Crystal structure of the anti-(carcinoembryonic antigen) single-chain Fv antibody MFE-23 and a model for antigen binding based on intermolecular contacts

Mark K. BOEHM^{*†}, Adam L. CORPER[‡], Tommy WAN[‡], Maninder K. SOHI[‡], Brian J. SUTTON[‡], Jeremy D. THORNTON[†], Patricia A. KEEP[†], Kerry A. CHESTER[†], Richard H. J. BEGENT[†] and Stephen J. PERKINS^{*1}

^{*}Department of Biochemistry and Molecular Biology, Royal Free Campus, Royal Free and University College Medical School, University College London, Rowland Hill Street, London NW3 2PF, U.K., [†]Department of Oncology, Royal Free Campus, Royal Free and University College Medical School, University College London, Rowland Hill Street, London NW3 2PF, U.K., and [‡]The Randall Centre, King's College London, New Hunt's House, Guy's Campus, London Bridge, London SE1 1UL, U.K.

MFE-23 is the first single-chain Fv antibody molecule to be used in patients and is used to target colorectal cancer through its high affinity for carcinoembryonic antigen (CEA), a cell-surface member of the immunoglobulin superfamily. MFE-23 contains an N-terminal variable heavy-chain domain joined by a (Gly₄Ser)₃ linker to a variable light-chain (V_L) domain (κ chain) with an 11-residue C-terminal Myc-tag. Its crystal structure was determined at 2.4 Å resolution by molecular replacement with an R_{cryst} of 19.0%. Five of the six antigen-binding loops, *L1*, *L2*, *L3*, *H1* and *H2*, conformed to known canonical structures. The sixth loop, *H3*, displayed a unique structure, with a β -hairpin loop and a bifurcated apex characterized by a buried Thr residue. In the crystal lattice, two MFE-23 molecules were associated back-to-back in a manner not seen before. The antigen-binding site displayed a large acidic region located mainly within the *H2* loop

and a large hydrophobic region within the *H3* loop. Even though this structure is unliganded within the crystal, there is an unusually large region of contact between the *H1*, *H2* and *H3* loops and the β -sheet of the V_L domain of an adjacent molecule (strands DEBA) as a result of intermolecular packing. These interactions exhibited remarkably high surface and electrostatic complementarity. Of seven MFE-23 residues predicted to make contact with antigen, five participated in these lattice contacts, and this model for antigen binding is consistent with previously reported site-specific mutagenesis of MFE-23 and its effect on CEA binding.

Key words: antibody-combining site, dimerization, humanization, immunoglobulin.

INTRODUCTION

The vertebrate immune system is able to produce high-affinity antibodies to protect the host against infection. In the same way, antibodies can be used in a wide variety of therapeutic and diagnostic clinical applications [1]. Thus antibodies raised against pathology-specific markers can be used for targeting purposes. For tumours, an important marker is carcinoembryonic antigen (CEA), which has been used as a target for therapeutic treatment and a serum marker to monitor tumour growth [2]. CEA is a membrane-bound cell surface glycoprotein that contains seven domains of the immunoglobulin superfamily and is 50% carbohydrate by weight. In solution, the seven domains form a highly elongated arrangement similar to that observed between the two domains of the homologous cell-surface protein CD2, from which the 28 carbohydrate chains of CEA are extended into solution [3]. This extended CEA structure would be significantly exposed on the cell surface and would present extensive protein surfaces that are thought to mediate its protein–protein adhesive functions and to be important for carcinogenesis [4].

MFE-23 is a high-affinity antibody used against colorectal tumours, being the first single-chain Fv (scFv) molecule to be

used in patients [5]. It binds to CEA with high affinity ($K_d = 2.5 \pm 1.3$ nM), which in principle confers a therapeutic advantage [6]. Thus MFE-23 was found to have a better biodistribution in murine tumour models than a murine anti-CEA scFv derived from the monoclonal antibody A5B7 [7]. Such scFv molecules possess a variable heavy-chain (V_H) immunoglobulin domain joined with a variable light-chain (V_L) domain by an oligopeptide linker [8] and are the smallest antibody fragments that retain full antibody–antigen binding functions. They generally have better tumour penetration and a lower immunogenicity than four-domain Fab fragments from antibodies or intact 12-domain antibodies; this is advantageous for targeting purposes [9]. MFE-23 was developed from a filamentous phage expression library of random murine V_H–V_L-domain pairings obtained from a CEA-immunized mouse [10], in which the domains are connected by a (Gly₄Ser)₃ linker (see Figure 1). The screening of the phage library for high-affinity binding to CEA resulted in the selection of MFE-23, which is currently undergoing trials on patients.

Improvements to the molecular design of MFE-23 as an anti-tumour agent, such as its humanization, would be greatly facilitated by the determination of its crystal structure. The lifetime of murine MFE-23 in human circulation would be

Abbreviations used: CEA, carcinoembryonic antigen; H1–H120 and L1–L106, sequence numbering of MFE-23 heavy and light chains respectively; H1–H3 and L1–L3, antigen-binding loops of MFE-23; r.m.s., root-mean-square; SDR, specificity-determining residue; scFv, single-chain Fv; V_H, variable heavy-chain domain; V_L, variable light-chain domain.

¹ To whom correspondence should be addressed (e-mail steve@rfhsm.ac.uk).

The identification code for the MFE-23 co-ordinates in the Protein Data Bank is 1qok.

improved by the rational design of a less immunogenic scFv structure. The crystal structure presented here is the first for an anti-tumour scFv molecule derived from phage technology. Even though five scFv crystal structures have been determined (those of Se155-4, NC10, L5MK16, CC49/212 and C219 [11–15]), the present MFE-23 structure determination exhibited four novel aspects. First, although canonical loop structures for five of the six antigen-binding loops, *H1*, *H2*, *L1*, *L2* and *L3*, in MFE-23 can be predicted by database approaches [16], and even though there is much recent progress in classifying and predicting *H3* loops [17–19], we find that MFE-23 has a unique *H3* structure. Secondly, the MFE-23 structure bears a close resemblance to a particular human antibody; this permits a humanization strategy to be proposed. Thirdly, the intermolecular contacts found within MFE-23 molecules in the crystal structure showed how an unusual mode of back-to-back association of Fv fragments can be achieved. Fourthly, and most importantly, the MFE-23 crystal structure fortuitously provided information on how this antibody might recognize CEA. Unusually extensive intermolecular contacts were found in the crystal structure involving residues of the *H1*–*H3* loops, which commonly participate in antigen binding. These correspond closely to predictions of specificity-determining residues (SDRs) in MFE-23 [20] and are supported by experimental mutagenesis data on MFE-23 [21].

MATERIALS AND METHODS

Cloning, expression and purification of MFE-23

The cloning, expression and purification of the murine scFv molecule MFE-23 has been described [10]. MFE-23 (GenBank reference 2299568) contained a pelB leader peptide at its N-terminus to permit its secretion from bacterial cells (after which this leader is cleaved), and a C-terminal 11-residue Myc-tag for identification during purification. A pUC119 plasmid containing this construct was used to transform *Escherichia coli* cells. MFE-23 was purified from cultures in 2TY medium (16 g/l bactotryptone, 10 g/l yeast extract, 5 g/l NaCl) containing ampicillin (100 µg/ml) and glucose (0.1%), to which a final concentration of 1 mM isopropyl β-D-thiogalactoside was added to induce expression after the cells had reached a D_{600} of 0.9. From the culture, the supernatant was collected by centrifugation, filtered (0.2 µm pore-size filter; Nalgene), concentrated 5-fold by using a spiral cartridge with a 10 kDa cut-off (S1Y10; Amicon), and dialysed against PBS [137 mM NaCl, 0.5 mM MgCl₂, 2.7 mM KCl, 8.1 mM Na₂HPO₄, 1.5 mM KH₂PO₄ (pH 7.4)] containing 0.02% NaN₃. After refiltration, MFE-23 was purified by affinity chromatography with CEA covalently attached to Sepharose-4B gel (Pharmacia), eluting it with 0.05 M diethylamine, pH 11, and buffering immediately with 1 M phosphate buffer, pH 7.5. The MFE-23 fractions were pooled, dialysed overnight against PBS and 0.02% NaN₃, concentrated by ultrafiltration (PM10 membrane; Amicon), subjected to gel filtration on Sephacryl S-100 (Pharmacia) and concentrated (YM10 membrane; Amicon). Samples used for crystallization were dialysed against 20 mM Tris/HCl, pH 6.5. The concentration of MFE-23 was determined from a $A_{1\text{cm}, 420}^{1\%}$ of 20.0 calculated from its sequence [22].

Crystallization and data collection

Crystals of MFE-23 were grown by the hanging-drop vapour-diffusion method at 18 °C. In brief, protein (2 mg/ml) was mixed 1:1 with 100 mM Tris/HCl buffer, pH 6.5, containing 45%-satd. (NH₄)₂SO₄. A 10 µl droplet of this mixture was equilibrated against 0.5 ml of 100 mM Tris/HCl, pH 6.5, in 45%-satd. (NH₄)₂SO₄. Crystals were mounted in quartz glass capillary

Table 1 Summary of data-collection and structure-refinement statistics for MFE-23

Stage	Variable	Value
Data collection	Resolution (Å)	15.0–2.4
	No. of unique reflections	
	All data	11 537
	2.53–2.40 Å shell	1613
	R_{merge}^* (%)	
	All data	8.2
	2.53–2.40 Å shell	18.9
	$I/\sigma(I)$	
	All data	4.4
	2.53–2.40 Å shell	3.2
	Completeness† (%)	
	All data	99.6
	2.53–2.40 Å shell	100.0
	Redundancy‡ (all data)	5.3
Structure refinement	No. of crystals	1
	Resolution (Å)	8.0–2.4
	No. of unique reflections	11 204
	No. of protein atoms with occupancy of 1	1697
	After simulated annealing:	
	$R_{\text{cryst}}^{\S}$ (%)	25.5
	$R_{\text{free}}^{\S}$ (%)	29.9
	No. of unique reflections used for $R_{\text{free}}^{\parallel}$	909
	Final R_{cryst} (91.5% of structure factors) (%)	20.5
	Final R_{free} (8.5% of structure factors) (%)	26.7
	Final R_{cryst} (all data) (%)	19.0
	R.m.s. bond deviation (Å)	0.010
	R.m.s. angle deviation (degrees)	1.7
	R.m.s. improper torsion angle deviation (degrees)	1.3

* $R_{\text{merge}} = 100 \sum_{hkl} \sum_n | \langle I \rangle - I_n | / \sum_{hkl} \sum_n I_n$, which is summed over all reflections, where $\langle I \rangle$ is the mean intensity of the reflection hkl and I_n is the intensity of the n th observation of a reflection hkl .

† Completeness = $100 \times (\text{number of independent reflections measured}) / (\text{theoretical maximum number})$.

‡ Redundancy = $(\text{number of measurements}) / (\text{number of independent reflections})$.

§ $R_{\text{cryst}} = 100 \sum |F_o - F_c| / \sum F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes within the set of reflections used for refinement.

|| $R_{\text{free}} = 100 \sum |F_o - F_c| / \sum F_o$ calculated for a randomly selected 8.5% set of structure factors throughout the resolution range and not used in refinement.

tubes cooled to 5 °C and diffraction data were recorded with a precession camera (for preliminary analysis) and an R-AXIS-IIC image plate mounted on an RU200 HB rotating-anode X-ray source (Cu Kα: $\lambda = 1.5418$ Å). Data from a single crystal were used to solve the MFE-23 structure (Table 1). Autoindexing and integration of reflections to 2.4 Å resolution were performed with DENZO [23]. By using the CCP4 suite of programs [24], the initial processing of the data with SCALA and AGROVATA showed that the crystal belonged to a trigonal space group with unit cell dimensions $a = 61.70$ Å, $b = 61.70$ Å, $c = 128.00$ Å, $\alpha = 90^\circ$, $\beta = 90^\circ$, $\gamma = 120^\circ$. Calculation of the Matthews constant [$V_m = \text{unit cell volume} / (n \times \text{molecular mass})$, where n is the number of protein molecules in the unit cell] indicated that six MFE-23 molecules (V_m 2.24 Å³/Da) best satisfied the protein content normally observed for protein crystals [25]. The solvent content of the MFE-23 crystal was estimated as 43% of the unit cell volume.

Crystal space group and structure determination by molecular replacement

Systematic absences in the data set along 001 indicated the presence of a screw axis in the crystal lattice. The data were

initially processed in the lowest trigonal space group $P3$; molecular replacement was used to determine which of the two alternative screw axes, 3_1 or 3_2 , was correct. With the use of the AMoRe program [26], the Fv fragment from a murine IgA Fab molecule J539 solved at 1.95 Å resolution (PDB code 2fbj) was used in cross-rotation function and translation function searches. The J539 search model contained residues equivalent in position to the first 120 residues of the MFE-23 V_H domain and the first 106 residues of the MFE-23 V_L domain. The J539 and MFE-23 sequences were identical at 59 of 120 positions in the V_H domains, and at 84 of 106 positions in the V_L domains (see Figure 1). Cross-rotation function and translation function searches were performed over the resolution range 15.0–3.5 Å. The cross-rotation function was performed on the data processed in space group $P3$ by using an integration radius of 24 Å; it identified two clear solutions. The two peaks for space group $P3$ were consistent with the estimate of six molecules per unit cell. The two best solutions from the rotation search were used to perform translation searches in space groups $P3$, $P3_1$ and $P3_2$ to test alternative trigonal rotation axes. Each produced a single clear solution for translation searches in space group $P3_2$ but failed to produce single clear solutions in space groups $P3$ and $P3_1$; this demonstrated that the crystal system has a 3_2 screw axis. Data reduction in $P3_2$ confirmed that this was the space group, with only one molecule in the asymmetric unit. Data were processed using 11 537 unique reflections in the resolution range 15–2.4 Å with an overall R_{merge} of 8.2% and an average multiplicity of 5.3. The molecular replacement analysis was repeated in $P3_2$; a single solution in the rotation and translation searches confirmed this space group. The rigid-body refinement function of AMoRe was used for data between 8 and 2.8 Å to refine the positional parameters of the molecular replacement solution and generate the initial model with 1724 atoms. The R_{cryst} was 51.3% for data between 8.0 and 2.4 Å with $F/\sigma(F) \geq 2$ (where F is the structure factor and $\sigma(F)$ is its standard deviation).

Crystallographic model building and refinement

The initial model was manually rebuilt with the program O and refined with X-PLOR version 3.1 [27,28]. The crystal packing showed that pairs of MFE-23 molecules were closely associated by the crystallographic two-fold symmetry axes. The 3_2 trigonal screw axis produces infinite helices of scFv pairs with large solvent pores between the helices. In the rigid-body refinement protocol of X-PLOR, only reflections with $F/\sigma(F) \geq 2$ were used. The model was treated as separate V_H and V_L domains. Note that the Kabat database sequence numbering is used throughout this report; however, this is compared with the MFE-23 sequence numbering in Figure 1. Atoms in the putative antigen-binding loop residues H26 to H35, H50 to H58 and H93 to H100c in the heavy chain, and L24 to L34, L50 to L56 and L89 to L97 in the light chain, were set to zero occupancies, and no atoms for the linker or the Myc-tag were included. After rigid-body refinement, $2F_o - F_c$ and $F_o - F_c$ electron density maps contoured at 1σ and $\pm 3\sigma$ respectively were visualized in O. Atoms were set to zero occupancies if they did not correspond to atoms in MFE-23 or did not fit into the electron density, a total of 630 in all. The model contained mainly the conserved residues and the ten β -strands that form the DEBA/A'GFCC'C'' β -sandwich of V-type Ig domains. Two further rounds of rigid-body refinement and manual inspection resulted in an R_{cryst} of 46.4% for 1094 atoms. After the removal of 8.5% unique reflections for the calculation of R_{free} , cycles of positional refinement and model rebuilding in O were used to convert the J539 sequence into that of MFE-23 and to build the antigen-

binding loops as density became apparent, adding a total of 584 atoms. The 1678 atoms in this model gave an R_{cryst} of 26.2%. The 'slowcool' simulated-annealing protocol in X-PLOR was then used to refine the model against reflections between 8 and 2.4 Å by using an initial temperature of 4000 K and a weight of 100000, and a further 18 atoms were added to the structure. This gave good correlation between the rates of reduction in R_{cryst} and R_{free} to values of 26.1% and 30.9% respectively [29]. Individual atomic B -factor refinement decreased these values to 25.5% and 29.9% respectively; 49 water molecules were now included by using the PEAKMAX and WATPEAK programs with peaks ($> 3.5\sigma$) found in the $F_o - F_c$ map at distances of 0.1 to 3.5 Å from the protein surface. A model with 49 water oxygen atoms resulted in an R_{cryst} of 20.5% and an R_{free} of 26.7%. Finally, with 97 water molecules included, positional and B -factor refinements against all reflections between 8 and 2.4 Å resulted in the final model with 1688 atoms and an R_{cryst} of 19.0% calculated from all data within this range. The secondary structure was assigned by using the DSSP program [30]. The model was stereochemically verified with the PROCHECK program [31]. Solvent accessibilities were calculated by using a probe of 1.4 Å in the COMPARE program [32,33]. The electrostatic surface charge was calculated with INSIGHT II and DELPHI software (Biosym/MSI, San Diego, CA, U.S.A.) on INDY Workstations (Silicon Graphics, Reading, Berks., U.K.).

RESULTS AND DISCUSSION

Refined molecular structure of MFE-23

An electron density map for MFE-23 was derived by using X-ray diffraction data at 2.4 Å resolution from a single well-formed crystal of MFE-23, as described in the Materials and methods section. The space group of the crystal was $P3_2$ and the structure of MFE-23 was solved by molecular replacement with a single molecule in the asymmetric unit. The search model was based on the Fv fragment from the murine IgA Fab molecule J539 (Figure 1; see the Materials and methods section). The initial model was refined to a final R_{cryst} of 19.0% (see the Materials and methods section). Good visual agreement was seen with a $2F_o - F_c$ electron density map contoured at 1σ , which is exemplified by views of the H3 antigen-binding surface loop in Figure 2. This was supported by a mean correlation coefficient of 0.9 for the real-space fit of the model to this map. Generally, the V_L domain with a mean B -factor of 21.2 Å² was better defined than the V_H domain with a mean B -factor of 26.9 Å². The MFE-23 structure displayed the typical V-type β -strand topology with two β -sheets DEBA and A'GFCC'C'' in both the V_H and V_L domains. Quantitative analyses with the DSSP secondary structure analysis program showed excellent correlations with the consensus β -sheet structures of other Fv structures (Figure 1).

In accord with general experience of antibody structures, the B -factors were lowest in the regions corresponding to β -strands. The N-terminal and C-terminal residues and several loop residues were generally more flexible and fitted the density less well than the β -strand residues. Within the V_H domain, the weaker density included the side chains of residues H40 to H44 on the loop between β -strands C and C', which form the back-to-back lattice contacts at the MFE-23 dimer interface. The B -factors of the nine residues of H3 were lower than those of the five other loops; this might be correlated with well-defined lattice contacts (see below). The final MFE-23 model yielded a good Ramachandran plot of the main chain torsion angles ϕ and ψ (Figure 3) [34]. The only outlier was Thr-L51 with $\phi = 68^\circ$ and $\psi = -46^\circ$ in a γ -turn, even though this was well defined in the electron density map; this is also true of many other Fab and Fv structures.

																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																</
--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	----

Figure 1 Sequence and secondary structure of the MFE-23 crystal structure

The MFE-23 sequence is also shown with Kabat numbering, together with the structural and hypervariable definitions of the six antigen-binding loops *H1*–*H3* and *L1*–*L3* and the consensus labelling of ten β -strands A to G. The six MFE-23 loops, the (Gly,Ser)₃ linker between the V_H and V_L domains and the C-terminal Myc-tag are underlined. Murine J539 was used for molecular replacement (see the Materials and methods section). MFE-23 resembles the human TR1.9 crystal structure most closely. The 29 residues in MFE-23 to be converted into TR1.9 residues to humanize MFE-23 are denoted by asterisks if they are chemically dissimilar and plus signs if they are similar (see the text). The 67 residues in TR1.9 to be converted into those of MFE-23 for MFE-23 antigen-binding loop grafts are denoted by asterisks (see the text). The secondary structure identified with DSSP are labelled as follows: E, β -strand; B, single residue β -ladders; T, turns; S, bends; G, 3₁₀ helix.

Figure 2 Structure of the H3 loop of MFE-23

(a) A stereo view of the $2F_o - F_c$ electron density map phased by the final MFE-23 model at 2.4 Å and contoured at 1σ is shown with the structure for residues H94–H102. (b) A stereo view of residues H94–H102 in *H3*, with the sequence identified as shown. Hydrogen bonds (broken lines) were assigned if the angle between the acceptor and donor was greater than or equal to 120° and their separation was within 2.5 Å for H–N or H–O bonds or 3.0 Å for N–O or O–O bonds.

Figure 3 Ramachandran plot of the main-chain angles ϕ and ψ for MFE-23

Allowed regions are shown in grey for α -helix (A), β -sheet (B), left-handed helix (L) and polyproline (P) main-chain conformations, with the most favoured conformational region indicated by increased darkness. Gly residues are indicated by triangles; all other residues are shown as squares. Thr-L51 is the only non-glycine residue that occupies a disallowed region of the plot.

In the final model, the atoms at the extremities of the side chains of 12 residues in the V_H domains (H1, H3, H5, H28, H30, H42, H43, H64, H66, H74, H85, H113; 37 atoms) and 3 residues in the V_L domains (L45, L56, L107; 5 atoms) in the MFE-23 map showed disorder and their occupancies were set to zero. In addition, it was not possible to model either the 15 linker residues between residues H113 and L1 or the 15 C-terminal residues after residue L107, which includes the Myc-tag (Figure 1); these were presumed to be flexible and solvent-exposed. Small fragments of density were apparent, perhaps indicative of contacts between the linker, the Myc-tag and the Fv structure, but these could not be modelled convincingly. The final MFE-23 model contained 1696 atoms, which is 98 % of those present in the V_H and V_L domains, and 93 % of the total in the MFE-23 sequence. A total of 97 water molecules were included.

Six antigen-binding loops of MFE-23

In antibodies, the six antigen-binding loops *H1–H3* and *L1–L3* can be defined by structural variability [16] or sequence hypervariability [35], as shown in Figure 1. Sequence hypervariability can also be used to identify SDRs for antigen binding by the correlation of loop residues that have a higher relative variability than others with residues that are known from crystal structures of antibody complexes to be important in antigen binding [20].

H3 has the most important role in antigen binding [36]. In MFE-23, this contains nine structurally defined loop residues (H96–H101) and eleven defined as hypervariable (H95–H102); it is largely hydrophobic. Eight contiguous *H3* residues, H95–

H100b, occupy SDR positions. The crystal structure showed that *H3* had clear density in a $2F_o - F_c$ map contoured at 1σ (Figure 2a). In the *H3* 'torso' (i.e. the two ends of the *H3* loop, proximal to the F and G β -strands), residues H94 and H102 form part of the β -ladder between β -strands F and G. The apex of *H3* is bifurcated; there is a large protrusion to one side of the β -hairpin loop structure at its tip (Figure 2b). An atypical feature of this *H3* loop is that the position of the β -carbon atom of Gly-H95 immediately before the loop deviated by 3.2 ± 0.4 Å from those found in nine other representative Fab crystal structures (PDB codes 1mlb, 1vfa, 2fbj, 1vge, 7fab, 1igm, 8fab, 2fb4 and 1dfb). Within the *H3* loop, Thr-H96 is buried and is centrally important for determining its conformation. Its side chain is sandwiched between the aromatic side chains of Tyr-H100a, Tyr-H100b and Phe-H100c (Figure 2b). Its OH group forms a hydrogen bond with the main-chain NH of Phe-H100c and its main-chain NH group forms one with the CO of Phe-H100c (Figure 2a). As the result of this topology, the side chains of Thr-H98, Pro-H100 and Tyr-H100b have high accessibilities to solvent.

The other loops, *H1*, *H2* and *L1–L3*, in MFE-23 (Figure 1) correspond to known canonical structures that are predictable from their sequences [16]. The structural *H1* loop (seven residues) corresponds to the single known canonical class of structures for this loop. The structurally important Arg residue normally found at H94 is replaced by the buried residue Glu-H94, whose carboxy group forms a hydrogen bond with the carbonyl oxygen of Ile-H29. The structural *H2* loop (six residues) belongs to canonical class 2 (type A). Its conformation includes a notable display of four acidic residues. The structural *L1* loop (six residues) belongs

to canonical class 1, as expected for κ V_L domains. The structural $L2$ loop (three residues) is a turn for which only one canonical class is known, in which Thr-L51 was the only outlier on a Ramachandran plot (Figure 3). The structural $L3$ loop (six residues) belongs to canonical class 1.

In confirmation of the predictability of these five canonical structures, the crystal structure was compared with two homology models of MFE-23 that were generated beforehand with an early version of the program AbM [21]. The two models were constructed by using a combination of canonical structures and database searches. These showed that the five loops were well predicted and gave low root-mean-square (r.m.s.) deviations of main-chain conformations, in agreement with experience [37].

Structural classification of the $H3$ loop in MFE-23

Given the importance of the $H3$ loop for antibody function and in view of recent developments that show that most $H3$ loops can be structurally classified and predicted [17–19], the uniqueness of the $H3$ structure in MFE-23 was investigated.

$H3$ torso

Even though 49 $H3$ loops have been classified [19], the start and end of the $H3$ loop in MFE-23 differs from many $H3$ structures. That involving the position of Gly-H95 has been noted above. In addition, a salt bridge commonly occurs between Arg(Lys)-H94 and Asp-H101; this leads to a C-terminal β -bulge in 41 of 49 $H3$ loops in what would otherwise be a β -hairpin [19]. In MFE-23, the conserved Arg-H94 residue is replaced by Glu-H94 and the C-terminal end of the $H3$ loop is accordingly predicted not to be bulged. The MFE-23 crystal structure confirmed the absence of the β -bulge (see Figure 2 of [17]). To compensate for the absence of Arg-H94, Asp-H101 is now buried to form hydrogen bonds with the buried side chains of Asn-H93, Tyr-H100a and Trp-H103. The hydrogen bond with Tyr-H100a is responsible for the burial of this side chain, whereas that with Trp-H103 is in agreement with the classification in [17].

$H3$ apex

Even though the apices of 62 $H3$ loops have been classified [19], that of MFE-23 is different. First, in all non-bulged $H3$ structures, an aromatic residue from $H1$ is packed with an aromatic residue at H96 or H99 in $H3$. This does not occur in MFE-23 because no aromatic residues occur in these $H3$ positions. Secondly, although bifurcated $H3$ apices have previously been identified in an irregular-type $H3$ structure Yst9.1 and for a β -bulged $H3$ structural class K+/K 18.5 [18,19], MFE-23 represents a new example of a bifurcated loop that occurs in a non-bulged loop. In both Yst9.1 (sequence C⁹²TRDPYGPAA¹⁰⁴WG¹⁰⁴) and the $H3$ canonical class K+/K 18.5, the bifurcated $H3$ structure is formed by a buried Asp residue that forms internal hydrogen bonds with other main-chain positions to divide $H3$ into two parts. In MFE-23, the buried Thr-H96 residue and a PXGP motif in the $H3$ loop have the same role.

Humanization strategies for murine MFE-23

The creation of a less immunogenic form of MFE-23 would improve its lifetime in circulation; the MFE-23 crystal structure showed how its humanization might be achieved. It is essential that the antibody framework does not change during humanization [38]. The crystallographic refinement of MFE-23 revealed that the V_H framework belonged to class 2 (Figure 1) [39]. The superimposition of the MFE-23 V_H framework on three

Figure 4 Stereo view of the Fv dimers in the MFE-23 crystal structure

Two Fv structures related by a two-fold axis perpendicular to the plane of the paper are shown. On the left, the positions of the six loops $H1$ – $H3$ and $L1$ – $L3$ and the N- and C-termini are shown. Note that the six loops are fully exposed on the surface of the dimer.

murine V_H crystal structures (PDB codes 1mlb, 1vfa and 2fbj) and all six known human V_H crystal structures (PDB codes 1vge, 7fab, 1igm, 8fab, 2fb4 and 1dfb) showed that the murine V_H domains gave lower r.m.s. deviations than the corresponding human ones, with the class 2 V_H domains giving the lowest r.m.s. deviation of 0.45–0.47 Å. The κ chain of the MFE-23 V_L structure was similarly superimposed on other murine and human V_L domains of both κ and λ types in these nine crystal structures. The κ domains gave lower r.m.s. deviations of between 0.36 and 0.56 Å in comparison with the λ domains with r.m.s. deviations of between 0.78 and 0.82 Å. It was concluded that the V_H and V_L domains from human Fab TR1.9 (PDB code 1vge) offered the best framework model for humanization, as both domains are in the same classes as those of MFE-23. Humanization can be achieved either by grafting the six MFE-23 loops on the human TR1.9 framework or by converting the MFE-23 framework into the human TR1.9 equivalent [36,39]. In the first approach, solvent accessibility calculations suggested that 57 MFE-23 loop and 10 MFE-23 framework residues were required in TR1.9, which is a net change of 44 residues in TR1.9 (see Figure 1). In the second approach, 116 residues in MFE-23 were determined to be over 30% solvent-accessible, of which it would only be necessary to alter 19 that were chemically different between MFE-23 and TR1.9, and 10 more that were chemically similar (see Figure 1).

Intermolecular back-to-back contacts in MFE-23

MFE-23 molecules form two sets of lattice contacts in the crystal structure. One of these corresponds to back-to-back contacts between two Fv molecules to form symmetrical dimers related by crystallographic two-fold axes (Figure 4). Such a mode of interaction has not been seen in any of the earlier scFv crystal structures.

The back-to-back contacts and the lack of electron density for the 15-residue linker mean that it is not possible to distinguish between a two-domain monomer or a four-domain dimer of MFE-23 in the crystal structure (Figure 4). If fully extended, the linker would be 53 Å in length. This is compatible with the 30 Å separation between the α -carbon atoms at the V_H C-terminus and the V_L N-terminus if a MFE-23 monomer is formed between the two GFC β -sheet faces of these domains [40]. It is also

Figure 5 Antigen-binding site of MFE-23 viewed face-on

(a) Parts of the six loops of MFE-23 (*H1*, residues H26–H35; *H2*, H50–H55; *H3*, H95–H101; *L1*, L24–L34; *L2*, L50–L57; *L3*, L89–L97). The 23 SDRs are coloured magenta (H33, H35, H50, H52, H53, H56, H58, H95, H96, H97, H98, H99, H100, H100a, H100b, L33, L50, L53, L55, L89, L91, L92 and L94) and the remaining loop residues are coloured blue. (b) The site is shown as a Connolly surface that is accessible to a probe of 1.4 Å radius. The loops are coloured magenta and blue as above, and the remainder of the MFE-23 structure is coloured yellow. (c) Electrostatic view of the site. Red represents a potential of less than $-4kT$ (acidic), blue a potential of more than $+4kT$ (basic) and white as $0kT$ (neutral). Linear interpolation of the colours represents potentials between $-4kT$ and $+4kT$.

compatible with the 21 Å separation required to form a domain-rearranged dimer scFv structure (sometimes referred to as a diabody), in which the V_H and V_L domains within a single molecule are separated from each other and then are reassociated through their GFC faces with a second scFv molecule to form a dimer (right or left pairs in Figure 4). The total side-chain surface area lost by the back-to-back interaction between the V_H and V_L domains is 170 Å² per molecule, which is comparable to that of 230 Å² per molecule lost between the two GFC β -sheets in a V_H – V_L domain interaction. The total accessible side chain area is 2560 Å² per molecule.

The back-to-back contacts involve eight residues. At the centre, Pro-H41 in the two adjacent V_H domains form hydrophobic contacts with each other, and Pro-L40 and Gly-L41 in the two adjacent V_L domains do so similarly. In fact, the Pro-H41

contacts correspond to the highest B -factors that were found in the H40–H44 surface loop. At the surface, other reciprocal contacts include hydrogen bonds between Thr-H108 and Glu-L105 on the β -strands G of a pair of opposing V_H and V_L domains, and hydrophobic contacts between Ala-H9, Leu-H11, Ala-L9 and Ile-L10 on the β -strand A' of a pair of opposing V_H and V_L domains.

Appearance of the CEA-binding site of MFE-23

The general appearance of the six antigen-binding loops in MFE-23 when viewed from the side is that of a convex surface whose apex is formed by the slight protrusion of the *H3* loop. When the loops are viewed face-on (Figure 5), the main chains of the six loops form a large cavity into which their side chains protrude. In antibody–antigen complex structures, residues that make contact with antigen generally lie on the inwards-facing ridges surrounding this cavity [20]. The 64 structural and hypervariable residues are shown in blue and magenta in Figures 5(a) and 5(b) and have a total side-chain surface area of 2340 Å², of which 750 Å² (32%) is solvent-exposed. Notably the electrostatic surface (Figure 5c) showed a large band of negative charge on one side that corresponded to Asp-H31 (70% exposed), Asp-H52 (16% exposed), Glu-H53 (64% exposed), Asp-H56 (65% exposed) and Glu-H58 (36% exposed) on the *H1* and *H2* loops.

Model for antibody–antigen interactions between MFE-23 and CEA

There is a second set of intermolecular contacts almost as extensive as the back-to-back interaction described above. This involves the three antigen-binding loops *H1*, *H2* and *H3*, where the part that involves the *H3* loop is visible in Figure 2(a). In Figure 6, molecule 1 interacts with molecules 2 and 3. Thus the *H1*, *H2* and *H3* loops in molecule 1 interact with the DEBA β -sheet of the V_L domain in molecule 3. This corresponds to a decrease of 150 Å² per molecule in the accessible surface area of 2560 Å² per molecule. A smaller lattice contact occurs between the *L1* and *L2* loops and the A' β -strand of the adjacent V_H domain from molecule 2 (a decrease of 80 Å² per molecule), whereas *L3* made no contact. The decrease in surface area involving the *H1*, *H2* and *H3* loops of MFE-23 is approximately half of that observed to be lost in crystal structures of antibody–antigen complexes, where six values between 316 and 456 Å² per molecule have been reported [41], and it is also known that the *H2* and *H3* loops (as well as the *L3* loops) mediate the largest contacts in antibody–antigen interactions [42]. Furthermore, a remarkable degree of surface and electrostatic complementarity was apparent within the contact area.

To examine the surface complementarity of the *H1*, *H2* and *H3* loops, Connolly surfaces were created for the MFE-23 molecules in the lattice, and INSIGHT II was used to inspect a section 3 Å thick as it was passed through the contact region (exemplified in Figure 6b). *H3* was inserted at an indentation formed between molecules 2 and 3, and *H2* was positioned on the other side of a protrusion from the V_L domain of molecule 3 that included Lys-L18. To test the electrostatic complementarity, 14 contact residues between the loop residues and the adjacent molecules were identified by using X-PLOR (Table 2). This showed that Asp-H52 and Glu-H53 on the *H2* loop of molecule 1 made salt bridge contacts with Lys-L18 and Arg-L77 on molecule 3. The complementary regions of oppositely charged surfaces can be seen by comparing the electrostatic maps of Figures 5(c) and 6(a).

Figure 6 Interaction between the antigen-binding loops of MFE-23 molecule 1 with two other MFE-23 molecules, 3 and 2, in the crystal lattice packing of MFE-23

(a) The MFE-23 structure (ribbon view) of molecule 1 is seen adjacent to the electrostatic surfaces of two adjacent MFE-23 molecules 2 and 3. The V_H and V_L domains are coloured green and cyan respectively, and the H1–H3 and L1–L3 loops are shown in magenta and yellow respectively. Asp-H52 and Glu-H53 (arrowed) are shown as spheres. These form contacts with Lys-L18 and Arg-L77 in the V_L domain (arrowed). (b) Two 3 Å-thick slices of the three MFE-23 molecules are shown in the same plane as that in (a). The upper and lower slices are separated by 5 Å. The surface of the H2 and H3 loops of molecule 1 is well matched to that of the V_L domain of molecule 3.

The list of contact residues (Table 2) showed that extensive hydrogen-bond and hydrophobic contacts were made between five residues in the H3 loop and the V_L domain and that six more residues in the H1 and H2 loops made direct contacts with the V_L domain. These contact residues can be compared with a prediction of the MFE-23 residues important for binding to CEA. This prediction involved the calculation of side-chain accessibilities to solvent, the frequency of contacts made with antigen in known antibody–antigen complex structures and their occurrence at SDR positions for the 64 residues in the six loops. The summary in Figure 7 shows that 10 residues (Tyr-H33, Glu-H53, Asp-H56, Glu-H58, Thr-H98, Pro-H100, Tyr-H100b, Arg-L91, Ser-L92 and Tyr-L94) all presented solvent-exposed areas of more than 10 Å², had at least 30 % probability of making contact with CEA and occurred at SDR positions. Tyr residues have

Table 2 Contact residues between the structural antigen-binding loops of MFE-23 and the adjacent MFE-23 molecule in the crystal lattice

In both columns, unless this is unambiguous, residue side chain atoms are specified in parentheses according to the standard Protein Data Bank nomenclature. In the last column, the distance between the atoms specified in the two columns is specified in parentheses, together with the assignment of the type of interaction present.

Loop	MFE-23 antigen-binding loop residues	Contact residues on adjacent molecule
H1	Lys-H30 main-chain C = O Asp-H31 CO ₂ [−] (OD) Ser-H32 main-chain C = O Tyr-H33 OH*	Lys-L18 (NZ) (2.8 Å; hydrogen bond) Thr-L74 (OG) (4.0 Å; hydrogen bond) Lys-L18 (NZ) (2.8 Å; hydrogen bond) Arg-L77 (NH2) (3.6 Å; hydrogen bond†)
H2	Asp-H52 CO ₂ [−] (OD1, OD2) Asp-H52 CO ₂ [−] (OD2) Glu-H53 CO ₂ [−] (OE1, OE2)	Lys-L18 (NZ) (2.5 Å, 3.8 Å; salt bridge) Arg-L77 (NH1, NH2) (3.6 Å, 3.5 Å; salt bridge) Arg-L77 (NH1) (2.9 Å; 3.6 Å; salt bridge)
H3	Pro-H97 (CB) Thr-H98 (CG) Thr-H98 OH Pro-H100 (CD) Pro-H100 (CG) Tyr-H100b (CD, CE) Tyr-H100b (CE, CZ) Tyr-H100b (CE, CZ) Tyr-H100b OH Phe-H100c (CE)	Lys-L18 (CB) (4.0 Å; hydrophobic) Ala-L13 (CB) (3.9 Å; hydrophobic) Lys-L18 main-chain C = O (2.6 Å; hydrogen bond) Met-L11 (CB) (3.7 Å; hydrophobic) Leu-H11 (CD1) (4.1 Å; hydrophobic) Ser-L7 (CB) (3.6 Å; 3.6 Å; hydrophobic) Pro-L8 (CD) (3.6 Å; 3.9 Å; hydrophobic) Thr-L22 (CG) (3.6 Å; 3.9 Å; hydrophobic) Thr-L20 main-chain C = O (2.7 Å; hydrogen bond) Thr-L20 (CG) (3.9 Å; hydrophobic)
L1	Tyr-L32 OH	Glu-H10 CO ₂ [−] (OE1) (3.4 Å; hydrogen bond†)
L2	Ser-L52 OH Asn-L53 (ND2)*	Arg-H13 main-chain NH (3.4 Å; hydrogen bond†) Leu-H11 main-chain C = O (3.1 Å; hydrogen bond)

* Adjacent to structural loop. All distances were derived by using the geomanal.inp script file in X-PLOR and confirmed with stereoscopic glasses.

† Three hydrogen-bond distances are greater than 3.3 Å but are included because these are at the limit of positional error in the co-ordinates.

been found to be commonly involved in antigen binding [43,44]. All these seven V_H and three V_L predicted contact residues are clustered in the centre of the binding site (magenta in Figures 5a and 5b). Five of the seven V_H predicted contact residues (Tyr-H33, Glu-H53, Thr-H98, Pro-H100 and Tyr-H100b) correspond to the observed lattice contacts in Table 2, supporting the hypothesis that the lattice contacts for the H1, H2 and H3 loops constitute a good model for the binding of MFE-23 to CEA.

Experimental mutagenesis studies also support this hypothesis. The interaction between MFE-23 and CEA has been examined by using four V_H-mutated forms of MFE-23 [21]. Given the extensive contacts made by Tyr-H100b (Table 2), it is highly significant that a Tyr-H100b → Pro mutation abolished MFE-23 binding to CEA. Note that, even though this mutation incorporated a Pro residue into the H3 loop, the main-chain structure of the H3 loop should be unaltered because the ϕ and ψ angles of Tyr-H100b are similar to those generally seen for Pro residues in proteins. Likewise a Glu-H53 → Lys mutation decreased MFE-23 binding to CEA, which is consistent with the electrostatic lattice contact involved with the H2 loop (Table 2). A Thr-H98 → Ala mutation caused a minor improvement in binding to CEA; this is consistent with the role of Thr-H98 as a contact residue. A Tyr-H100a → Ala mutation had almost no effect on binding to CEA; this agrees with the absence of this residue from the lattice contacts of Table 2.

Conclusions

MFE-23 is a therapeutically valuable molecule [5]. The structure reported here has demonstrated several interesting and novel

Figure 7 Summary of residue properties in the six loops of MFE-23

The total side-chain surface accessibility ($\text{\AA}^2$) determined with a probe of $1.4 \text{ \AA}$ is shown for each residue (continuous line). The frequency of each loop residue making contact with antigen in antibody–antigen complexes (bars) was taken from Table 1 in [20]. The 23 SDR positions highlighted in (a) and (b) are denoted by their single-letter amino acid codes. The actual contact residues observed in the intermolecular contacts (Table 2) are denoted by asterisks.

aspects. First, whereas five of the six antigen-binding loops possessed expected structures, the *H3* loop (Figure 2) adopted a unique structure that could not be predicted. Secondly, the murine MFE-23 structure showed a close framework similarity with the human TR1.9 structure, which led to the proposal of humanization strategies. Thirdly, despite the theoretical possibility that a back-to-back dimer of two V_H – V_L pairs can be readily formed in any scFv structure, the MFE-23 structure is the first in which this has been observed experimentally.

Finally, the MFE-23 crystal structure suggests how MFE-23 might use the *H1*–*H3* loops to bind with its target molecule CEA. Antibodies possess complementarity in shape and electrostatic surface with their antigen [45]; both criteria are well satisfied by the *H1*, *H2* and *H3* loop lattice contacts, which are unusually extensive. Five of the seven predicted antigen contact residues in *H1*, *H2* and *H3* participate in these lattice contacts, and the effects of four experimentally tested MFE-23 V_H mutations are consistent with this model, including two contact residues predicted to be the most important. The crystal structure of MFE-23 might thus fortuitously reveal information on how it recognizes its antigen.

We thank Professor K. D. Bagshaw for his interest and for useful discussions. The work described here formed part of a Ph.D. thesis by M.K.B. for the University of London. M.K.B. acknowledges a Clement Wheeler-Bennett Trust studentship, and A.L.C. a Biotechnology and Biological Sciences Research Council studentship. This work was supported by the Cancer Research Campaign (J.D.T., P.A.K.) and the Ronald Raven Chair in Oncology Trust (K.A.C., R.H.J.B.).

REFERENCES

- Waldmann, T. A. (1991) *Science* **252**, 1657–1662
- Gold, P. and Freedman, S. O. (1965) *J. Exp. Med.* **121**, 439–462
- Boehm, M. K., Mayans, M. O., Thornton, J. D., Begent, R. H. J., Keep, P. A. and Perkins, S. J. (1996) *J. Mol. Biol.* **259**, 718–736
- Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shiota, K. and Stanners, C. P. (1989) *Cell* **57**, 327–334
- Begent, R. H. J., Verhaar, M. J., Chester, K. A., Casey, J. L., Green, A. J., Napier, M. P., Hope-Stone, L. D., Cushen, N., Keep, P. A., Johnson, C. J. et al. (1996) *Nature Med.* **2**, 979–984
- Schlom, J., Eggensperger, D., Colcher, D., Molinolo, A., Houchens, D., Miller, L. S., Hinkle, G. and Siler, K. (1992) *Cancer Res.* **52**, 1067–1072
- Verhaar, M. J., Chester, K. A., Keep, P. A., Robson, L., Pedley, R. B., Boden, J. A., Hawkins, R. E. and Begent, R. H. J. (1995) *Int. J. Cancer* **61**, 497–501
- Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M.-S., Novotný, J., Margolies, M. N., Ridge, R. J., Brucoleri, R. E., Haber, E., Crea, R. and Oppermann, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 5879–5883
- Yokota, T., Milenic, D. E., Whitlow, M. and Schlom, J. (1992) *Cancer Res.* **52**, 3402–3408
- Chester, K. A., Begent, R. H. J., Robson, L., Keep, P. A., Pedley, R. B., Boden, J. A., Boxer, G., Green, A., Winter, G., Cochet, O. and Hawkins, R. E. (1994) *Lancet* **343**, 455–456
- Zdanov, A., Li, Y., Bundle, D. R., Deng, S.-J., MacKenzie, C. R., Narang, S. A., Young, N. M. and Cygler, M. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6423–6427
- Perisic, O., Webb, P. A., Holliger, P., Winter, G. and Williams, R. L. (1994) *Structure* **2**, 1217–1226
- Kort, A. A., Malby, R. L., Caldwell, J. B., Gruen, L. C., Ivancic, N., Lawrence, M. C., Howlett, G. J., Webster, R. G., Hudson, R. G. and Colman, P. M. (1994) *Eur. J. Biochem.* **221**, 151–157
- Raag, R. and Whitlow, M. (1995) *FASEB J.* **9**, 73–80

- 15 Hoedemaeker, P. J., Signorelli, T., Johns, K., Kuntz, D. A. and Rose, D. R. (1997) *J. Biol. Chem.* **272**, 29784–29789
- 16 Al-Lazikani, B., Lesk, A. M. and Chothia, C. (1997) *J. Mol. Biol.* **273**, 927–948
- 17 Shirai, H., Kidera, A. and Nakamura, H. (1996) *FEBS Lett.* **399**, 1–8
- 18 Oliva, B., Bates, P. A., Querol, E., Aviles, F. X. and Sternberg, M. J. E. (1998) *J. Mol. Biol.* **279**, 1193–1210
- 19 Morea, V., Tramontano, A., Rustici, M., Chothia, C. and Lesk, A. M. (1998) *J. Mol. Biol.* **275**, 269–294
- 20 Padlan, E. A., Abergel, C. and Tipper, J. P. (1995) *FASEB J.* **9**, 133–139
- 21 Read, D. A., Chester, K. A., Keep, P. A., Begent, R. H. J., Pedersen, J. T. and Rees, A. R. (1995) *Br. J. Cancer* **71**, Suppl. XIV, 57 (abstr. P132)
- 22 Perkins, S. J. (1986) *Eur. J. Biochem.* **157**, 169–180
- 23 Otwinowski, Z. and Minor, W. (1997) *Methods Enzymol.* **276**, 307–326
- 24 Collaborative Computational Project Number 4 (1994) *Acta Cryst. D* **50**, 760–763
- 25 Matthews, B. W. (1968) *J. Mol. Biol.* **33**, 491–497
- 26 Navaza, J. (1994) *Acta Cryst. A* **50**, 157–163
- 27 Jones, T. A., Zou, J. Y., Cowan, S. W. and Kjeldgaard, M. (1991) *Acta Cryst. A* **47**, 110–119
- 28 Brünger, A. T. (1992) X-PLOR, Version 3.1. Yale University, New Haven, CT, U.S.A
- 29 Kleywegt, G. J. and Jones, T. A. (1997) *Methods Enzymol.* **277**, 208–230
- 30 Kabsch, W. and Sander, C. (1983) *Biopolymers* **22**, 2577–2637
- 31 Laskowski, R. A., MacArthur, M. W., Moss, D. S. and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- 32 Lee, B. and Richards, F. M. (1971) *J. Mol. Biol.* **55**, 379–400
- 33 Šali, A. and Blundell, T. L. (1990) *J. Mol. Biol.* **212**, 403–428
- 34 Ramachandran, G. N. and Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437
- 35 Wu, T. T. and Kabat, E. A. (1970) *J. Exp. Med.* **132**, 211–249
- 36 Padlan, E. A. (1994) *Mol. Immunol.* **31**, 169–217
- 37 Rees, A. R., Searle, S. J., Henry, A. H., Whitelegg, N. and Pedersen, J. (1996) in *Protein Structure Prediction* (Sternberg, M. J. E., ed.), pp. 141–172, Oxford University Press, Oxford
- 38 Bajorath, J., Harris, L. and Novotny, J. (1995) *J. Biol. Chem.* **270**, 22081–22084
- 39 Saul, F. A. and Poljak, R. J. (1993) *J. Mol. Biol.* **230**, 15–20
- 40 Chothia, C., Novotny, J., Brucoleri, R. and Karplus, M. (1985) *J. Mol. Biol.* **186**, 651–663
- 41 Braden, B. C. and Poljak, R. J. (1995) *FASEB J.* **9**, 9–16
- 42 Wilson, I. A. and Stanfield, R. L. (1994) *Curr. Opin. Struct. Biol.* **4**, 857–867
- 43 Padlan, E. A. (1990) *Proteins* **7**, 112–124
- 44 Mian, I. S., Bradwell, A. R. and Olson, A. J. (1991) *J. Mol. Biol.* **217**, 133–151
- 45 Braden, B. C., Goldman, E. R., Mariuzza, R. A. and Poljak, R. J. (1998) *Immunol. Rev.* **163**, 45–57

Received 10 September 1999/12 November 1999; accepted 6 December 1999