Reprint from

RECENT ADVANCES IN DOPING ANALYSIS (4)

W. Schänzer
H. Geyer
A. Gotzmann
U. Mareck-Engelke
(Editors)

Sport und Buch Strauß, Köln, 1997

P. LAIDLER, D.A. COWAN, R.C. HIDER, A.T. KICMAN:
Qualitative Analysis of hCG for Confirmatory Purposes
In: W. Schänzer, H. Geyer, A. Gotzmann, U. Mareck-Engelke (eds.) Recent advances in doping analysis (4). Sport und Buch Strauß, Köln, (1997) 285-304

P. Laidler[1], D. A. Cowan[1], R. C. Hider[2] and A. T. Kicman[1]

Qualitative analysis of hCG for confirmatory purposes

[1]Drug Control Centre, King's College London, London, England
[2]Department of Pharmacy, King's College London, London, England

Abstract

Analysis of untreated hCG revealed signals corresponding to the molecular ions of the intact hCG heterodimer and both its non-covalently linked subunits. Due to carbohydrate heterogeneity, the peaks are broad which makes accurate mass assignment, and consequently identification, difficult. The mass spectrum generated by MALDI-TOF MS, following the tryptic digestion of reduced and S-carboxymethylated hCG, contained mass signals corresponding to peptides covering 59 % and 52 % of the α- and β-subunits respectively. This was sufficient to provide unambiguous identification of hCG. The validation of this conclusion was achieved by searching peptide-mass databases with experimentally determined masses. Peptides corresponding to only 21 % and 19 % of the α- and β-subunits, respectively, were required to identify hCG as the protein best matched to the experimental data.

Introduction

Male athletes may administer hCG either to stimulate testosterone production or to prevent testicular atrophy during chronic anabolic-androgenic steroid abuse [1]. Consequently the administration of hCG is banned by the IOC. Immunometric assays incorporating highly specific antibodies are sufficiently sensitive to detect heterodimeric hCG in urine collected after hCG administration. The specificity of the antibodies is assessed by cross-reaction studies with molecules which exhibit similar biological properties or possess structures in common with the analyte of interest, *e.g.* the other glycoprotein hormones FSH, LH and TSH, free hCGβ and hCGβ core fragment (hCGβcf). Such methodological validation characterizes the specificity of the assay only by the process of excluding likely interferents.

In contrast, mass spectrometry (MS) of a protein may immediately yield specific information; in ideal circumstances showing the molecular ion of the analyte or, in the case of hCG, ions corresponding to the heterodimer and its individual subunits. Traditional mass spectrometric

techniques are incompatible with the analysis of large macromolecules due to their inability to produce intact, isolated molecular ions in the gas phase. However, the development, in particular, of two soft ionization techniques, electrospray ionization (ESI) [2] and matrix-assisted laser desorption ionization (MALDI) [3] which facilitate the conversion of solution or solid crystalline-phase macromolecules into gas-phase intact molecular ions, has made the analysis of hCG by MS possible [4]. Although molecular weight data can provide convincing evidence for the identification of a protein, this may be ambiguous due to the close structural homology between a series of proteins, *e.g.* the glycoprotein hormones LH, FSH, TSH and hCG. However, unequivocal identification of a protein can be achieved by peptide mapping. Peptide mapping is based on the analysis of a pattern or 'map' created by the separation of peptide fragments resulting from the hydrolysis of a protein [5]. All amino acids, with the exception of Leu and Ile, have different molecular masses. Consequently, the masses of the peptides produced by proteolytic cleavage are dependent on both the specificity of the enzyme/reagent used to perform the cleavage and the primary sequence of the polypeptide. The endoproteinase trypsin cleaves specifically only on the carboxy-terminal side of arginine and lysine residues except if they are adjacent to the amino acid proline [5]. MS analysis of the resulting fragments generates a mass spectrum containing a series of characteristic ions which is sufficiently discriminating to allow unequivocal identification of the analysed protein. It has been reported previously that peptide-mass information is as discriminating as linear peptide sequences for the identification of proteins [6, 7]. Therefore, despite the structural homology of the glycoprotein hormones, minor variations in primary sequence between each hormone will result in a characteristic peptide map that will allow discrimination between the hormones. Here we describe the use of MALDI time-of-flight (TOF) MS analysis of untreated hCG and a tryptic digest of hCG in the discriminatory analysis of hCG. Such a strategy can be considered ideal for the unequivocal identification of hCG. Furthermore, we report the use of peptide-mass database searching as a means to validate the discriminatory value of a mass spectrum generated by MS of a tryptic digest hCG (see Figure 1).

To confirm the uniqueness of the 'peptide-mass map' to the analysed protein, experimentally determined masses may be used to search peptide mass databases. These databases contain the masses of all the possible peptides that would be produced by theoretical fragmentation, according to specific enzyme or chemical cleavage rules [5], of the sequences present in existing protein databases. A number of research groups have developed programs which locate the protein sequences, or nucleotide sequences encoding for proteins, that are best matched to the set of experimentally determined masses [6-10]. The identity of a protein can be confirmed by searching the databases with as few as two, typically three to five, experimentally derived peptide masses [6, 7].

Experimental

Preparation of a tryptic digest of derivatized hCG

Desalted pharmaceutical hCG (Profasi; Serono Laboratories, Welwyn Garden City, UK) was reconstituted in 65 μL of denaturing buffer (8 M urea, 0.36 M Tris, 3 mM ethylene diamine tetracetic acid (EDTA), pH 8.6) to separate the subunits. Following denaturation, the disulphide bonds were reduced by mixing the sample with 10 μL of 1 M dithiothreitol and incubating for 30 min at 37° C. The resulting free thiols were acetylated by adding 25 μL of freshly prepared 1 M iodoacetic acid in 1 M NaOH. Reduction of the disulphide bonds was deemed essential, as digestion of the non-reduced subunits results in two large tryptic cores which are of little diagnostic value. S-carboxymethylation of the free thiols was performed to prevent reformation of incorrect disulphide pairs which would have complicated mass assignments. Following 30 min in the dark at room temperature, the sample was diluted to 1 mL with distilled deionized water and desalted by centrifugal ultrafiltration. The sample was transferred to a Centricon-10 micro-concentrator containing a 10 kDa molecular mass cut-off membrane (Amicon Ltd., Stonehouse, UK). The micro-concentrator was spun at 2000 *g* in a fixed angle rotor (45°) for 1 h to remove excess reagents. The reduced and S-carboxymethylated hCG (RCM-hCG) retained on the filter was washed by adding 1 mL of distilled deionized water to the microconcentrator and centrifuging for a further 1 h under the same conditions. The RCM-hCG was recovered into a retentate cup by inverting the microconcentrator (after removing the filtrate cup) and centrifuging for 10 min. Trypsin, 20 μL (Type XIII from bovine pancreas; Sigma, Poole, UK) (100 mg/L) in 50 mM NH_4HCO_3 (pH 8.5), treated previously with L-1-(tosylamido)-2-phenylethyl chloromethyl ketone (TPCK) a specific inhibitor of α-chymotrypsin activity [5], was added to 100 μg of RCM-hCG. The digest was incubated at room temperature for 24 h; a second 20 μL aliquot of trypsin being added after 4 h to give a final protein:protease ratio of 25:1 (w/w). Following digestion, the sample was lyophilized exhaustively and then stored desiccated at +4° C until analysis.

Mass spectrometry

Samples were reconstituted to a concentration between 0.1-1 μg/μL in 0.1% trifluoroacetic acid (TFA). A 1 μL aliquot was then mixed with 1 μL of either sinapinic acid or α-cyano-4-hydroxycinnamic acid matrix solution (10 g/L in 70 % acetonitrile/30 % water). A 1 μL aliquot of this mixture was applied to the target of a sample slide, allowed to crystallize at room temperature then analysed using a LASERMAT 2000 MALDI-TOF mass spectrometer

(Thermo Bioanalysis Hemel Hempstead, UK). The spectra were generated by summing the signals from 10-30 laser shots.

Mass assignment to peaks in the spectra is by computer-controlled determination of the peak centroids following either internal or external calibration of the instrument. The spectrum of untreated hCG was calibrated externally using data generated from a separate analysis of cytochrome c (horse heart). Calibration of the mass/charge axis of the spectrum resulting from analysis of the tryptic digestion of RCM-hCG was initially performed using the instrument's own default calibration. Following computer-controlled mass assignment, signals corresponding closely to the calculated masses of two fragments expected from the digestion were selected as calibrants. These masses were corrected, then used to perform an accurate two point internal calibration.

Peptide fragment nomenclature

Peptide masses were calculated manually using average isotopic masses and values were confirmed using a CH109 chemical formulae calculator (St. Andrews Electronics, Fife, Scotland). Fragments were designated according to the subunit from which they were generated, the digesting enzyme, T for trypsin and numerically from the amino terminal, *e.g.* tryptic fragments of the α-subunit are designated αT1, αT2, αT3 etc. The nomenclature was extended to include partial cleavage products, where '+' indicates incomplete digestion of a potential cleavage site between adjacent peptides, and fragments showing cleavages not expected from the original enzyme, *e.g.* αT1C1 and αT1C2 are peptides produced following an α-chymotryptic-like cleavage of the tryptic peptide αT1. In addition, peptide βT5 is cleaved by an enzyme, designated G, termed gonadotropin β-subunit nicking enzyme [11].

Peptide-mass database searches

Experimentally determined fragment masses were used to search the MOWSE (Molecular weight search) database [7] and the facilities of the Computational Biochemistry Research Group (CBRG) in Switzerland [8]. Both facilities are accessible on the InterNet via e-mail at mowse@dl.ac.uk and cbrg@inf.ethz.ch respectively. Information about the databases, the available search facilities and procedures for performing a search may be accessed by sending the message, help or help all, to the e-mail addresses given above.

Searches were performed using masses, determined experimentally, corresponding to the peptide signals that matched with the calculated masses of fragments expected from total proteolytic cleavage of RCM-hCG. Entered masses are recognized as the unprotonated fragments and both databases interpret cysteine residues as free thiols. Therefore, raw signal data was adjusted accordingly. However, the CBRG facility allows the user to define the mass of chemically modified amino acids, *e.g.* the cysteine residue mass was defined as 161.1 Da, *i.e.* the mass of the carboxymethylated derivative of cysteine. This enabled the discrimination of the searches to be increased. The only other parameter that must be specified to run a search is the protease or the reagent used for cleavage, *i.e.* trypsin. Additional information that limits the search to a restricted molecular weight range, reduces the influence of partially cleaved fragments and/or reduces the acceptable mass accuracy error of experimental mass determinations, may be input in order to increase the discrimination of MOWSE searches. Matches of the theoretically derived database peptides was limited to the peptides whose masses were within ± 1 Da of the experimental data (see Tables 4 and 5). Furthermore, matches with peptides derived by incomplete or partial cleavage of the database sequences were given the lowest possible significance because only masses resulting from complete digestion of RCM-hCG were used to search the database.

Results and discussion

MALDI-TOF MS of hCG, purified by size-exclusion chromatography on Sephadex G-100, produces a mass spectrum that is potentially unique to hCG (Figure 2). The spectrum shows peaks corresponding to the protonated molecules of both the intact hCG heterodimer (*m/z* 35 140) plus the free α- (*m/z* 13 408) and β-subunits (*m/z* 21 446). Additional signals at *m/z* 17 668 and 6619 are indicative of the doubly charged species of the heterodimer and free α-subunit respectively, whereas the peak at *m/z* 69 545 represents the singly charged hCG dimer and the signal at *m/z* 48 142 presumably corresponds to a non-covalent complex formed between hCG and a free α-subunit.

The ability to detect both the heterodimer and its composite subunits enhances the discriminatory power of an identification using molecular weight information alone. Furthermore, the presence of a signal corresponding to the hCG heterodimer is indicative of the presence of molecules capable of eliciting a biological response. However, the microheterogeneous nature of hCG, due primarily to variations in glycosylation, and in particular sialic acid content, results in broad peaks which makes accurate molecular weight determination extremely difficult. Because individual glycoforms cannot be resolved masses are assigned to the centroids of these broad peaks. Consequently, the unambiguous

identification of hCG using this approach, due to its homology with the structurally related glycoprotein hormones LH, FSH and TSH [12], is difficult.

Unequivocal identification can be achieved by peptide mapping [6-10, 13, 14]. Despite the structural homology between the glycoprotein hormones, minor variations in primary sequence will result in mass spectra containing characteristic fragment signals that allow discrimination between the hormones. Mass spectrometric analysis of the fragments generated by tryptic digestion of hCG produces a mass spectrum which permits unambiguous identification.

The mass spectrum generated by MALDI analysis of a tryptic digestion of RCM-hCG is presented in Figure 3. Of the 25 tryptic peptides expected from complete tryptic digestion eight, αT2, αT7, αT8, βT2, βT5, βT7, βT8 and βT12 could be assigned signals in the mass spectrum (Table 1). Further peaks could be assigned to fragments resulting from partial cleavages, *e.g.* βT1+2. In addition, a number of signals corresponding to peptides that are not generated by trypsin alone are also observed (Table 2). Two peaks appearing in the cluster between 4400-5100 Da can be assigned, albeit tentatively due to low signal-to-noise ratios, to glycoforms of the peptide βT4 and the signal at 1832.1 Da may correspond to the fully sialylated variant of the βT13 glycopeptide (Table 3). Observed fragments correspond to 59 % and 52 % of the primary sequences of hCGα and hCGβ respectively. Importantly, the signal corresponding to the peptide βT5 relates to an amino acid sequence that comprises part of a surface orientated loop, formed by the linkage between cysteines 38 and 57, which plays a major role in the stimulation of steroidogenesis by hCG [15, 16]. Detection of this fragment provides persuasive evidence that the material analysed was biologically active hCG.

To validate that the information contained in the peptide-mass map, produced following tryptic cleavage of RCM-hCG, would allow unequivocal identification and, in particular, discrimination from the other glycoprotein hormones, experimentally determined masses were used to perform searches of two peptide-mass databases (see Scheme 1). Only masses assigned to the eight expected fragments (see above and Table 1) were used to perform the searches. Database entries for both hCGα and hCGβ were identified, individually, as the protein(s) best matched to experimental data using masses corresponding to just two tryptic fragments (see Tables 4 and 5). However, searches should be performed using as many masses as are available from the experimental data in order to increase their discriminatory power (Table 6). No entry for the hCG heterodimer exists in either database, each subunit being regarded as a separate protein.

Selection of the most likely candidate sequences using the CBRG facility is based upon an algorithm which determines the probability that the similarity of the experimental masses

against those computed in the database occurs at random [8]. The lower the probability of a random match, the greater the likelihood that the result is significant. The lowest probabilities are computed to give similarity scores of greatest magnitude. The significance of a match is determined by performing a set of searches using a series of random masses, equal in number to the experimental masses, generated uniformly within the range of experimental masses. The limit above which a similarity score is considered significant is calculated from the mean and variance of 5 searches performed with randomly generated masses.

The MOWSE scoring scheme is based on the frequency that a given fragment mass is observed in a protein of a defined molecular weight range [7]. Protein sequences in an existing database were grouped into 10 kDa intervals according to their molecular weight. Subsequently, for each protein in a given 10 kDa molecular weight range the expected peptide molecular weights produced by theoretical digestion, using a specific cleavage reagent, were assigned to molecular weight bands of 100 Da. This produced a distribution indicating the number of times a particular fragment mass occurs in a protein of any given size and a frequency value was calculated for each 100 Da band by dividing the cell value by the total number of peptides in each 10 kDa intact protein weight range. When a search is performed, matching peptide molecular weights are assigned a 'score' according to the frequency value of the cells in which they are distributed. If two or more peptides masses are matched in a single protein, the values contributed by these hits are multiplied to give a final score. Greater scoring power is attached to matches with larger peptides and the likelihood of non-random matches occurring in large proteins is compensated for by normalizing all scores to an average protein size of 50 kDa. Nevertheless, the matching of peptides in proteins of lower molecular weight does appear to carry greater scoring power (see Tables 4 and 5). Sequences selected by matches with less than 30 % of the experimental data are not considered to be relevant.

The results presented here show that only two peptide masses are required to identify either hCGα of hCGβ as the protein being analyzed. Although the certainty of the matches can be improved by increasing the number of masses used in the search, the ability to identify a protein using a minimum number of masses is an important issue. Firstly, the identity of glycosylated proteins such as hCG can be confirmed without recourse to using masses corresponding to glycopeptides. This means that carbohydrate heterogeneity will not affect the identification of hCG using this technique. Nevertheless, the ability to characterise the carbohydrate microheterogeneity may be useful in order to identify whether the material originated from pregnancy urine, was produced recombinantly, or was secreted by a tumour. Secondly, difficulties encountered in the extraction and handling of sub-picomole amounts of material may preclude the generation of a complete peptide map. For the purposes of detecting hCG abuse by athletes the small amounts of material available for analysis are close

to the absolute minimum required for MALDI MS. If an athlete produces a urine sample containing a concentration of hCG just above the recommended decision limit for positive findings, *i.e.* >10 IU/L [17], a 10 mL sample of urine, which is a realistic estimate of the volume likely to be made available for analysis, would contain approximately 300 fmol of hCG. Consequently, it may not be possible to generate a MALDI mass spectrum containing signals corresponding to all the peptides expected from total proteolysis. A procedure incorporating both centrifugal ultrafiltration [18] and immunoaffinity extraction [19] should provide the means by which to isolate selectively the hCG for analysis.

It is worth noting that caution should be exercised when trying to verify the uniqueness of a peptide-mass map by database searching. This is primarily due to a number of idiosyncracies that exist in the available databases. A number of database sequences are derived from DNA sequences because the sequencing of a particular genome proceeds at a much faster rate than sequencing of the proteins translated from it [9]. Consequently, 'hypothetical' protein sequences are produced by translating the corresponding DNA sequences. For many such proteins it is the sequences of their precursors, including NH_2-terminal signal peptides, that are entered into the databases. Unfortunately, the cleavage site of the signal peptide rarely corresponds to a cleavage site of the enzyme used to generate the peptide map [20]. The hCGα and hCGβ sequences in both the databases used are inferred from complementary DNA. Consequently, all entries contain signal peptides and the search algorithms fail to match peptides corresponding to the actual NH_2-terminal peptide of either subunit, *e.g.* αT1 and βT1, with fragments generated by theoretical digestion of the database sequences [7, 20]. If the approach of peptide mapping followed by database searching is to be utilized to identify known proteins it is prudent to avoid the use of masses corresponding to the NH_2-terminal peptides in the searches if possible.

A further problem that can arise was illustrated effectively by searches using masses from hCGβ (see Table 6). hCGβ is encoded by a cluster of six genes/pseudogenes of which the gene designated hCGβ5 is expressed predominantly [21, 22]. The MOWSE database contains sequences corresponding to the expected products of genes hCGβ5 and hCGβ3. Although the database is non-redundant, *i.e.* when the database is upgraded new sequences are deleted if they are found to be present in the database under an alternative name, the sequences which differ by only two amino acid residues are considered sufficiently different to warrant individual database entries (D. J. C. Pappin, personal communication). The SwissProt database only contains the protein sequence encoded by the hCGβ3 gene. When searches of both the MOWSE and SwissProt databases, using the masses corresponding to the fragments βT2 and βT5 were performed, the search algorithms ranked only one of the entries in the MOWSE 'hit' list and totally failed to identify the entry in the SwissProt database. The entry

that was identified in the MOWSE database corresponds to the product of the hCGβ5 gene. The unidentified entries in both the MOWSE and the SwissProt databases are inferred from the hCGβ3 gene. This sequence contains an amino acid substitution at the position β4 where the residue encoded for is a methionine residue compared to a proline residue in the hCGβ5 gene sequence [21]. As a result the mass corresponding to the peptide βT5 is not matched even though it is derived experimentally from hCG and is expected to be present from the known amino acid sequence of hCGβ [23]. Furthermore, it has been reported that sequence data in the database is occasionally incorrect which can significantly affect the chances of achieving a correct match [9]. This data also indicates that it may be judicious to perform searches of more than one database! Therefore, if searches are performed to validate the identity of known proteins from their peptide-mass maps, the failure of a search program to identify the sequence in a database should not necessarily be interpreted to mean that the peptide map produced is not unique to the protein. This is especially the case where only a few peptide masses are used to search a database; statistically, large proteins have a high chance of being matched accidentally to a set of experimental masses [8, 9].

Conclusions

The mapping of hCG by proteolytic digestion with trypsin, followed by MALDI-TOF MS, generates a peptide-mass map that is unique to hCG and which allows its identity to be ascertained unambiguously. Detection of the βT5 peptide in a validated peptide map provides strong evidence to indicate that the material analysed was capable of producing a biological response. This data when considered in parallel with the molecular mass data, showing the hCG heterodimer, would preclude an athlete, found positive for hCG using such a confirmatory procedure, from mounting a defence that a) there was no heterodimer in the sample, the peptide signals being generated by the digestion of free α- and β-subunits, or b) the heterodimer is not biologically active due to nicking in the β-subunit.

Considering that the current limits of detection of MALDI-TOF MS may preclude the production of a peptide map covering the entire sequences of the two subunits of hCG, the proposed confirmatory procedure would include MALDI MS analysis of untreated material prior to proteolytic digestion. This will not only provide molecular weight information, which is characteristic for hCG, but also show the presence of molecules that are capable of eliciting a biological response in the sample.

References

1. Kicman A. T., Brooks R. V. and Cowan D. A. (1991). Human chorionic gonadotropin and sport. *Br J Sports Med*, **25** (2): 73-80.

2. Meng C. K., Mann M. and Fenn J. B. (1988). Of protons or proteins. *Z Phys D*, **10**: 361-8.

3. Hillenkamp F. and Karas M. (1988). Laser desorption ionization of proteins with molecular masses exceeding 10 000 daltons. *Anal Chem*, **60**: 2299-301.

4. Laidler P., Cowan D. A., Hider R. C., Keane A. T. and Kicman A. T. (1995). Tryptic mapping of human chorionic gonadotropin by matrix-assisted laser desorption/ionization mass spectrometry. *Rapid Commun Mass Spectrom*, 9: 1021-6.

5. Carrey E. A. (1990). Peptide mapping. In: *Protein structure: A practical approach* (ed. Creighton T.E.), Oxford University Press, New York. 117-44.

6. Henzel W. J., Billeci T. M., Stults J. T., Wong S. C., Grimley C. and Watanabe C. (1993). Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA*, **90**: 5011-5.

7. Pappin D. J. C., Højrup P. and Bleasby A. J. (1993). Rapid identification of proteins by peptide-mass fingerprinting. *Current Biol*, **3**(6): 327-32.

8. James P., Quadroni M., Carafoli E. and Gonnet G. (1993). Protein identification by mass profile fingerprinting. *Biochem Biophys Res Commun*, **195** (1): 58-64.

9. Mann M., Højrup P. and Roepstorff P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol Mass Spectrom*, **22**: 338-45.

10. Yates III J. R., Speicher S., Griffin P. R. and Hunkapiller T. (1993). Peptide mass maps: A highly informative approach to protein identification. *Anal Biochem*, **214** (2): 397-408.

11. Kardana A. and Cole L. A. (1994). Human chorionic gonadotropin β-subunit nicking enzymes in pregnancy and cancer patient serum. *J Clin Endocrinol Metab*, **79** (3): 761-7.

12. Pierce J. G. and Parsons T. F. (1981). Glycoprotein hormones: Structure and function. *Ann Rev Biochem*, **50**: 465-95.

13. Ling V., Guzzetta A. W., Canova-Davis E., Stults J. T., Hancock W. S., Covey T. R. and Shushan B. I. (1991). Characterization of the tryptic map of recombinant DNA derived tissue plasminogen activator by high-performance liquid chromatography-electrospray ionization mass spectrometry. *Anal Chem*, **63** (24): 2909-15.

14. Billeci T. M. and Stults J. T. (1993). Tryptic mapping of recombinant proteins by matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem*, **65**: 1709-16.

15. Keutmann H. T., Charlesworth M. C., Mason K. A., Ostrea T., Johnson L. and Ryan R. J. (1987). A receptor-binding region in the human choriogonadotropin/lutropin β-subunit. *Proc Natl Acad Sci USA*, **84**: 2038-42.

16. Cole L. A., Kardana A., Andrade-Gordon P., Gawinowicz M. A., Morris J. C., Bergert E. R., O'Connor J. and Birken S. (1991). The heterogeneity of human chorionic gonadotropin (hCG). III. The occurrence and biological and immunological activities of nicked hCG. *Endocrinology*, **129** (3): 1559-67.

17. Laidler P., Cowan D. A., Hider R. C. and Kicman A. T. (1994). New decision limits and quality-control material for detecting human chorionic gonadotropin misuse in sports

18. Cowan D. A., Kicman A. T., Walker C. J. and Wheeler M. J. (1991). Effect of administration of human chorionic gonadotropin on criteria used to assess testosterone administration in athletes. *J Endocrinology*, **131**: 147-54.

19. Liu C. and Bowers L.D. (1996). HPLC/MS confirmation of urine human chorionic gonadotropin. *J Chromatography*, [in press].

20. Mørtz E., Vorm O., Mann M. and Roepstorff P. (1994). Identification of proteins in polyacrylamide gels by mass spectrometric peptide mapping combined with database search. *Biol Mass Spectrom*, **23**: 249-61.

21. Policastro P., Ovitt C. E., Hoshina M., Fukuoka H., Boothby M. R. and Boime I. (1983). The β-subunit of human chorionic gonadotropin is encoded by multiple genes. *J Biol Chem*, **258** (19): 11492-9.

22. Jameson J. L. and Hollenberg A. H. (1993). Regulation of chorionic gonadotropin gene expression. *Endocrine Revs*, **14** (2): 203-21.

23. Morgan F. J., Birken S. and Canfield R. C. (1975). The amino acid sequence of human chorionic gonadotropin. *J Biol Chem*, **250** (13): 5247-58.

Table 1. Tryptic peptides of reduced and S-carboxymethylated hCG observed by MALDI-TOF MS.

Amino acid residue no from	to	Tryptic[a] peptide	Calculated (M+H)[b]	Observed (M+H)[c]	Δ[d]	Amino acid sequence[ef]
hCGα						
1	35	αT1	4125.7			APDVQDCPECTLQEDPFFSQPGAPI -LQCMGCCFSR
36	42	αT2	818.0	817.5	-0.5	AYPTPLR
43	44	αT3	234.3			SK
45		αT4	147.2			K
46	51	αT5	720.0			TMLVQK
52	63	αT6	Glyco (N-linked)			**N**VTSESTCCVAK
64	67	αT7	539.6	539.4	-0.2	SYNR
68	75	αT8	839.1	838.7	-0.4	VTVMGGFK
76	91	αT9	Glyco (N-linked)			VE**N**HTACHCSTCYYHK
92		αT10	106.1			S
hCGβ						
1	2	βT1	234.3			SK
1	8	βT1+2	983.2	983.0	-0.2	SKEPLRPR
3	8	βT2	767.9	767.8	-0.1	EPLRPR
9	20	βT3	Glyco (N-linked)			CRPI**N**ATLAVEK
21	43	βT4	Glyco (N-linked)			EGCPVCITV**N**TTICAGYCPTMTR
44	60	βT5	1928.4	1928.4[g]	0	VLQGVLPALPQVVCNYR
61	63	βT6	389.4			DVR
64	68	βT7	651.8	651.3	-0.5	FESIR
69	74	βT8	700.9	700.7	-0.2	LPGCPR
69	95	βT8+9+10	3066.7	3067.1	0.4	LPGCPRGVNPVVSYAVALSCQCAL CRR
75	94	βT9	2228.6			GVNPVVSYAVALSCQCALCR
95		βT10	175.2			R
95	104	βT10+11	1080.2	1081.0	0.8	RSTTDCGGPK
96	104	βT11	924.0			STTDCGGPK
105	114	βT12	1227.3	1227.3[g]	0	DHPLTCDDPR
115	122	βT13	Glyco (O-linked)			FQDSSS**S**K
123	133	βT14	Glyco (O-linked x 2)			APPP**S**LPSP**S**R
134	145	βT15	Glyco (O-linked)			LPGP**S**DTPILPQ

[a]Tryptic peptides are designated either α or β according to the subunit of origin and numerically from the amino terminal of each subunit and '+' indicates an undigested clip.

[b]Calculated from average isotopic masses.

[c]Data taken from Figure 3.

[d]Difference between calculated and observed masses.

[e]Glycosylation sites are in bold type and are underlined.

[f]Amino acid

[g]Calibrant for mass assignment.

Table 2. Additional peptide fragments observed by MALDI-TOF MS following the tryptic digestion of reduced and S-carboxymethylated hCG.

Amino acid residue no. from	to	Peptide[a]	Calculated[b] (M+H)	Observed[c] (M+H)	Δ[d]	Amino acid sequence
α1	α17	αT1C1[e]	2023.9	2025.1	1.2	APDVQDCPECTLQEDPF
α18	α35	αT1C2	2120.4	2121.0	0.6	FSQPGAPILQCMGCCFSR
β48	β59	βT5C	1374.5	1374.0	-0.5	VLPALPQVVCNY
β48	β60	βT5G[f]	1530.7	1530.4	-0.3	VLPALPQVVCNYR
β75	β93	βT9C1[g]	2072.2	2073.0	0.8	GVNPVVSYAVALSCQCALC
β83	β94	βT9C2	1412.5	1412.8	0.3	AVALSCQCALCR
β87	β94	βT9C3	1058.0	1058.9	0.9	SCQCALCR

[a]Peptides are designated either α or β according to the subunit of origin and numerically from the amino terminal of each subunit.

[b]Calculated from average isotopic masses.

[c]Data taken from Figure 3.

[d]Difference between calculated and observed masses

[e]α-Chymotrypsin-like cleavage.

[f]Gonadotropin β-subunit nicking enzyme clip of βT5 fragment.

[g]Not a true α-chymotrypsin-like cleavage.

Table 3. Tryptic glycopeptides of reduced and S-carboxymethylated hCG observed by MALDI MS.

Tryptic peptide	Glycoform[a]	Calculated (M + H)[b]	Observed (M + H)[c]
βT4	$NeuNAc_2Gal_2Man_3GlcNAc_4Fuc$	5018.2	5020.3
βT4	$NeuNAcGal_2Man_3GlcNAc_4Fuc$	4727.0	4732.1
βT13	$NeuNAc_2GalNAcGal$	1832.8	1832.1

[a]Fuc, Fucose; Gal, Galactose; GalNAc, N-acetylgalactosamine; GlcNAc, N-acetylglucosamine; Man, Mannose; NeuNAc, N-acetylneuraminic acid.

[b]Calculated from average isotopic masses.

[c]Data from Figure 3.

Table 4. The results of a MOWSE database (constructed from OWL version 20.1) search using fragment masses 838 Da and 817 Da corresponding to fragments αT2 and αT8 observed by MALDI MS of a tryptic digest of reduced and S-carboxymethylated hCG[ab].

Score[c]	Mass[d]	Accuracy[e]	Description[f]
30.3	13075	100	**Glycoprotein hormones pre-alpha chain. Human**
28.4	13956	100	Thioredoxin H. Common tobacco
26.8	14778	100	Myelin P2 protein. Human
26.8	14804	100	Myelin P2 protein. Mouse
21.3	18568	100	β-Phycoerythrin beta chain. *Cryptomonas phi*
20.9	18981	100	Lincomycin resistance protein linA'. *Staph. aureus*
20.8	19041	100	Lincosamide resistance protein. *Staph. aureus*
13.5	10547	100	High mobility group protein HMG-Y. Human
13.5	10550	100	High mobility group protein isoform Y. Human
13.4	10511	50	60S acidic ribosomal protein P2-A. *Trypanoso. cruzi*

[a]Masses entered in the search correspond to non-protonated fragments and the mass of cysteine was adjusted to that of the free thiol in peptides containing half-cystine residues.

[b]Data taken from Figure 3 and Table 1.

[c]The database score achieved by limiting the mass range of the search between 10-20 kDa. The permitted mass error for correct matches was set at 1 Da and fragments resulting from partial cleavage were given only 10% of the scoring weight of completely digested fragments.

[d]Mass of the identified protein in the database. It should be noted that database sequences often contain signal peptides and take no account of known glycosylation.

[e]Calculated as a percentage of the number of experimental masses used to perform the search that were actually matched in the database for the identified protein.

[f]Database description of the identified protein indicating the species of origin.

Table 5. The results of a MOWSE Database (Constructed from OWL ver. 20.1) search using peptide fragment masses 1869 Da and 642 Da corresponding to fragments βT5 and βT8 observed by MALDI MS of a tryptic digest of reduced and S-carboxymethylated hCG[ab].

Score[c]	Mass[d]	Accuracy[e]	Description[f]
49.3	17728	100	**Chorionic gonadotropin beta subunit precursor. Human**
49.2	17738	100	**Choriogonadotropin pre-beta subunit. Human**
38.1	10556	50	3C protein. Avian infectious bronchitis virus
37.7	10675	50	Hypothetical protein 91. Garden pea chloroplast
36.0	11180	50	Apolipoprotein B. Pig
34.7	11589	50	Ig kappa chain V-VI region (NQ2-48.2.2). Mouse
34.6	11616	50	Ig lambda chain V-region. Human
34.5	11664	50	60S acidic ribosomal protein P2. Human
34.5	11665	50	Ig kappa chain V region (DP12VK). Mouse
34.4	11678	50	Ig kappa chain V-VI region (NQ5-78.2.6). Mouse

[a]Masses entered in the search correspond to non-protonated fragments and the mass of cysteine was adjusted to that of the free thiol in peptides containing half-cystine residues.

[b]Data taken from Figure 3 and Table 1.

[c]The database score achieved by limiting the mass range of the search between 10-20 kDa. The permitted mass error for correct matches was set at 1 Da and fragments resulting from partial cleavage were given only 10% of the scoring weight of completely digested fragments.

[d]Mass of the identified protein in the database. It should be noted that database sequences often contain signal peptides and take no account of known glycosylation.

[e]Calculated as a percentage of the number of experimental masses used to perform the search that were actually matched in the database for the identified protein.

[f]Database description of the identified protein indicating the species of origin.

Table 6. Output rankings for the entries corresponding to hCGβ in two peptide mass databases following searches using peptide fragment masses observed in the MALDI mass spectrum of the tryptic digest of reduced and S-carboxymethylated hCG. Only masses corresponding to those expected from complete digestion were used to perform the searches.

Peptide fragment masses combinations used to search the MOWSE[ab] and (CBRG[cd]) databases.	Listing position	
	MOWSE[e]	CBRG[f]
1869, 642 (1927.4, 699.7)	2 & 3	1
1869, 650 (1927.4, 650.3)	7 & 8	1
1869, 767 (1927.4, 766.8)	2	-
1869, 1168 (1927.4, 1226.3)	1 & 2	1
1168, 767 (1226.3, 766.8)	7	-
1168, 650 (1226.3, 650.3)	18 & 19	7
1168, 642 (1226.3, 699.7)	9 & 10	2
767, 650 (766.8, 650.3)	13	-
767, 642 (766.8, 699.7)	7	-
650, 642 (650.3, 699.7)	21 & 22	-
1869, 1168, 767 (1927.4, 1226.3, 766.8)	1 & 2	1
1869, 1168, 650 (1927.4, 1226.3, 650.3)	1 & 2	1
1869, 1168, 642 (1927.4, 1226.3, 699.7)	1 & 2	1
1869, 767, 650 (1927.4, 766.8, 650.3)	1 & 9	8
1869, 767, 642 (1927.4, 766.8, 699.7)	1 & 4	6
1869, 650, 642 (1927.4, 650.3, 699.7)	1 & 2	1
1168, 767, 650 (1226.3, 766.8, 650.3)	1 & 26	-
1168, 767, 642 (1226.3, 766.8, 699.7)	2 & 16	8
1168, 650, 642 (1226.3, 650.3, 699.7)	1 & 2	1
767, 650, 642 (766.8, 650.3, 699.7)	1	-
1869, 1168, 767, 650 (1927.4, 1226.3, 766.8, 650.3)	1 & 2	1
1869, 1168, 767, 642 (1927.4, 1226.3, 766.8, 699.7)	1 & 2	1
1869,1168, 650, 642 (1927.4, 1226.3, 650.3, 699.7)	1 & 2	1
1869, 767, 650, 642 (1927.4, 766.8, 650.3, 699.7)	1 & 2	1
1168, 767, 650, 642 (1226.3, 766.8, 650.3, 699.7)	1 & 3	5
1869, 1168, 767, 650, 642 (1927.4, 1226.3, 766.8, 650.3, 699.7)	1 & 2	1

[a]Searches were performed using version 5.2 of the MOWSE database (constructed from the non-redundant OWL database - release 26th March 1996).

[b]Cysteine residues were calculated as the free thiols and all masses correspond to unprotonated fragments.

[c]Sequences in the SwissProt database (release 32) were searched using the CBRG facility.

[d]Cysteine residues were calculated as the carboxymethylated derivative (C=161.1 Da) and all masses correspond to unprotonated fragments.

[e]Sequences corresponding to both the hCGβ5 and the hCGβ3 genes exist in the MOWSE database.

[f]Only one sequence corresponding to the hCGβ3 gene exists in the SwissProt database (release 32).

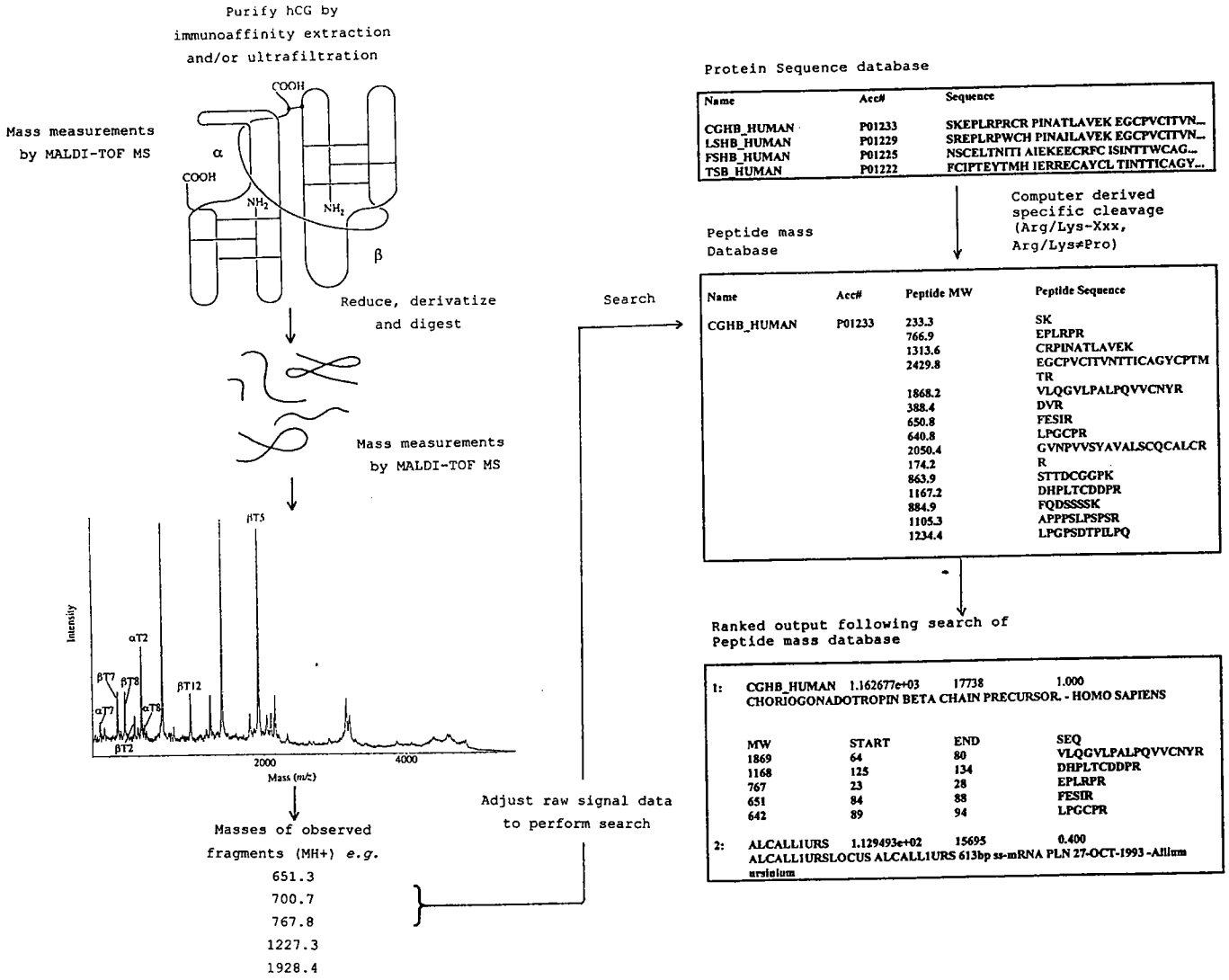

Figure 1. A flow diagram showing a suggested procedure for the confirmatory analysis of hCG. Mass spectrometric analysis of a tryptic digest of hCG produces a characterisitic spectrum. Validation of the spectrum's ability to identify hCG unambiguously is achieved by searching peptide mass databases using experimentally determined fragment masses. Peptide mass databases can be generated by computation of the peptide fragment masses expected from the theoretical digestion, according to specific cleavage rules (*e.g.* digestion with trypsin occurs after Arg or Lys except when followed by Pro), of sequences present in protein sequence databases. A search outputs a ranked list of the proteins best matched to the experimental data.

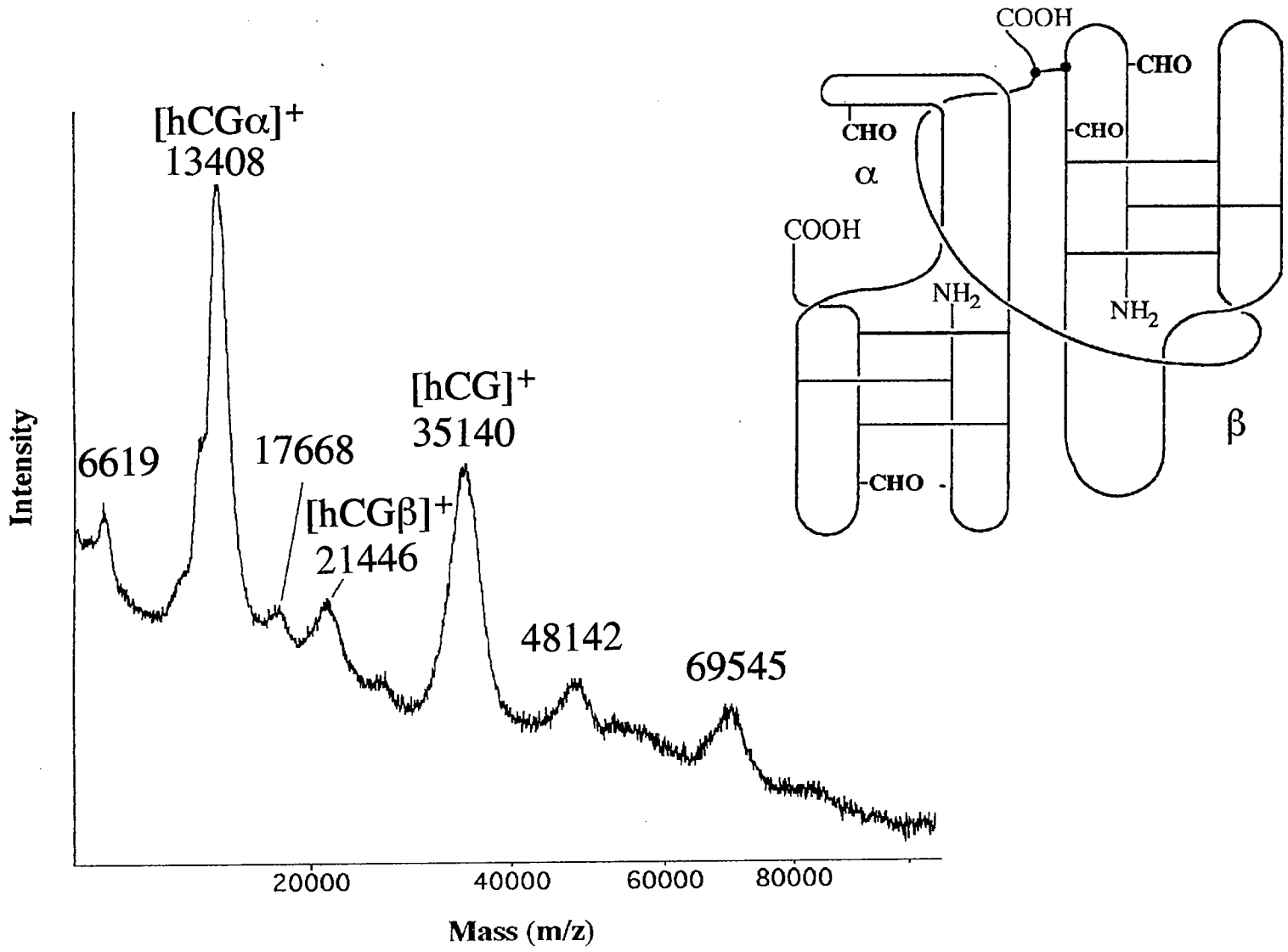

Figure 2. MALDI-TOF mass spectrum of pharmaceutical hCG in sinapinic acid. Adapted from Laidler *et al.*, 1996. *Rapid Commun Mass Spectrom* with permission from John Wiley & Sons.

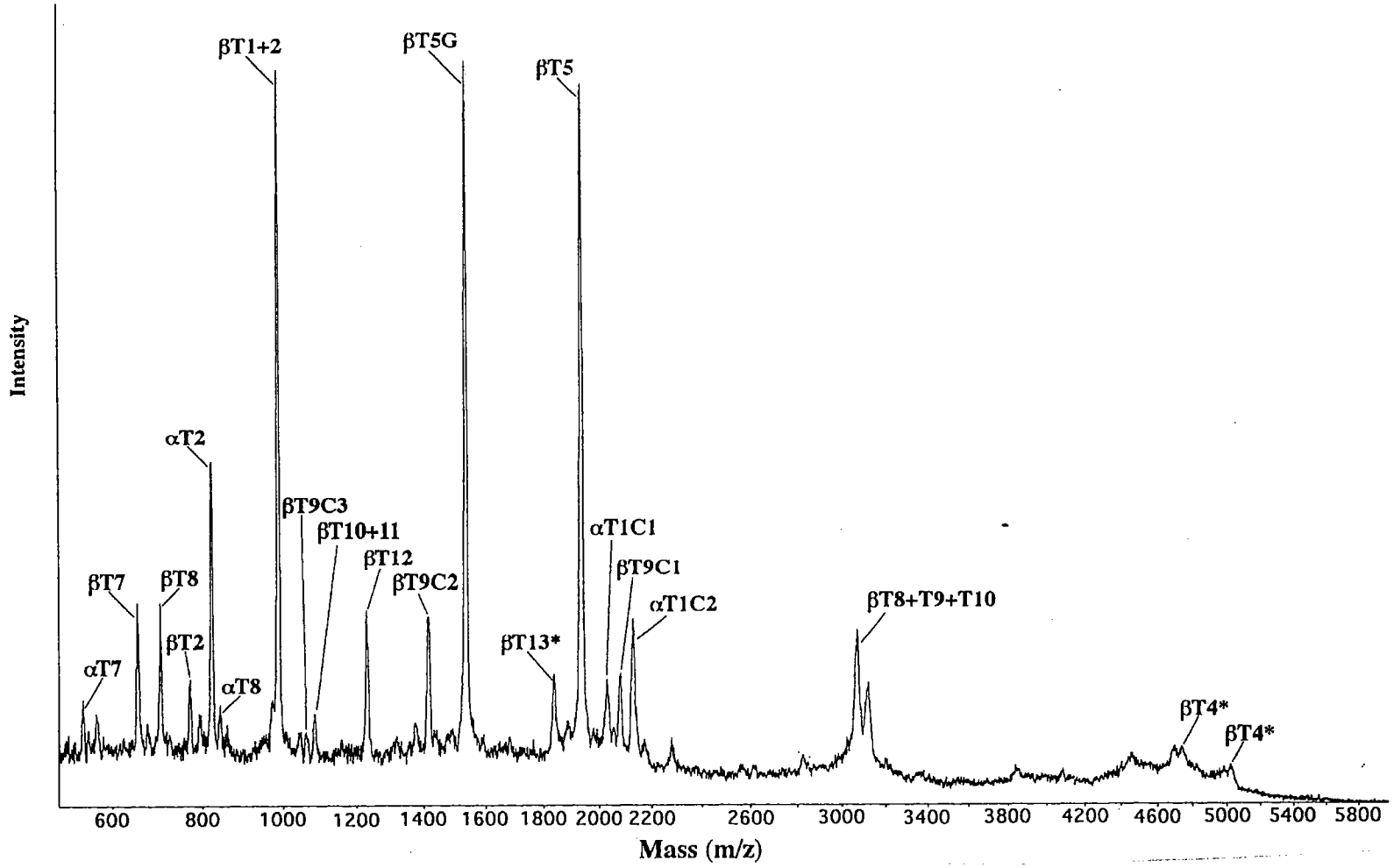

Figure 3. MALDI-TOF mass spectrum of a tryptic digest of reduced and *S*-carboxymethylated hCG in α-cyano-4-hydroxycinnamic acid matrix. Fragments are labelled according to the subunit from which they were generated and numerically from the amino terminal. T represents the digesting enzyme trypsin and fragments generated by additional cleavage are labelled either C or G corresponding to α-chymotrypsin and gonadotropin β-subunit nicking enzyme respectively. Glycopeptides are designated by *. Signals from 30 laser shots were summed to generate the spectrum and the *m/z* axis was calibrated internally using fragment masses 1227.3 Da and 1928.4 Da. Reproduced from Laidler *et al.*, 1996. *Rapid Commun Mass Spectrom* with permission from John Wiley & Sons.