15. Mosmann T. Rapid colorimetric assay for cellular growth and survival. Application to proliferation and cytotoxicity assays. *J Immunol Meth* 1983, **65**, 55–64.
16. Pieters R, Loonen AH, Huismans DR, *et al. In vitro* drug sensitivity of cells from children with leukemia using the MTT assay with improved culture conditions. *Blood* 1990, **76**, 2327.
17. Campling BG, Pym J, Galbraith PR, Cole SP. Use of the MTT assay for rapid determination of chemosensitivity of human leukemic blast cells. *Leukemia Res* 1988, **12**, 823–831.
18. Finlay GJ, Wilson WR, Baguley BC. Comparison of *in vitro* activity of cytotoxic drugs towards human carcinoma and leukaemia cell lines. *Eur J Cancer Clin Oncol* 1986, **22**, 655–662.
19. Tanaka M, Yoshida S. Formation of cytosine arabinoside-5'-triphosphate in cultured human leukemia cell lines correlates with nucleoside transport capacity. *Jpn J Cancer Res* 1987, **78**, 851–857.
20. Abe I, Saito S, Hari K, Suzuki M, Sato H. Role of dephosphorylation in accumulation of 1-β-D-arabinofuranosylcytosine 5'-triphosphate in human lymphoblastic cell lines with reference to their drug sensitivity. *Cancer Res* 1982, **42**, 2846–2851.
21. Gandhi V, Plunkett W. Modulation of arabinosylnucleoside metabolism by arabinosylnucleotides in human leukemia cells. *Cancer Res* 1988, **48**, 329–334.
22. Plunkett W, Liliemark JO, Estey E, Keating MJ. Saturation of Ara-CTP accumulation during high dose Ara-C therapy: pharmacologic rational for intermediate-dose Ara-C. *Semin Oncol* 1987, **14**, 159–166.
23. Chiba P, Tihan T, Szekeres T, *et al.* Concordant changes of pyrimidine metabolism in blasts of two cases of acute myeloid leukemia after repeated treatment with Ara-C *in vivo*. *Leukemia* 1990, **4**, 761–765.
24. Capizzi RL, White JC, Powell BL, Perrino F. Effect of dose on the pharmacokinetic and pharmacodynamic effects of cytarabine. *Semin Hematol* 1991, **28**, 54–69.
25. Liliemark J. Pharmacokinetic studies on Ara-C. *Scand J Haematol* 1986, **34**, (suppl 44), 41–50.
26. Jamieson GP, Snook MB, Wiley JS. Saturation of intracellular cytosine arabinoside triphosphate accumulation in human leukemic blasts cells. *Leuk Res* 1990, **14**, 475–479.
27. Bhalla K, MacLaughlin W, Coli J, *et al.* Deoxycytidine preferentially protects normal versus leukemic myeloid progenitor cells from cytosinearabinoside-mediated cytotoxicity. *Blood* 1987, **70**, 568–571.
28. Liliemark JO, Plunkett W. Regulation of 1-β-D-arabinofuranosylcytosine 5'triphosphate accumulation in human leukemia cells by deoxycytidine 5'-triphosphate. *Cancer Res* 1986, **46**, 1079–1083.
29. Plagemann PGW, Marz R, Wohlhuter RM. Transport and metabolism of deoxycytidine and 1-β-D-arabinofuranosylcytosine into cultured Novikoff rat hepatoma cells, relationship to phosphorylation and regulation of triphosphate syntheses. *Cancer Res* 1978, **38**, 978–983.
30. Kees UR, Ford J, Dawson VM, Piall E, Aherne GW. Development of resistance to 1-D-arabinofuranosylcytosine after high-dose treatment in childhood lymphoblastic leukemia: analysis of resistance mechanism in established cell lines. *Cancer Res* 1989, **49**, 3015–3019.
31. Dollinger MR, Burchenal JH, Kreis W, Fox JJ. Analogs of 1-β-D-arabinofuranosylcytosine. Studies on mechanisms of action in Burkitt's cell culture and mouse leukemia, and *in vitro* deamination studies. *Biochem Pharmacol* 1967, **16**, 689–706.
32. Chou TC, Hutchison DJ, Schmid FA, Philips FS. Metabolism and selective effects of 1-β-D-arabinofuranosylcytosine in L1210 and host tissues *in vivo*. *Cancer Res* 1975, **35**, 225–236.
33. Ohno Y, Spriggs D, Matsukage A, Ohno T, Kufe D. Effects of 1-β-D-arabinofuranosylcytosine incorporation on elongation of specific DNA sequences by DNA polymerase β. *Cancer Res* 1988, **48**, 1494–1499.

Acknowledgements—This research is supported by Messerstiftung, Kinderhilfestiftung e.V., Dr Erna Ludwig-Stiftung and Arthur and Margarete Ebert-Stiftung.

Pergamon

European Journal of Cancer Vol. 31A, No. 2, pp. 214–221, 1995

0959–8049/95 $9.50+0.00

0959-8049(94)00431-5

Primary Sequence Determination and Molecular Modelling of the Variable Region of an AntiMUC1 Mucin Monoclonal Antibody

G. Denton, G. M. Davies, M. J. Scanlon, S. J. B. Tendler and M. R. Price

Polymerase chain reaction (PCR) products representative of the DNA sequence coding for the variable heavy (V_H) and the variable light (V_L) chains of an antiMUC1 mucin monoclonal antibody, C595, have been produced. These products were cloned, sequenced, and the primary amino acid sequences of the V_H and V_L regions deduced. The hypervariable complementarity determining regions (CDRs) and framework regions in the heavy and light chains were located, and homologies with canonical forms for the CDR loops L_1, L_2, L_3, H_1 and H_2 were identified by database searching. The structure for the H_3 loop was calculated directly. Computational molecular modelling was accomplished using the fully automated AbM package (Oxford Molecular, Oxford, U.K.). Energy minimisation was performed using the program InsightII (Biosym, San Diego, California, U.S.A.). The investigation provides a basis for the molecular analysis of the antigen binding site of the C595 antibody with the aim to identify key residues and interactions involved in the immune recognition of the C595 antibody defined epitope, which is expressed in the majority of breast and ovarian carcinomas.
Eur J Cancer, Vol. 31A, No. 2, pp. 214–221, 1995

INTRODUCTION

MUC1 MUCINS ARE highly glycosylated glycoproteins expressed on the luminal surfaces of glandular epithelia [1, 2]. In breast carcinomas, their expression is frequently up-regulated and they may be secreted into the circulation. Determination of the levels of MUC1 antigen in the blood has been exploited as a measure of tumour burden, and changing levels reflect the response to therapy [3–6]. The MUC1 glycoprotein is a complex molecule with a protein core containing a large domain of variable numbers of a highly conserved 20 amino acid repeat sequence (P D T R P A P G S T A P P A H G V T S A) [7]. Many murine antibodies reactive with the MUC1 mucin have now been produced by immunisation with diverse materials including milk fat globule membranes, tumour cells and isolated mucin preparations. It has been determined that most, if not all, antiMUC1 antibodies reactive with the protein core identify epitopes of three, four or five amino acids within the hydrophilic region, A P D T R P A P, of the 20 amino acid repeat. To date, all antibodies examined require the presence of the arginine within their epitope [2, 8]. The antibody C595 (IgG3, kappa light chain) is one such antibody. This antibody has been proved to be a reagent of clinical utility. It has been employed in immunoassays for the quantitation of circulating mucin in breast cancer patients [9, 10], and has been used for *in vivo* diagnostic tests in the identification of malignant ovarian tumours by immunoscintigraphy [11, 12].

The present investigation was initiated to further explore the molecular basis for the identification of malignant cells by the C595 antibody. The epitope for the antibody has been mapped to the tetra peptide, R P A P [13], and biophysical studies have been performed to define the major structural features of the peptide core in the immunodominant hydrophilic epitope region of the mucin [14, 15]. This report describes the results of primary sequence analysis of the C595 antibody, and progress in the development of a model for the three-dimensional structure of its antigen binding site.

MATERIALS AND METHODS

Monoclonal antibody production

Monoclonal antibody, C595, was originally prepared by conventional hybridoma technology [16] using spleen cells from a BALB/c mouse immunised against purified MUC1 urinary mucin [17]. The C595 monoclonal antibody also has the alternative designation NCRC-48.

Reverse transcription and PCR

C595 hybridoma cells (1×10^7) were harvested by centrifugation (3000 *g* for 5 min) and RNA extracted using RNAzol™B solution (Biogenesis Ltd, Bournemouth, U.K.), according to the manufacturer's instructions. cDNA was produced by conventional means using random primers (N_6-random hexamers; Pharmacia, Uppsala, Sweden) and Superscript™ RNase H⁻ reverse transcriptase (Gibco BRL, Middlesex, U.K.) as previously described [18].

PCR was initiated by adding AmpliTaq® DNA polymerase (Perkin Elmer/Cetus, California, U.S.A.) and the appropriate primers to 1 μl aliquots of generated cDNA [18]. Appropriate negative control samples were included. The primers used for amplification of the heavy and light chain DNA are given in Table 1.

Agarose gel electrophoresis

Agarose gels (1%) were prepared by dissolving NuSieve™ agarose (Flowgen, Kent, U.K.) in TBE buffer (0.09 M Tris-borate, 0.002 M EDTA; [19]) with ethidium bromide incorporated at 0.2 mg/ml. Samples were diluted 4 : 1 with DNA loading buffer (0.25% bromophenol blue, 0.25% xylene cyanol FF, 15% Ficoll Type 400 in water, Pharmacia [19]). Electrophoresis was performed at a constant voltage of 100 V for 45 min. DNA was visualised using a Spectroline TM-312A ultraviolet transilluminator and photographs recorded using a Polaroid DS-34 direct screen instant camera. Molecular weight markers were non-methylated bacteriophage ΦX174 DNA digested with restriction endonuclease HaeIII (NBL).

Cloning of PCR products and restriction digest analysis

Ligations were performed using a TA Cloning™ kit (Invitrogen, San Diego, California, U.S.A.). Transformations were performed using TA OneShot™ supercompetent *E. coli* supplied with the TA Cloning™ kit. Transformed cells were selected and transferred to Luria-Bertani (LB) medium containing kanamycin (5 ml; [19]). The cultures were grown overnight on a gyratory shaker (200 rpm, 37°C). Plasmid DNA was purified from 1.5 ml of the bacterial culture by minipreparation using a modified alkaline lysis method [19].

Purified plasmid DNA was tested for incorporation of the appropriate PCR fragment by digestion with EcoRI (5U; USB, Cambridge Bioscience, Cambridge, U.K.) in High Salt Buffer (USB). The plasmid (1 μl from a 50 μl stock obtained by minipreparation) was digested in a volume of 10 μl for 2 h at 37°C and the restriction digest visualised on a 1% agarose gel.

DNA sequencing and autoradiography

Di-deoxy sequencing reactions [20], incorporating ^{35}S labelled dATP were performed using a Sequenase® Version 2.0 DNA sequencing kit (USB). Single stranded plasmid DNA was prepared [21]. Sequence analysis was achieved through polyacrylamide gel electrophoresis of the sequencing reactions (5.7% w/v acrylamide, 0.3% w/v bis-acrylamide and 48% w/v urea in TBE) using a Bio-Rad Sequi-Gen® nucleic acid sequencing cell. Autoradiography was performed on dried gels using Kodak XAR-5 X-ray film.

Sequence analysis

Autoradiographs of sequencing gels were evaluated independently by at least two individuals and the primary sequence data were loaded on to the SERC SEQNET Computer (Daresbury, Warrington, U.K.) Similar DNA sequences were identified using the program FASTA [22]. Sequences were aligned using the program Gap [23]. After translation, similar protein sequences contained in the Swissprot database were identified with the program PFASTA [22].

N-terminal protein sequencing

Affinity purified C595 monoclonal antibody (200 μg) was added to an equal volume of reducing buffer (0.0625 M Tris/HCl pH 6.8, 10% v/v glycerol, 2% w/v sodium dodecyl sulphate

Correspondence to G. Denton.
G. Denton, G.M. Davies, M.J. Scanlon and M.R. Price are at the Cancer Research Laboratory and Department of Pharmaceutical Sciences; S.J.B. Tendler is at the Department of Pharmaceutical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, U.K.
M.J. Scanlon is presently at the Centre for Drug Design and Development, The University of Queensland, Brisbane, Australia.
Revised 4 Aug. 1994; accepted 5 Sep. 1994.

Table 1. *Table* 1. *Murine immunoglobulin PCR primers*

Heavy chain variable region	
VH1 FOR-2 5′–	TGAGGAGACGGTGACCGTGGTCCCTTGGCCCC- 3′
VH1 BACK 5′–	AGGTCAAACTGCAGCAGTCAGG- 3′
	----GC-G------G----T--
Kappa light chain (V_L and C_L)	
3′ Primer 5′–	GCGCCGTCTAGAATTAACACTCATTCCTGTTGAA- 3′
5′ Primers 5′–	CCAGTTCCGAGCTCGTTGTGACTCAGGAATCT- 3′
	----------GT----G---CCGC-C
	----------GC-C--C---TCTC-A
	----------AGA----C---TCTC-A
	----------GA----C---ACTC-A
	----------CA----C---TCTC-A
	----------GA----A---TCTC-A

(SDS), 0.00125% w/v bromophenol blue, 0.625% v/v 2-mercaptoethanol) and subjected to SDS-polyacrylamide gel electrophoresis (SDS-PAGE) using a 15% homogeneous gel in an LKB Bromma 2050 Midget electrophoresis unit (200 V for 60 min), followed by blotting for 1 h in blotting buffer (10 mM 3-(cyclohexylamino)-1-propanesulphonic acid (CAPS) pH 11, 10% methanol) onto ProBlott™ membrane (Applied Biosystems, ABI, Warrington, U.K.) using the Phastgel Western blotting apparatus (Pharmacia). The membrane was then stained with amido black, the light and heavy chains excised and sequenced independently using an ABI 473A automated peptide sequencer.

The N-terminally blocked C595 light chain was treated with pyroglutamate aminopeptidase (Boehringer) after SDS-PAGE [24]. Briefly, the light chain band was excised and pretreated with 200 μl of 0.5% polyvinylpyrrolidone (PVP-40) in 100 mM acetic acid at 37°C for 24 h. The membrane was washed thoroughly with water and soaked in 0.1 M sodium phosphate buffer (pH 8) containing 5 mM dithiothreitol and 10 mM ethylenediaminetetraacetic acid. Pyroglutamate aminopeptidase (25 μg; enzyme : substrate, 1 : 10) was added and the reaction mixture incubated at 30°C for 24 h. The membrane was washed with water, dried *in vacuo* and sampled by an ABI 473A automated peptide sequencer.

Antibody modelling and minimisation

The structure of the variable region of the C595 antibody was modelled on an Iris Indigo XS24Z Workstation (Silicon Graphics, Reading, Berkshire, U.K.) using the program AbM (Oxford Molecular) which incorporates many of the features of the modelling algorithms of refs [25] and [26]. The modelled structure was subjected to energy minimisation using the module Builder within the program InsightII (Biosym). Ramachandran plot analyses were obtained using the program Procheck [27].

RESULTS

PCR product and cloning analysis

The DNA coding for the light chain (kappa) and the heavy chain variable domain (V_H) of C595 was obtained by PCR amplification. The PCR products were ligated into the PCR™II vector and used to transform *E. coli*. Colonies selected were tested for incorporation of the appropriately sized PCR fragment by restriction digest analysis. Figure 1 shows the digests of plasmids into which either the V_H domain (Figure 1; upper) or the kappa chain (Figure 1; lower) PCR products have been incorporated. The agarose gels confirm that the DNA inserted into the vector was of the same size as the corresponding PCR

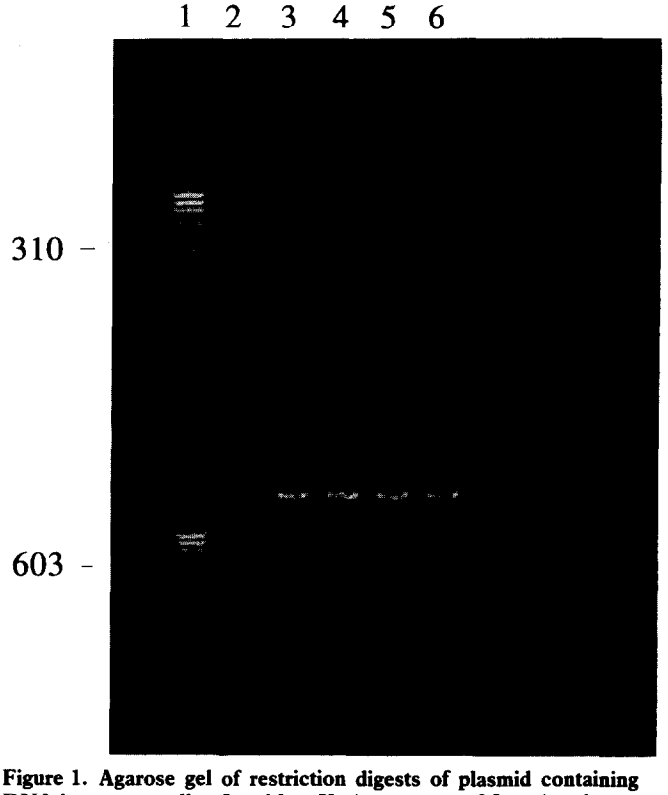

Figure 1. Agarose gel of restriction digests of plasmid containing DNA inserts encoding for either V_H (upper part of figure) or kappa light chain (lower part of figure). In both cases, lane 1 contains the molecular weight standards (ΦX174 DNA–HaeIII digest with the 603 and 310 bp markers indicated to the left), lane 2 contains the appropriate PCR product, and the remainder of the lanes contain corresponding digests of plasmid preparations.

product suggesting that the correct DNA fragment had been inserted.

Sequence analysis

Autoradiographs of the sequencing reactions performed on numerous clones obtained through independently obtained PCRs yielded the primary DNA sequence and deduced protein sequences of the V_H and V_L domains of C595 antibody as shown in Table 2.

Because of the degenerate nature of the primers used to generate PCR products (Table 1), N-terminal protein sequencing was performed to evaluate the precise amino acid composition of the heavy and light chains encoded for at the priming sites. Amplification of C595 cDNA using the various kappa

Table 2. cDNA and deduced protein sequence of C595 (CDRs are underlined)

(a) C595 VL

DNA sequence								CCA	GCA	ATC	ATG	TCT	GCA	TCT	CCA
Protein sequence	Q	I	V	L	T	Q	S	P	A	I	M	S	A	S	P
	1														
	GGG	GAG	AAG	GTC	ACC	ATG	ACC	TGC	AGT	GCC	AGC	TCA	AGT	GTA	AGT
	G	E	K	V	T	M	T	C	S	A	S	S	S	V	S
						21									
	TAC	ATG	CAC	TGG	TAC	CAG	CAG	AAG	TCA	GGC	ACC	TCC	CCC	AAA	AGA
	Y	M	H	W	Y	Q	Q	K	S	G	T	S	P	K	R
											41				
	TGG	ATT	TAT	GAC	ACA	TCC	AAA	CTG	GCT	TCT	GGA	GTC	CCT	GCT	CGC
	W	I	Y	D	T	S	K	L	A	S	G	V	P	A	R
	TTC	AGT	GGC	AGT	GGG	TCT	GGG	ACC	TCT	TAC	TCT	CTC	ACA	ATC	AGC
	F	S	G	S	G	S	G	T	S	Y	S	L	T	I	S
	61														
	AGC	ATG	GAG	GCT	GAA	GAT	GCT	GCC	ACT	TAT	TAC	TGC	CAG	CAG	TGG
	S	M	E	A	E	D	A	A	T	Y	Y	C	Q	Q	W
						81									
	AGT	AGT	AAC	CCA	CCC	ACG	TTC	GGA	GGG	GGG	ACC	AAG	CTG	GAA	
	S	S	N	P	P	T	F	G	G	G	T	K	L	E	I
											101				
	K	R	A												
			108												

(b) C595 VH

DNA sequence									GGA	GGC	TTA	GTG	CAG	CCT	GGA
Protein sequence	E	V	Q	L	V	E	S	G	G	G	L	V	Q	P	G
			111												
	GGG	TCC	CTG	AAA	CTC	TCC	TGT	GCA	GCC	TCT	GGA	TTC	ACT	TTC	AGT
	G	S	L	K	L	S	C	A	A	S	G	F	T	F	S
								131							
	AGC	TAT	GGC	ATG	TCT	TGG	GTT	CGC	CAG	ACT	CCA	GAC	AAG	AGG	CTG
	S	Y	G	M	S	W	V	R	Q	T	P	D	K	R	L
													151		
	GAG	TTG	GTC	GCA	ACC	ATT	AAT	AGT	AAT	GGT	GGT	AGC	ACC	TAT	TAT
	E	L	V	A	T	I	N	S	N	G	G	S	T	Y	Y
	CCA	GAC	AGT	GTG	AAG	GGC	CGA	TTC	ACC	ATC	TCC	AGA	GAC	AAT	GCC
	P	D	S	V	K	G	R	F	T	I	S	R	D	N	A
			171												
	AAG	AAC	ACC	CTG	TAC	CTG	CAA	ATG	AGC	AGT	CTG	AAG	TCT	GAG	GAC
	K	N	T	L	Y	L	Q	M	S	S	L	K	S	E	D
								191							
	ACA	GCC	ATG	TAT	TAC	TGT	GCA	AGA	GAT	CGG	GAT	GGT	TAC	GAC	GAG
	T	A	M	Y	Y	C	A	R	D	R	D	G	Y	D	E
													211		
	GGA	TTT	GAC	TAC	TGG	GGC									
	G	F	D	Y	W	G									
						219									

chain 5′ primers given in Table 1, yielded PCR products of the correct size in each case (data not shown), although the DNA sequences of these seven primers encode seven different peptides. The significance of this observation, in the context of modelling, is outlined by the effect of residue 2 of the light chain. In order for L_1 to form a canonical loop, this residue must be an isoleucine. Although this residue does not reside within the CDR, it is important in the stabilisation of the loop structure [25]. The presence of any other residue at this position results in L_1 not forming a canonical loop structure. In this situation, the algorithm used in the AbM program would calculate *de novo* the structure of this CDR.

N-terminal protein sequencing revealed the residues present at positions 1 to 5 for the light chain and positions designated 109 to 118 for the heavy chain (Table 2). Initial studies found the light chain to be N-terminally blocked, as no sequence information was obtained from over 200 pM of protein. Subsequent treatment with pyroglutamate aminopeptidase led to the generation of sequence data for the first five residues of the light chain (inclusive of the cleaved glutamine), indicating the presence of pyroglutamate at the N-terminus. The sequence obtained (Table 2) was that expected by comparison of the remainder of the sequence by germline DNA homology searching. The remaining residues encoded by the PCR primers (at

positions 6 and 7) were deduced from the murine light chain germline DNA database.

Protein sequence data for the N-terminal sequenced heavy chain is presented in Table 2 as residues 109–118. Residues at positions 116–118 were all glycine, in agreement with the protein sequence deduced from data generated by di-deoxy DNA sequencing.

Sequence homology analysis

The five protein sequences with greatest homology to the identified sequence for the V_H and V_L domains of C595 are listed in Table 3. Of 50 similar protein sequences identified by database searching, all of the homologous sequences coded for immunoglobulin variable regions of the correct chain type. In the case of both the V_H and V_L domains, the most similar sequences were murine immunoglobulin. The majority of mismatches between C595 antibody and identified homologous sequences occur within the CDRs. Furthermore, sequence data obtained for the C_L domain of C595 showed complete homology with a number of other murine C_L domains.

Modelling analysis

Primary sequence data obtained for C595 V_H and V_L regions were entered into the AbM program and the modelling routines initiated. The majority of the backbone of the variable region of the C595 antibody was constructed using simple sequence homology comparisons to known crystallographic structures. For C595, the sequences with the highest homology identified by these database searches were the V_H of the murine antibody 17/9 (an anti-influenza virus peptide antibody) and the V_L of the murine antibody HyHel-5 (an anti-lysozyme antibody). These were used as the backbone templates for the model.

Comparison of the primary sequences of the CDRs identified for C595 with the primary sequences of known CDR loops L_1, L_2, L_3, H_1 and H_2 (for which canonical forms have been identified, [25, 28]), whose three-dimensional structure has been resolved, is presented in Table 4. Analysis by reference to information presented by Chothia and coworkers [25, 28] and by computational database interrogation by AbM, revealed canonical loop structures for C595 L_1, L_2, L_3, H_1 and H_2 regions. Loops L_1, L_2 and L_3 showed a high degree of homology with type 1 canonical structures for these loops, whereas for the heavy chain, H_1 and H_2 were homologous to type 1 and type 3 structures, respectively [25, 28]. These canonical loop structures were edited on to the backbone templates.

Table 3. Protein sequences with highest homology to C595 V-regions

Protein	Swissprot accession number	% identity
(a) Mouse Ig heavy chain V region	hv55-m	90.8
Mouse Ig heavy chain V region	hv54-m	89.8
Mouse Ig heavy chain V region	hv53-m	90.8
Mouse Ig heavy chain V region	hv57-m	87.8
Mouse Ig heavy chain V region	hv58-m	84.7
(b) Mouse Ig kappa chain V-VI region	kv6f-m	93.5
Mouse Ig kappa chain V–VI region	kv6h-m	92.5
Mouse Ig kappa chain V–VI region	kv6i-m	92.5
Mouse Ig kappa chain V–VI region	kv6g-m	91.6
Mouse Ig kappa chain V–VI region	kv6j-m	91.6

Table 4. Comparison of primary sequence of C595 CDRs with known sequences of canonical forms. After [25]

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
	C595	**S**	**S**	**S**	**V**	**S**	—	—	—	—	—	—	—	**Y**	**I**	**A**	**M**	**Y**

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
	C595	**D**	**T**	**S**	**I**	**G**

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
	C595	**W**	**S**	**S**	**N**	**P**	**P**	**Q**

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
	C595	**G**	**F**	**T**	**F**	**S**	**S**	**Y**	**M**	**Y**

H2 Regions

Canonical Structure	Protein	52a	b	c	53	54	55	71
3	KOL	D	—	—	D	G	S	R
	J539	P	—	—	D	S	G	R
	NQ10	S	—	—	G	S	S	R
	C595	**S**	—	—	**N**	**G**	**G**	**R**

CDR H_3 does not have any comparable forms as far as loop structures are concerned, so putative loops were built using information from the Brookhaven database. This database contains the interatomic distances for carbon atoms in proteins whose three-dimensional structure has been resolved. The amino acid sequence of the CDR H_3 was compared with protein sequences in the database and translated into the three-dimensional coordinates of 425 candidate H_3 loops. The ideal generated

loop would have the same sequence and conformation as a section of a protein in the database. This is unlikely, so a method of ranking candidate H_3 loops is employed, based on the residue similarity between generated and database loops, and by identifying key structurally determining residues (SDRs) within these loops. The 200 highest scoring loops were selected as the preferred conformations.

The central portions of these loops were reconstructed using CONGEN, a program which searches and relieves unsatisfactory van der Waals contacts between amino acids by altering their conformations. The energies of individual atoms in these revised loops, based on the relationships between neighbouring amino acids, were analysed and the five lowest energy conformers for the H_3 loop selected and retained.

One H_3 CDR loop was finally selected from these five low energy conformers, chosen on the basis of homology with the database structure from which it was taken. The model of the C595 antibody variable region produced in this way is essentially a non-energy minimised structure.

Energy minimisation

Energy minimisation was performed on the AbM generated C595 model using the Optimize option within the computer program InsightII (Biosym), incorporating the Combined Valence Forcefield [29]. The model was subjected to 691 steepest descents iterations, followed by 4910 conjugate gradients iterations (Figure 2).

Ramachandran plot analysis

In the polypeptide backbone, the amount of rotation at the bond between nitrogen and the α-carbon atom is designated phi (ϕ) and the rotation between the α-carbon and carbonyl atoms is designated psi (ψ). The conformation of the main chain is completely defined when ϕ and ψ are specified for each residue in the chain. Certain combinations are not acceptable because of steric hindrance. Allowed ranges can be readily predicted and visualised in steric contour diagrams known as Ramachandran plots. The Ramachandran plot of the minimised C595 model is displayed in Figure 3 and the plot statistics illustrate the quality of the model constructed.

Several representations of the model for the structure of the variable region of the C595 antibody are shown in Figure 4. A ribbon style view of the β-barrel assembly of this domain is

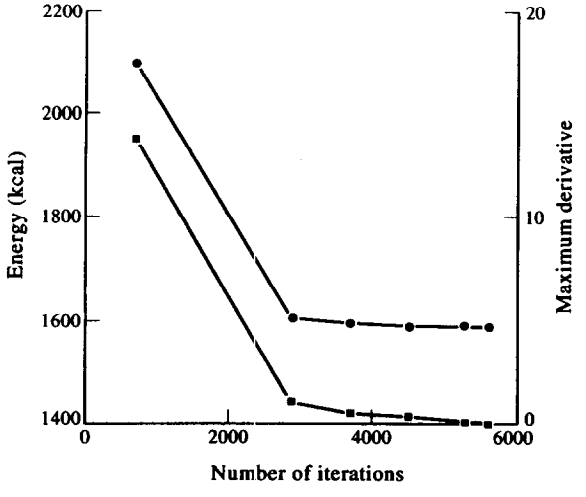

Figure 2. The number of iterations performed on the AbM created model of C595 to produce the minimal energy conformation for the model (—●—). The maximum derivative is also indicated (—■—).

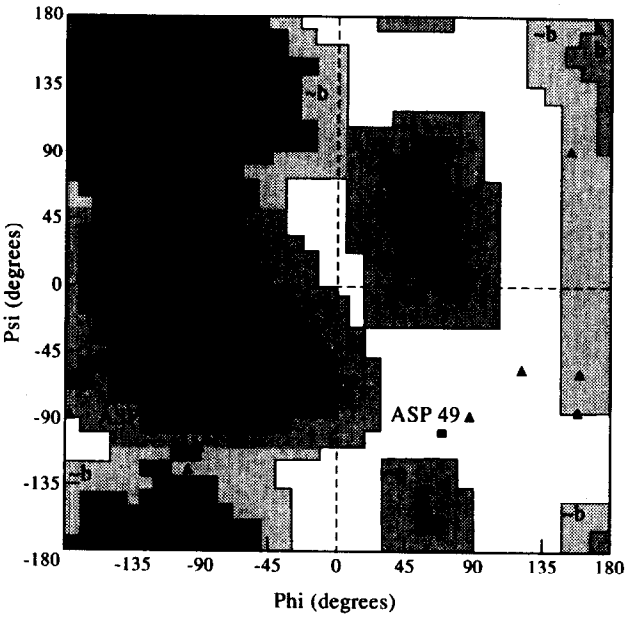

Figure 3. Ramachandran plot analysis of the minimised molecular model of the variable region of the C595 antibody (for non-glycine and non-proline residues). This analysis examines to what extent ϕ and ψ angles about Cα calculated in the model are allowable on the basis of steric hindrance.

Plot statistics:

Residues in most favoured regions [A,B, L]	**150**	**81.1%**
Residues in additionally allowed regions [a, b, 1, p]	**32**	**17.3%**
Residues in generously allowed regions [~a, ~b, ~1, ~p]	**2**	**1.1%**
Residues in disallowed regions	**1**	**0.5%**
Number of non-glycine and non-proline residues	**185**	**100.0%**
Number of end residues	**3**	
Number of glycine residues (shown as triangles)	**22**	
Number of proline residues	**9**	
Total number of residues	**219**	

shown in Figure 4a with the hypervariable loops clearly located at the upper end of the structure. A view from above the CDR region illustrates the compact array of loops surrounding a cleft or depression within this region (Figure 4b). The compact packing around this cleft may also be visualised by examination of Figure 4c showing a CPK (Corey-Pauling-Koltun) representation of the same view of the CDR domains shown in the ribbon style model in Figure 4b.

DISCUSSION

PCR technology has been utilised in a number of studies to clone and sequence immunoglobulin variable regions without having to construct cDNA libraries [18]. Sequences coding the whole of the mature variable domain are amplified using sense primers annealing to the leader sequences of heavy and light chains, and antisense primers annealing near to the variable-constant region junctions. This strategy, incorporating N-terminal protein sequencing, has been used to identify the V domain sequences of the C595 antibody.

The primary sequence of the C595 antibody V_H and V_L regions has been determined (Table 2), and shows a high level of homology with respect to other murine immunoglobulin sequences.

It is possible that hybridoma DNA, encoding the unexpressed alleles of C595 heavy and light chains, could have been isolated, cloned and sequenced instead of those representative of the

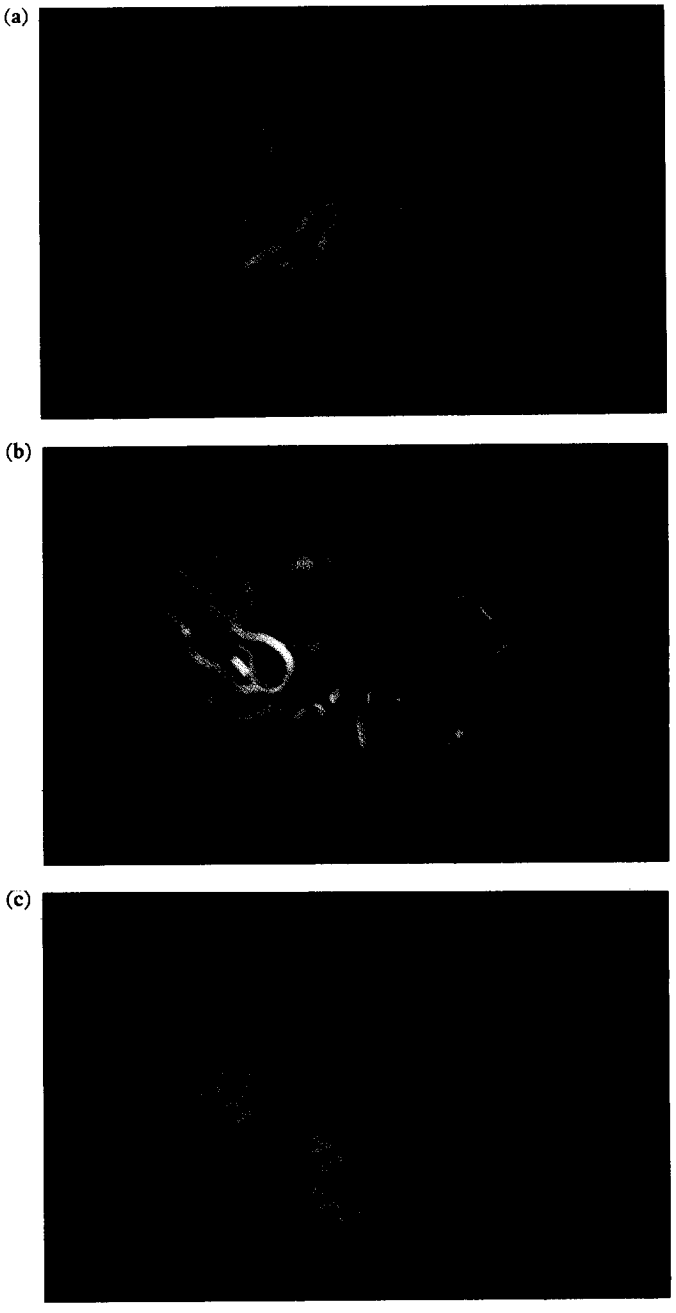

Figure 4. Molecular models of the variable region of the C595 antibody. Figures 4a and 4b show a ribbon style presentation of the protein backbone. Figure 4c shows the same view as in Figure 4b, over the C595 CDR loops in a CPK (Corey-Pauling-Koltun) representation. In each figure, the light chain is depicted in white and the heavy chain is gold. V_L CDR loops are coloured green and V_H CDR loops are purple.

monoclonal antibody. This seems unlikely for three reasons. Firstly, at the time of harvest, the C595 hybridoma was growing at log phase and analysis of the supernatant derived from these cells by ELISA showed the secretion of a high titre of C595 antibody. This would indicate the presence of a large amount of antibody encoding mRNA within the hybridoma. With a substantial excess of message for the secreted antibody, the chances of an amplifying message from the normally unexpressed alleles present in these cells would be remote. Secondly, the data generated through molecular biological techniques are in agreement with the data originating through N-terminal protein sequencing, relating the transcribed and translated DNA messages to the expressed protein. Finally, PCR products representative of C595 variable regions have been incorporated into appropriate vectors for the production of a recombinant single chain variable fragment scFv (rscFv) and recombinant Fab (rFab) fragments using phage display technology [30]. Preliminary examination of the binding of C595 rFab expressed on the phage head by ELISA indicates an antigen specificity essentially identical to the native C595 antibody. These findings suggest that the PCR products used in DNA sequencing represent the DNA encoding for the C595 antibody in the hybridoma.

It is possible that errors of transcription could have occurred during cDNA synthesis and that proof reading errors could have been introduced during the PCR amplification. This seems improbable since numerous independently cloned and sequenced PCR products originating from separate RNA extractions yielded the same sequences for each particular V_H or V_L encoding DNA fragment. The only differences in sequence occurred at the 3′ ends of the PCR products where the polymerase enzyme had incorporated differing numbers of adenosine nucleosides. The addition of these nucleosides is a natural function of the enzyme, does not affect the determination of the DNA sequence for C595 variable regions and is a property resourcefully utilised in the TA cloning™ strategy.

The sequences of five of the six C595 CDR loops were comparable with those in libraries of canonical structures, whereas no canonical structures exist for H_3. The epitope for the C595 antibody is the simple tetra peptide R P A P, and it has been determined that the arginine residue is an essential amino acid for recognition by the C595 antibody. Replacement of this residue with any of the other natural amino acids is not permissible for antibody binding [8]. A putative arginine binding site is centred around the H_3 loop since three of the first four aspartic acid residues on the H_3 loop are surface accessible, and their side chains may interact with the charged arginine residue in the epitope. As combining site residues need to be positioned precisely in order to take full advantage of electrostatic, hydrogen bonding and van der Waals interactions [31], additional evidence is required to define a putative site. Clearly, new methodologies are required for the characterisation of antigen–antibody interactions at the molecular and atomic levels and such techniques should have the capacity to accommodate the dynamic aspects of the interactions as they occur in solution. At present, high field nuclear magnetic resonance (NMR) spectroscopy would appear to offer many advantages in this respect although the combination of NMR with molecular modelling and X-ray crystallography should prove the most effective. Structural studies incorporating these techniques, on rscFv and rFab fragments of C595 produced through PCR and phage display technology, should assist in resolving the three-dimensional structure of the binding site of this antibody.

The relationships between the amino acid sequences of immunoglobulins and the structures of their antigen binding sites are critical for understanding the molecular basis of antigen recognition. Antibody modelling is an important approach in the assembly of structural information relevant to these interactions. These data are invaluable for future work directed towards the development of specifically designed antibodies engineered to be preferentially reactive with human tumour cells. Site specific mutagenesis of C595 encoding DNA, expression of mutant C595 rFab or rscFv fragments and subsequent binding studies should give an insight into the key residues in the CDRs involved in antigen recognition. In addition, primary sequence analysis of

the variable region of a murine antibody, supported by the results of molecular modelling, provides the foundation for further manipulations involving CDR grafting onto a human immunoglobulin framework to produce an antibody which would be considerably less immunogenic when administered into humans. These approaches have already met with success in the case of the development of a humanised HMFG1 antibody which identifies a peptide epitope on the MUC1 antigen [32] distinct from that defined by the C595 antibody. The findings obtained in this study will also assist in the production of engineered C595 antibody fragments (Fab, scFv) to be exploited for tumour localisation by radioimmunoscintigraphy and have the potential to be utilised as the preferred reagents for radioimmunotherapy.

In summary, the present investigation has been developed to provide a better understanding of the nature of antibody interaction of C595 with products of the MUC1 gene. Our intention is to utilise the information presented here to establish a basis for future rational design of engineered antibodies and their respective fragments with increased or modified affinity for antigens aberrantly expressed as a consequence of malignant transformation and increased potential usage *in vivo*. The nature of the experimentation outlined here, facilitates rapid and efficient parallel studies to be performed on other antibodies reactive with MUC1 mucins, as well as antibodies recognising other tumour related antigens.

1. Gendler SJ, Spicer AP, Lalani E-N, *et al.* Structure and biology of a carcinoma-associated mucin, MUC1. *Am Rev Resp Disease* 1991, **144**, S42–S47.
2. Price MR, Tendler SJB. Polymorphic epithelial mucins (PEM): molecular characteristics and association with breast cancer. *The Breast* 1993, **2**, 3–7.
3. Kenemans P, Bast RC, Yemeda CA, Price MR, Hilgers J. Mucins and mucin-like antigens as circulating tumour markers. In Hilgers J, Zotter S, eds. *Cancer Reviews*, Volumes 11–12. Copenhagen, Munksgaard, 1988, 119–144.
4. Kufe D, Hayes D, Abe M. Monoclonal antibody assays for breast cancer. In Kupchik HZ, ed. *Cancer Diagnosis In Vitro Using Monoclonal Antibodies*, Immunology Series Volume 39. New York, Dekker, 1988, 67–100.
5. Linsley P, Brown J, Magnani J, Horn D. Monoclonal antibodies reactive with mucin glycoproteins found in sera from breast cancer patients. *Cancer Res* 1988, **48**, 2138–2148.
6. Gion M, Mione R, Nascimben O, *et al.* The tumour associated antigen CA15.3 in primary breast cancer. Evaluation in 667 cases. *Br J Cancer* 1991, **63**, 809–813.
7. Gendler S, Taylor-Papadimitriou J, Duhig T, Rothbard J, Burchell J. A highly immunogenic region of a human polymorphic epithelial mucin expressed by carcinomas is made up of tandem repeats. *J Biol Chem* 1988, **263**, 12820–12823.
8. Briggs S, Price MR, Tendler SJB. The specificity of antibody recognition of carcinoma-associated epithelial mucins: antibody binding to synthetic peptide epitopes. *Eur J Cancer* 1993, **29A**, 230–237.
9. Dixon AR, Price MR, Hand CW, Sibley PEC, Selby C, Blamey RW. Epithelial mucin core antigen (EMCA) in assessing therapeutic response in advanced breast cancer—a comparison with CA15.3. *Br J Cancer* 1993, **68**, 947–949.
10. Price MR, Briggs S, Scanlon MJ, Tendler SJB, Sibley PEC, Hand CW. The mucin antigens—what are we measuring? *Disease Markers* 1992, **9**, 205–212.
11. Symonds IM, Price MR, Pimm MV, *et al.* Preliminary report of tumour localisation and imaging of ovarian neoplasia with a new monoclonal antibody raised against urinary mucin. In Klapdor R, ed. *Tumor Associated Antigens, Oncogenes, Receptors, Cytokines in Tumor Diagnosis and Therapy at the Beginning of the Nineties*. W Zuckschwerdt Verlag, Munchen, 1992, 572–577.
12. Perkins AC, Symonds IM, Pimm MV, Price MR, Wastie ML, Symonds EM. Immunoscintigraphy of ovarian carcinoma using a monoclonal antibody (In-111-NCRC-48) defining a polymorphic epithelial mucin (PEM) epitope. *Nucl Med Commun* 1993, **14**, 578–586.
13. Price MR, Hudecz F, O'Sullivan C, Baldwin RW, Edwards PM, Tendler SJB. Immunological and structural features of the protein core of human polymorphic epithelial mucin. *Mol Immunol* 1990, **27**, 795–802.
14. Tendler SJB. Elements of secondary structure in a human epithelial mucin core peptide fragment. *Biochem J* 1990, **267**, 733–737.
15. Scanlon MJ, Morley SD, Jackson DE, Price MR, Tendler SJB. Structural and computational investigations on the conformation of antigenic fragments of human polymorphic epithelial mucins. *Biochem J* 1992, **284**, 137–144.
16. Kohler G, Milstein C. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 1975, **256**, 495–497.
17. Price MR, Pugh JA, Hudecz F, *et al.* C595–a monoclonal antibody against the protein core of human urinary epithelial mucin commonly expressed in breast carcinomas. *Br J Cancer* 1990, **61**, 681–686.
18. Orlandi R, Gussow DH, Jones PT, Winter G. Cloning immunoglobulin variable domains for expression by the polymerase chain reaction. *Proc Natl Acad Sci USA* 1989, **86**, 3833–3837.
19. Sambrook J, Fritsch EF, Maniatis T. *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Press, Cold Spring Harbor, 1989.
20. Sanger F, Nicklen S, Coulsen AR. DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 1977, **74**, 5463–5467.
21. Wong WM, Au DMY, Lam VMS, Tam JWO, Cheng LYL. A simplified and improved method for the efficient double stranded sequencing of mini-prep plasmid DNA. *Nucleic Acid Res* 1990, **18**, 5573.
22. Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 1988, **85**, 2444–2448.
23. Needleman SB, Wunsch CD. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 1970, **48**, 443–453.
24. Hirano H, Komatsu S, Kajiwara H, Takagi Y, Tsunasawa S. Microsequence analysis of the N-terminally blocked proteins immobilised on polyvinylidene difluoride membrane by Western blotting. *Electrophoresis* 1993, **14**, 839–846.
25. Chothia C, Lesk AM, Tramontano A, *et al.* Conformations of immunoglobulin hypervariable regions. *Nature* 1989, **342**, 877–883.
26. Martin ACR, Cheetham JC, Rees AR. Modelling antibody hypervariable loops: a combined algorithm. *Proc Natl Acad Sci USA* 1989, **86**, 9268–9272.
27. Laskowski RA, MacArthur MW, Moss SD, Thornton JM. Procheck: a program to check the stereochemical quality of protein structures. *J Appl Crys* 1993, **26**, 283–291.
28. Chothia C, Lesk AM. Canonical structures of the hypervariable regions of immunoglobulins. *J Mol Biol* 1987, **196**, 901–917.
29. Bruccoleri RE, Karplus M. Prediction of the folding of short polypeptide segments by uniform conformational sampling. *Biopolymers* 1987, **26**, 137–168.
30. Clackson T, Hoogenboom HR, Griffiths AD, Winter G. Making antibody fragments using phage display libraries. *Nature* 1991, **352**, 624–628.
31. Mian IS, Bradwell AR, Olson AJ. Structure, function and properties of antibody binding sites. *J Mol Biol* 1991, **217**, 133–151.
32. Verhoeyen ME, Saunders JA, Price MR, *et al.* Construction of a reshaped HMFG1 antibody and comparison of its fine specificity with that of the parent mouse antibody. *Immunology* 1993, **78**, 364–370.

Acknowledgements—These studies were supported by the Cancer Research Campaign (Project Grant Number SP 2168/0101). The SERC is thanked for the provision of a studentship for MJS and a CASE Award with Unilever for GMD.

Thanks are expressed to Dr P. Jones and Dr P. Scotting (Department of Biochemistry, Queen's Medical Centre, Nottingham) and Dr L. Wilson (Yamanouchi Research Institute, Oxford) for helpful discussions on molecular biological techniques. Thanks are also expressed to John Keyte and Kevin Bailey (BSAU, Department of Biochemistry, Queen's Medical Centre, Nottingham) for protein sequencing and Kathryn Measures (Cancer Research Laboratory, Nottingham) for discussion and advice on molecular modelling. V_H and kappa chain primers were kindly provided by Dr R. Spooner (Hammersmith Hospital, London) and Dr M. Walker (Wolfson Research Laboratories, University of Birmingham, Birmingham), respectively.