with residues 26–19 and 32 packed against the framework in the same conformation[8]. The remaining residues form a turn or loop on the surface (Figs 1 and 4). The ends of the long loops have some flexibility. There are another 25% of the human sequences and 20% of the mouse sequences that have one more residue than structure 2, or one fewer than structure 4, and whose sequences satisfy the requirements listed above. It is expected that these differ only in the conformations of the tips of the surface loops.

‡ The H1 hypervariable regions with canonical structure 1 have very similar conformations: the r.m.s. differences in the coordinates of their main-chain atoms are 0.3–0.8 Å. The H1 regions in NEW and HyHEL-10 only partly satisfy the sequence requirements for structure 1 and have a distorted version of its conformation.

§ The H2 region here comprises residues 52a–55. The region with high sequence variation is 50–65 (ref. 26). In the known structures the main-chain conformation of 50–52 and 56–63 do not differ significantly[8] (Fig. 1*b*). (*), The residues at positions 55 or 54 in the canonical structures 2, 3 and 4 have residues with positive values for ϕ and ψ, and usually, but not in all cases, Gly, Asn or Asp is found at these sites. For a sequence to match that of canonical structure 2, 3 or 4 the presence of these residues at sites 54 or 55 is required.

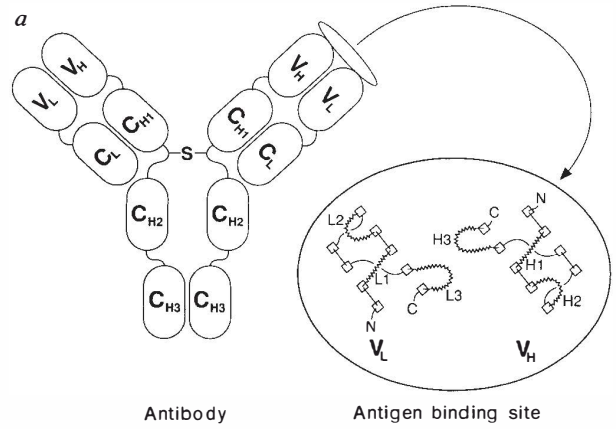
cal structures. We then compared the predicted structures of these hypervariable regions with the subsequently determined structures. Another immunoglobulin structure, 4-4-20 (ref. 19) has recently been reported. We did not have the opportunity to predict the structure of 4-4-20 before its experimental determination, and we discuss here only how its hypervariable regions have the conformations expected from the known canonical structures. Also, we report that the refined conformation of D1.3 (ref. 20) corresponds more closely to the predicted structure.

Model building procedure

The main-chain conformations of the hypervariable regions in the Vκ and VH domains of known structure are shown in Fig. 1. The residues responsible for these conformations are listed in Table 1. Each hypervariable region in the immunoglobulins of unknown structure was examined to determine (1) whether it has the same size as any homologous hypervariable region of known structure and (2) whether its sequence contains the set of residues responsible for a known conformation. Except for L3 in HyHEL-5, all the light-chain regions correspond to a known canonical structure, as do all the H1 regions and the H2 region in HyHEL-10 (Table 1). The conformation predicted for the H2 regions in NC41 and HyHEL-5 was based on the analysis of the H2 region in the preliminary structure of J539 (ref. 8). In all three of these antibodies the H2 region is a four-residue turn with Gly at the fourth position and the predicted conformation is that almost always found for such turns[21]. (Below we present a more accurate analysis of H2 regions.) For H3 regions in HyHEL-5, HyHEL-10, NC41 and NQ10, no prediction of

conformation could be made on the basis of the known canonical structures.

The sequences of the VL and VH domains were compared to see which of the known framework structures have sequences close to those of the unknown structures. From the comparisons of the hypervariable and framework regions, one known VL and one known VH structure were taken as the starting points—parents—for the model of the predicted structure. If the conformation predicted for a hypervariable region was not present in the parent, but was present in another known structure, the hypervariable region in the parent was replaced by that in the other structure. Side chains in the parent that were different from those in the unknown structure were replaced[22,23] and the resulting model subjected to a very limited energy refinement[24].



HyHEL-5, HyHEL-10 and NC41 hypervariable regions

The atomic structures of Fab fragments of immunoglobulins HyHEL-5, HyHEL-10 and NC41 in complexes with their protein antigens were determined by X-ray crystallography[16–18]. The resolution of the X-ray data used to determine the structures and value of the residual (R) after refinement are (complex, resolution, R value): HyHEL-5–lysozyme, 2.5 Å, 20%; HyHEL-10–lysozyme, 3.0 Å, 24%; and NC41–neuraminidase, 2.9 Å, 19%. These structures have been determined at medium resolution. The tracing of the polypeptide chain of the hypervariable regions is unambiguous, although the orientation of some of

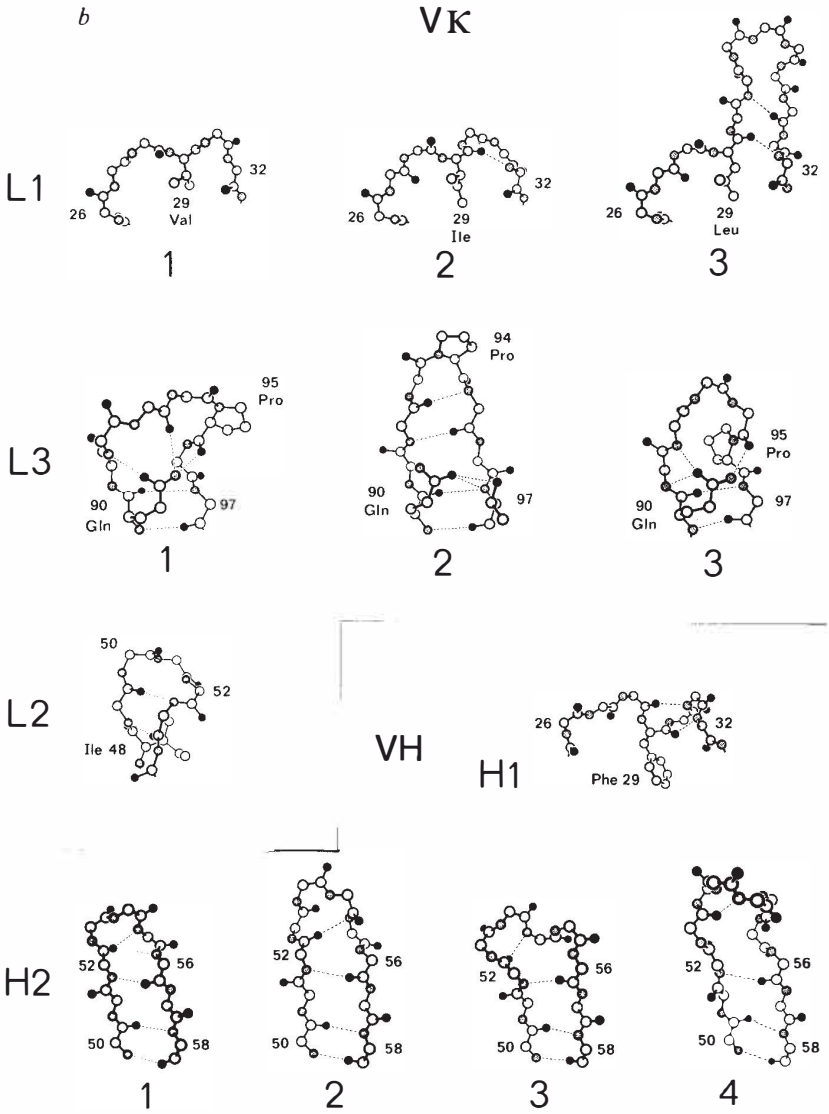


FIG. 1 *a*, Antigen-binding sites of immunoglobulins are formed by six loops of polypeptide, three from the VL domain L1, L2 and L3 and three from the VH domain H1, H2 and H3 (wavy lines). These loops are attached to strands (□) of a conserved β-sheet. *b*, Canonical structures for the hypervariable regions of Vκ and VH domains. In each drawing the region is viewed so that the accessible surface is at top and the framework region below. The main-chain conformation and some of the side chains that determine this conformation are shown. For the definition of the hypervariable regions used here, see Table 1. The immunoglobulins in which the different canonical structures occur are listed in Table 1.

the peptide groups is uncertain. Most side chains are unequivocally placed.

Figure 2 shows the predicted and observed structures of each of the hypervariable regions, superposed by a least-squares fit of their main-chain atoms. Table 2*a* gives, for each predicted and observed hypervariable region, the r.m.s. difference in position of the main-chain atoms.

The main-chain conformations of the predicted and observed hypervariable regions are very similar (Fig. 2, Table 2*a*). The only exception is the H1 region of HyHEL-5. Although the observed and predicted conformations of residues 26–29 and 32 are the same, residues 30 and 31 are in quite different positions. The recently refined structure of Fab J539 (T. N. Bhat, E.A.P. and D.D., manuscript in preparation) shows that these differences were inherited as a result of an error in the original determination of the J539 structure used as the parent for this region. Rebuilding the predicted model with the refined J539 structure puts residues 30 and 31 in the correct position and gives an r.m.s. difference between the predicted and observed H1 regions of 0.6 Å.

Given the medium resolution of the structures used to derive the models and of the experimental structures, the agreement of the predicted and observed loop conformations is excellent.

Relative positions of the hypervariable regions

Figure 3 shows the positions of the hypervariable regions relative to each other and to the framework for the predicted and observed structures. To produce this figure the predicted and observed structures were superposed by a least-squares fit of framework residues. In Table 2*b*, the differences in position of the hypervariable regions are reported.

Small differences in the relative positions of the hypervariable regions in the predicted and observed structures might be expected because of two factors not corrected for in the model building. First, the predicted structures were built using parent V domains that have some residues in the framework and VL–VH interface that are different from those in the final predicted structure. These differences have small but significant effects on the main-chain structure of the individual domains and the way they pack together[2,25,26]. Second, differences between the predicted and observed structures could arise from the effects of the association with the antigen[18]. In other proteins, ligand binding can result in the movement of close-packed segments of polypeptide relative to each other by 1–2 Å, and the ends of loops are able to move somewhat more[27].

The differences in the positions of the predicted and observed H2 regions in HyHEL-5 and NC41 (Table 2*b*) are larger than expected from these factors. The interactions that the H2 regions make with the rest of the VH domain were therefore examined.

Residue 71 and position and conformation of H2

At the same time as the structures of HyHEL-5 and NC41 became available, the refinement of the atomic structure of J539 was completed (T. N. Bhat, E.A.P. and D.D., manuscript in preparation). The conformation of H2 in the refined structure is not like that in HyHEL-5 and NC41 but is the same as that in KOL. This was quite unexpected. The main determinant of the conformation of small turns is usually the position of glycine residues[21]: in KOL, Gly occurs at position 54, and in J539, HyHEL-5 and NC41, Gly occurs at position 55.

An examination of the environments of the H2 regions[28] shows that in KOL and J539 the side chain of framework residue Arg 71 packs between, and forms hydrogen bonds to, H1 and H2. In HyHEL-5, residue 71 is Ala, and here the cavity that would be created by this smaller side chain is filled by the insertion of a residue from the H2 region—Pro 52a. In KOL and J539 the side chain at position 52a is on the surface. The relative movement of position 52a involves a change in the

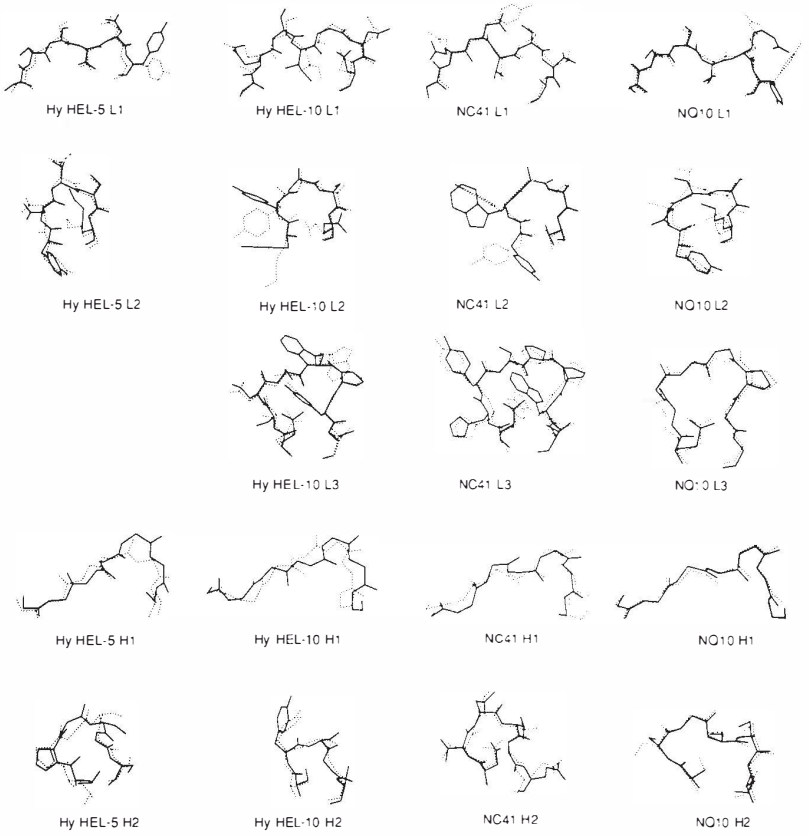


FIG. 2 The predicted (broken line) and observed (continuous line) conformations of the hypervariable regions. The structures have been superposed by a least-squares fit of their main-chain atoms. Residue numbers and the r.m.s. difference in position of the superposed atoms are given in Table 2*a*. Predicted and observed side-chain conformations are shown for all regions except H1 and L3 in NQ10 where they obscure the main chain. After our prediction of the NC41 structure several revisions were made to the sequence and some of the differences can be seen here.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.