Sequence analysis of a monoclonal antibody specific for the preS2 region of hepatitis B surface antigen, and the cloning, expression and characterisation of its single-chain Fv construction

Maria Passafiume, Brigitte Vulliez-le Normand, Marie-Madeleine Riottot, Graham A. Bentley*

Unité d'Immunologie Structurale (CNRS URA 1961), Institut Pasteur, 25 rue du Dr. Roux, 75724 Paris cedex 15, France

Received 14 November 1998

Abstract The nucleotide sequence of the monoclonal antibody F124, specific for the preS2 region of the surface antigen of hepatitis B virus, has been determined and an single-chain Fv fragment (scFv) recombinant construction has been cloned and expressed into the periplasmic region of *Escherichia coli*. The recombinant antibody fragment contains a $(Gly_4Ser)_3$ linker connecting the C-terminus of the heavy chain variable region (V_L) domain to the N-terminus of the light chain variable region (V_L) domain. A 23-residue peptide segment, containing a c-*myc* marker for immunochemical detection of the scFv and hexahistidine tag to facilitate its purification, was added C-terminal to the V_L domain. The scFv mimics the antibody in binding to the native antigen in the form of recombinant hepatitis B surface antigen (HBsAg) particles as well as to peptide fragments carrying the viral epitope.

Key words: Antibody engineering; c-*myc* tag; Hepatitis B virus; PreS2; Single chain Fv fragment

1. Introduction

The major humoral immune response to hepatitis B virus (HBV) is mounted against the hepatitis B surface antigen (HBsAg), which is composed of three related membrane-bound proteins, S, M and L, located in the lipid bilayer of the viral envelope [1]. The proteins are translation products from an open reading frame with three in-phase initiation codons placed such that the M and L polypeptides share the S sequence (224 residues) at their C-terminal moieties. The M protein carries an additional 55 residues, known as the preS2 region (residues 120–174), in a position N-terminal to the S sequence, while the L protein carries the preS1 region, comprising 108 or 119 residues depending on the serotype, in a position N-terminal to the M sequence.

F124 is an anti-HBsAg monoclonal antibody (isotype IgG1/κ) which was obtained by immunisation of BALB/c mice with HBV particles, serotype ay, purified from the plasma of a chronic carrier [2]. The antibody recognises an epitope that includes residue 126 of the preS2 region since it binds less well to the peptide segment 120–153 derived from the adw serotype, where Thr-126 from the immunising ay serotype is substituted by alanine [3]. Furthermore, the epitope includes the preS2 glycan N-linked to Asn-123 since removal of carbohydrate from HBsAg particles carrying proteins S and M significantly reduces F124 binding [4]. Finally, the epitope is restricted essentially to the N-terminal region of preS2 since its binding to the peptide fragment 120–145 is not compromised by the presence of F376, a monoclonal antibody recognising the segment 132–145 [5]. As a part of our programme of structural studies on the antigenic recognition of HBsAg, we report here the nucleotide sequence of the gene segments encoding the variable domains of F124, and describe the cloning and expression of a single-chain Fv construction of the antibody. We also compare the antigen-binding properties of single-chain Fv fragment (scFv) F124 with those of the native antibody.

2. Materials and methods

2.1. Sources of antibody and antigens

The monoclonal anti-HBV antibody, F124 (CNCM I-470, 03/07/1985), was purified from ascitic fluids by ammonium sulfate precipitation and chromatography on DEAE-Sephacel, and the Fab was obtained by papain digestion. Recombinant lipoprotein particles, r-HBsAg, expressed in Chinese hamster ovary cells [6] and bearing the preS2 and S domains (serotype *ayw*), were purchased from Pasteur Mérieux. PreS2 peptides used in competitive binding studies were 120–132 (subtype ayw): MQWNSTTFHQTLQ (Neosystem); 120–145 (subtype ayw): MQWNSTTFHQTLQDPRVRGLYFPAGG; and 148–174 (subtype adw2): SGTVNPAPNIASHISSISARTGDVALN (provided by O. Siffert, Institut Pasteur).

2.2. RNA extraction and cDNA synthesis

Total RNA was isolated from 10^7 F124-expressing hybridoma cells by the acid guanidinium thiocyanate-phenol-chloroform method [7]. cDNA synthesis was performed using oligo dT 15-mer primer (Boehringer), Moloney murine leukemia virus reverse transcriptase (Gibco), RNase inhibitor (Boehringer) and a mixture of the four deoxyribonucleotides.

2.3. Amplification of the variable domain genes

PCR amplification was carried out using *Taq* DNA polymerase (Cetus Perkin Elmer). The heavy chain variable region (V_H) gene was amplified with the primers V_H IIA (5′-GAG GTC CAG CTC GAG CAG TCT GGA CC-3′) and C_H (5′-GGC AGC GAT CCA GGG GCC AGT GG-3′) (G. Orsanoudakis and P. Lafaye, personal communication), while the light chain variable region (V_L) gene was amplified with the primers $V_κ$ LC7 (5′-CCA GTT CCG AGC TCG TGA TGA CAC AGT CTC CA-3′) and $C_κ$ (5′-GCG CCG TCT AGA ATT AAC ACT CAT TCC TGT TGA A-3′) [8]. PCR products were gel purified and ligated into the pMOSBlue T-vector (Amersham). MOS Blue *E. coli* bacteria were transformed and selected by

*Corresponding author. Fax: (33) (1) 45 68 86 39.
E-mail: bentley@pasteur.fr

Abbreviations: CDR, complementarity determining region; FR, framework region; HBV, hepatitis B virus; HBsAg, hepatitis B surface antigen; r-HBsAg, recombinant hepatitis B surface antigen particles; scFv, single-chain Fv fragment; SDS-PAGE, sodium dodecyl sulfate polyacrylamide gel electrophoresis; V_L, light chain variable region; V_H, heavy chain variable region

PII: S0014-5793(98)01581-6

α-complementation screening, and the plasmid DNA from several positive clones was analysed by restriction digestion. Double-stranded nucleotide sequencing was performed by the dideoxy chain-termination method (Promega) from multiple independent clones. In order to correct the 5′ end of V_H and V_L for the possible sequence differences with the primers, the N-terminal amino acid sequence was determined for both chains (Laboratoire de Microséquençage des Protéines, Institut Pasteur, Paris). Analysis of the nucleotide sequence was performed with FASTA 3.1 [9] and DNAPLOT (http://www.genetik.uni-koeln.de/dnaplot/). The V_L and V_H nucleotide sequences of F124 have been deposited at the EMBL Nucleotide Sequence Database with accession codes AJ012372 and AJ012373, respectively.

2.4. Expression vector and construction of scFv F124

The expression vector chosen for scFv F124 was pAB1 [10] which is based on pUC119. It contains a multiple cloning site preceded by a *pelB* leader sequence, and followed by a c-*myc* epitope tag and a six-histidine tag for immobilised metal affinity chromatography purification.

V_H and V_L cDNA isolated from the F124 hybridoma were joined by a two step fusion PCR using an oligonucleotide coding for the $(Gly_4Ser)_3$ linker. The V_H and V_L sense primers were corrected on the basis of the amino acid sequences. Accordingly, V_H was amplified with the primers *Sfi*I (5′-TAC TCG CGG CCC AAC CGG CCA TGG CCG AGG TVC AGC TGC ARC ART CTG G-3′) and V_H-linker (5′-AGA GCC ACC TCC GCC TGA ACC GCC TCC ACC TGC AGA GAC AGT GAC CAG-3′). V_L was amplified with the primers V_L-linker (5′-GGC GGA GGT GGC TCT GGC GGT GGC GGA TCG GAC ATG GTG CTV ACM CAR TCT-3′) and *Not*I (5′-GAT ATG AGA TAC TGC GGC CGC CCG TTT TAT TTC CAA CTT TGT CCC-3′) (Genset). The first PCR step introduced a 3′-terminal $(Gly_4Ser)_2$-coding sequence into V_H with the V_H-linker primer, and a 5′-terminal $(Gly_4Ser)_2$-coding sequence into V_L with the V_L-linker primer. The purified amplification products of V_L and V_H were used for the second step fusion PCR with the primers *Sfi*I and *Not*I. The resulting PCR fragment coding for the single chain molecule was subcloned (using restriction sites *Sfi*I and *Not*I) into the bacterial expression vector pAB1 and sequenced.

2.5. Expression, purification and oligomerisation studies of scFv F124

E. coli TG1 cells transformed with pAB1-F124 were grown in 2YT, 0.1% glucose and 100 μg/ml ampicillin in a shaking flask at 37°C to an optical density of 0.8 at 600 nm, and induced by the addition of 1 mM IPTG (ICN). After 3 h of induction at 24°C, cells were harvested and resuspended in 20 mM Tris-HCl at pH 8 and 0.7 M sucrose. Solutions of lysozyme (Sigma) and EDTA were added to give final concentrations of 0.22 mg/ml and 10 mM, respectively. After incubating for 30 min at 4°C, cells were harvested and the supernatant was dialysed against PBS. The protein was loaded onto a Ni^{2+}-charged His-Bind Resin (Novagen), washed with 20 mM Tris at pH 7.9, 0.5 M NaCl and 5 mM imidazole, and then eluted with 20 mM Tris at pH 7.9, 0.5 M NaCl and 200 mM imidazole. The protein was further purified using gel filtration chromatography on a Superdex 200 HR 10/30 column (Pharmacia).

Monomeric and multimeric species of the purified scFv were separated by gel filtration on a Superdex 200 HR 10/30 column in PBS, at a flow rate of 0.5 ml/min. The column was calibrated with Chromatography Standards (Bio-Rad). The stability of purified monomer and dimer species was monitored by size exclusion HPLC on a Superdex 200 column.

2.6. Competitive binding of scFv to r-HBsAg

Inhibition of binding of the F124 Fab or IgG to r-HBsAg in the presence of scFv was measured by ELISA. Wells of an ELISA plate (Nunc-immunomodules) were coated with recombinant particles (0.1 μg/100 μl in PBS) and incubated overnight at 4°C. A mixture of a constant amount of Fab (10 μg/ml) or IgG (0.1 μg/ml) and various amounts of scFv diluted in PBS, supplemented with 0.02% BSA and 0.05% Tween 20, was then added to the wells for 1 h of incubation at 37°C. Residual binding of Fab or IgG to r-HBsAg was measured from the absorbance at 405 nm after the addition of alkaline-phosphatase-conjugated anti-mouse Fab-specific (Sigma) or Fc-specific (ICN) IgG, respectively, followed by *p*-nitrophenyl phosphate disodium (Sigma).

2.7. Competitive binding of scFv and IgG to peptides

Inhibition of binding of scFv or entire IgG to r-HBsAg in the presence of three peptides spanning the preS2 region (see Section 2.1) was analysed by ELISA. Wells were coated with 0.1 μg of r-HBsAg dissolved in PBS and incubated 2 h at 37°C. Each peptide was then mixed at various concentrations with a constant amount of scFv (0.1 μg/ml) or IgG (0.05 μg/ml) in PBS supplemented with 0.02% BSA and 0.05% Tween 20. After 30 min of incubation at 37°C, 100 μl of each mixture were transferred to the wells of the ELISA plate previously coated with the recombinant antigen. The monoclonal antibody anti c-*myc* 9E10 was added after a further 20 min at 37°C for the scFv trials and incubated for 1 h at 37°C. Binding to antigen-coated wells by 9E10-linked scFv or IgG was then detected by the addition of biotin-conjugated anti-mouse IgG (Vector), then streptavidin-horseradish peroxidase (Vector), followed by 2,2′-azinodi[3-ethylbenzthiazoline sulfonate] (Kierkegaard and Perry Laboratories) as colour-development reagent. The absorbance was measured at 405 nm.

3. Results

3.1. Analysis of V_H and V_L sequences

The nucleotide sequence obtained from the rearranged V region genes of F124 is shown in Fig. 1. Sequence analysis of the V_H region (Fig. 1A) showed that it belongs to the mouse H-chain subgroup IIA of Kabat [11]. It shares 97.5% sequence identity with the germline gene, M104E, a member of the J558 V_H subfamily. Seven base pair changes are observed: four are silent mutations located in the framework region (FR)-H1, one is placed within the complementarity determining region (CDR)-H2, resulting in the replacement Gly-58 → Ser, and two are located in the 3′-terminal codon 94 of the V_H gene segment, giving rise to the change Arg-94 → Asn. The CDR-H3 comprises 10 codons contributed by the D gene segment DFL16.1, used in reading frame 3, and the 5′ end of the J_H3 segment (Fig. 1B). Three nucleotides are changed within CDR-H3, resulting in the following changes: Asp-95 → Tyr and Thr-99 → Asn, from mutations within the D gene segment, and Gly-100A → Trp, from a mutation in J_H.

Sequence analysis of the V_L region showed that it belongs to the mouse κ chain subgroup III of Kabat [11]. The $V_κ$ gene segment of F124 exhibits 97.9% homology with the MMY15968 germline sequence, belonging to the $V_κ21$ family (Fig. 1C) and is joined to the $J_κ4$ segment (Fig. 1D). Six base-pair differences were observed with respect to the germline sequence, two of which are silent mutations. This leads to amino acid replacements in FR-L1 (Met-2 → Ile), CDR2-L2 (Val-50 → Ala and Lys-55 → Glu) and CDR-L3 (Gln-90 → Arg). No mutations are present in the sequence contributed by $J_κ4$.

3.2. Cloning, bacterial production and purification of scFv

The scFv was constructed by introducing the 15-residue linker, $(Gly_4Ser)_3$, connecting the C-terminus of V_H to the N-terminus of V_L. The recombinant protein also carries a 23-residue peptide segment, AAAEQKLISEEDLNGAAHHHHHH, in position C-terminal to V_L. This sequence contains a c-*myc* marker (underlined residues) recognised by monoclonal antibody 9E10 [12] and a hexahistidine tail to facilitate purification. A diagram of the construction is shown in Fig. 2. Since expression of scFv F124 was toxic for bacteria, different strains and induction times were tested. Optimum results were obtained using TG1 cells in the presence of 0.1% glucose during the growth period and an induction of 3 h at 24°C with 1 mM IPTG. Fig. 3A shows the response of

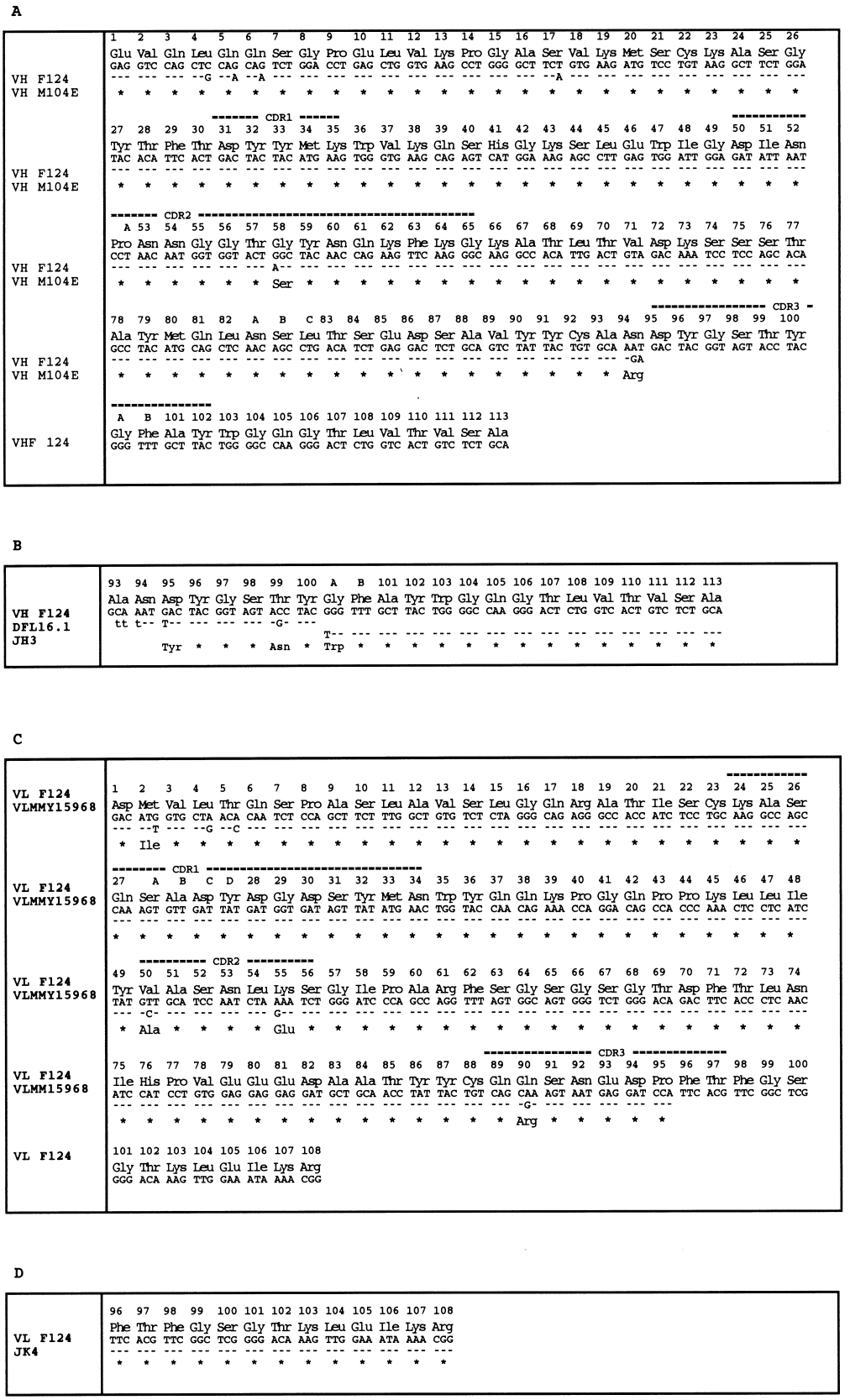

Fig. 1. Nucleotide and encoded amino acid sequences for the variable domain genes of F124. The amino acid residue numbering and the CDR limits are according to Kabat [11]. Differences with germline sequences and the corresponding encoded amino acid sequences are also shown. A: V_H domain compared with V_H germline M104E. B: CDR-H3 and FR-H4 compared with DFL16.1 and J_H3 gene segments. The three 5′-terminal nucleotides of DFL16.1, which are assumed to be deleted during somatic recombination, are shown in small letters. C: V_L domain compared with germline V_κ MMY15968 germline. D: FR-L4 compared with $J_\kappa4$.

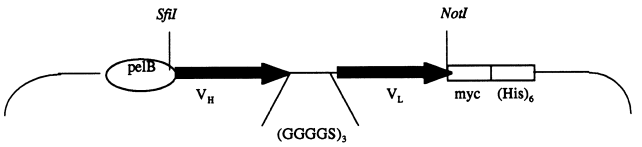

Fig. 2. Schematic presentation of the genetic construction of ScFv F124.

bacteria to different incubation times in the presence of IPTG. In the presence of glucose and absence of IPTG no expression of scFV occurred. During the first 4–5 h of induction at 24°C, scFv was recovered in the periplasm. After 24 h, however, the cells were weakened and scFv was released into the growth medium. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) analysis of fractions coming from IMAC chromatography (Fig. 3B) showed that elution of the protein began at 60 nM imidazole in quite pure form (verified with the anti-c-*myc* antibody 9E10). Gel filtration chromatography on a Superdex HR eliminated all impurity. The yield of purified recombinant scFv was 0.8 mg per litre of culture medium. Although the molecular weight was calculated

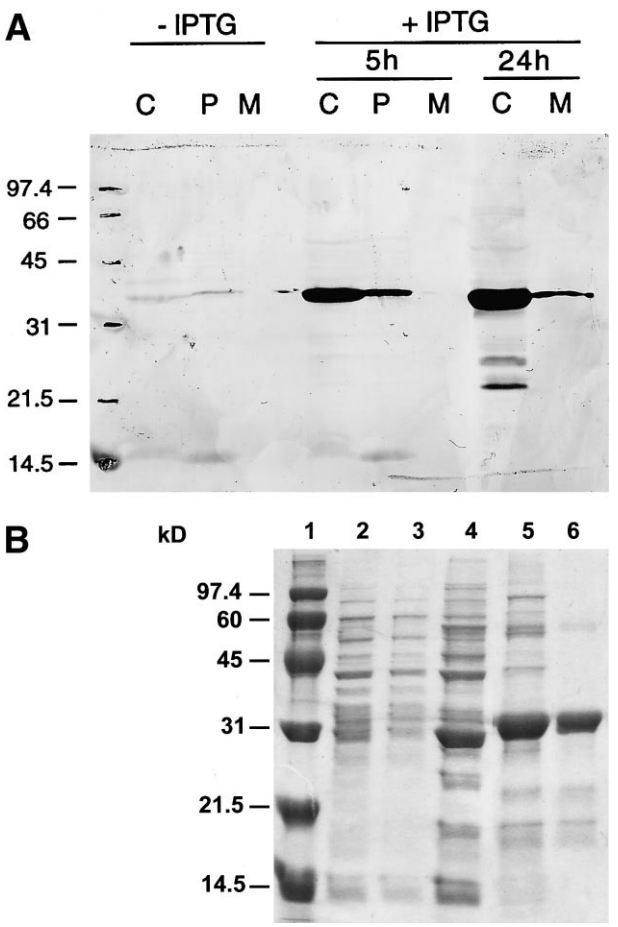

Fig. 3. Production and purification (using a Ni^{2+}-charged column) of scFv F124 carrying a c-*myc* marker and a hexahistidine tag. A: Western blot of cytoplasmic expression of scFv in the absence of (−IPTG) or presence (+IPTG) of induction. The scFv was detected using the 9E10 anti-c-*myc* antibody and an alkaline phosphatase-conjugated anti-mouse IgG antibody (substrate BCIP/NBT). B: 12% SDS-PAGE revealed by Coomassie staining of purified scFv F124. Lanes: 1, molecular standards; 2, periplasmic fraction containing the scFv; 3, flow-through; 4, washing with 5 mM imidazole; 5, elution with 60 mM imidazole; 6, elution with 200 mM imidazole.

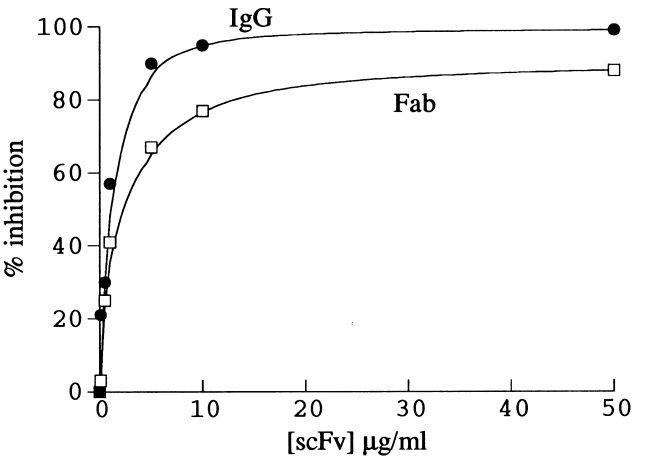

Fig. 4. Inhibition of the binding of Fab and IgG to r-HBsAg by scFv.

from the sequence to be 28.7 kDa, SDS-PAGE analysis gave an apparent molecular weight of approximately 33 kDa.

3.3. scFv monomer-dimer equilibrium

Equilibrium between the monomer and the dimer, the principal species present in the solution, was studied by size-exclusion HPLC (data not shown). The monomer and the dimer were first separated and concentrated, and then analysed separately by gel filtration for interconversion between monomer and dimer at various time intervals in PBS at 4°C. At a protein concentration lower than 200 µg/ml, the distribution was analysed at times 0, 6, 15 and 28 days, while at higher protein concentrations (1.5 mg/ml and 2.2 mg/ml), the analysis was made at 0, 2, 6 and 15 days. At low protein concentrations, the monomer showed no tendency to aggregate over 28 days, while the dimer slowly dissociated. At high protein concentrations equilibrium was achieved at a monomer/dimer ratio of 3:1. Thus, while dilute solutions of the monomer are stable, aggregation becomes more important at higher concentrations.

3.4. Competitive binding of scFv

F124 recognises an epitope localised at the N-terminal part of the M protein and containing the preS2 residue 126 [3]. The recombinant scFv F124 construction we describe here mimics the antigenic specificity of the native antibody since it inhibits the binding of both the Fab and IgG to r-HBsAg in a dose-dependent manner (Fig. 4). Although the conjugated anti-mouse Fc-specific IgG did not bind to scFv F124, the conjugated anti-mouse Fab-specific IgG did show a small non-specific binding to the construction (data not shown), thus explaining the small differences observed between the two curves obtained with the Fab and the entire IgG.

Competitive binding to r-HBsAg by scFv or IgG in the presence of preS2 peptides were also measured (Fig. 5). Peptides 120–132 and 120–145 inhibit the binding of scFv or IgG to r-HBsAg (coated on the ELISA plate) in a dose-dependent manner, while the control peptide 148–174 exhibits no inhibition of binding. Both the IgG and scFv bound 120–145 more effectively than 120–132, suggesting the larger peptide maintains the correct conformation of the epitope more efficiently for recognition by F124. A similar observation has been made elsewhere for a monoclonal antibody recognising the preS2

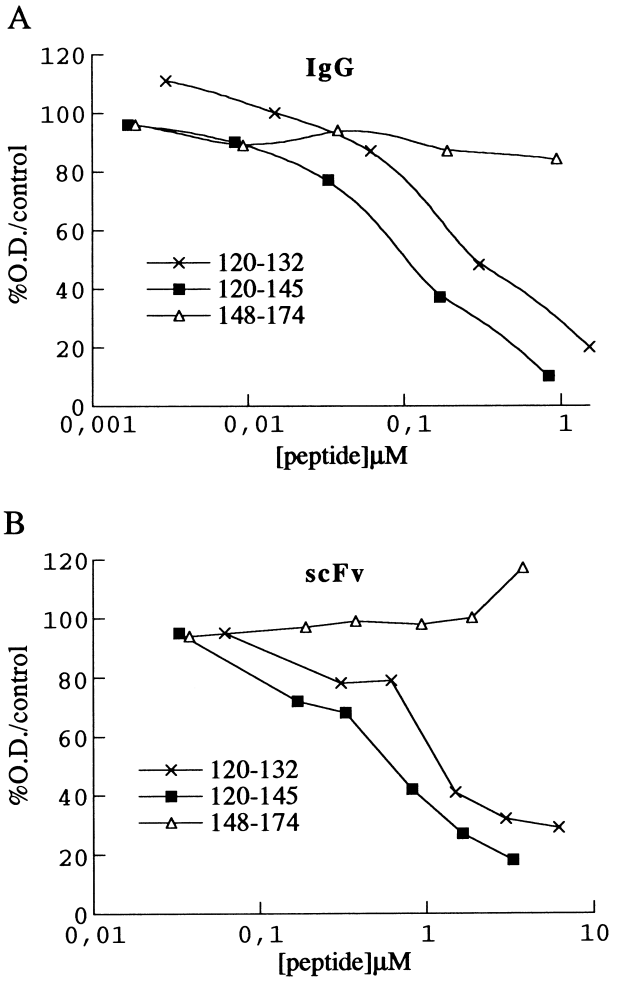

Fig. 5. Competitive inhibition of (A) IgG and (B) scFv to r-HBsAg binding by preS2 peptides 120–132, 120–145 and 148–174. The peptide concentration is shown on a logarithmic scale. Competition inhibition is expressed as a percentage with respect to a control made in the absence of peptide.

region 130–145; 120–145 was more effectively bound than 132–145 [13].

4. Discussion

The scFv construction of F124, which was recovered from the periplasm of *E. coli* with good yields, is functional since it competes with F124 IgG and Fab for binding to antigen. As is frequently observed with scFv constructions, that of F124 forms dimers and higher multimeric species, even at moderate protein concentrations.

The sequence of F124 has allowed us to identify the V, D and J gene segments used during somatic recombination to create genes coding for the complete V_H and V_L domains of the antibody, and to locate nucleotide mutations with respect to the germline sequence. Of the 16 nucleotide changes, six are silent mutations: four in the V_H domain and two in the V_L domain. The remaining nucleotide mutations give rise to five amino acid changes in the V_H domain and four amino acid changes in the V_L domain. All but two of these (Asn-94 in V_H and Met-2 in V_L) lie within the CDR defined by Kabat [11]. The heavy chain mutation Ile-2→Met is the only amino acid change in F124 that is buried in known 3-dimensional structures of closely homologous V_H domains. This conservative change should be easily accommodated, however, since the side-chain volumes of these two hydrophobic, aliphatic residues are essentially identical [14]. The amino acid change Arg-94→Asn in V_H might have arisen during somatic recombination since codon 94 shares nucleotides from the V_H and D gene sequences (Fig. 1B). The codon sequence could thus have arisen from imprecise joining between the two gene segments, the first nucleotide being contributed by the V_H gene, and the second and third nucleotides by DFL16.1. The amino acid mutation, Gly-100A→Trp, might also have arisen during somatic rearrangement of the D and J_H segments (a deletion and an N-insertion), but hypermutation appears an equally probable alternative given that only one nucleotide is changed (Fig. 1B).

Crystallographic studies on immunoglobulins have shown that, with the exception of CDR-H3, the structure of the hypervariable loops that form the antigen-binding site is limited to a small number of canonical conformations [15,16]. Although the number of possible permutations of CDR canonical conformations is significant (300), analysis of immunoglobulins in the Kabat Data Base [11] has shown that the immune system exploits only a small fraction of this potential repertoire [17]. Indeed, just 10 CDR permutations from all possible combinations were found to account for 86.9% of the antibodies with known antigen specificity in the data base. Furthermore, a correlation was detected between the particular CDR combination and the nature of the specific antigen (protein, peptide, polysaccharide, nucleic acid or hapten). Based on the amino acid sequence of F124, assignments can be made for the CDR canonical conformations; CDR-H1: canonical conformation 1; CDR-H2: 2; CDR-L1: 3; CDR-L2: unique; CDR-L3: 1 [15]. Antibodies using this permutation are specific exclusively for proteins or surface antigens [17] and thus corresponds to the expected character of the epitope on HBsAg recognised by F124.

Acknowledgements: We particularly thank Pierre Martineau and Otto Pritsch for their frequent and precious advice. We also thank Pierre Lafaye for supplying the PCR primers. This work was supported by funds from the Institut Pasteur, the Centre National de la Recherche Scientifique, the Ligue Nationale contre le Cancer and the Association pour la Recherche sur le Cancer.

References

[1] Neurath, A.R. and Kent, S.B.H. (1988) Adv. Virus Res. 34, 65–142.
[2] Budkowska, A., Riottot, M.-M., Dubreuil, P., Lazizi, Y., Brechot, C., Sobczak, E., Petit, M.-A. and Pillot, J. (1986) J. Med. Virol. 20, 111–125.
[3] Neurath, A.R., Kent, S.B.H., Strick, N., Parker, K., Courouce, A.-M., Riottot, M.M., Petit, M.A., Budkowska, A., Girard, M. and Pillot (1987) J. Mol. Immunol. 24, 975–980.
[4] Budkowska, A., Bedossa, P., Groh, F., Louise, A. and Pillot, J. (1995) J. Virol. 69, 840–848.
[5] Budkowska, A., Dubreuil, P., Riottot, M.M., Braintais, M.-J. and Pillot (1987) J. Immunol. Methods 97, 77–85.
[6] Michel, M.-L., Pontisso, P., Sobczak, E., Malpièce, Y., Streek, R.F. and Tiollais, P. (1984) Proc. Natl. Acad. Sci. USA 81, 7708–7712.
[7] Chomczynski, P. and Sacchi, N. (1987) Anal. Biochem. 162, 156–159.
[8] Kang, S.A., Burton, D.R. and Lerner, R.A. (1991) Methods: A Companion to Methods in Enzymology 2, 111-118.

[9] Pearson, W.R. and Lipman, D.J. (1988) Proc. Natl. Acad. Sci. USA 85, 2444–2448.
[10] Martineau, P., Jones, P. and Winter, G. (1998) J. Mol. Biol. 280, 117–127.
[11] Kabat, E.A., Wu, T.T., Perry, H.M., Gottesman, K.S. and Foeller, C. (1991) in: Sequences of Proteins of Immunological Interest, 5th Edn., US Department of Health and Human Services, NIH, Bethesda, MD.
[12] Evan, G.I., Lewis, G.K., Ramsay, G. and Bishop, J.M. (1985) Mol. Cell. Biol. 5, 3610–3616.
[13] Lee, M.K., Kim, K.L. and Hahm, K.-S. (1997) Mol. Cells 7, 340–346.
[14] Chothia, C. (1984) Annu. Rev. Biochem. 53, 537–572.
[15] Chothia, C. and Lesk, A.M. (1987) J. Mol. Biol. 196, 901–917.
[16] Chothia, C., Lesk, A.M., Tramontano, A., Lewit, M., Smith-Gill, S.J., Air, G., Sheriff, S., Padlan, E.A., Davies, D., Tulip, W.R., Colman, P.M., Spinelli, S., Alzari, P.M. and Poljak, R.J. (1989) Nature 342, 877–883.
[17] Vagras-Madrazo, E., Lara-Ochoa and Almagro, J.C. (1995) J. Mol. Biol. 254, 497–504.