ELSEVIER

Biophysical Chemistry 68 (1997) 9–16

Biophysical Chemistry

Antibody structure, prediction and redesign

V. Morea [a,b], A. Tramontano [a,*], M. Rustici [c], C. Chothia [d], A.M. Lesk [e]

[a] *IRBM P. Angeletti, Via Pontina Km 30.600, 00045 Pomezia, Italy*
[b] *Dipartimento di Studi Farmaceutici, Universita' La Sapienza, 00100 Rome, Italy*
[c] *Dipartimento di Chimica, Universita' degli Studi di Sassari, Via Vienna, 2, 07100 Sassari, Italy*
[d] *MRC Laboratory of Molecular Biology, MRC Centre, Hills Road, Cambridge CB2 2QH, UK*
[e] *Department of Haematology, University of Cambridge, Clinical School, MRC Centre, Hills Road, Cambridge CB2 2QH, UK*

Received 15 November 1996; accepted 5 December 1996

Abstract

So far the difficulty to predict the structure of the third hypervariable loop of the heavy chain of antibodies has represented the main limitation in modelling the complete antigen binding site. We carefully analysed all available structures of immunoglobulins searching for rules relating the loop conformation to its amino acid sequence. Here, we analyse the conformation of this loop and show that we are able to predict the conformation of the ten residues proximal to the framework. The conformation of the remaining residues of loops longer than 10 residues can also be predicted in many cases. This, combined with the previously defined canonical structures for the other five hypervariable loops, is an important step toward the prediction of the complete immunoglobulin antigen-binding site. We exemplify our prediction protocol using three known immunoglobulin structures as test cases. © 1997 Elsevier Science B.V.

Keywords: Immunoglobulin structure; Hypervariable loops; Canonical structures; Structure prediction; Protein design; Hair-pin loops

1. Introduction

Immunoglobulins are multidomain proteins whose basic domain is formed by two β-sheets packed face to face, pinned together by a disulphide bridge. Two of the domains, the variable domains of a light and a heavy chain (VL and VH), are packed together to create a scaffolding of relatively conserved structure, on which the antigen-binding site is formed by six loops, named complementarity-determining regions or CDRs. The three CDRs from the light chain are called L1, L2 and L3 in order of appearance in the sequence, and the three CDRs from the heavy chain are called correspondingly H1, H2 and H3.

It has been previously shown that, in spite of their high sequence variability, five of these six loops, the three of Vk chains, and the first two of VH chains, can assume just a small number of main-chain conformations, called canonical structures and that these are determined by the loop length and by the presence of a few key residues in the loop and/or in the framework [1–3]. The main chain conformation of these loops can therefore be predicted from the antibody sequence, leading to successful predictions of antibody structures in 'blind' tests [2].

We took advantage of our understanding of the conformation of immunoglobulin antigen binding

* Corresponding author.

PII S0301-4622(96)02266-1

loops to design a molecule de novo, called Minibody. We used a portion of the heavy chain variable domain of an immunoglobulin as a template, obtaining a molecule with a β-sheet scaffold of novel fold and two regions corresponding to the hypervariable loops H1 and H2. The design was successful and the molecule was used as a scaffold for displaying random sequences from which ligands with specific activities could be selected [4–6].

The third CDR of VH chains (H3), is much more variable in length, sequence and structure than the other antigen-binding loops. The lack of a method to predict the structure of H3, which is located at the centre of the antigen binding site and therefore plays a central role in the antigen recognition process (Fig. 1), has impaired the prediction of the complete immunoglobulin binding site.

Here, we analyse the conformation of the H3 loops and demonstrate that recurring structures also exist for this loop. We also show that we are able to predict the conformation of the ten residues proximal to the framework. The conformation of the remaining residues of loops longer than 10 residues can also be predicted in many cases. We illustrate our prediction protocol using three examples.

2. Methods

Structural analysis was performed using Insight [7] and Pinq [8]. Energy minimisation was carried out using Discover (MSI) with default parameters.

The GCG package [9] was used for all sequence analyses.

Kabat numbering [10] is used throughout the manuscript.

3. Results and discussion

3.1. The torso and the head of the H3 loops

Residues Cys 92 and Gly 104, strictly conserved in immunoglobulin sequences, were used here to define the boundaries of the H3 loop.

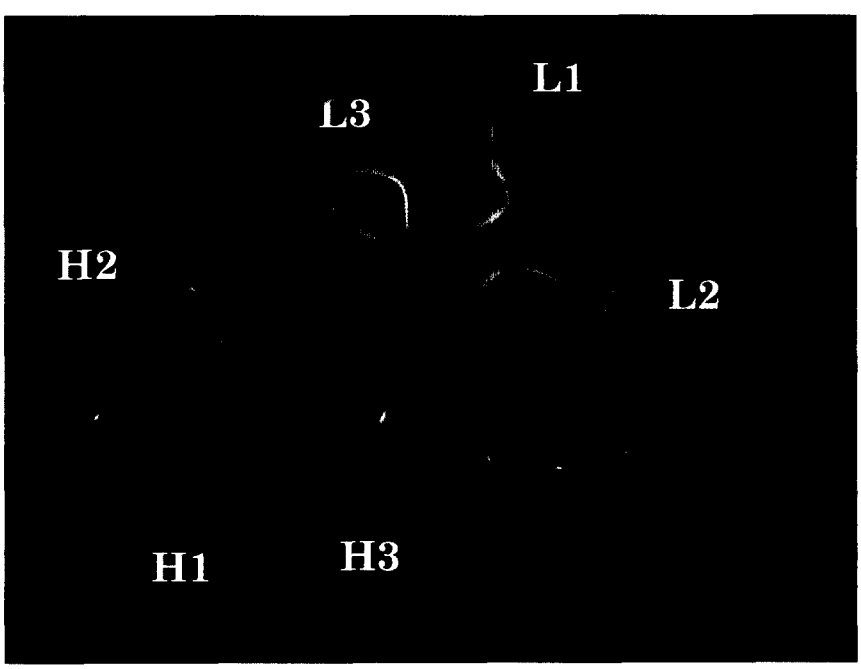

Fig. 1. Ribbon representation of the D1.3 immunoglobulin variable domain showing the position of the H3 loop in the middle of the antigen binding site. The light and heavy chain frameworks are shown in green and cyan respectively. The light and heavy chain antigen binding loops are in yellow and orange respectively.

We selected from Rel. 72 of the Brookhaven Data Bank [11] a set of well refined immunoglobulin structures (Table 1), with non-redundant H3 loop conformations: when different structures of the same antibody are available, with the same conformation of the H3 loop (rms deviation of the backbone atoms lower than 1.0 Å), we only kept the structure with the better resolution. We found that the H3 region could be separated into a torso region, defined as the first 4 and the last 6 residues of the loop, and the apex of the loop, which we call the head. Indeed, two major classes of H3 torsos can easily be identified, one showing a regular pattern of β-sheet hydrogen bonds, the other characterised by the presence of a β-bulge at residue 101 (Fig. 2).

We visually inspected the interactions within each

Table 1
List of analysed immunoglobulin structures

Ig	PDB code[a]	Resolution (Å)	Length of H3	Sequence
Cha255	1ind	2.2	10	CASHR----------FVHWG
NQ10/12.5	NQ10	2.8	10	CARDA----------GAYWG
50.1	1ggi_m1	2.8	10	CVQEG----------YIYWG
50.1	1ggb	2.8	10	CVQEG----------YIYWG
4-4-20	1flr	1.85	12	CTGSY----------YGMDYWG
JEL103	1mrd	2.3	12	CANLR----------GYFDYWG
HyHEL-5	1bql	2.6	12	CLHGN----------YDFDGWG
TE33	1tet	2.3	12	CARRS----------WYFDVWG
NC6.8	2cgr	2.2	12	CTRGY----------SSMDYWG
D1.3	1vfa	1.8	13	CARER---------DYRLDYWG
Yst9.1	1mam	2.45	13	CTRDP---------YGPAAYWG
8F5	1bbd	2.8	14	CDGYYS--------YYDMDYWG
D11.15	1jhl	2.4	14	CTRDDN--------YGAMDYWG
J539	2fbj	1.95	14	CARLHY--------YGYNAYWG
NEW	7fab	2.0	14	CARNLI--------AGGIDVWG
SE155-4	1mfa	1.7	14	CTRGGH--------GYYGDYWG
JE142	1jel	2.8	14	CARVMG--------EQYFDVWG
R6.5	1rmf	2.8	15	CARGGWL-------LLSFDYWG
17-Ia	1for	2.75	15	CARSGNY-------PYAMDYWG
26-10	1igj_m1	2.5	15	CAGSSGN-------KWAMDYWG
BV04-01	1nbv	2.0	15	CVRDQTG-------TAWFAYWG
BV04-01	1cbv	2.66	15	CVRDQTG-------TAWFAYWG
DB3	1dbb	2.7	15	CTRGDYV-------NWYFDVWG
B13I2	1igf_m1	2.8	15	CTRYSSD-------PFYFDYWG
4D5	1fvc_m1	2.2	16	CSRWGGD------GFYAMDYWG
4D5	1fvd_m1	2.5	16	CSRWGGD------GFYAMDVWG
4D5	1fvd_m2	2.5	16	CSRWGGD------GFYAMDVWG
McPC603	1mcp	2.7	16	CARNYYG------STWYFDVWG
NC41	1nca	2.5	16	CARGEDN------FGSLSDYWG
26/9	1frg	2.8	16	CARRERY------DEKGFAYWG
17/9	1ifh	2.8	16	CARRERY------DENGFAYWG
17/9	1hil_m1	2.0	16	CARRERY------DENGFAYWG
POT	1igm	2.3	17	CAKHRVS-----YVLTGFDSWG
HIL	8fab_m1	1.8	17	CARDPDI-----LTAFSFDYWG
HIL	8fab_m2	1.8	17	CARDPDI-----LTAFSFDYWG
36-71	6fab	1.9	17	CARSEYY-----GGSYKFDYWG
HC19	1gig	2.3	19	CARDFYDYD---VFYYAMDYWG
R19.9	1fai	2.7	20	CARSFYGGS--DLAVYYFDSWG
KOL	2fb4	1.9	22	CARDGGHGFCSSASCFGPDYWG
3D6	1dfb	2.7	22	CVKGRDYYDSGGYFTVAFDIWG
R454511	1ikf	2.5	22	CTRHTLYDTLYGNYPVWFADWG

[a] The suffixes m1 and m2 indicate different antibody molecules in the same unit cell.

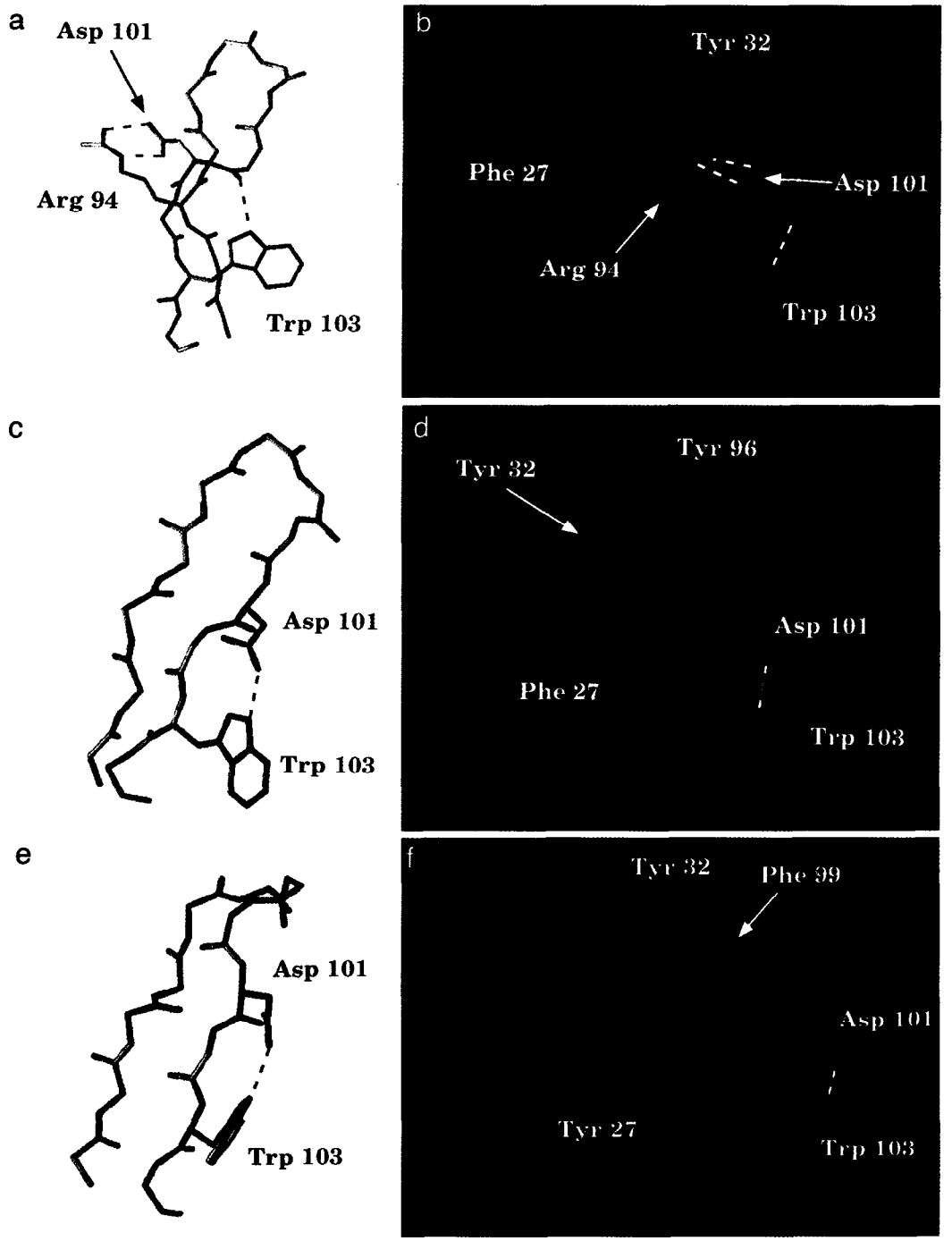

Fig. 2. Canonical structures of the H3 loop. Typical bulged (a, b) and non-bulged (c–f) conformations. (b, d, f) Packing of the H3 loops with the conserved aromatic residues of H1 (cyan). When H3 has a bulged conformation (b), the H1 loop packs with the hydrophobic portion of an Arg or Lys in position 94 of H3. When H3 has a non-bulged conformation, the H1 loop packs with an aromatic residue of H3 either in position 96 (d) or in position 99 (f). Antibodies shown: D1.3 (a, b), 4-4-20 (c, d), JEL103 (e, f).

of these two classes and noticed that the bulged structure is always present when a positively charged residue is present in position 94. This conformation is stabilised by the packing of the hydrophobic portion of the basic residue with the H1 loop and by a hydrogen bond between the conserved Trp 103 and the carbonyl oxygen of the residue preceding position 101. Moreover, when a negatively charged residue is present in position 101, this always forms a salt bridge with the basic residue 94 (Fig. 2a–b). On the other hand, when a negatively charged residue is present in position 101 but no positively charged residue is present in position 94, the H3 torso assumes a non-bulged conformation with the acidic side chain that, deprived of the salt-bridge partner, forms a hydrogen bond with the side chain of the conserved Trp 103 (Fig. 2c–e).

For non-bulged H3 loops another relevant interaction is the packing of an aromatic residue of H3 against residues H27 and H32 (in the H1 loop). The position in the loop of the aromatic residue, which is observed in either position 96 or 99, determines two different non-bulged main chain conformations (Fig. 2d–f).

These observations are summarised in Table 2.

There are other relevant interactions of the torso of the H3 loop both within itself and with other residues but their description goes beyond the scope of this analysis and will be discussed elsewhere [12].

Table 2
Possible conformations for the torso regions of H3 loops and their diagnostic sequence patterns

Main chain conformation of the H3 torso	Key residues			
	94	101	96	99
Bulged	Arg/Lys	Asp		
Bulged	Arg/Lys	non-Asp		
Non-bulged type I	non-Arg/Lys	Asp	Aromatic	Non-aromatic
Non-bulged type II	non-Arg/Lys	Asp	Non-aromatic	Aromatic

For the purpose of this discussion, it is instead relevant to address the point of whether the conformation of the head of the loop can be predicted using known rules relating sequence and structure in hairpin loops [1,13–17].

All non-bulged known structures have quite short H3 loops, so that the problem of predicting the structure of this loop is reduced to that of predicting the structure of the torso (which depends upon the sequence pattern of the torso itself) and that of a short hairpin. In all the structures in our data base, the head of non-bulged H3 loops does indeed follow the rules relating sequence and structure in short hairpins.

A more complicated pattern arises for bulged loops. In these cases, the conformation of the head does not conform to the aforementioned rules, most likely because of the distortion in the beta sheet introduced by the presence of the bulge.

We attempted to predict these regions by using data base search techniques, often described and used as loop modelling tools [18–20].

We use the backbone of the residues flanking the head (two residues before and the three after, in order to take into account the particular hydrogen bond pattern of the β-bulge) to search the data base of known protein structures for loops having similar flanking regions and the appropriate number of residues between them. When it exists, we select as template a loop having an rms deviation between the backbone atoms of the flanking regions lower than 0.6 Å and a similar pattern of glycines and prolines as the target loop. Our weighting scheme for the data base search is such that the residue immediately preceding the loop and the subsequent residue have weight 1.0, and the weight of each successive residue is decreased by a factor of 0.8.

We used this protocol on all analysed H3 loops up to 16 residue long and can conclude that:

(1) In 16 out of 32 cases a loop with the desired characteristics is found (fit of the backbone of the stems with rms lower than 0.6 Å and similar location in the sequence of glycines and/or prolines).

(2) In all these cases, the fit of the loop obtained with the data base search is similar within 1.0 Å to the immunoglobulin loop.

(3) In no case we select a loop with the incorrect conformation.

These results are quite encouraging especially because they strongly suggest that there are rules underlying the conformation of the heads of these loops and that the continuous growth of the data base of known structures will improve our ability to predict them.

3.2. Prediction protocol

To model an immunoglobulin of unknown structure [2,21], we select parent domains from among the corresponding domains of known structure. These are in general the VL and VH domains sharing the highest sequence identity with the corresponding target domain unless another known structure that has been solved to significantly higher resolution, has only moderately lower sequence identity (not more than 5% less than the most similar one).

If the templates for the two domains are not from the same immunoglobulin, we pack the parent VL and VH domains together by a least-squares fit of the main chain atoms of residues conserved in the VL–VH interface [22].

If the rms deviation between the corresponding sets of atoms is above 1.0 Å, we delete from the set the atoms that are farthest apart, recalculate the rms deviation and superpose the interface, repeating the procedure if necessary until the rms deviation is below 1.0 Å.

The canonical structures of the torso of the H3 loops and of each of the other loops is identified by checking the antibody sequence for the particular sets of residues that form the signature of each canonical structure and, if the canonical structure identified for a loop is different from that of the parent structure, a loop from another known immunoglobulin structure having the same canonical structure is grafted into the model.

The residues adjacent to the loop in the model and in the structure with the selected canonical structure are superposed by a weighted least-squares fit [19] and the loop is transferred to the model. We use the four residues before the N-terminus of the loop and the four residues after its C-terminus to superpose the structures; the residue immediately preceding the loop and the subsequent residue have weight 1.0, and

Table 3
Summary of the prediction protocol used

Example	4-4-20	DB3	NC41
VL framework template	1mrd	1igj	1iai
(% identity)	(97.3)	(91.9)	(83.9)
L1 template	1nbv	1igj	1iai[a]
Canonical structure	4	4	2
L2 template	1nbv	1igj	1iai[a]
Canonical structure	1	1	1
L3 template	1nbv	1igj	1iai[a]
Canonical structure	1	1	1
VH framework template	1ifh	1tet	1iai[a]
(% identity)	(62.5)	(83.5)	(87.1)
H1 template	1ifh	1tet	1iai[a]
Canonical structure	1	1	1
H2 template	1nbv	1tet	1iai[a]
Canonical structure	4	2	2
H3 torso	1bbd	1vfa	1iai[a]
Canonical structure of the torso	Non-bulged	Bulged	Bulged
H3 head	Standard[b]		
Data base search hit		1lan	None

Canonical structure numbers refer to the definitions of Chothia et al. [2] and of Lesk and Tramontano [21].

[a] There are two antibody molecules in the PDB entry, here we used molecules l (light chain) and h (heavy chain).

[b] The head of this loop was modelled by using the standard angles of hairpin loops with the sequence pattern XXGX.

Table 4
Comparison between the predicted and experimental structures for the three selected immunoglobulins[a]

	4-4-20	DB3	NC41
VL framework	0.615	0.636	0.684
VH framework	0.624	0.651	0.646
VL, VH framework	1.133	0.687	0.952
L1	0.161	0.491	0.594
L2	0.069	0.214	0.177
L3	0.199	0.308	0.523
H1	0.257	0.325	0.501
H2	0.380	0.428	0.186
H3	0.869	0.779	0.414
L1, L2, L3	0.268	0.502	0.787
H1, H2, H3	1.276	0.839	0.567
L1, L2, L3, H1, H2, H3	1.981	0.916	1.088
VL	0.617	0.631	0.750
VH	0.892	0.742	0.693
VL, VH	1.393	0.755	0.989

[a] The values represent the rms deviations of the backbone atoms excluding the carboxyl oxygen.

the weight of each successive residue is decreased by a factor of 0.8.

The conformations of the side chains are then modelled by retaining the conformation of the parent structures at positions where the parent structure and the model have the same residue. If the side chain is different, its conformation is taken from an immunoglobulin having the same residue in the corresponding position whenever possible [21]. If the residue is part of a hypervariable loop, then the side chain conformation is taken only from a loop with the same canonical structure. In the remaining cases the conformation of the side chains in the parent structure is retained as far as the relative length of the side chains permits.

The model is subjected to 100 cycles of energy refinement to tidy up the stereochemistry.

3.3. Examples of predictions

We will illustrate the ability of our protocol to predict immunoglobulin variable domain structures by using three examples: 4-4-20 (PDB code 1flr), DB3 (PDB code 1dbb) and NC41 (PDB code 1nca). They were selected because they are representative for three possible scenarios: the case of a non-bulged H3 torso (4-4-20), and two of a bulged torso with (DB3) and without (NC41) hits in the data base search.

Table 3 summarises the details of the prediction protocols used while the results of the rms fit between the resulting predictions and the known structure are shown in Table 4. We show the separate contribution of each loop and of the framework to the final value of rms. The average accuracy of the predictions for the whole immunoglobulin and for H3 is quite satisfactory. In all cases the rms deviation between the predicted and experimental structure of each loop is below 0.9 Å, and the final rms deviation value for the complete variable domains is always below 0.9 Å (Fig. 3). The somewhat higher value for the complete structure compared with the values for each individual domain suggests that more effort should be devoted to the improvement of the protocol we use to pack the domains and we are addressing this problem.

In our test cases, we did not make any use of the knowledge of the three-dimensional structures of the

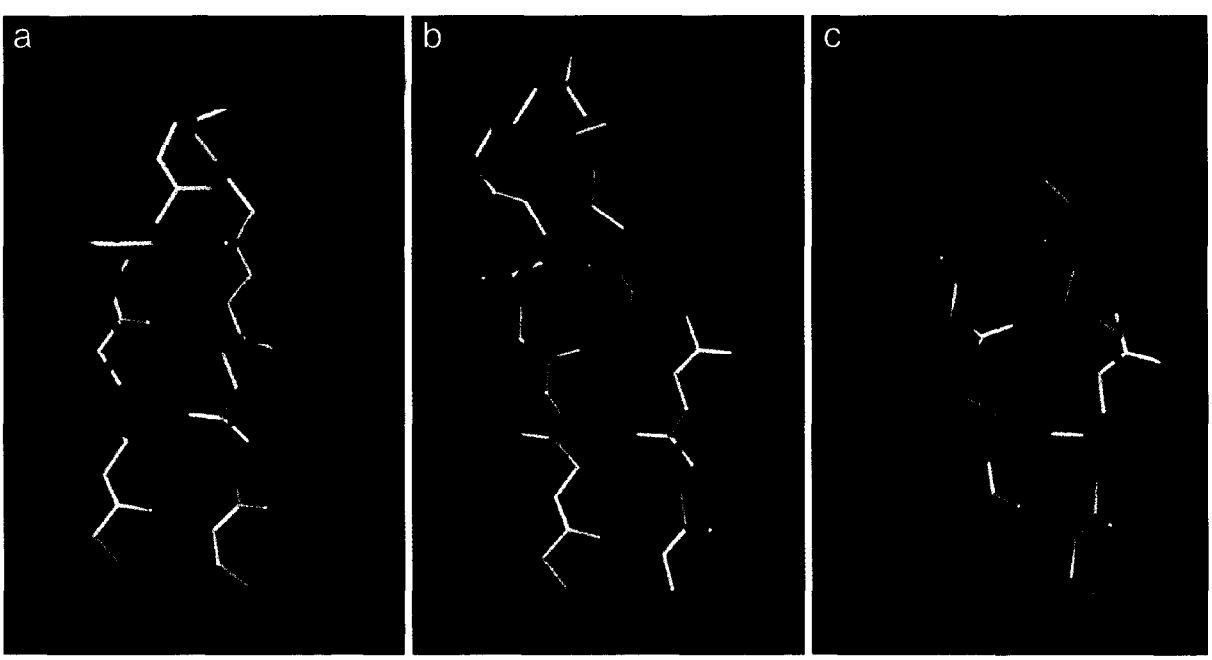

Fig. 3. Comparison between experimental and modelled H3 loop for antibodies 4-4-20 (a), DB3 (b) and NC41 (c). Modelled structures are shown in yellow, crystal structures in green.

three immunoglobulins, however strictly speaking, the predictions cannot be described as blind tests and should be considered as examples of our protocol. Nevertheless, the immunoglobulin structures released after our analysis was concluded conform to the rules described here with very few exceptions.

Acknowledgements

AML gratefully acknowledges support from the Wellcome Trust (Grant No. 039046/Z/96/A). VM thanks IRBM for the hospitality.

References

[1] C. Chothia, A.M. Lesk, J. Mol. Biol. 196 (1987) 901.
[2] C. Chothia, A.M. Lesk, A. Tramontano, M. Levitt, S.J. Smith-Gill, G. Air, S. Sheriff, E.A. Padlan, D. Davies, W. Tulip, P. Colman, Nature, 342 (1989) 877.
[3] A. Tramontano, C. Chothia, A.M. Lesk, J. Mol. Biol. 215 (1990) 175.
[4] A. Pessi, A. Bianchi, A. Crameri, S. Venturini, A. Tramontano, M. Sollazzo, Nature, 362 (1993) 367.
[5] A. Tramontano, E. Bianchi, S. Venturini, F. Martin, A. Pessi, M. Sollazzo, J. Mol. Recognition 7 (1994) 9.
[6] F. Martin, C. Toniatti, A.L. Salvati, S. Venturini, G. Ciliberto, R. Cortese, M. Sollazzo, EMBO J. 13 (1994) 5303.
[7] H.E. Dayringer, A. Tramontano, S.R. Sprang, R.J. Fletterick, J. Mol. Graphics 4 (1986) 82–87.
[8] A.M. Lesk, in: C. Saccone (Ed.), Integrated access to sequence and structural data (Biosciences: Perspectives and user services in Europe), EEC, Brussels, 1986.
[9] J. Devereux, P. Haeberli, O. Smithies, Nucl. Acids Res. 12 (1984) 387.
[10] E.A. Kabat, E.T. Wu, H.M. Perry, K.S. Gottesman, C. Foeller, Sequences of proteins of immunological interest, US Department of Health and Human Service NIH, Bethesda, MD, 1991.
[11] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F.J. Meyer, M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanochi, M. Tasumi, J. Mol. Biol. 112 (1977) 532.
[12] Morea et al., in preparation.
[13] C. Venkatachalam, Biopolymers 6 (1968) 1425.
[14] B.L. Sibanda, J.M. Thornton, Nature 316 (1985) 170.
[15] G.D. Rose, L.M. Gierasch, J.A. Smith, Adv. Protein Chem. 37 (1985) 1.
[16] C. Wilmot, M.J. Thornton, J. Mol. Biol. 203 (1988) 221.
[17] B.L. Sibanda, T.L. Blundell, J.M. Thornton, J. Mol. Biol. 206 (1989) 759.
[18] T.A. Jones, S. Thirup, EMBO J. 5 (1986) 819.
[19] A. Tramontano, A.M. Lesk, Proteins, 13 (1992) 231.
[20] M. Rustici, A.M. Lesk, J. Comp. Biol. 1 (1994) 121.
[21] A.M. Lesk, A. Tramontano, in: C. Borrebaeck (Ed.) Antibody Structure and Structural Predictions Useful in Guiding Antibody Engineering (Antibody Engineering: a practical guide), Freeman, New York, 1990.
[22] C. Chothia, J. Novotny, R. Bruccoleri, M. Karplus, J. Mol. Biol. 186 (1985) 755.