ELSEVIER

Journal of Pharmaceutical and Biomedical Analysis
21 (2000) 1099–1128

JOURNAL OF
PHARMACEUTICAL
AND BIOMEDICAL
ANALYSIS

www.elsevier.com/locate/jpba

Validation of a peptide mapping method for a therapeutic monoclonal antibody: what could we possibly learn about a method we have run 100 times?☆

Jacob Bongers [a,*], John J. Cummings [a], Mary Beth Ebert [a], M. Marcia Federici [c], Linden Gledhill [a], Dipti Gulati [a], George M. Hilliard [a], Brian H. Jones [a], Kwan R. Lee [d], Jacek Mozdzanowski [a], Michael Naimoli [b], Sudhir Burman [a]

[a] *Department of Analytical Sciences, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406-0939, USA*

[b] *Department of Biopharmaceutical Production, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406-0939, USA*

[c] *Department of Pharmaceutical Technologies, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406-0939, USA*

[d] *Department of Statistical Sciences, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, PA 19406-0939, USA*

Accepted 2 August 1999

Abstract

Peptide mapping is a key analytical method for studying the primary structure of proteins. The sensitivity of the peptide map to even the smallest change in the covalent structure of the protein makes it a valuable 'finger-print' for identity testing and process monitoring. We recently conducted a full method validation study of an optimised reverse-phase high-performance liquid chromatography (RP–HPLC) tryptic map of a therapeutic anti-CD4 IgG1 monoclonal antibody. We have used this method routinely for over 1 year to support bioprocess development and test production lots for clinical trials. Herein we summarize the precision and ruggedness of the testing procedure and the main findings with respect to 'coverage of amino acid sequence' and limits-of-detection for various hypothetical

Abbreviations: AAA, amino acid analysis; DTT, dithithreitol; Glu-C, endoproteinase Glu-C; IgG1, immunoglobulin G1; LC/ESI-MS, liquid chromatography electrospray-ionization mass spectrometry; LOD, limit-of-detection; Lys-C, endoproteinase Lys-C; MALDI-TOF MS, matrix-assisted laser-desorption mass spectrometry; Met(O), methionine sulfoxide; PCA, principal component analysis; pGlu, pyroglutamic acid; PTH, phenylthiohydantoin; RP–HPLC, reverse-phase high-performance liquid chromatography; S/E, substrate-to-enzyme ratio (weight-to-weight); TFA, trifluoroacetic acid; TPCK, Tosyl-L-phenylalanine chloromethyl ketone.

☆ This work was presented on January 7, 1999 at the 3rd Symposium on the Analysis of Well Characterized Biotechnology Pharmaceuticals in Washington, DC.

* Corresponding author. Tel.: +1-610-270-6903; fax: +1-610-270-5830.

E-mail address: jacob_s_bongers@sbphrd.com (J. Bongers)

PII: S0731-7085(99)00181-8

structural variants. We also describe, in more detail, two unanticipated insights into the method gained from the validation study. The first of these is a potentially troublesome side-product arising during the reduction/alkylation step. Once the cause of this side-product was identified, it was easily prevented. We also report on subtle changes to the peptide map upon extended storage of the digest in the autosampler. These findings helped us to develop a 'robust' method for implementation in a quality control laboratory.

Keywords: Peptide mapping; Tryptic mapping; Recombinant protein; Monoclonal antibody; Identity testing; Method validation

1. Introduction

Rapid advances in analytical methods and instrumentation have made it incumbent on workers in biotechnology to provide more rigorous structural characterization and quality control of recombinant protein drugs [1–3]. The expression by the host cells of a homogenous protein with the desired amino acid sequence is a fundamental requirement. Moreover, the gene sequence is not sufficient to predict the final product since the nascent protein is often subject to various chemical modifications following translation of the gene, many of which can be affected by the particular cell line and the conditions under which the cells are maintained. Proteolysis, phosphorylation, N-terminal acetylation, and pyroglutamate formation are common examples of post-translational modifications and the use of mammalian cell culture has added glycosylation to the picture. All these elements of the structure can be experimentally verified with modern analytical methods.

Peptide mapping [4–10] is an essential technique for studying the primary structure of proteins. For recombinant protein pharmaceuticals, peptide mapping is used for the initial 'proof of structure' characterization, i.e. to confirm expression of the desired amino acid sequence and to characterize any post-translational modifications. Further, peptide mapping is employed for subsequent lot-to-lot identity testing ('finger-printing') in support of bioprocess development and clinical trials. Peptide mapping is also the current method of choice for monitoring the 'genetic stability' of recombinant cell lines [11,12].

In recent times, peptide mapping has become a much more rapid and convenient method [8] and this greatly expands its utility for process monitoring and other higher volume applications. For an analyst to reduce, alkylate, and digest a dozen or more samples in parallel over the course of a working day, well nigh impossible by the older procedures, is now fairly routine. We recently completed a full validation study of an optimized RP–HPLC tryptic mapping method for a therapeutic anti-CD4 IgG1 monoclonal antibody [13] produced by large-scale mammalian cell culture. The method is diagrammed in Fig. 1 and a typical example of the data shown in Fig. 2. Large savings in time and labor can be gained by the use of gel-filtration instead of dialysis for removing the excess reagents and exchanging the carboxymethylated protein into the digestion buffer. Efficient sample 'clean-up' (desalting) enhances the efficiency of the trypsin digestion. It is especially critical to remove guanidinium ions prior to digestion as guanidinium is a competitive binding inhibitor of trypsin. Traditionally, ammonium bicarbonate is used as a digestion buffer because of its advantage of being 'volatile'. Unfortunately, this widely used digestion buffer is not very favorable to the activity of trypsin. The Tris/HCl buffer in Fig. 1 and other digestion buffers [7,9] support considerably more rapid proteolysis than ammonium bicarbonate for a given amount of trypsin and digestions can thus be run over several hours rather than overnight. In this report, we provide an overview of the method validation study for this rapid tryptic mapping method and focus on two findings that were of special practical relevance.

2. Experimental

The anti-CD4 IgG1 monoclonal antibody [13] was produced in recombinant CHO cells by

SmithKline Beecham. Trypsin (TPCK-treated) was from Worthington Biochemical Corporation unless stated otherwise. Endoproteinase Lys-C (Achromobacter protease I) was from Wako Pure Chemical Industries, Limited. Synthetic peptides were made by standard solid-phase methods (American Peptide Company) and characterized by RP–HPLC, amino acid analysis, and MALDI-TOF MS. Dithiothreitol and sodium iodoacetate were from Aldrich Chemical Company. Water was purified on a Milli-Q UV Plus system (Millipore). Disposable prepacked Bio-Gel P-6 (Bio-Rad Econo-Pac 10DG) and G-25 gel columns (Pharmacia NAP-5) were used for buffer exchanges.

RP–HPLC was performed, unless stated otherwise, on a Hewlett Packard 1090M with DR5 pumping system, static mixer (HP PN 79835-873307), thermostated column oven, and diode-array UV detector (DAD) with 4 mm slit and standard 6 mm flow-cell. The eluents used throughout were: (A) 0.1% v/v TFA in water; and (B) 0.1% v/v TFA in 80% acetonitrile. Liquid-phase acid hydrolysis of the carboxymethylated protein and selected peptides was in 6 M HCl at 110°C overnight and compositions and recoveries determined quantitative amino acid analysis (norleucine internal standard) was performed by ion-exchange/postcolumn-derivatization (ninhydrin) on a Beckman 6300 amino acid analyzer. Amino

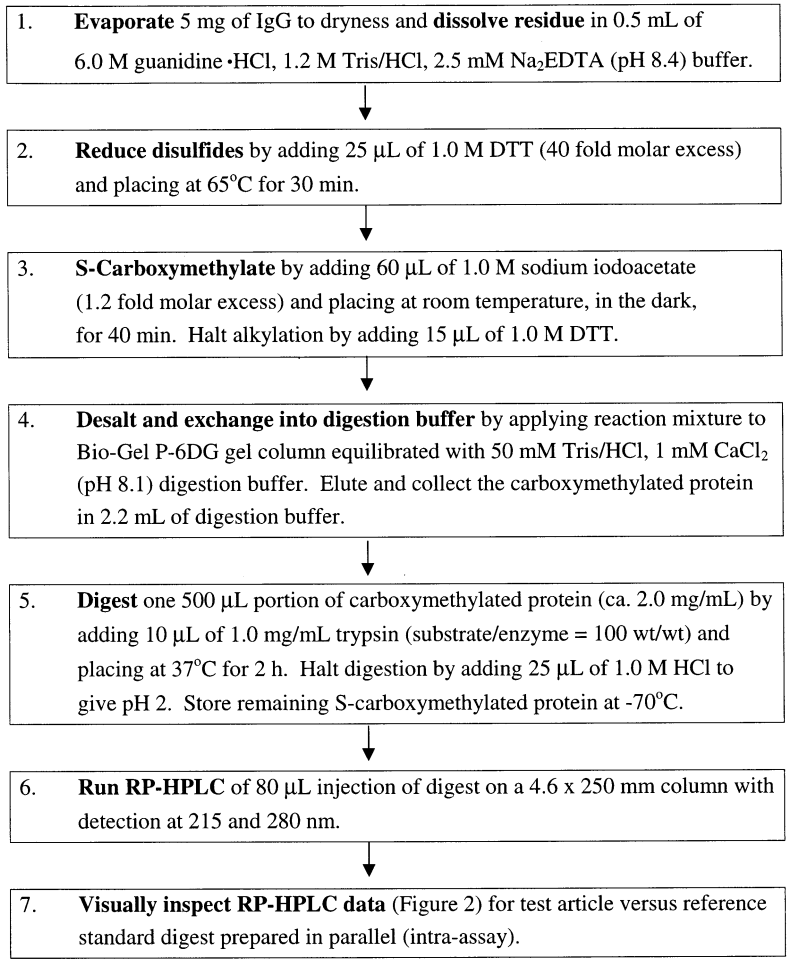

Fig. 1. Procedure for rapid tryptic mapping of the anti-CD4 IgG1.

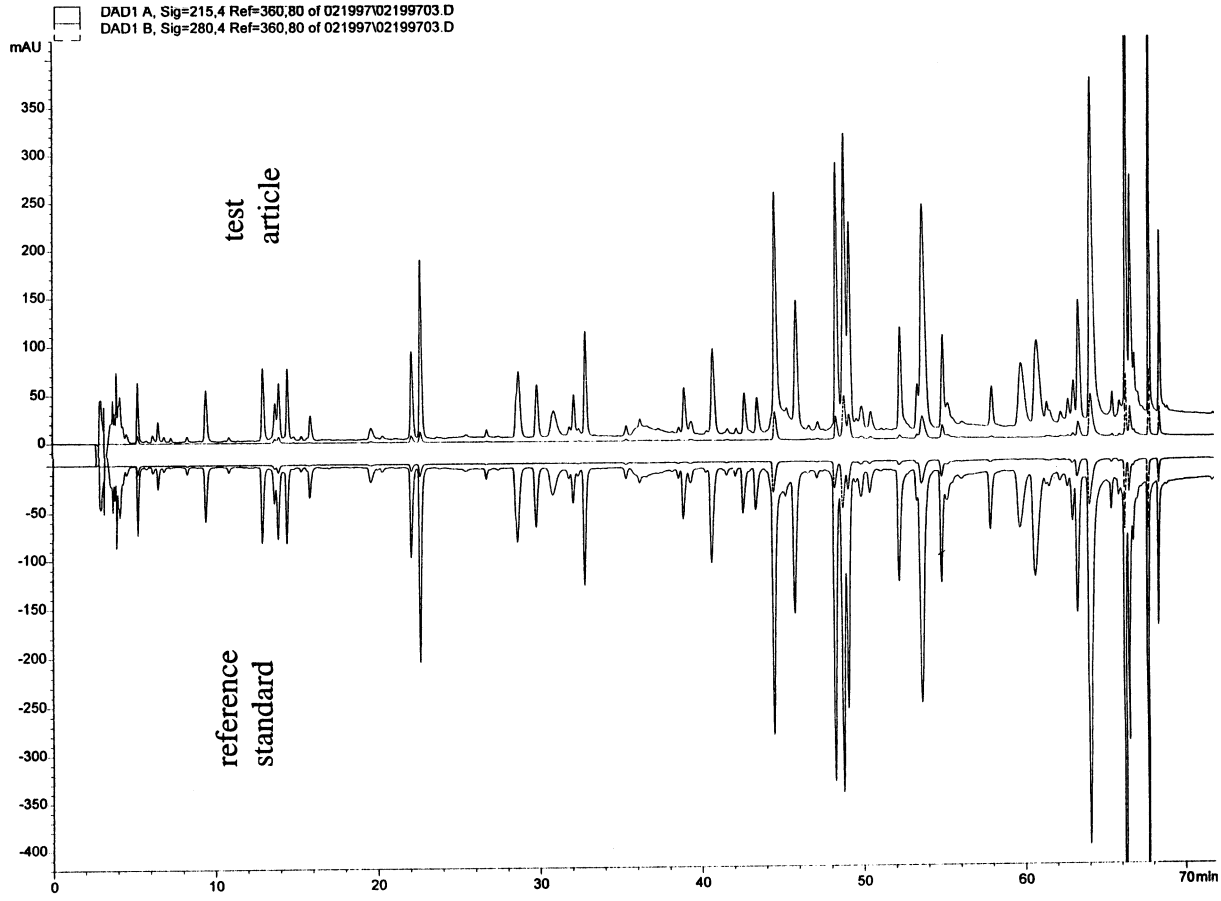

Fig. 2. Typical RP–HPLC tryptic mapping data for anti-CD4. Detection at 215 and 280 nm.

acid compositions were also determined for large sets of isolated peaks by batch-wise vapor-phase acid hydrolysis by using the apparatus of Meltzer et al. [14] and amino acid analysis by the DABS-Cl precolumn-derivatization/RP–HPLC method of Knecht and Chang [15]. SDS-PAGE (4–20% tris-glycine) in the presence or absence of reducing agent (dithiothreitol) was by the Laemmeli method. Ten micrograms of each of the samples were applied to the gel following incubation (10 min at 70°C) in the appropriate sample buffer. Immediately following electrophoresis samples were visualized with Coomassie blue R-250 stain.

Sequencing was performed on a Hewlett-Packard G1005A Protein Sequencer as described earlier [16]. Peptides were immobilized on HP biphasic sequencer columns by the standard procedure, i.e. wetting the column with 1 ml of methanol, followed by 1 ml of dilute aqueous TFA, and then applying the sample. On-line RP–HPLC detection and quantitation of the PTH amino acids was performed by the standard HP method with one modification; i.e. *S*-carboxymethylcysteine (Sigma) was added to the standard mixture of PTH amino acids LC/ESI-MS of the tryptic digest was performed using the standard LC conditions below on a Hewlett Packard 1090M LC interfaced, via a flow-splitter, to a Perkin-Elmer/Sciex API III electrospray ionization mass spectrometer. MALDI-TOF MS was performed in sinapinic acid or α-cyano-4-hydroxycinnamic acid matrix on a Hewlett Packard 2025A LD-TOF mass spectrometer using external MW standards.

Reduction/alkylations were carried out by evaporating the antibody sample to dryness, dissolving the residue to 10 mg ml^{-1} in 6 M guanidine hydrochloride, 1.2 M Tris/HCl buffer (pH 8.1), reducing with 50 mM DTT at 65°C for 30 min, alkylating with 120 mM sodium iodoacetate at room temperature for 40 min in the dark, and then desalting the carboxymethylated antibody into 50 mM Tris/HCl, 1 mM $CaCl_2$ (pH 8.1) digestion buffer by use of a pre-packed disposable gel-filtration column (Bio-Rad Econo-Pac 10DG Bio-Gel P-6 column).

Tryptic digestions of the resulting 2.0 mg ml^{-1} carboxymethylated antibody in 50 mM Tris/HCl, 1 mM $CaCl_2$ (pH 8.1) were carried out with 0.02 mg ml^{-1} trypsin (Worthington Biochemical, bovine pancreatic, TPCK-treated) at 37°C for 2 h. Digestions were halted by acidifying to $\sim$pH 2 with 1.0 M HCl. RP–HPLC of the tryptic digest (RP–HPLC mapping) was performed using the above eluents, and the following: Vydac 218TP54 C18 column (4.6 × 250 mm) thermostated at 35°C; gradient: 5–40% B in 60 min, 60–100% B in 20 min at 1.0 ml min^{-1}; detection at 215 and 280 nm.

Carboxymethylated heavy and light chains of anti-CD4 were isolated by RP–HPLC (Fig. 3) on a PerSeptive Biosystems Poros R/H (4.6 × 100 mm) column at 75°C; flow: 1.0 ml min^{-1}; detection: 215 and 280 nm; gradient: 30–50% B in 7 min. The purified chains were then buffer-exchanged into the 50 mM Tris/HCl, 1 mM $CaCl_2$ (pH 8.1) digestion buffer and trypsin digestion (100/1 S/E, 37°C, 2 h) and RP–HPLC mapping performed according to the standard method (Fig. 1).

Fc and Fab were prepared [17] by trypsin digestion (100/1 wt:wt S/E) of 10 mg ml^{-1} native anti-CD4 in 50 mM Tris/HCl, 1 mM $CaCl_2$ (pH 8.1) at room temperature for 24 h and purified by RP–HPLC on a PerSeptive Biosystems Poros R/H (4.6 × 100 mm) column at 75°C; flow: 2.0 ml min^{-1}; detection: 215 and 280 nm; gradient: 38–52% B in 6 min at 1.0 ml min^{-1}. The purified Fc and Fab were evaporated to dryness and reduction/carboxymethylation, trypsin digestion, and RP–HPLC mapping performed according to the standard method (Fig. 1).

Oxidation of Met residues in anti-CD4 to Met(O) was done by dissolving a 5 mg portion of anti-CD4 in 250 μl of 5% v/v acetic acid/HCl (pH 1.5) buffer, adding 3 μl of 30% hydrogen peroxide, allowing to vial to stand at room temperature for 60 min, and then evaporating to dryness.

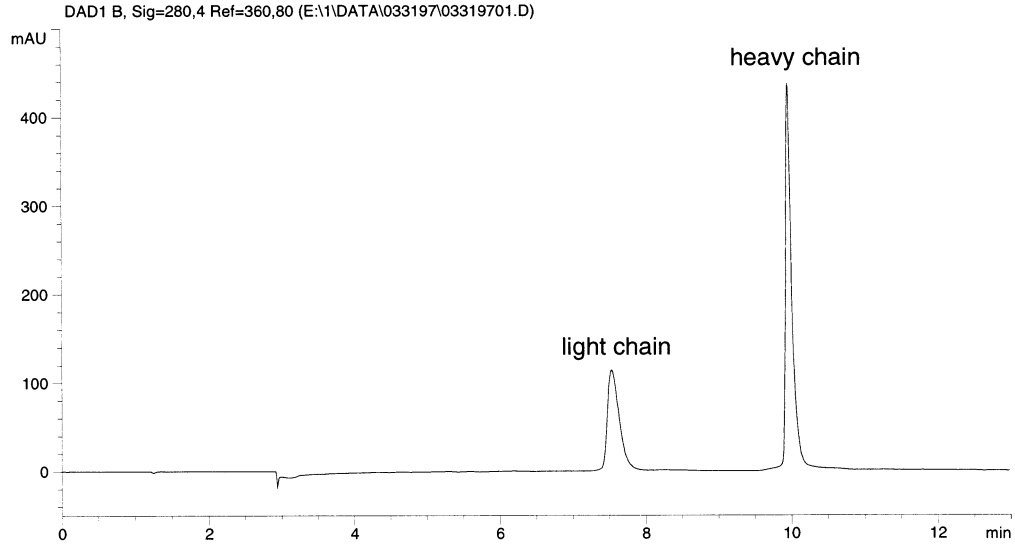

Fig. 3. RP–HPLC analysis of carboxymethylated anti-CD4. The light and heavy chains were isolated and analyzed by MALDI-TOF MS and tryptic mapping. Detection at 280 nm.

Tryptic mapping of the oxidized anti-CD4 residue was by the standard method (Fig. 1).

3. Results and discussion

3.1. Characterization and recoveries of carboxymethylated antibody and tryptic peptides

Efficient carboxymethylation was demonstrated for three separate preparations of reference standard. Amino acid analysis (AAA) gave an average *S*-carboxymethylcysteine content of 87% of the theoretical 32 moles per mole IgG1. The recovery of the carboxymethylated protein immediately prior to adding trypsin was found by quantitative AAA, using norlecuine as an internal standard, to be 90% versus the starting material. The Coomassie-Blue stained SDS-PAGE gel showed heavy and light chain bands in good purity ($>$ 95%), with the correct apparent molecular weights and the correct 2-to-1 ratio of staining intensities for the heavy versus the light chain. Purity of the carboxymethylated protein was also assessed by RP–HPLC (Fig. 3). The method of Gill and von Hippel [18], for calculating UV extinction coefficients for proteins at 280 nm based on their amino acid compositions, was used to calculate a theoretical ratio of peak areas for the heavy versus the light chain of 2.10 and we found a value of 2.09 from the data in Fig. 3. MALDI-TOF MS of the isolated peaks gave the expected m/z values for the light chain (m/z in kDa of 22.90 obs, 22.73 theor for MH^+; 11.47 obs, 11.37 theor for MH_2^{2+}) and heavy chain (m/z in kDa of 51.15 obs, 52.34 theor for MH^+; 25.66 obs, 26.17 theor for MH_2^{2+}).

Shown in Fig. 4 is a full set of peak assignments for the RP–HPLC tryptic map that account for 651 of the 663 residues (98%) in the IgG1 primary structure (Fig. 5). These assignments are based on comparative tryptic mapping of IgG1 and purified light chain, heavy chain, Fc, and Fab domains (Figs. 6 and 7); endopeptidase Lys-C mapping of the IgG1 (Fig. 8); electrospray ionization mass spectral detection of the map (LC/ESI-MS); and extensive AAA and N-terminal sequencing (Edman degradation) of isolated peaks. The tryptic map peak assignments in Fig. 4 are based on combinations of the following data:

1. The m/z value(s) from LC/ESI-MS.
2. Amino acid composition.
3. N-Terminal sequence.
4. UV absorption at 280 nm?
5. Same peak in endoproteinase Lys-C map?
6. Same peak in tryptic map of purified heavy or light chain?
7. Same peak in tryptic map of purified Fc or Fab domain?

The combinations of the above seven parameters measured for each peak are summarized in Tables 1 and 2. Parameters 4–7 above were obtained for all the peaks in the map.

LC/ESI-MS was obtained for nearly all the peaks with the exception of those small peptides (<250 Da) that fell below the spectral range. Amino acid compositions were obtained for nearly all the peaks. N-terminal sequencing was used only selectively. Comprehensive characterization of peptide maps has traditionally been accomplished by the microchemical methods of N-terminal sequencing (Edman degradation) and AAA and, more recently, mass spectrometry.

LC/ESI-MS is very valuable in that it gives a large amount of 'm/z information' across the entire map from a single instrument run. However, sifting through all this MS data can be laborious and identifications by MS alone can sometimes be difficult or ambiguous without some other independent data. Some combination of MS and microchemical methods is usually sufficient to identify nearly all peptide species in the map. The information contained in criteria 4–7 can be useful in a different way. Consider, for example, the H260-278 peak eluting at 45.8 min (Fig. 4). Figs. 6–8 reveal that this peak lacks 280 nm signal, belongs to the heavy chain and the Fc domain, and is not present in the Lys-C map. This narrows down the possible assignments for the 45.8 min peak from the 53 total theoretical tryptic fragments in anti-CD4 to the 12 theoretical tryptic fragments 16–18, 20, 24, 28–31, 36, and 38 in the Fc portion of the heavy chain (Table 2). Several of these 12 theoretical tryptic fragments are amino acids or small oligopeptides and can thus be eliminated from consideration on this basis.

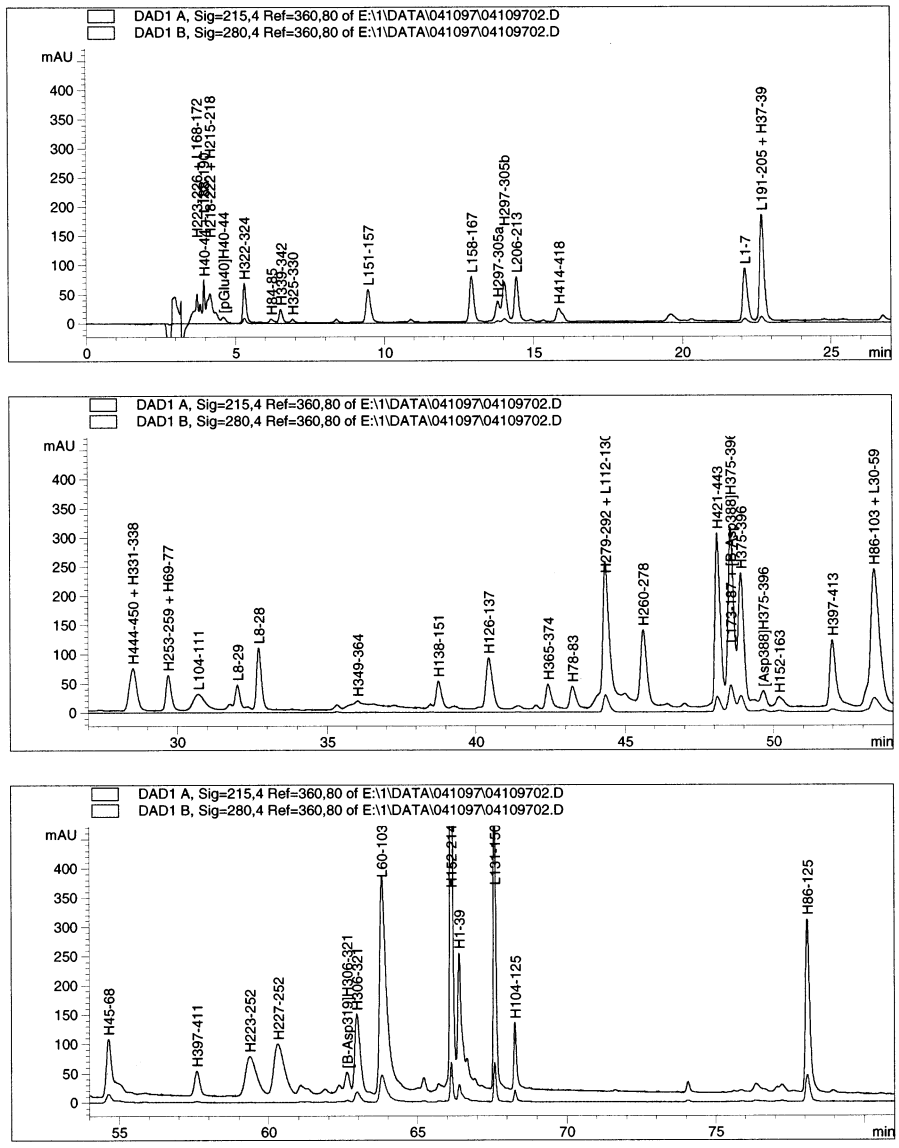

Fig. 4. Peak assignments for the RP–HPLC tryptic map of carboxymethylated anti-CD4 IgG1. The peptides are identified by residue numbers and 'H' for heavy chain or 'L' for light chain (refer to Fig. 5 for amino acid sequences).

Selected tryptic peptides were collected from the standard RP–HPLC of the digest and their recoveries determined by quantitative AAA. Briefly, the effluent from the UV detector was manually collected in borosilicate hydrolysis tubes for the four representative peptides, H322–324 (5.3 min), L1–7 (22.1 min), H421–443 (48.1 min), and L131–150 (67.6 min). These tubes were then spiked with norleucine internal standard, flame-sealed, hydrolyzed, and the hydrolysate run on a Beckman 6300 amino acid analyzer. The peptide recoveries versus the known amount of digest injected onto the column, as determined above by AAA of the carboxymethylated protein were 126, 112, 91, 100% for H322–324, L1–7, H421–443, and L131–150, respectively. These recoveries are reasonably high and close to equimolar.

Anti-CD4 is a glycoprotein of the immunoglobulin G1 (IgG1) structural type and contains two identical copies each of the light and heavy chains (Fig. 5). All 32 cysteines in anti-CD4 are disulfide bonded. The N-linked oligosaccharides are attached to asparagine 301 of the heavy chain. The two main glycoforms of the H297–305 glycopeptide, which differ by a single terminal galactose [19], are partially resolved in the RP–HPLC tryptic map (Fig. 4). The heavy chain contains a pyroglutamate residue (pGlu) which is generated post-translationally from an N-terminal glutamine. The gene also encodes for a C-terminal lysine 451 residue in the heavy chain which is nearly completely absent in the final protein product. Presumably, this lysine is removed by a carboxypeptidase. These three post-translational modifications to anti-CD4 were all confirmed and are typical for an IgG1 [20–26].

3.2. Incomplete reduction of C_H3 disulfide-loop due to insufficient sample dissolution

Early in the validation exercise, we observed an unusual peak at 65.5 min that occurred only sporadically. This was identified as a mixture of unreduced/undigested H343–450, H345–450, H343–443, and H345–443, i.e. essentially an entire C_H3 immunoglobulin fold (Fig. 9) with 'ragged ends'. This identification was based on analysis of the isolated peak by N-terminal sequencing and MALDI-TOF MS (MW in kDa for H343–450: 12.15

anti-CD4 heavy chain trypsin map

```
pGlu-VQLQESGPG LVKPSETLSL TCSVSGGSIS GDYYWFWIRQ SPGKGLEWIG   50
    YIYGSGGGTN YNPSLNNRVS ISIDTSKNLF SLKLRSVTAA DTAVYYCASN  100
    ILKYLHWLLY WGQGVLVTVS SASTKGPSVF PLAPSSKSTS GGTAALGCLV  150
    KDYFPEPVTV SWNSGALTSG VHTFPAVLQS SGLYSLSSVV TVPSSSLGTQ  200
    TYICNVNHKP SNTKVDKKAE PKSCDKTHTC PPCPAPELLG GPSVFLFPPK  250
    PKDTLMISRT PEVTCVVVDV SHEDPEVKFN WYVDGVEVHN AKTKPREEQY  300
    NSTYRVVSVL TVLHQDWLNG KEYKCKVSNK ALPAPIEKTI SKAKGQPREP  350
    QVYTLPPSRD ELTKNQVSLT CLVKGFYPSD IAVEWESNGQ PENNYKTTPP  400
    VLDSDGSFFL YSKLTVDKSR WQQGNVFSCS VMHEALHNHY TQKSLSLSPG  450
```

anti-CD4 light chain trypsin map

```
    YELSQPRSVS VSPGQTAGFT CGGDNVGRKS VQWYQQKPPQ APVLVIYADS   50
    ERPSGIPARF SGSNSGNTAT LTISGVEAGD EADYYCQVWD STADHWVFGG  100
    GTRLTVLGQP KAAPSVTLFP PSSEELQANK ATLVCLISDF YPGAVTVAWK  150
    ADSSPVKAGV TTTPSEKQSN NKYAASSYLS LTPEQWKSHR SYSCQVTHEG  200
    STVEKTVAPT ECS  213
```

Fig. 5. The peptides detected in the RP–HPLC tryptic map of the CM-Cys anti-CD4 IgG1 in Fig. 4 account for 651 of the 663 residues (98%) in the amino acid sequence. Those residues not detected in the map are underlined.

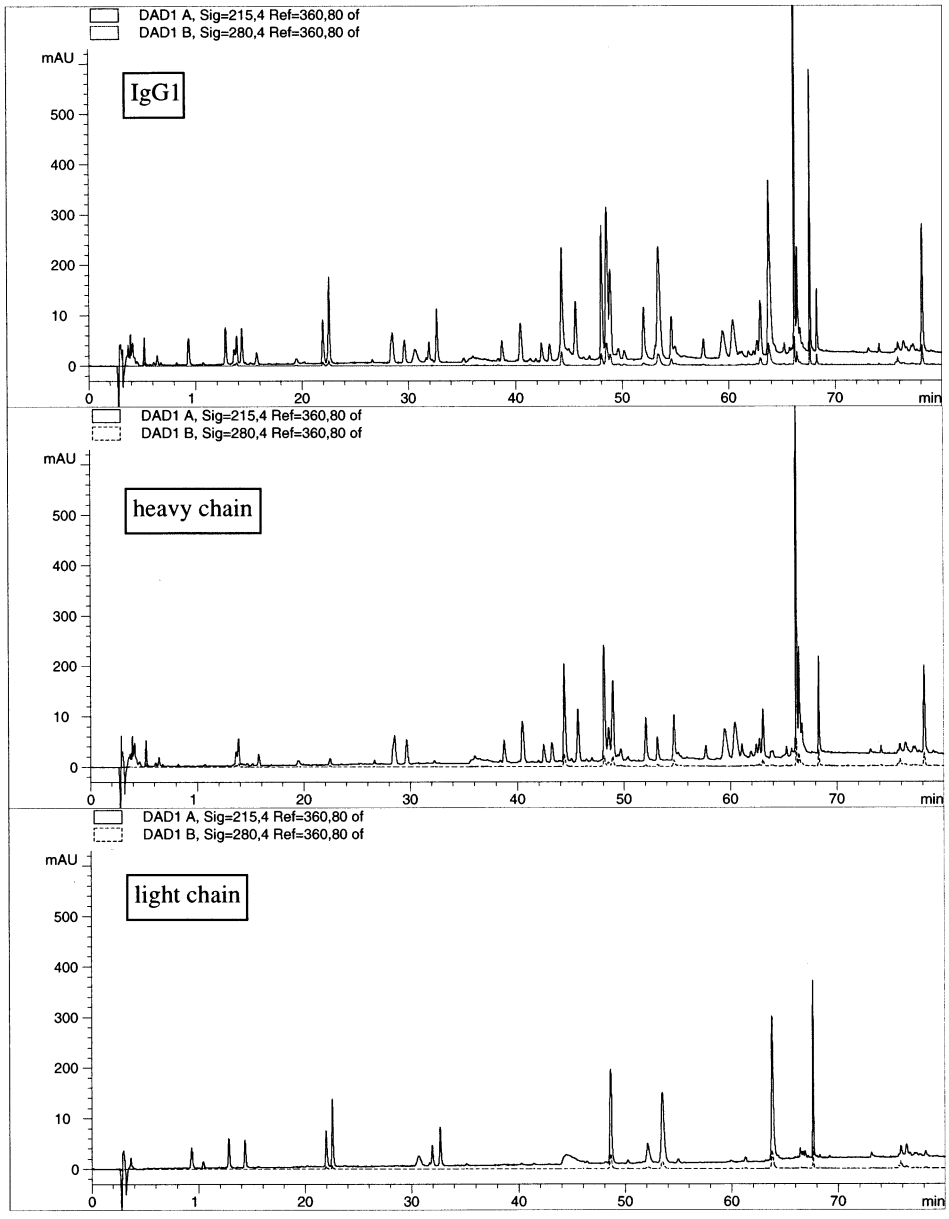

Fig. 6. Tryptic mapping of the purified heavy and light chains of CM-Cys anti-CD4 IgG1. The chains were purified by RP–HPLC as shown in Fig. 3.

obs, 12.15 theor; H345–450: 11.96 obs, 11.95 theor; H343–443: 11.52 obs, 11.51 theor; H345–443: 11.31 obs, 11.31 theor). This species, when present, was typically at 15–20 mole percent. Essentially the same C_H3 disulfide-loop species was identified by LC/MS tryptic mapping of a similar humanized IgG1 several years ago [20]. This C_H3 disulfide-loop portion of the protein is apparently quite resistant to proteolysis. The extreme reluctance of the C_H3 domain to unfold may be due, at least in part, to the very high number of proline residues throughout this region of the human sequence.

Some IgG1 samples formed a dense glassy pellet upon evaporation at the very beginning of the procedure (Fig. 1). We suspected that the rather slow dissolution of this glassy pellet in the guanidine hydrochloride buffer might be causing the incomplete reduction of the C_H3 disulfide-loop and proved this to be true by a definitive experiment wherein pellets were purposefully left undissolved and partially dissolved (Fig. 10). In conclusion, the analyst should take particular care

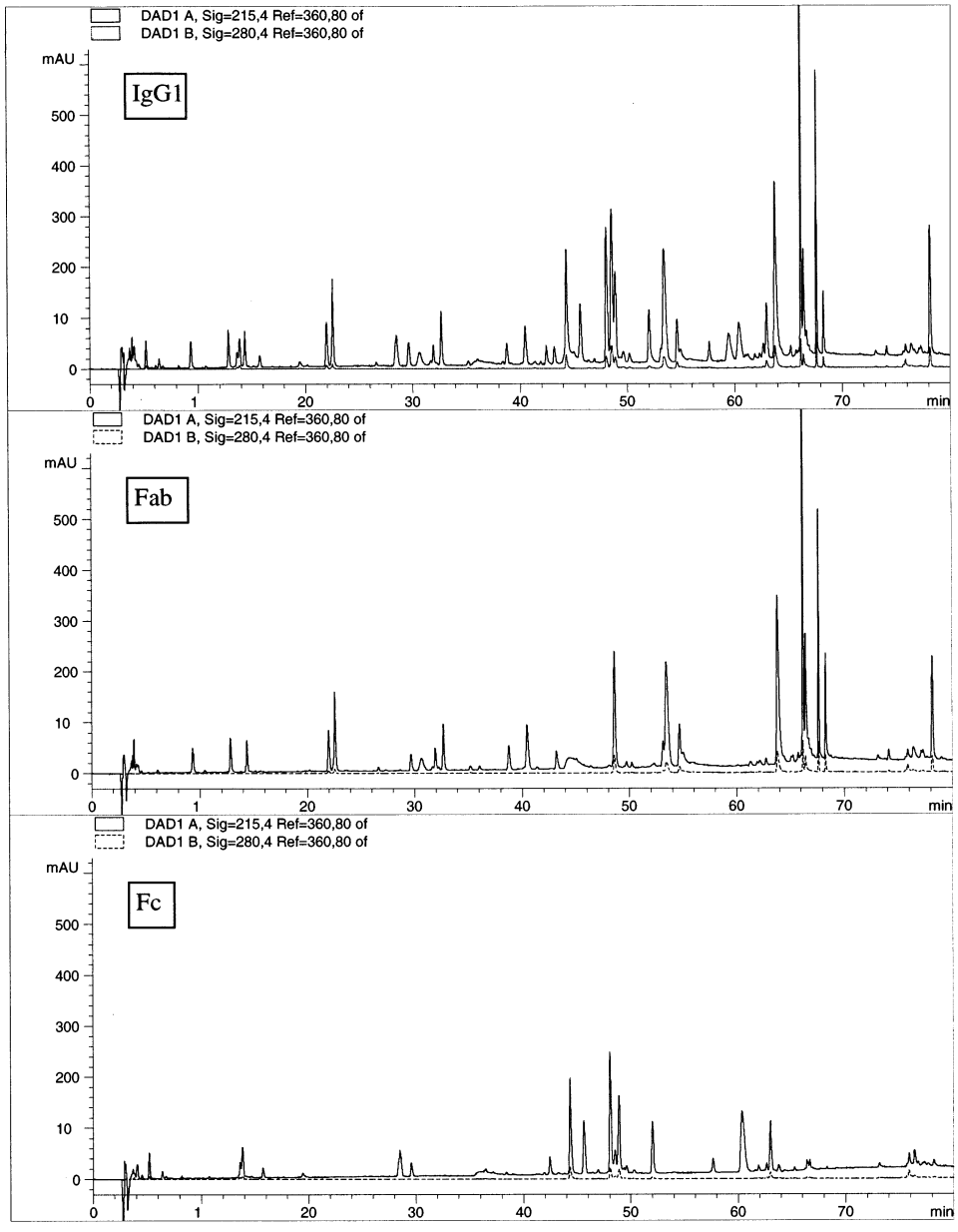

Fig. 7. Tryptic mapping of separate Fc and Fab domains of anti-CD4 IgG1 prepared by specific trypsin cleavage at the Lys^{226}–Thr^{227} in the 'hinge' region of the native IgG1.

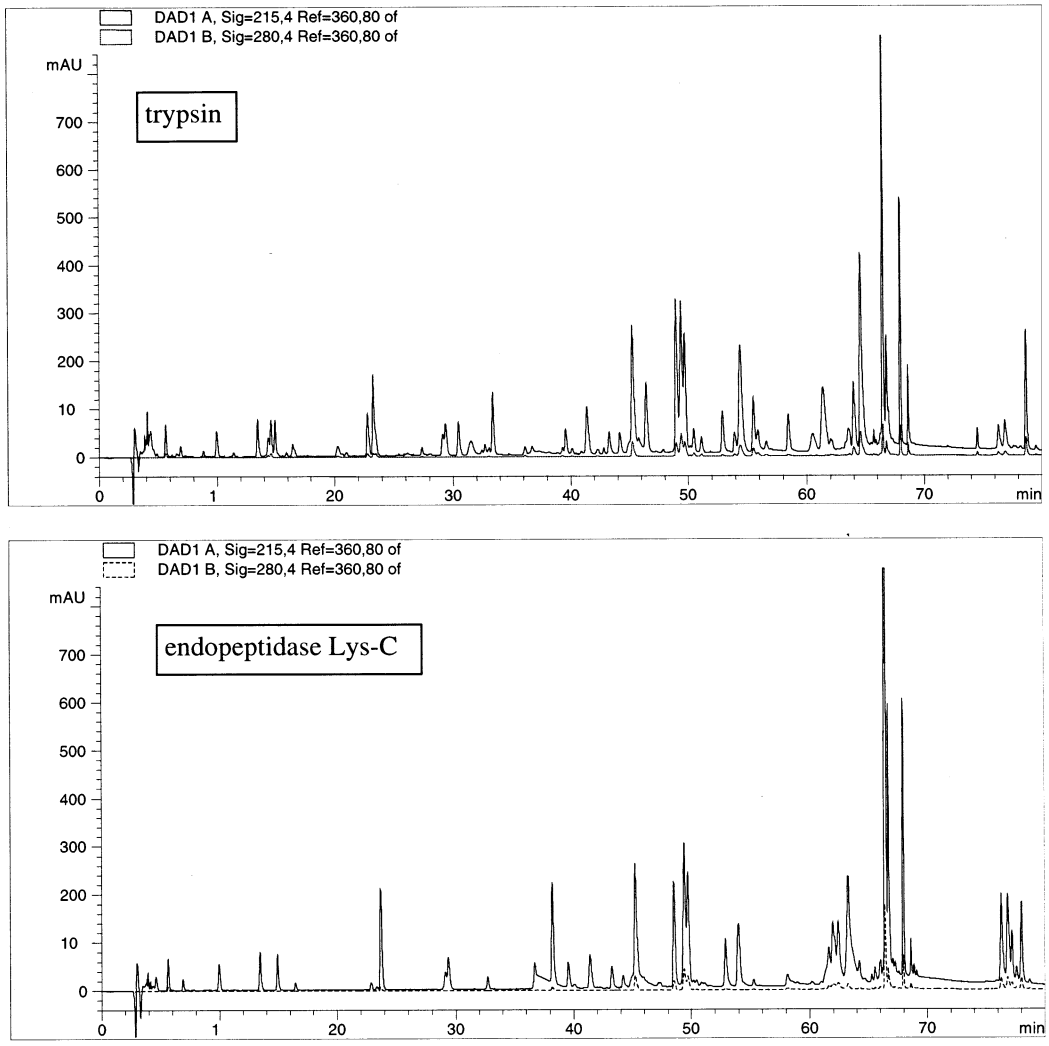

Fig. 8. Comparative peptide mapping of anti-CD4 IgG1 with endoproteinase Lys-C and trypsin.

to ensure that the sample pellet dissolves completely before proceeding any further.

3.3. Tryptic digest stability

The tryptic digest for this IgG1 is a complex mixture of ~ 50 peptides with unique solubility and other chemical properties. A number of realistic storage conditions were tested, including room temperature, 4°C, and -70°C. Fig. 11 shows expanded overlay plots revealing two subtle changes to the map occurring upon extended storage in the polypropylene autosampler vial at room temperature. A decrease in the H40–44 peak (3.9 min) correlated well with the appearance of a new peak at 4.6 min. LC/ESI-MS indicated that this new peak is the pyroglutamate (pGlu) derivative [29] of H40–44 formed by loss of ammonia from the N-terminal glutamine on this fragment (Fig. 11). This formation of pyroglutamate from N-terminal glutamine residues generated by trypsin digestion under these RP–HPLC mapping conditions is well known in the literature [29,30].

The second change was a decrease in the L60–103 peptide eluting at 62.9 min, which was assigned as L60–103. The disappearance of this peptide was not associated with any corresponding appearance of a new peak or peaks elsewhere in the map and is probably due to precipitation or adsorption to the vial walls. As long as the digest is not stood longer than a typical working day at room temperature, the map does not change significantly.

The digest is also extremely stable when stored at −70°C and is not affected by up to five freeze/thaw cycles. The ability to store the tryptic digests and the carboxymethylated protein for long periods at −70°C can be useful. A 'frozen archive' can be maintained so that samples can be retrieved for further analysis at a future date. For example, it is sometimes desirable to rerun the RP–HPLC in order to isolate a selected peak or peaks for further investigation without having to repeat the entire procedure.

3.4. Selectivity: limits-of-detection (LOD) for hypothetical structural variants

The purpose of this section was to test the capability of the method to detect hypothetical modifications to the antibody structure. Any conceivable change, no matter how small, will cause a new peak to appear in the tryptic map at the expense of the 'parent' peak with the notable exception a pre-existing hydrolysis ('clip') at a Lys–Xaa or Arg–Xaa bond. We have chosen examples of rather small changes that are of some practical relevance to the production and storage of the protein and the use of the map to monitor genetic stability. The results of this section are summarized in Table 3.

Ideally, a digest of pure selectively modified protein is mixed with a control digest in varying proportions. One then observes a decrease in peak response for the unmodified ('parent' or 'wild-type') tryptic peptide and a corresponding in-

Table 1
Summary of characterization of the light chain (L) tryptic peptides in the RP–HPLC tryptic map of anti-CD4 IgG1 in Fig. 1[a]

Chain	#	N	C	Comments	rt (min)	280 nm	Lys-C	Fc/Fab	LC-MS	AAA	SEQ
L	1	1	7		22.10	+		Fab	+	+	+
L	2	8	28		32.72			Fab	+	+	
L	3	29	29		nf			Fab			
L	2–3	8	29	Partial cleavage	32.00			Fab	+	+	
L	4	30	59		53.39	+		Fab	+	+	
L	5	60	103		63.81	+		Fab	+	+	
L	6	104	111		30.69			Fab	+	+	
L	7	112	130		44.35		+	Fab	+	+	
L	8	131	150		67.60	+	+	Fab	+	+	
L	9	151	157		9.45		+	Fab	+	+	
L	10	158	167		12.93		+	Fab	+	+	
L	11	168	172		3.72		+	Fab	+	+	
L	12	173	187		48.56	+	+	Fab	+	+	+
L	13	188	190		3.95		+	Fab	+	+	
L	14	191	205		22.67	+	+	Fab	+	+	+
L	15	206	213		14.43		+	Fab	+	+	+

[a] Abbreviations: #, number of tryptic peptide fragment (versus N-terminus of light chain); N, sequence position of N-terminal residue; C, sequence position of C-terminal residue; rt, RP–HPLC retention time; 280 nm, + indicates that peptide contains Tyr and/or Trp and is thus predicted to have UV absorption at 280 nm; Lys-C, + indicates that peptide is predicted to be common to both tryptic and endoproteinase Lys-C maps; LC-MS, + indicates that peak was identified by LC/ESI-MS; AAA, + indicates that amino acid composition was determined by amino acid analysis of isolated peak; SEQ, + indicates that N-terminal sequence was determined by Edman degradation of isolated peak.

Table 2
Summary of characterization of the heavy chain (H) tryptic peptides in the RP–HPLC tryptic map of CM-Cys anti-CD4 IgG1 in Fig. 1[a]

Chain	#	N	C	Comments	rt (min)	280 nm	Lys-C	Fc/Fab	LC-MS	AAA	SEQ
H	1	1	39		nf			Fab			
H	1	1	39	pGlu1	66.41	+		Fab	+	+	+
H		37	39	Chymotryptic clip	22.67	+		Fab	+	+	+
H	2	40	44		3.95			Fab	+	+	
H	2	40	44	pGlu40	4.60			Fab			
H	3	45	68		54.65	+		Fab	+	+	
H	4	69	77		29.69			Fab	+	+	
H	5	78	83		43.24		+	Fab	+	+	
H	6	84	85		6.20			Fab		+	
H	7	86	103		53.39	+		Fab	+	+	
H	8	104	125		68.27	+	+	Fab	+	+	
H	7–8	86	125	Partial cleavage	78.09	+		Fab	+	+	
H	9	126	137		40.44		+	Fab	+	+	
H	10	138	151		38.75		+	Fab	+	+	
H	11	152	214		66.14	+	+	Fab	+	+	
H		152	163	Clip	50.19	+		Fab	+	+	+
H		164	214	Clip	62.63	+		Fab	+	+	+
H	12	215	217		nf			Fab	+	+	
H	13	218	218		nf			Fab	+	+	
H	12–13	215	218		4.30			Fab	+	+	
H	14	219	222		4.30			Fab	+		
H	15	223	226		3.72			Fab	+	+	
H	16	227	252		60.31			Fc	+	+	
H	15–16	223	252	Partial cleavage	59.37			Fc	+	+	
H	17	253	259		29.69			Fc	+	+	
H	18	260	278		45.61			Fc	+	+	
H	19	279	292		44.35	+	+	Fc	+	+	
H	20	293	296		nf			Fc			
H	21	297	305	Glycan a	13.81	+		Fc	+	+	+
H	21	297	305	Glycan b	14.03	+		Fc	+	+	+
H	22	306	321		62.96	+		Fc	+	+	+
H	22	306	321	β-Asp321	62.63	+		Fc	+	+	+
H	23	322	324		5.30	+	+	Fc	+	+	
H	24	325	326		nf			Fc			
H	24–25	325	330		6.91		+	Fc	+	+	
H	26	331	338		28.51		+	Fc	+	+	
H	27	339	342		6.51		+	Fc	+	+	
H	28	343	344		nf			Fc			
H	29	345	348		nf			Fc			
H	30	349	359		nf			Fc			
H	31	360	364		nf			Fc			
H	30–31	349	364		36.03			Fc	+	+	
H	32	365	374		42.42		+	Fc	+	+	
H	33	375	396		48.90	+	+	Fc	+	+	+
H	33	375	396	β-Asp388	48.56	+	+	Fc	+	+	+
H	33	375	396	Asp388	49.65	+	+	Fc	+	+	+
H	34	397	413		51.98	+	+	Fc	+	+	
H		397	411	Chymotryptic clip	57.61	+		Fc	+	+	
H	35	414	418		15.84		+	Fc	+	+	
H	36	419	420		nf			Fc			
H	37	421	443		48.11	+		Fc	+	+	
H	38	444	451		nf		+	Fc			
H	38	444	450	desLys451	28.51		+	Fc	+	+	+

[a] Abbreviations: #, number of tryptic peptide fragment (versus N-terminus of heavy chain); N, sequence position of N-terminal residue; C, sequence position of C-terminal residue; rt, RP–HPLC retention time; 280 nm, + indicates that peptide contains Tyr and/or Trp and is thus predicted to have UV absorption at 280 nm; Lys-C, + indicates that peptide is predicted to be common to both tryptic and endoproteinase Lys-C maps; LC-MS, + indicates that peak was identified by LC/ESI-MS; AAA, + indicates that amino acid composition was determined by amino acid analysis of isolated peak; SEQ, + indicates that N-terminal sequence was determined by Edman degradation of isolated peak.

crease for the modified (or 'mutated') peptide. We estimated LOD's for the oxidation of the two methionine residues in anti-CD4 in this way. Briefly, the methionine residues in anti-CD4 were quantitatively converted to methionine sulfoxide via a mild selective oxidation of native anti-CD4 with hydrogen peroxide. The oxidized anti-CD4 was then reduced, carboxymethylated, and digested with trypsin according to the method along with a control sample of the unmodified anti-CD4.

It was not obvious to us at the outset that methionine sulfoxide residues would be stable with respect to the DTT reduction step of the method (Fig. 1). The reduction of Met(O) to Met by thiol reductants is known [31] and, in fact, this redox chemistry has been employed for the reversible protection of Met side chains [32]. Moreover, Met(O) is well known to be partially converted to Met during amino acid analysis and Edman sequencing [33]. In other words, if Met(O) were to be completely or partially reduced to Met during the procedure then this type of peptide mapping method might be of little or no use for detecting Met(O) in proteins. However, the quantitative recovery of the [Met(O)^{432}]H421–443 peptide (41.3 min) and corresponding complete disappearance of the parent H421–443 peak (48.2) in the map of the oxidized anti-CD4 (Fig. 12) shows that Met(O) is not reduced to Met by this tryptic mapping method.

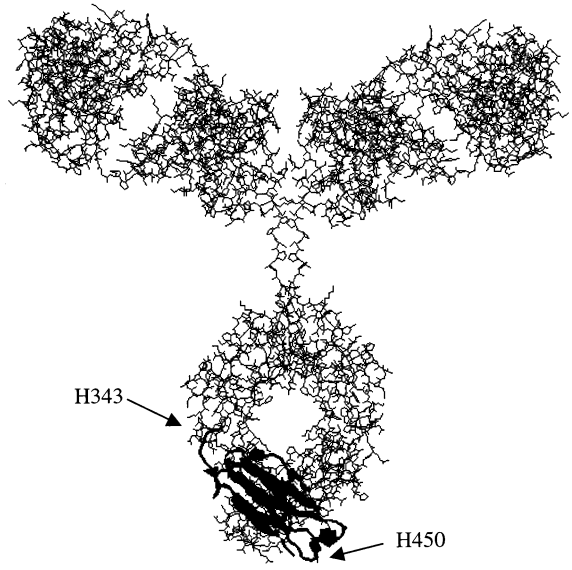

Fig. 9. Structure of an IgG1 with residues 343–450 in one heavy chain indicated by the 'ribbon' representation. This 'unreduced/undigested C_H3 disulfide-loop species' comprises a complete immunoglobulin fold. Drawing was created using the program Rasmol [2] and 'composite' IgG atomic coordinates [28].

The two methionine-containing peptide fragments in the anti-CD4 IgG1 map are H421–443 (48.2 min) and H253–259 (29.8 min). The modified form of the H421–443 peptide, [Met(O)^{432}]H421–443, appears in the map of the oxidized sample at 41.4 min as a 'doublet' or 'split peak'. The [Met(O)^{432}]H421–443 peak is a 'doublet' because oxidation of methionine to methionine sulfoxide results in a new chiral center at the sulfur atom and thus a racemic mixture of diastereomeric peptides [33]. The pair of diastereomeric peptide sulfoxides is sometimes resolved or partially resolved by RP–HPLC but not in all cases [34]. The other methionine-containing peptide, H253–259 (29.7 min), is oxidized to [Met(O)^{256}]H253–259 eluting at 25.1 min in the map of the oxidized protein (Fig. 12).

An approximate LOD for oxidation of Met^{432} to Met(O)^{432} was determined by spiking of the standard tryptic digest of anti-CD4 IgG1 (control) with 0, 2, 5, 10, 15, and 20% of the digest of the oxidized anti-CD4 IgG1 as shown in Fig. 13. From these data we estimate the LOD to be $\sim$5% oxidation of this methionine as evidenced by the small but clearly noticeable 'doublet' appearing at 41.4 min for the 5% spike. Although oxidation Met^{432} manifests itself both by the increase of a new peak for [Met(O)^{432}]H421–443 at 41.4 min and by a corresponding decrease of the parent H421–443 peak at 48.2 min, the increase of the new peak is much easier to detect for samples near the LOD. Similarly, the LOD for oxidation of Met^{256} to Met(O)^{256} as detected by the step-wise increase of the relatively small [Met(O)^{256}]H253–259 peak was determined to be $\sim$10%. The LOD for Met(O)^{256} is twice that for Met(O)^{432} simply because it happens to reside in a tryptic peptide, [Met(O)^{256}]H253–259, that is considerably shorter than [Met(O)^{432}]H421–443 and thereby absorbs less at 215 nm (the 215 nm UV response is roughly proportional to the number of

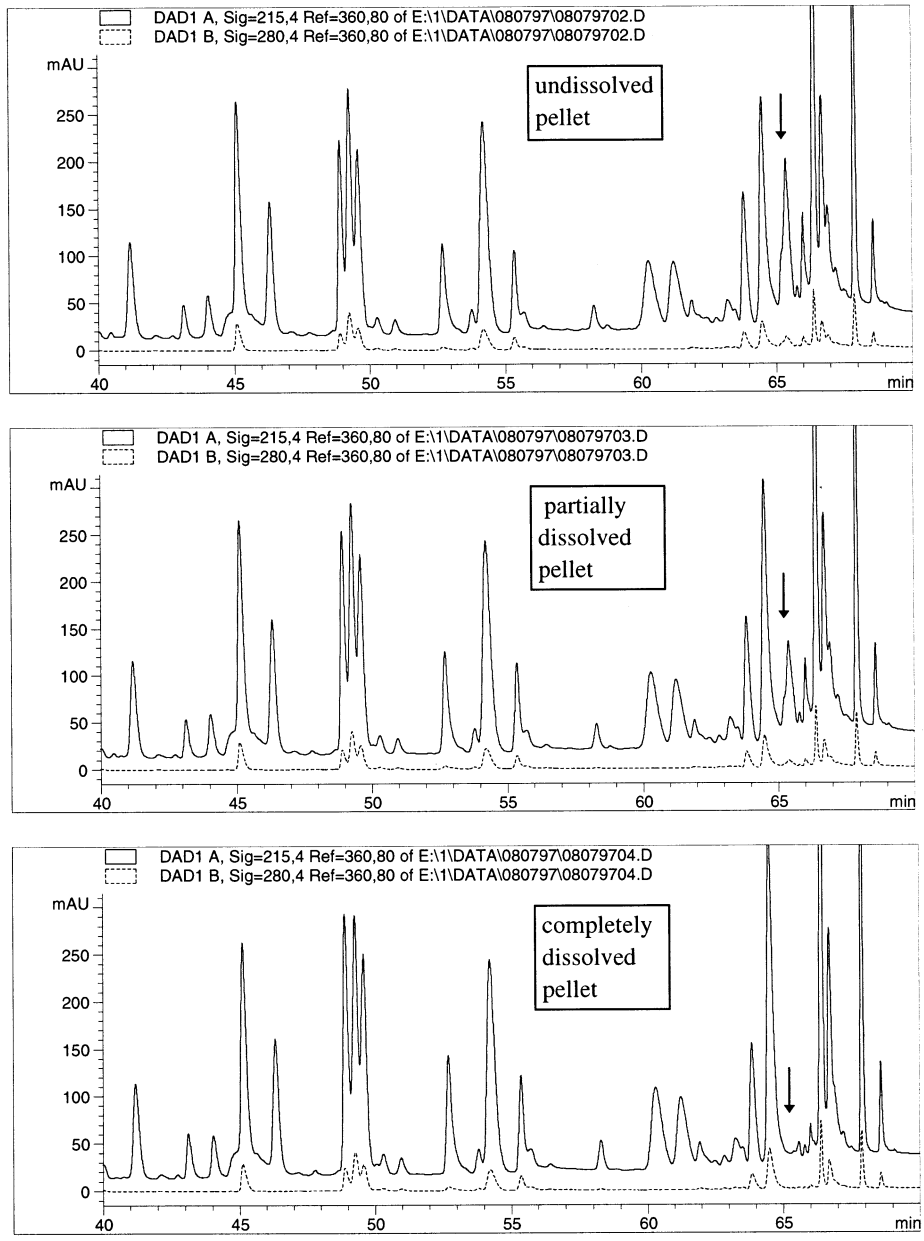

Fig. 10. Experiment showing that the peak at 65.5 min for the unreduced/undigested C_H3 disulfide-loop species (Fig. 9) is caused by incomplete dissolution of the glassy sample pellet at the very start of the procedure.

peptide bonds). In summary, this experiment determined the exact retention times and profiles for peaks in the map resulting from oxidation of the Met residues to Met(O) as well as approximate LOD values.

Not all chemical modifications of interest can be carried out in the laboratory as readily as the above selective oxidation of methionine residues. For example, a hypothetical 'mutation' at single amino acid residue by use of recombinant DNA

techniques would be difficult to justify merely for the purpose of validating its detection in a tryptic map. Thus we have taken a different approach to testing the method against such modifications (real and hypothetical) employing the use of synthetic peptides. For example, a peptide corresponding to a hypothetical Tyr^{H51}-to-Gln^{H51} mutation, [Gln^{51}]H45–68, was prepared by solid-phase synthesis and then spiked into the standard tryptic digest of anti-CD4 IgG1 as shown in Fig. 14. The only difference between the data obtained for the synthetic peptide in Fig. 14 and those for an actual Tyr^{H51}-to-Gln^{H51} mutation is that the latter would manifest itself by a concurrent de-

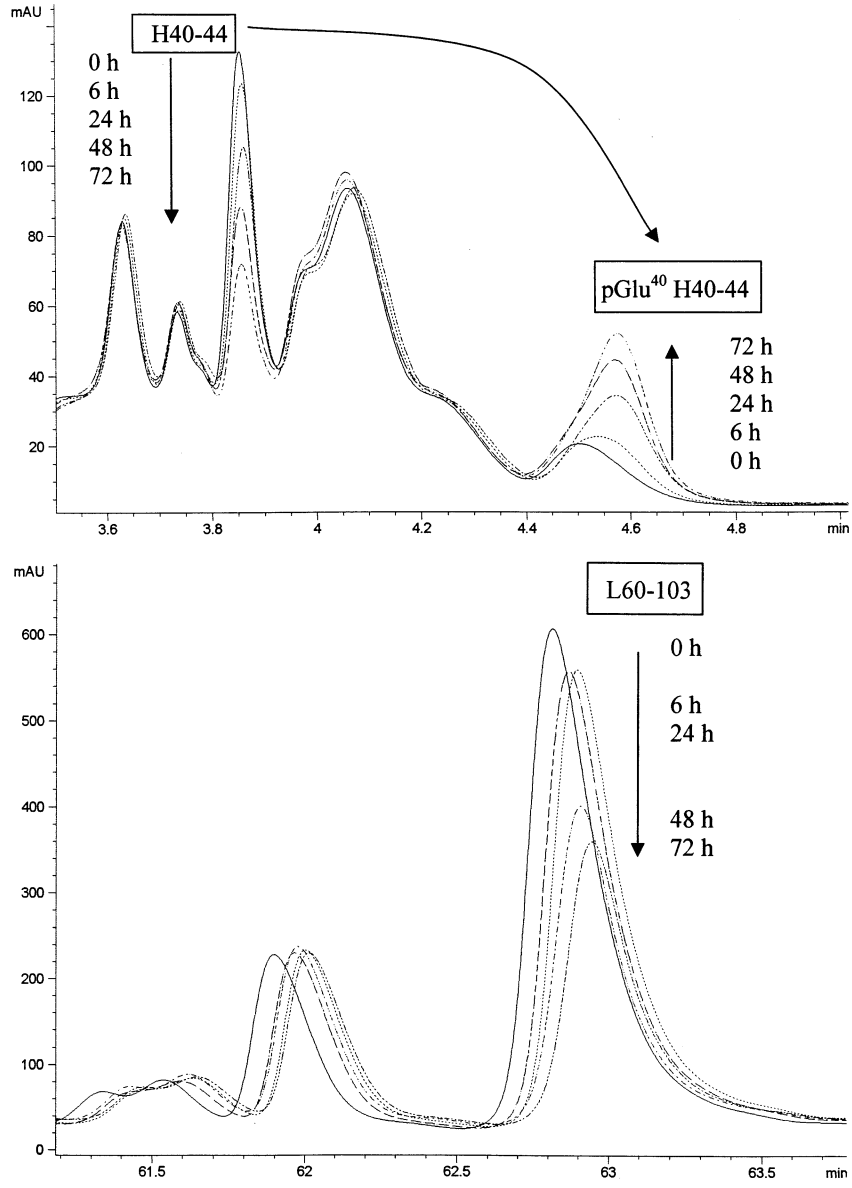

Fig. 11. Overlay plots (215 nm) of two regions of the map revealing changes to the digest upon extended storage at room temperature in polypropylene autosampler vials.

Table 3
Limits-of-detection (LOD) for structural variants of anti-RSV IgG1 as determined by tryptic mapping

Modification	Parent	rt (min)	Modified	rt (min)	LOD (mole%)
Air oxidation of Met H256	H253–259	29.7	[Met(O)^{256}]H253–259	25.1	5
Air oxidation of Met H432	H421–443	48.2	[Met(O)^{432}]H421–443	41.2/41.5	10
Tyr-to-Gln mutation at H51	H45–68	55.3	[Gln^{51}]H45–68	51.8	2
Tyr-to-Gln mutation at H411	H397–413	52.0	[Gln^{411}]H397–413	49.2	15
Nonglycosylated heavy chain	H297–305 glycans	13.8/14.0	H297–305 aglycan	15.9	5
Residual C-terminal Lys H451	H444–450	28.3	H444–451	26.8	10
Deamidation at Asn H319	H306–321	62.9	[β-Asp^{319}]H306–321	62.6	5

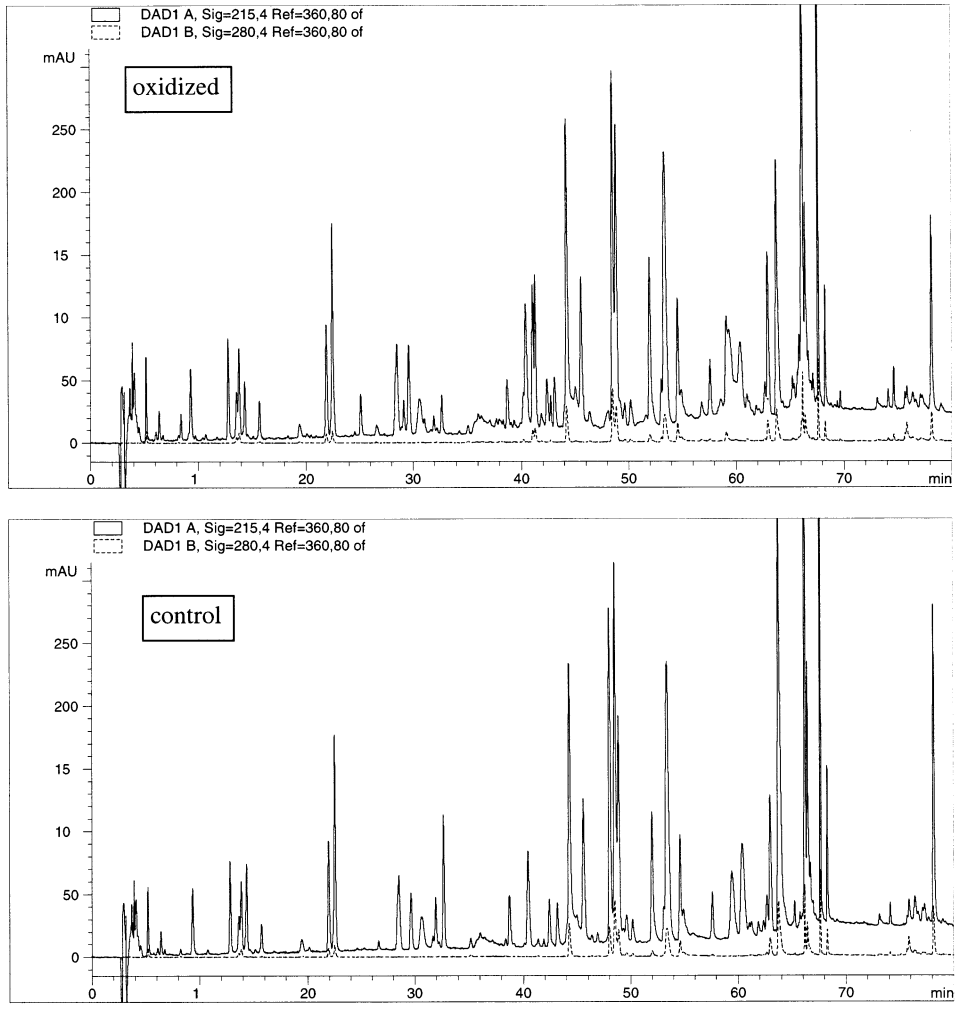

Fig. 12. Tryptic mapping of anti-CD4 following mild hydrogen peroxide oxidation to convert methionine (Met) residues to methionine sulfoxide residues (Met(O)). Detection at 215 and 280 nm.

crease in the 55.3 min peak for 'wild-type' H45–68 in addition to the increase in the [Gln[51]]H45–68 peak at 51.8 min.

Despite this slight imperfection in the synthetic peptide approach to the problem, we still obtain the desired retention time and profile for the peak

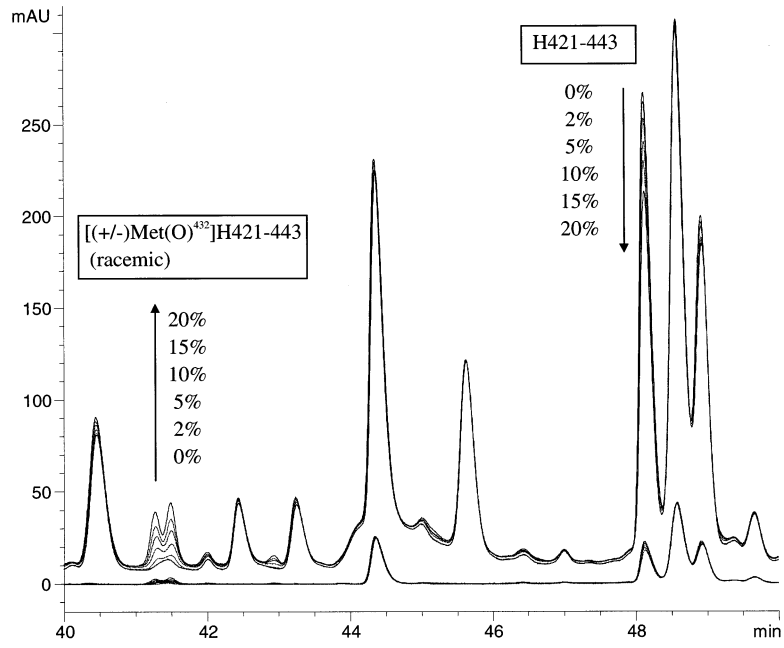

Fig. 13. Estimation of the limit-of-detection (LOD) for oxidation of methionine H432 to methionine sulfoxide. The overlayed chromatograms are of mixtures of the tryptic digest of oxidized anti-CD4 IgG1 with that of a control. The numbers refer to the percentage of the oxidized anti-CD4 IgG1 in the mixture. The [Met(O)[432]]H421–443 peptide is a racemic mixture of diastereomers (chiral sulfur atom). A LOD of 5% was estimated from these data.

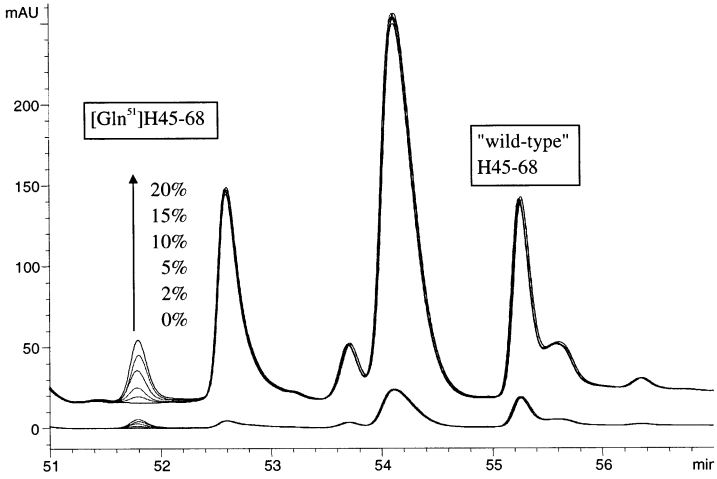

Fig. 14. Estimation of the limit-of-detection (LOD) for the hypothetical single-site mutation, Tyr[H51]-to-Gln[H51], as determined by spiking the standard tryptic digest with the specified mole percentages of the synthetic peptide, [Gln[51]]H45–68. An LOD of 2% was estimated from these data (see summary of LODs in Table 3).

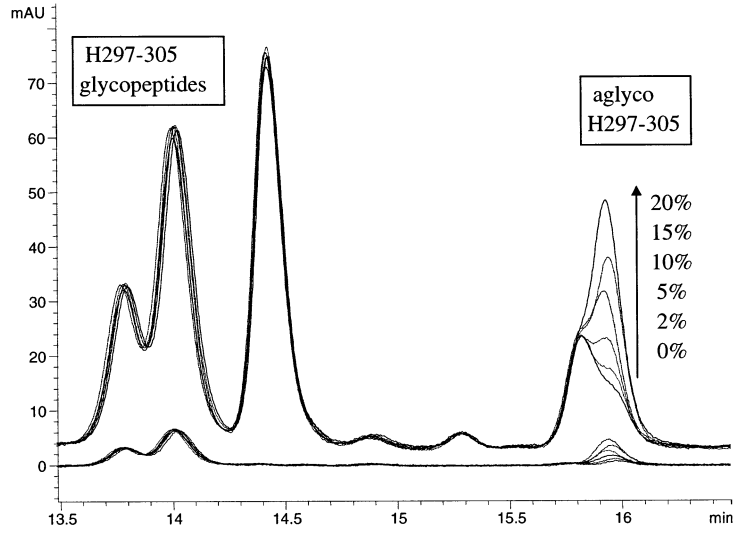

Fig. 15. Estimation of the limit-of-detection (LOD) for nonglycosylated heavy chain (NGHC), as determined by spiking the standard tryptic digest with the specified mole percentages of the synthetic peptide, H297–305 (aglyco). An LOD of 5% was estimated from these data (see summary of LODs in Table 3).

of the modified peptide and an estimate for the LOD. Based on the data in Fig. 14, an LOD of ~ 2% does not seem unreasonable for the hypothetical Tyr^{H51}-to-Gln^{H51} mutation. This LOD is relatively low simply because the peak for [Gln^{51}]H45–68 happens to occur in a favorably flat portion of the map. In contrast, we also tested the method for another hypothetical Tyr-to-Gln mutation at residue H411 by the use of a synthetic [Gln^{411}]H397–413 and estimate the LOD at ~ 15%. The much higher LOD for the hypothetical Tyr-to-Gln mutation at residue H411 is simply due to the fact that [Gln^{411}]H397–413 happens to elute as a shoulder on the descending slope of an existing peak in the map for H375–396.

We also used synthetic peptides to measure LOD's for nonglycosylated heavy-chain (NGHC) shown in Fig. 15, residual C-terminal lysine H451, and deamidation via cyclic imide of asparagine H319. The results in this section are summarized in Table 3. Note that, of all the modified tryptic fragments studied, the [β-Asp^{319}]H306–321 is the only one that normally occurs to a detectable amount in the standard (unspiked) map. It's worthwhile to point out that the presence of 20% [β-Asp^{319}]H306–321 in tryptic map is not because this residue is deamidated in the original protein reference standard. The Asn^{H319} residue is actually deamidated by the application of the method to the extent of ~ 20%. In other words, the method itself causes partial deamidation of certain asparagine residues (primarily at Asn–Gly sites) and ~ 20% [β-Asp^{319}]H306–321 is normally present in the map as a 'method artifact'. The partial deamidation of asparagine residues during the peptide mapping procedure is a well known phenomenon [30]. Deamidation at Asn^{H319}, then, is detected not by the appearance of a new peak but by the increase in an existing peak for this modified peptide. Several of the synthetic peptides happen to coelute with existing peaks in the map but [β-Asp^{319}]H306–321 is the only one tested that is actually present to begin with in the unspiked map.

The particular chemical modifications that one may want to validate the detection of in a peptide map will depend, of course, on the particular protein and expression host. The above example of nonglycosylated heavy chain, i.e. the occupancy of a glycosylation site, is relevant to the mammalian cell culture system used to produce anti-CD4. Expression in *E. coli*, on the other hand, might warrant the detection of residual N-terminal methionine [4] or incorporation of norleucine in place of methionine by mistranslation of the gene [35].

3.5. Method precision (repeatability) and its system precision component

Method precision was tested by running the entire method for five separate replicate sets of test articles, both lyophilized dosage form (Lyo) and bulk biological substance (BBS), and reference standard (Ref Std) in parallel (15 samples total). Injections were in the sequence: blank, Ref Std #1, BBS #1, Lyo #1, Ref Std #2, BBS #2, Lyo #2, Ref Std #3,... Lyo #5. System precision was measured separately by running six consecutive RP–HPLC injections from six replicate portions of a single pooled digest of the reference standard. The system precision is measured with all sample variability eliminated and is thus the instrument or system component of the overall method precision. The system repeatability is by definition, then, always less than or equal to the method repeatability. The method precision is the important limiting parameter to be evaluated in the method validation. However, it is often useful to measure the system repeatability separately in order develop system suitability criteria.

Fig. 16 shows an example of an integrated chromatogram from the system precision study juxtaposed with a scatter plot of peak area %RSD ($n = 6$) for the 44 individual integrated peaks. Relative retention times (versus H126–127 at 41.4 min) and relative peak areas (percentage of total peak area) were also compiled for these 44 peaks. Retention time precision is almost entirely a function of the particular HPLC system. These precision data were acquired with a sampling interval

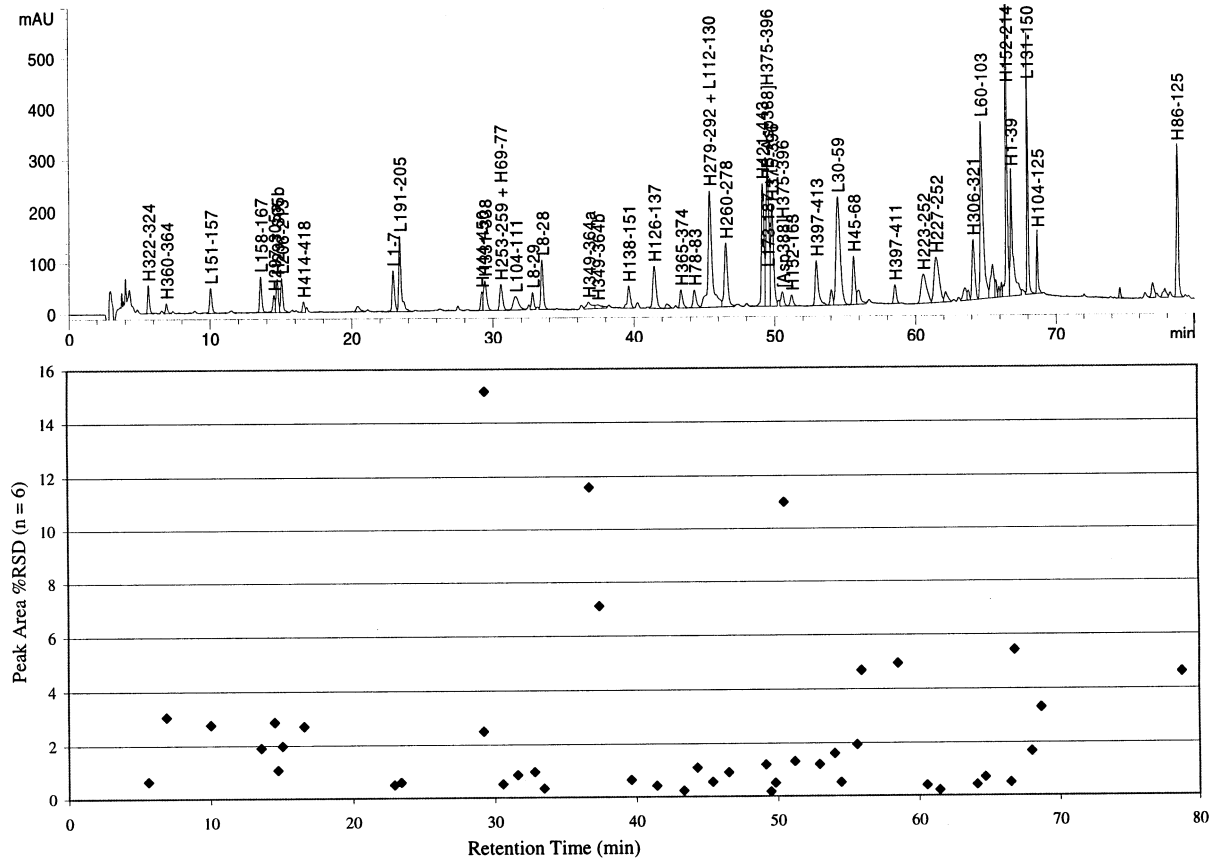

Fig. 16. System precision: Peak area %RSD ($n = 6$) by peak. Integrated chromatogram for reference standard replicate #1 in system precision study. Integration parameters (HP Chemstation): Threshold = −0.300, Area reject = 160.00, Peak width = 0.200.

of 0.64 s on a Hewlett Packard 1090M LC system equipped with the DR5 fluidics. A water blank was placed in the sequence immediately before the first of the six replicate sample injections to eliminate any 'first run' effects and thus make the operation of the six consecutive sample runs as homogeneous as possible. A water blank was run first for all analyses as a standard practice stipulated in the written method.

Average retention times (rt_avg) and standard deviations (SD) across the six replicate injections ($n = 6$) were tabulated for each of the 44 individual integrated peaks. The SD values covered the range from 0.0075 min for the broad tailing peak for the H349–364 peptide (rt_avg = 37.4 min) to 0.0011 min for the H126–137 peak (rt_avg = 41.4 min) with an average SD across all 44 peaks ($n = 44$) of 0.0041 min (0.25 s). These highly precise values are a direct measure of the gradient formation provided by the diaphragm pumping system of the HP 1090M instrument. These values are quite good when one considers that the gradient program is 105 min in duration. Often it is useful to convert absolute retention times to relative retention times in order to minimize variability due to gradual drifts in operating conditions and/or instrument performance. Relative retention times (rt_rel) were calculated for the current data set by subtracting each retention time from that of the H126–137 peak (rt_avg = 41.43 min). Note that the average SD ($n = 44$) for the rt_rel values is 0.0039 min versus the average SD ($n = 44$) of 0.0041 min found for the absolute retention time values (Table 6). This demonstrates the advantage of using relative values, which is actually very minimal in this case due to the very precise raw data from this instrument. Scatter plots of SD versus rt and SD versus rt_rel did not show any apparent trends, i.e. they were random distributions.

The parameter of peak area in Table 6 is quite a bit more complex than that of retention time. The overall system precision of all the peak areas in the map is influenced primarily by the precision of the instrument sample injection (80 μl) and the computer algorithm for integration of the peak areas (data processing software). The first of these factors, the variability in injection volume, can be effectively eliminated by converting absolute peak areas to relative peak areas (%A). In addition to these two general factors, the precision for a given peak in the map depends on its particular profile, i.e. area, width, tailing, etc. For example, note that the peak with the least precise percent area (%RSD = 14.8) is that of the H331–338 peptide eluting at 29.4 min (Fig. 16). This high variability in area for the H331–338 peak is due to its being partially fused to (i.e. coeluting with) the preceding peak for H444–445 (Fig. 16). Another example of very imprecise peaks are the pair of extremely broad and tailing peaks for the H349–364 peptide centered at $\sim$37 min (Fig. 16). The H349–364 peptide (sequence: EPQVYTLPPSRDELTK) contains a Pro–Pro bond that undergoes a slow *cis-trans* isomerization. The broadening and 'splitting' of RP–HPLC peaks as a result of this slow *cis-trans* isomerization of some Pro–Pro and Pro–Xaa bonds is well documented in the literature [36,37].

The peak with the most precise %A (%RSD = 0.60) is the largest peak in the map, that for the H152–214 peptide (the largest tryptic fragment) eluting at 66.5 min. It stands to reason that the H152–214 peak area is the most precise as it has a favorable chromatographic profile and the large area and narrowness combine to minimize errors resulting from variability in setting the base-line, i.e. setting the 'lift-off' and 'set-down' points. The scatter-plot of %RSD versus A%_avg in Fig. 17 shows a general trend for the precision of the peak area to increase with increasing peak area. This trend is due to the larger relative error (%RSD) contributed by variability in setting the base-line for smaller peaks. Although the average %RSD for %A_avg across all 44 peaks is 2.59 (Table 6), this value is 1.76 if one excludes the extremely high %RSD values for H331–338, H349–364a, H349–364b, and [Asp388]H375–396. Excluding these four peaks, an average %RSD of 1.76 is a more typical value for the map as a whole as can be seen in the scatter-plot in Fig. 17.

The method repeatability was evaluated by running five separate analyses ($n = 5$) of both lyophile and BBS versus the reference standard (15 vials total run in parallel. Exactly the same compilation

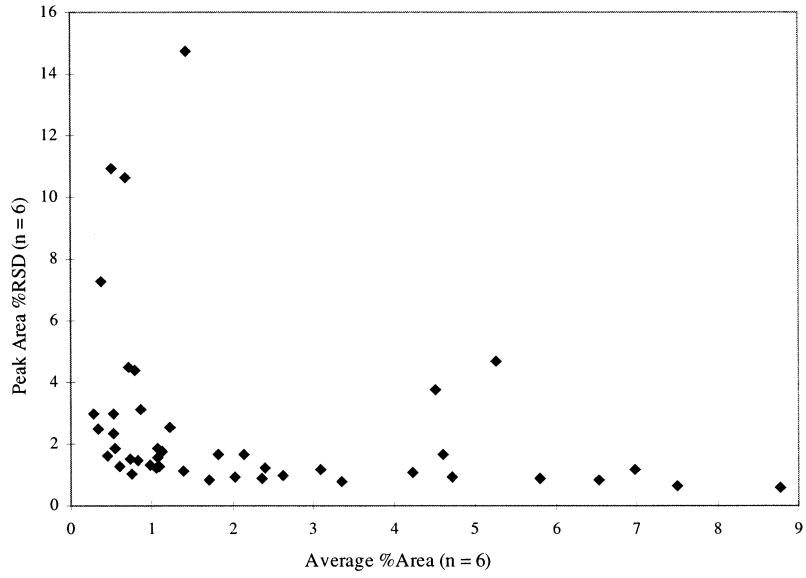

Fig. 17. Scatter-plot of the percent relative standard deviation (%RSD) versus percent area average (%*A*_avg) for the system precision peak data summarized in Table 6.

of data as was done for the system precision data in Table 5 was also done for these 15 chromatograms. Summaries for the reference standard and the lyophile and BBS test articles are shown in Tables 7–9. It is interesting to compare the 'bottom-line' average values ($n = 44$) of SD and %RSD for the reference standard from the method repeatability study in Table 7 to those from the system repeatability study in Table 5. The most obvious observation is that all four average values from the method repeatability study are higher than the corresponding values from the system repeatability study (SD: 0.0190 versus 0.0041 for rt_avg, 0.0081 versus 0.0039 for rt_rel_avg; %RSD: 11.05 versus 2.44 for *A*_avg, and 3.16 versus 2.59 for %*A*_avg). This illustrates the basic principle that method precision can never be better than its system precision component. Notice also the much greater improvement in the averages for the method precision results when relative (normalized) values are used for peak areas and retention times (%*A* and rt_rel) instead of the absolute values (*A* and rt). The use of %*A* values removes the trivial variability in the total area under the chromatogram (sum of absolute peak areas) due to minor differences in overall sample recovery. Note also that the difference between the %RSD for %*A*_avg and that for *A*_avg gives an indication of this variability in overall sample recovery. Essentially the same average results can be seen for the BBS and lyophile samples (Table 4)

Consider these method precision data in a different way. Rather than analyze the data as three separate sets of five chromatograms for reference standard, BBS, and lyophile, it is perhaps more to the point to consider the data as five replicate analyses of lyophile and BBS consisting of three chromatograms each. This is the routine performance of the method for a BBS and lyophile replicated five times. The analysis of the data for the test articles (BBS and lyophile) then involves a close visual comparison of the chromatogram to that of the reference standard to ascertain identity. In other words, do the two peptide maps or 'finger-prints' match each other within the limits of experimental error? The answer to that question for these five replicate analyses, by visual inspection, was yes (pass) for all ten test articles. Therefore, the intra-assay repeatability of the method, for BBS and lyophile, as defined by the actual acceptance criteria for the method, is established.

Having thus satisfied this element of the validation, let us go somewhat beyond what is required, for the purposes of this study, and try to make some kind of quantitative measure of how closely two peptide maps are matched based on integrated peak areas, for example. Firstly, however, it is important to point out that close human visual inspection (preferably by a highly trained analyst) is currently the very best means of comparing two peptide maps to find differences between them. Quantitative or numerical methods to analyze the data should only be used to augment the visual inspection.

For the purposes of this study, let us define the quantity, $|\Delta\%A|$, as the absolute value of the difference in percent area ($\%A$) for a peak in the map of the test article from the $\%A$ for that peak in the map of the reference standard. Average values of $|\Delta\%A|$, averaged across the five replicate analyses ($n = 5$) were tabulated for all 44 peaks. Summing these average $|\Delta\%A|$ values across the 44 peaks for the lyophile and BBS test articles gave values 7.7, and 8.4%, respectively. This value, $\Sigma|\Delta\%A|$, then, is the sum total of all 44 values for $|\Delta\_\%A|$ for the 44 integrated peaks in the test article map versus reference standard map. The value, $\Sigma|\Delta\_\%A|$, then, is a crude quantitative measure ('score') of how closely the map ('finger-print') of test article 'matches' that of the reference standard.

How much, if any, of $\Sigma|\Delta\%A|$ represents real differences between the two maps and how much is simply due to spurious differences caused by trivial variations in the machine integration? An estimate of the base-line contribution to $\Sigma|\Delta\%A|$ from such trivial errors ('noise') can be estimated, for instance, by calculating for the system precision data set by taking the difference between successive replicates of reference standard. Since these samples are replicate injections of exactly the same pooled digest the resulting value for $\Sigma|\Delta\%A|$ represents pure 'noise' absent of any real differences between the replicates. For the system repeatability data, we obtained a $\Sigma|\Delta\%A|$ value of 2.1% for the replicate pairs, $\#2-\#1$, $\#3-\#2$, $\#4-\#3$, $\#5-\#4$, and $\#6-\#5$. The same type of calculation for the separately prepared reference standard replicates from the method precision data yielded an only slightly higher value of 2.2%. Thus, a portion of $\Sigma|\Delta\%A|$ representing $\sim 2\%$ of the total peak area under the map ($n = 44$) can be attributed simply to trivial variations in the integrated peak areas ('noise'). Any portion of $\Sigma|\Delta\%A|$ above 2%, then, represents the sum of real differences between test article and reference standard maps.

It should be apparent that a single numerical score, such as $\Sigma|\Delta\%A|$, however interesting and perhaps even useful from theoretical and statistical standpoints, is nonetheless grossly inferior to human visual inspection as a means of spotting differences between two maps and judging identity. Our main purpose in defining this quantity is for use as a response factor for the fractional factorial study of method robustness in a latter section of the validation study.

Table 4
Summary of results for system precision and method precision

Experiment	Sample(s)	n	%RSD peak area[a]		SD retention time (min)[a]	
			Absolute (area)	Relative ($\%A$)	Absolute (rt)	Relative (rt_rel)
System precision:	Ref Std (pooled)	6	2.44	2.59	0.0041	0.0039
Method precision:	Ref Std	5	11.05	3.16	0.0190	0.0081
	BBS	5	6.49	3.09	0.0068	0.0067
	Lyo	5	6.84	4.02	0.0073	0.0065

[a] The percent relative standard deviation (%RSD) standard deviation (SD), for the specified number of chromatograms (n), were computed using average values (across 44 integrated peaks in Fig. 16) are for absolute and relative peak areas and retention times. Relative peak area ($\%A$) expresses the area as a percentage of the total area under the 44 peaks in the same chromatogram. The relative retention time (rt_rel) expresses the retention time as a difference versus the H126–137 peak in the same chromatogram.

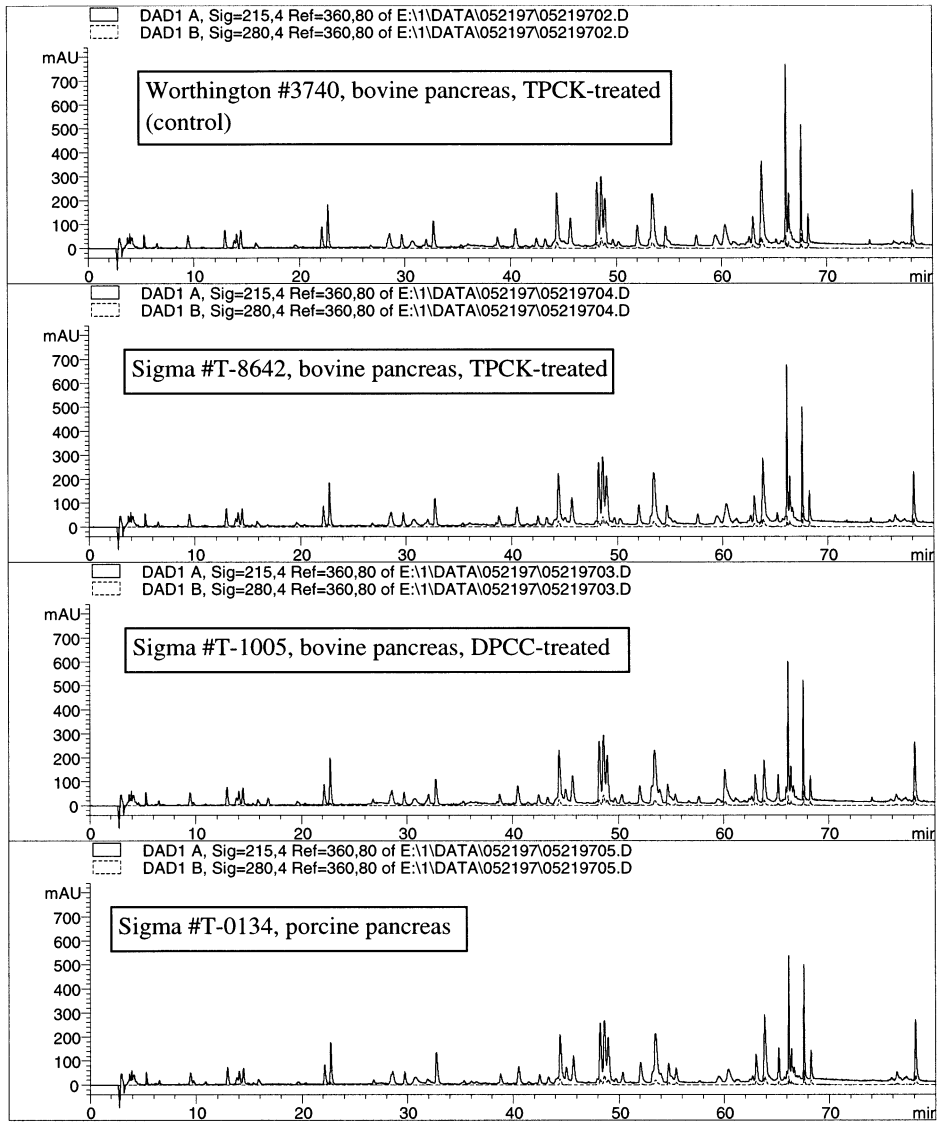

Fig. 18. Preliminary evaluation of commercial trypsin preparations as possible alternative vendors to the Worthington # 3740 trypsin specified by the method.

3.6. Ruggedness and robustness

The 'ruggedness' of the method was evaluated for day-to-day, analyst-to-analyst, laboratory-to-laboratory reproducibility, etc. An inter-laboratory study was also conducted with the quality control laboratory at the manufacturing site. The lot-to-lot consistency of the column and the trypsin were also studied and suitable alternative vendors for these two critical materials identified.

A preliminary evaluation of three different trypsin preparations from Sigma (Fig. 18) revealed that the bovine TPCK-Treated bovine pancreatic trypsin from Sigma (cat # T-8642) was the most equivalent of these to the TPCK-treated bovine pancreatic trypsin from Worthington (cat # 3740) specified by the method. Note that while we have found the Sigma # T-8642 trypsin

to the most suitable alternative trypsin preparation, the method could almost certainly be run successfully with the other two commercial trypsin preparations that were tested and probably others that were not tested. The differences between the RP–HPLC profiles for most of these different commercial trypsin preparations are relatively minor (Fig. 18). Moreover, while we have taken some pains in the development of this method to choose the highest quality commercial trypsin preparation from a reliable vendor, this reagent is not so critical to the method that it could not be replaced by an equivalent reagent if necessary.

The vendor-to-vendor and lot-to-lot variability of the trypsin was evaluated for two bottles each from two different lots of trypsin from Worthington and two bottles each from one lot of trypsin from Sigma. A pooled aliquot of carboxymethylated anti-CD4 was divided into six 0.5 ml portions and digested exactly according to the method using the above six trypsin preparations. All of the resulting RP–HPLC maps compared favorably to the historical profile and to each other. The slight differences in the RP–HPLC profiles for these different samples are very subtle and apparently related primarily to slight differences in the extent of digestion, which is directly proportional to the specific activity of the particular lot of trypsin. The method is sufficiently rugged that typical lot-to-lot variation in specific activity of the trypsin does not significantly affect the results. Thus the method requires only that 10 μl of 1 mg ml^{-1} trypsin be added to 0.5 ml of carboxymethylated anti-CD4 and makes no explicit mention of specific activity. However, the analyst should be aware of the specific activity of the trypsin employed and document the lot numbers. In summary, these experiments indicate that the method is sufficiently rugged that variability in the commercial trypsin should not affect the routine results significantly and we have identified a suitable alternative vendor.

We carried out preliminary experiments to identify a suitable alternative vendor for the Vydac 218TP54 RP–HPLC column (4.6 × 250 mm) specified by the method. Both of the 4.6 × 250 mm columns tested, a Zorbax 300SB-C8 column and a Phenomenex Jupiter C-18 column, gave suitable results. The Zorbax column was chosen as the alternative vendor for the ruggedness study. As seen in Fig. 19, the Zorbax column gives essentially the same overall RP–HPLC map but with significant specific differences. If it is ever necessary to replace the standard column with an alternative 'equivalent' column, these various differences between the two RP–HPLC profiles could be reconciled ('correlated') readily via LC/ESI-MS analysis. In other words, LC/ESI-MS allows a rapid and complete reassignment of the peak identities without repeating the N-terminal Edman sequencing and other analyses used for the initial comprehensive characterization the original map. The vendor-to-vendor and lot-to-lot variability was evaluated for two columns each from two different lots from Vydac and two columns each from one lot from Zorbax with satisfactory results.

We validated the 'robustness' of the method when challenged with the perturbations of selected key operating parameters from the reduction/carboxymethylation/digestion portion in Table 5. We tried to choose factors important to the practice of the method and ranges that are practical and appropriate. To estimate the main effects of all seven factors, a fraction (1/16) of two level full factorial design [38] was used including 'center runs' at standard conditions. The full experiment is shown Table 6. Duplicate RP–HPLC runs of the nine samples were made in the order #1–1, #2–2, #3–3, etc. followed by #1–2, #2–2, #3–2, etc. The highly multivariate nature of peptide mapping data makes it exceedingly difficult to define a 'response factor' or factors needed to interpret such statistically-designed validation experiments. We began by determining values of $\Sigma|\Delta\%A|$, as defined in the precision section, for #1–1 versus the control #2–1 run (center-point), #3–1 versus #2–1, #4–1 versus #2–1, etc. and also #1–2 versus #2–2, #3–2 versus #2–2, #4–2 versus 2–2, etc. These values of $\Sigma|\Delta\%A|$ are listed in Table 6 and range from a low of 2.2% for runs #7–1 and #9–1 to a high of 7.6% for #4–1. The low value of 2.2% is virtually the same as that found for the system precision data, i.e. it represents

almost pure 'noise' with no significant differences between the test chromatogram and the control. the high value of 7.6% is not any higher than was found for the method precision data (7.7 and 8.4%, respectively for the lyophile and BBS test articles). Thus, we conclude that the peptide map is largely unaffected by the various combinations of the seven factors A–G, within the limits tested (Table 5). This conclusion was verified by close visual inspection of the RP–HPLC data.

One interesting observation on the robustness study is that the two digests that showed the highest $\Sigma|\Delta\%A|$ responses are runs #6 and #3, with a combination of factors E–G, substrate/enzyme ratio, digestion temperature, and digestion time, that would be predicted to give the greatest and least extents of proteolysis, respectively (Fig. 20). The characteristic changes in the RP–HPLC tryptic map as a function of the extent of digestion were studied in considerable detail early in the development and optimization of this method. The peaks in the map that change with extended digestion are those associated with very slow tryptic cleavages ('partial' cleavages) and the minor 'chymotryptic-like' activity that only begins to assert itself after the tryptic cleavage is far ad-

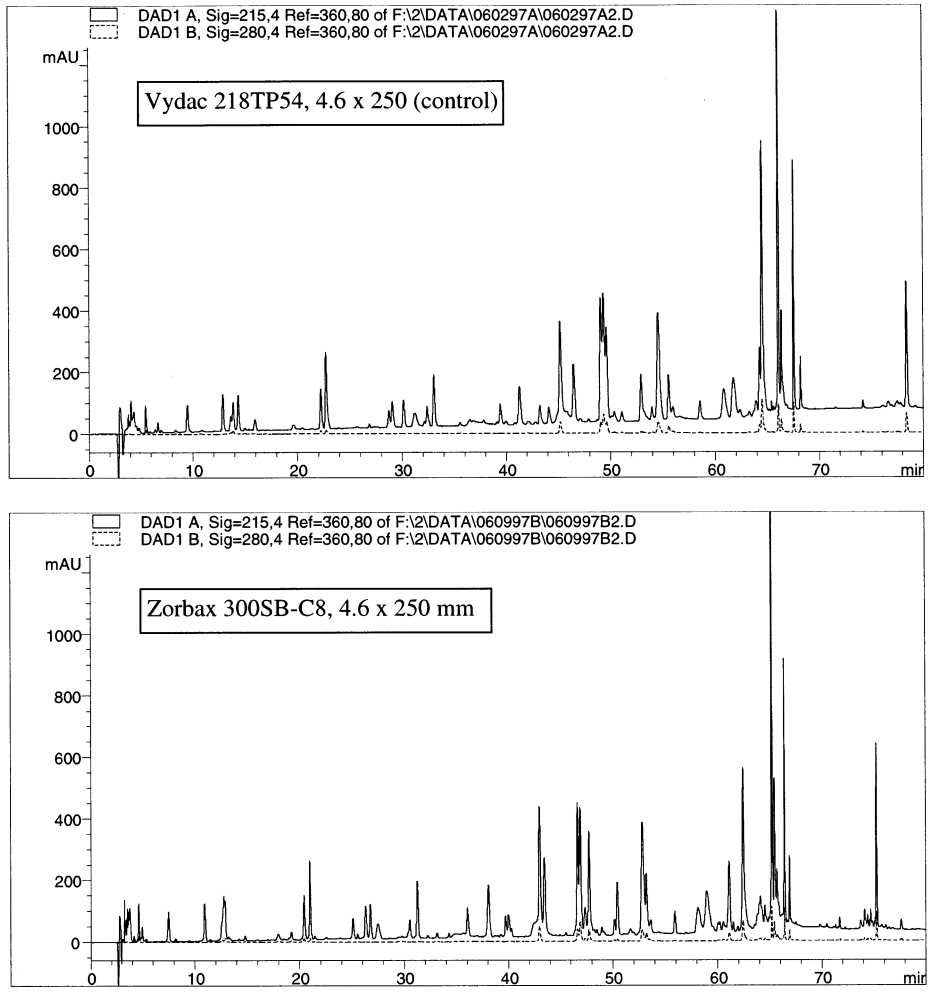

Fig. 19. Preliminary evaluation of a Zorbax RP–HPLC column as a possible alternate vendor to the Vydac column specified by the method.

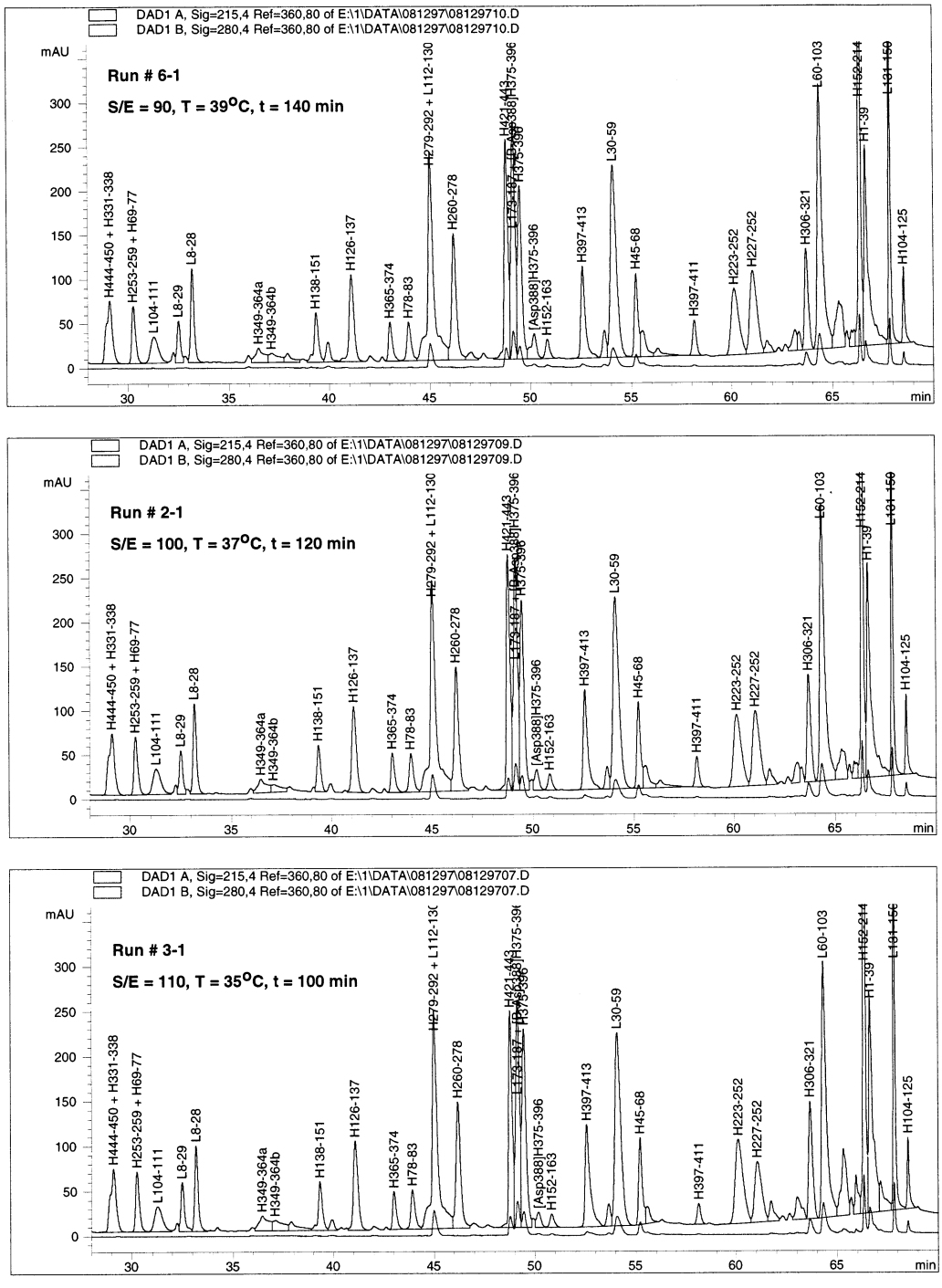

Fig. 20. Integrated chromatograms for Runs # 1, # 2 (center-point), and # 4 from fractional factorial robustness study in Table 6. Note differences in extent of trypsin digestion for the very slowly-cleaved Lys^{H226}–Thr^{H227} bond as seen by the relative peak areas for H223–252 (60.3 min) and H227–252 (61.2 min).

Table 5
List of factors and ranges for ruggedness study

Factor	Unit	Low–high (%)
A: dithiothreitol	μmol	22.5–27.5 (±10%)
B: sodium idoacetate	μmol	54–66 (±10%)
C: time of reduction	Min	20–40 (±30%)
D: time of alkylation	Min	30–50 (±20%)
E: substrate/enzyme (weight/weight)		90–110 (±10%)
F: temperature of digestion	°C	35–39 (±5%)
G: time of digestion	Min	100–140 (±20%)

vanced. The best gauge of the extent of digestion for humanized antibodies, such as anti-CD4, is the ratio of peak areas for H223–252 (60 min) and H227–252 (59 min) arising from the very slow cleavage of the Lys^{226}–Thr^{227} bond. The most prominent 'chymotryptic' clip in the map is at the Leu^{410}–Tyr^{411} bond in the H397–413 peptide (52 min) to generate the H397–411 peak (58 min). These two cleavages are sensitive indicators of the extent of digestion.

The chromatograms in Fig. 20 also illustrate the limitations of machine integration for certain 'problem peaks' that were alluded to earlier. Note the inconsistent results for L8–29 (32 min), the extremely broad H349–346 peaks (36 min), H138–152 (39 min), and others. The troublesome and laborious nature of peak integration for 'non-ideal' and complex chromatographic data, such as this peptide mapping data, has given us impetus to look into alternative modes of data processing. Recently, we have begun to explore the use of principle component analysis (PCA) for analysing the unintegrated chromatographic raw data itself [39–41]. The chromatogram is treated in a very general statistical sense as a data set of n points, where n is the number of samplings of intensity across the chromatogram. Values for n of several

Table 6
Fractional factorial ruggedness study with respect to factors A–G in Table 5[a]

Run	A	B	C	D	E	F	G	$\Sigma\lvert\Delta\ \%A\rvert$	PCA t [1]	PCA t [2]
9–1	22.5	54	20	50	110	39	100	2.2	695.7	−27.2
9–2	22.5	54	20	50	110	39	100	2.6	116.4	31.0
6–1	27.5	54	20	30	90	39	140	5.3	440.8	−21.6
6–2	27.5	54	20	30	90	39	140	5.0	−407.4	−2.7
3–1	22.5	66	20	30	110	35	140	3.4	471.0	204.0
3–2	22.5	66	20	30	110	35	140	3.5	−643.9	123.6
4–1	27.5	66	20	50	90	35	100	7.6	301.8	68.2
4–2	27.5	66	20	50	90	35	100	7.4	−464.5	57.0
5–1	22.5	54	40	50	90	35	140	3.6	230.8	−445.3
5–2	22.5	54	40	50	90	35	140	3.8	−400.4	20.9
7–1	27.5	54	40	30	110	35	100	2.2	619.3	27.3
7–2	27.5	54	40	30	110	35	100	3.0	−209.4	66.2
8–1	22.5	66	40	30	90	39	100	3.0	560.2	2.6
8–2	22.5	66	40	30	90	39	100	4.3	37.9	38.7
1–1	27.5	66	40	50	110	39	140	5.8	−35.1	−80.9
1–2	27.5	66	40	50	110	39	140	5.4	−775.1	−81.7
2–1	25.0	60	30	40	100	37	120	0	183.7	73.5
2–2	25.0	60	30	40	100	37	120	0	−721.8	−53.5

[a] Factors A–G are defined in Table 5. Response values are given for the integrated peak areas as $\Sigma|\Delta\%A|$, i.e. sums ($n = 44$ peaks) of absolute differences versus the control runs, 2–1 and 2–2, respectively. Response values from the principal component analysis (PCA) of the chromatographic raw data ($n = 3188$ points per 85 min chromatogram, i.e. 0.63 Hz sampling rate) are given as projections on the two principal component axes, t [1] and t [2], for the data as plotted in Fig. 21. See text for explanations of these response factors.

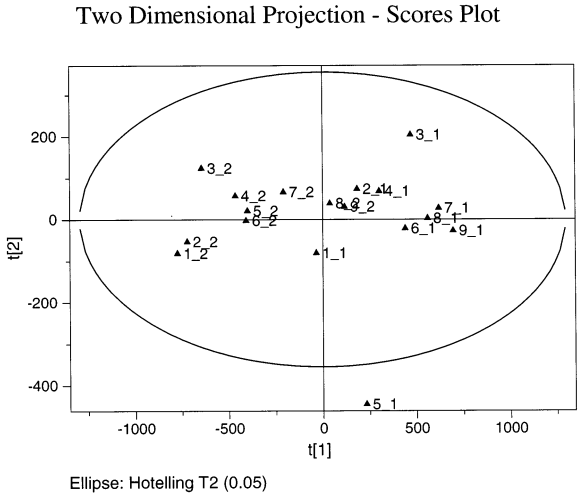

Fig. 21. Ellipse plot (95% confidence limits) of ruggedness data in Table 6 on the two principal component axes (PCA).

thousand can readily be processed by the modern software on a 'Pentium-class' microcomputer. This permits the use of the original raw data, or a slightly reduced form of it, eliminating the need to integrate peaks. This type of multivariate approach, which has been most extensively applied to spectroscopic data, may have potential benefits for chromatographic and other types of multichannel data.

Preliminary results of our principal component analysis of the ruggedness data are summarized in Table 6 and Fig. 21. The variability of the ruggedness data was not very much greater than that for the method precision data plotted onto the same space. Thus we confirmed that the method is quite robust within the limits in Table 5. Note that the first duplicate set of runs #1–1 through #9–1 form a distinct cluster of points separated along the first principal component axis from those for the second duplicate set of runs #1–2 through #9–2. According to the PCA, then, the largest source of variability for these samples was probably from a slight 'instrumental drift' in the data acquisition over the course this automated set of injections. In fact, this high sensitivity of PCA to slight instrumental changes is a problem with this approach [39]. The main advantage of this type of multivariate approach, aside from its versatility, is its speed and ease of application for complex data.

4. Conclusions

The rapid peptide mapping of many samples in parallel is now much more practical. For example, the 15 samples for the method precision study above were prepared in parallel by a single analyst within 1 working day. The RP–HPLC step is, of course, fully automated. The carboxymethylated protein and the tryptic digest can be stored frozen at −70°C and a 'frozen archive' can thus be maintained and samples readily retrieved and reanalyzed upon demand.

Some subtle changes to the tryptic digest upon standing in the autosampler, such as the slow partial conversion of N-terminal Gln to pGlu, are practically unavoidable. However, for a validated method, it is advisable to study the stability of the digest and account for any observed changes.

As observed in this study, and independently [20], the C_H3 disulfide-loop region in a humanized IgG1 is extremely reluctant to unfold. As a consequence of this disulfide remaining intact, this entire immunoglobulin domain emerges intact from the trypsin digestion; i.e. it is not proteolysed. In our hands, the incomplete reduction of this disulfide resulted from incomplete dissolution of the sample pellet in the denaturing buffer (6 M guanidine buffer). Once this potential problem is understood it can easily be prevented [27].

Acknowledgements

Jacob Bongers thanks Thomas Schreitmüller of F. Hoffmann-La Roche AG, Maureen Costello of Copernicus Gene Systems, William S. Hancock of Hewlett Packard, and Andrew Guzzetta of Scios for helpful discussions.

References

[1] R.L. Garnick, N.J. Solli, P.A. Papa, Anal. Chem. 60 (1988) 2546–2557.

[2] M.M. Federici, Biologicals 22 (1994) 151–159.
[3] S.A. Carr, M.E. Hemling, M.F. Bean, G.D. Roberts, Anal. Chem. 63 (1991) 2802–2824.
[4] J.J. Dougherty Jr., L.M. Snyder, R.L. Sinclair, R.H. Robins, Anal. Biochem. 190 (1990) 7–20.
[5] E.R. Hoff, R.C. Chloupek, Methods Enzymol. 271 (1996) 51–68.
[6] M.D. Jones, L.A. Merewether, C.L. Clogston, H.S. Lu, Anal. Biochem. 216 (1994) 135–146.
[7] D. Allen, R. Baffi, J. Bausch, J. Bongers, M.A. Costello, J. Dougherty Jr., et al., Biologicals 24 (1996) 255–275.
[8] K. Kannan, M.G. Mulkerrin, M. Zhang, R. Gray, T. Steinharter, K. Sewerin, et al., J. Pharm. Biomed. Anal. 16 (1997) 631–640.
[9] M.W. Dong, Adv. Chromatogr. 32 (1992) 21–51.
[10] M.W. Dong, A.D. Tran, J. Chromatogr. 499 (1990) 125–139.
[11] C.R. Bebbington, K. Lambert, in: F. Brown, A.S. Lubinecki (Eds.), Genetic Stability and Recombinant Product Consistency, Dev. Biol. Stand., Karger, Basel, 1994, pp. 183–184.
[12] R.J. Harris, A.A. Murnane, S.L. Utter, K.L. Wagner, E.T. Cox, G.D. Polastri, et al., Bio/Technol. 11 (1993) 1293–1297.
[13] R. Newman, J. Alberts, D. Anderson, K. Carner, C. Heard, et al., Biotechnology 10 (1992) 1455–1460.
[14] N.M. Meltzer, G.I. Tous, S. Gruber, S. Stein, Anal. Biochem. 160 (1987) (1987) 356–361.
[15] R. Knecht, J.-Y. Chang, Anal. Chem. 58 (1986) 2375–2379.
[16] J. Mozdzanowski, J. Bongers, K. Anumula, Anal. Biochem. 260 (1998) 183–187.
[17] J. Bongers, J.J. Cummings, J. Mozdzanowski, manuscript in preparation.
[18] S.C. Gill, P.H. von Hippel, Anal. Biochem. 182 (1989) 319–326.
[19] K.R. Anumula, S.T. Dhume, Glycobiology 8 (1998) 685–694.
[20] D.A. Lewis, A.W. Guzzetta, W.S. Hancock, M. Costello, Anal. Chem. 66 (1994) 585–595.
[21] R.J. Harris, K.L. Wagner, M.W. Spellman, Eur. J. Biochem. 194 (1990) 611–620.
[22] P. Rao, A. Williams, A.A. Baldwin-Ferro, E. Hanigan, D. Kroon, M. Makowski, et al., BioPharmacy 4 (1991) 38–43.
[23] D.J. Kroon, A. Baldwin-Ferro, P. Lalan, Pharm. Res. 9 (1992) 1386–1393.
[24] J.W. Ellison, B.J. Berson, L.E. Hood, Nucleic Acids Res. 10 (1982) 4071–4079.
[25] G.M. Edelman, B.A. Cunningham, W.E. Gall, P.D. Gottlieb, U. Rutishauser, M.J. Waxdal, Proc. Natl. Acad. Sci. USA 63 (1969) 78–85.
[26] G.D. Roberts, W.P. Johnson, S. Burman, K. Anumula, S.A. Carr, Anal. Chem. 67 (1995) 3613–3625.
[27] R. Sayle, E.J. Milner-White, Trends Biochem. Sci. (TIBS) 20 (1995) 374–376.
[28] E.A. Padlan, Mol. Immunol. 31 (1994) 169–217.
[29] G.N. Abraham, D.N. Podell, Mol. Cell. Biochem. 38 (1981) 181–190.
[30] J. Bongers, E.P. Heimer, T. Lambros, Y.-C. Pan, R.M. Campbell, A.M. Felix, Int. J. Pept. Protein Res. 39 (1992) 364–374.
[31] R.A. Houghten, C.H. Li, Method Enzymol. 91 (1983) 549–559.
[32] M.A. Rogero, C. Servis, G. Corradin, FEBS Lett. 408 (1997) 285–288.
[33] N. Brot, H. Weissbach, Arch. Biochem. Biophys. 223 (1982) 271–281.
[34] R.G. Keck, Anal. Biochem. 236 (1996) 56–62.
[35] G. Bogosian, B.N. Violand, E.J. Dorward-King, W.E. Workman, P.E. Jung, J.F. Kane, J. Biol. Chem. 264 (1989) 531–539.
[36] J. Jacobson, W. Melander, G. Vaisnys, C. Horvath, J. Phys. Chem. 88 (1984) 4536–4542.
[37] J.C. Gesquiere, E. Diesis, M.T. Cung, A. Tartar, J. Chromatogr. 478 (1989) 121–129.
[38] P.D. Haaland, Experimental Design in Biotechnology, Marcel Dekker, 1989, p. 37.
[39] G. Malmquist, R. Danielsson, J. Chromatogr. 687 (1994) 71–88.
[40] G. Malmquist, J. Chromatogr. 687 (1994) 89–100.
[41] K.R. Lee, J. Bongers, B.H. Jones, S. Burman, manuscript in preparation.