ELSEVIER

Journal of Chromatography B, 726 (1999) 25–35

JOURNAL OF CHROMATOGRAPHY B

Composition of the peptide fraction in human blood plasma: database of circulating human peptides

Rudolf Richter[a,*], Peter Schulz-Knappe[b], Michael Schrader[b], Ludger Ständker[a], Michael Jürgens[b], Harald Tammen[a], Wolf-Georg Forssmann[a]

[a]*Lower Saxony Institute for Peptide Research, Feodor-Lynen-Strasse 31, D-30625 Hannover, Germany*
[b]*BioVisioN, Feodor-Lynen-Strasse 5, D-30625 Hannover, Germany*

Received 7 September 1998; received in revised form 28 December 1998; accepted 28 December 1998

Abstract

A database was established from human hemofiltrate (HF) that consisted of a mass database and a sequence database, with the aim of analyzing the composition of the peptide fraction in human blood. To establish a mass database, all 480 fractions of a peptide bank generated from HF were analyzed by MALDI-TOF mass spectrometry. Using this method, over 20 000 molecular masses representing native, circulating peptides were detected. Estimation of repeatedly detected masses suggests that approximately 5000 different peptides were recorded. More than 95% of the detected masses are smaller than 15 000, indicating that HF predominantly contains peptides. The sequence database contains over 340 entries from 75 different protein and peptide precursors. 55% of the entries are fragments from plasma proteins (fibrinogen A 13%, albumin 10%, β2-microglobulin 8.5%, cystatin C 7%, and fibrinogen B 6%). Seven percent of the entries represent peptide hormones, growth factors and cytokines. Thirty-three percent belong to protein families such as complement factors, enzymes, enzyme inhibitors and transport proteins. Five percent represent novel peptides of which some show homology to known peptide and protein families. The coexistence of processed peptide fragments, biologically active peptides and peptide precursors suggests that HF reflects the peptide composition of plasma. Interestingly, protein modules such as EGF domains (meprin Aα-fragments), somatomedin-B domains (vitronectin fragments), thyroglobulin domains (insulin like growth factor-binding proteins), and Kazal-type inhibitor domains were identified. Alignment of sequenced fragments to their precursor proteins and the analysis of their cleavage sites revealed that there are different processing pathways of plasma proteins in vivo.

Keywords: Human hemofiltrate; Peptides; Circulating human peptides

1. Introduction

Progress in genome sequencing results in an increasing demand for sequence data of the processed gene products. There are different means to identify the consecutive products of a gene, the mRNA, the protein, and the processed protein fragments. mRNA analysis is used to identify open reading frames, transcription starts, signal sequences and stop codons and is therefore useful to postulate translated proteins. The translated protein itself is of interest to verify a postulated protein. Processed

*Corresponding author.

0378-4347/99/$ – see front matter
PII: S0378-4347(99)00012-2

protein and peptide fragments often bear the biological activity. To elucidate these products of a gene, different strategies are applied.

Approaches to identify protein coding regions (exons) from the genome are performed by generating a large number of expressed sequence tags (ESTs) resulting from the partial sequencing of cDNAs, using bioinformatic computer programs [1–3], or exon trapping/in vivo processing of genomic DNA [4,5].

Most of these methods consider neither post-translational modification of proteins and peptides (e.g., glycosylation, amidation) nor proteolytic processing of proteins and peptides to biologically active peptides. Many peptide hormones show their biological activity after specific proteolytic processing. Examples of this are insulin that is produced after cleavage of the C-peptide and proglucagon which is processed to at least three different hormones: glucagon, and glucagon-like peptides 1 and 2. Furthermore, the primary structure of a peptide deduced from a mRNA/cDNA sequence does not give final information concerning the disulfide bridges of the resulting protein or peptide. For definitive structural and functional studies, it is necessary to isolate the native proteins or peptides of interest.

Therefore, we started to identify peptides from human blood. To obtain sufficient amounts of plasma peptides, extracts of human hemofiltrate (HF) from patients with chronic renal failure were produced. Using chromatographic methods, a reproducible large-scale procedure to generate a peptide bank from up to 10 000 l of HF was developed [6]. The proteins and peptides of each of 480 fractions from this peptide bank are characterized by their molecular mass using matrix-assisted laser desorption ionization time-of-flight mass spectometry (MALDI–TOF-MS). These data were used to generate a mass database. Furthermore, using MS–MS sequencing and during our attempts to isolate circulating hormones, we also determined the amino acid sequences of many of these peptides. These data were accumulated in our sequence database, which now contains more than 340 entries. Mass- and sequence databases form our peptide database. Here we present an analysis of these data, giving insight into the composition of the peptide fraction in human blood.

2. Experimental

2.1. Hemofiltrate

Hemofiltrate was obtained from the Nephrologisches Zentrum Niedersachsen, Hannoversch-Münden, Germany, in quantities of 1600 to 2000 l per week. Patients with chronic renal failure were subjected to routine arterio–venous hemofiltration three times per week. The following hemofiltration equipment was used routinely: Hemoprozessor (Sartorius, Göttingen, Germany) and AK 10 HFM (Gambro, Hechingen, Germany). The filters used were Hemoflow F 60S and Hemoflow HF 80S (Fresenius, Bad Homburg, Germany), Hemofilter FH 77 H and Hemofilter FH 88 H (Gambro, Martinsried, Germany). All ultrafilters used had a specified molecular-mass cut-off around 20 000. For Hemoflow F 60S, and Hemoflow HF 80S polysulphone membranes, and for Hemofilter FH 77 H, and Hemofilter FH 88 H polyamide membranes are used. The effective membrane surfaces were 1.3 m^2, 1.8 m^2, 1.4 m^2, and 2.0 m^2, respectively. Filtration was driven by a transmembranous pressure gradient of 60 to 100 mmHg at a blood flow-rate of 250 to 350 ml/min (1 mmHg=133.322 Pa). 20 to 30 l of filtrate were recovered per patient and treatment.

2.2. Peptide extraction

The sterile filtrate was immediately cooled to 4°C and acidified to pH 3 to prevent bacterial growth and proteolysis. Extraction of the peptides from HF was either performed using alginic acid or a strong cation exchanger (Fractogel TSK SP 650(M), Merck, Darmstadt, Germany; 25×10 cm, Vantage VA 250 column, Amicon, Witten, Germany) [6]. Alginic acid extraction was performed using a modified method described by Mutt [7]. Briefly, batches of 400 l HF are adjusted to pH 2.7, 2.5 kg alginic acid were added to the bath and stirred for 8 to 12 h. Then, alginic acid was sedimented and separated from the HF, washed with ethanol (10 l) and 0.005 *M* hydrochloric acid on a Büchner funnel. The peptides were eluted with 0.2 *M* hydrochloric acid (10 l). The eluate was adjusted to pH 4.0 and a peptide precipitation was performed with 5.5 *M* sodium chloride at

4°C for 20 h. Peptide extraction using a strong cation-exchange column (Fractogel SP650 (M)) was performed as described by Schulz-Knappe et al. [6]. Briefly, 1000 l HF were conditioned to pH 2.7. These batches were applied onto the strong cation exchanger using an Autopilot chromatography system (PerSeptive Biosystems, Wiesbaden, Germany). Then, batch elution was performed with 10 l 0.5 *M* ammonium acetate (two column volumes). The eluate was stored at −20°C or lyophilized until further use.

2.3. Preparation of a peptide bank from human blood

The peptide bank was produced as described by Schulz-Knappe et al. [6]. Briefly, for the first separation step the extracts of 5000 l HF were pooled and loaded on a 10-l cation-exchange column (Fractogel SP 650(M)). Bound peptides were eluted using seven buffers with increasing pH. The seven buffers were composed as follow: I: 0.1 *M* citric acid monohydrate, pH 3.6; II: 0.1 *M* acetic acid+0.1 *M* sodium acetate, pH 4.5; III: 0.1 *M* malic acid, pH 5.0; IV: 0.1 *M* succinic acid, pH 5.6; V: 0.1 *M* sodium dihydrogenphosphate, pH 6.6; 0.1 *M* disodiumhydrogenphosphate, pH 7.4; VII: 0.1 *M* ammonium carbonate, pH 9.0. The seven pools (pH pools) were collected and each of them was loaded onto a 12.5×10 cm reverse-phase column (Source RPC, 15 μm, Pharmacia, Freiburg, Germany) and eluted in a 8 l gradient from 100% A (0.01 *M* HCl) to 60% B (80% acetonitrile, 0.01 *M* HCl). Fractions of 200 ml were collected (see Fig. 1).

2.4. MALDI–TOF-MS

Lyophilized aliquots of 1 l HF-equivalent of the peptide bank fractions were dissolved in 0.1 to 4 ml acetonitrile–0.1% aqueous trifluororacetic acid (TFA) (1:1, v/v) according to their absorption profile in reversed-phase chromatography (optical density at 214 nm). The resulting analyte solutions contained an equivalent of 0.25 to 10 ml HF/μl. A 1-μl sample solution was applied on a stainless steel multiple sample tray as admixture to either sinapinic acid (SIN) or α-cyanohydroxycinamic acid (CHC). The matrix solution was composed of 5 mg/ml fucose, and 5 mg/ml SIN or CHC in acetonitrile–0.1% TFA (1:1, v/v). The crystallization process was performed with accelerated ambient temperature air-drying using a microventilator. Measurements were performed in linear mode with a LaserTec RBT II MALDI–TOF-MS system (Perseptive/Vestec, Houston, TX, USA). The instrument was equipped with a 1.2 m flight tube and a 337 nm nitrogen laser. Positive ions were accelerated at 25 kV and up to 30 laser shots were automatically accumulated per sample position. The Voyager RP BioSpectrometry Workstation Version 3.07.1 (PerSeptive Biosystems, Framingham, MA, USA) was used as control software. The automatic measurement included a search pattern of 18 spots per sample position. The laser intensity was adjusted to signal intensity in a preset mass range. From the 18 spots per sample position, only the best measurement, i.e., that with the highest signal intensity, was saved to the hard disk. Values for laser intensity, signal intensity and preset mass range had to be differently adjusted for CHC and SIN according to their specific properties [8]. The time-of-flight data were externally calibrated for each sample plate and sample preparation. Calibration and further data processing were performed with the Voyager RP BioSpectrometry Workstation processing software (Version 3.07.1 PerSeptive Biosystems, Framingham, MA, USA) based on Grams/386 Version 3.0 (Galactic Industries, Salem, NH, USA).

2.5. Isolation of peptides

Peptides were isolated with a variety of liquid chromatographic methods such as cation-exchange and reversed-phase chromatography. For cation-exchange chromatography, Fractogel TSK SP 650 S (Merck, Darmstadt, Germany) or Parcosil PepKat or ProKat material (5 μm, 300 Å, Biotek, Östringen, Germany) with sodium chloride gradients were used. For preparative reversed-phase purification Source RPC material (15 μm, Pharmacia, Freiburg, Germany) was used and for analytical chromatography RP-C_4 and -C_{18} material (5 μm, 300 Å, Vydac, Hesperia, USA) for example were used with standard gradients of acetonitrile. Detailed information for

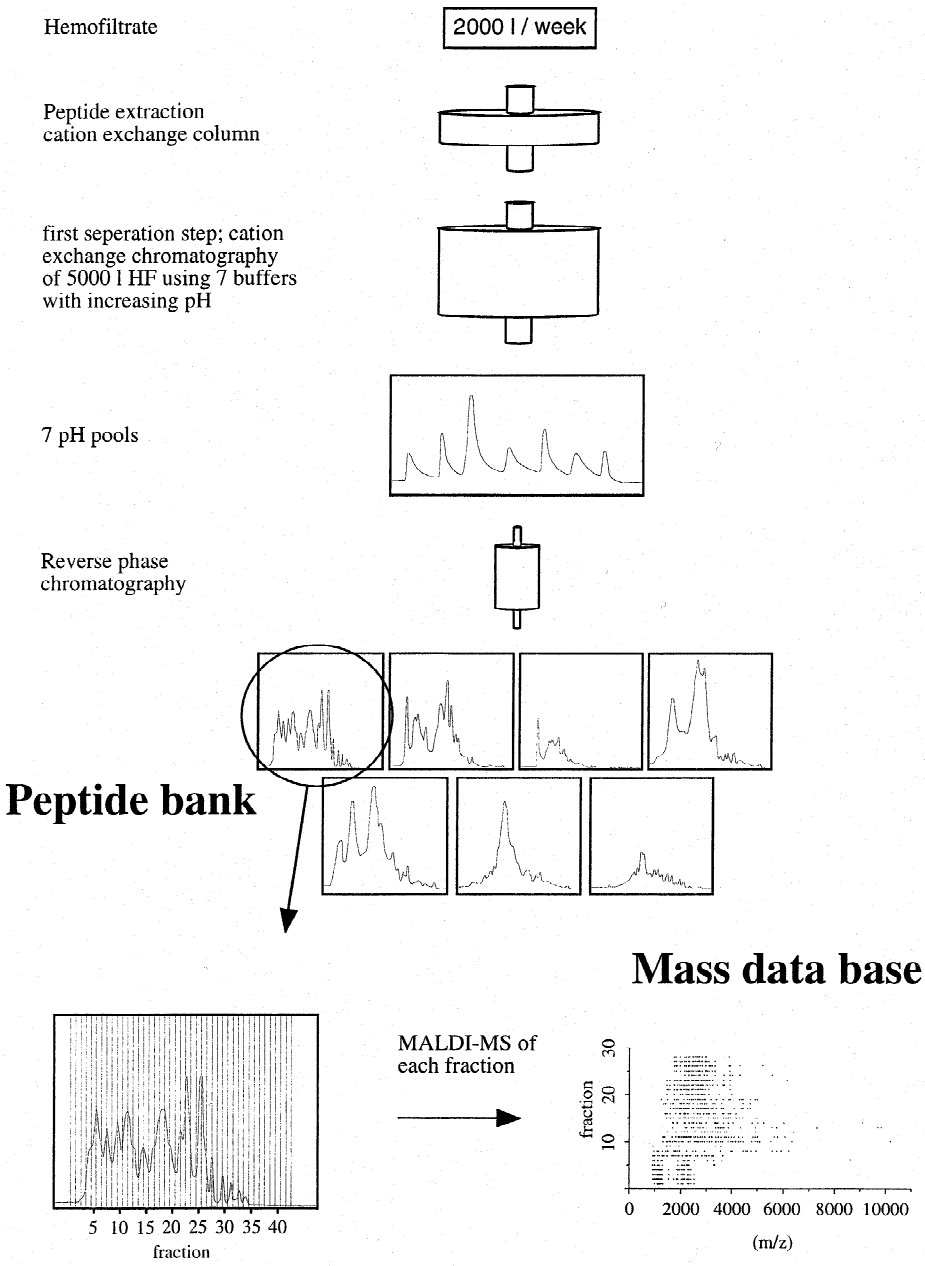

Fig. 1. Production of a human peptide bank and mass database from hemofiltrate. 2000 l hemofiltrate was collected per week. Extraction of the peptides was performed using a strong cation exchanger. Subsequently extracts of 5000 l HF were pooled and loaded on a 10-l cation-exchange column. Bound peptides were eluted using seven buffers with increasing pH. Each of the seven pH pools was rechromatographed on a reversed-phase column. Lyophilized aliquots of the fractions were analyzed on a MALDI–TOF-MS system. The MALDI–MS data of the fractions of each chromatography were accumulated in a one two-dimensional (2D) peptide map with the m/z values in the x-dimension and the fractions in the y-dimension. For each reversed-phase chromatography one 2D peptide map is produced.

different isolation procedures is given by Bensch et al. [9], Hess et. al. [10], Kuhn et al. [11], Schepky et al. [12], Schulz-Knappe et al. [13], and Ständker and co-workers [14,15].

2.6. Sequence analysis

Mass spectrometric sequencing from complex peptides mixtures was performed on an API III+ triple quadropol-MS system (PE Sciex Instruments, Toronto, Canada). The MS–MS experiments were performed using loop injection. Argon was used as collision gas in the collision-induced dissociation (CID) experiments.

The amino acid sequence of purified peptides was determined by automated Edman degradation using a gas-phase sequencer, Model 473 A or 494 (Applied Biosystems).

2.7. Production of the mass- and sequence databases

The peptide mass database was generated as described by Jürgens et al. [16]. It was established from the 480 fractions of the peptide bank. Each fraction was analyzed by MALDI–TOF-MS using CHC and SIN as matrices. The 2D maps were generated from MALDI–TOF-MS spectra of pH pool RP-fractions and contain the molecular mass in the *x*-dimension and the fractions in the *y*-dimension. Peak tables are processed in spreadsheet programs (Microsoft Excel and Microcal Origin) to result in a 2D peptide map for every pH pool.

The sequence database is the list of isolated peptides. The amino acid sequence and the N- and C-terminally adjacent amino acids of a peptide are given. The peptides are arranged according to their molecular mass (average mass). If the peptide was isolated from the HF peptide bank [6], the fraction used and the pH pool is shown. The functions of the isolated peptides are given in the last column as indicated by the SwissProt database [17].

2.8. Data interpretation and database research

Identification and assessment of the function of sequenced peptides was performed using the SwissProt database [17]. cDNAs of unknown peptides were searched in the dbEST-data bank. Databases were searched using the basic blast mode [18–20]. Multiple alignment studies of peptide fragments from proteins were performed using the MacMolly software package (Soft Gene, Berlin, Germany). Predicted signal peptide processing sites from cDNA sequences were determined by the use of the SignalP V1.1 WorldWideWeb Server [21].

3. Results and discussion

3.1. Peptide mass database

The peptide mass database contains one 2D peptide map for each of the seven pH pool fractions (Fig. 2). Each map contains approximately 3500 entries. Isolation and analysis of diverse molecular masses suggest that the detected masses predominantly represent proteins and peptides. The standard deviation for detected molecular masses using LaserTec RBT II MALDI–TOF-MS is 2‰. Considering this mass deviation, most molecules are repeatedly detected in consecutive fractions. Peaks are checked for multiply charged and multimeric ions. CHC has a tendency to produce more multiply charged ions than SIN. However, in these complex mixtures, doubly and rarely triply charged species are observed only for larger (approx. >7000) and/or abundant peptides. The most abundant peptides sometimes form multimeric species. These findings suggest that approximately 5000 different peptides are detected in HF subjecting 0.25 ml to 10 ml HF-equivalent to MALDI–TOF-MS. The detection limit of different pure peptides (e.g., insulin, GLP-1, CDD/ANP-99-126, HCC-1) is between 30 fmol/μl and 3 pmol/μl. This suggests minimal concentrations for the detection of these molecules in a range from 3 to 300 p*M*, but detection of a peptide by MALDI–TOF-MS is dependent on the concomitant components in the analyte which may dramatically increase the detection limit for such peptides. Therefore detection of molecules such as insulin, CDD/ANP-99-126, or GLP-1 by their molecular mass, and in consideration of their chromatographic characteristics in the peptide bank, was unsuccessful.

Ninety-five percent of the detected masses are $<15\,000$. This reflects that hemofiltrate contains

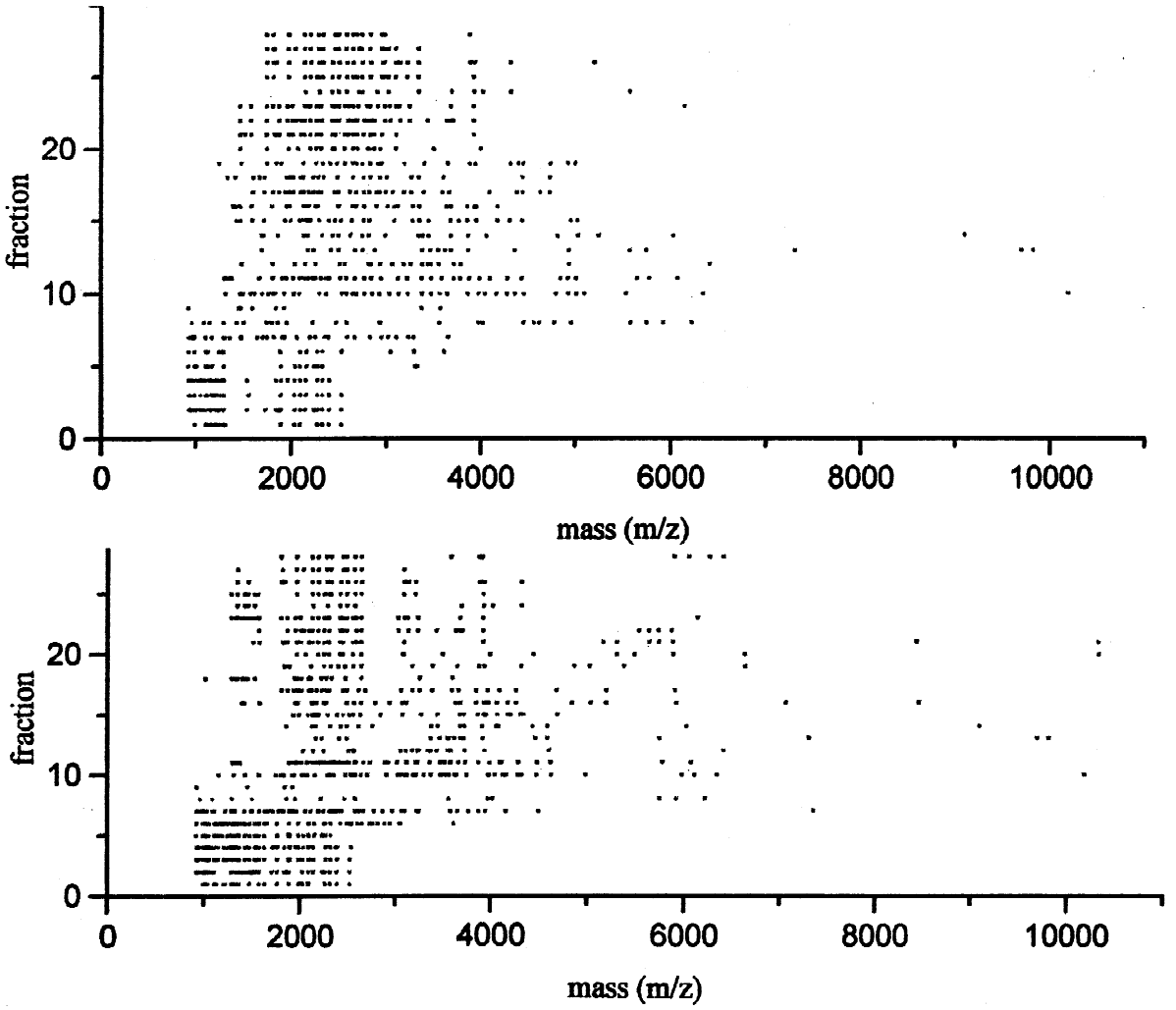

Fig. 2. Peptide maps from pH pool 1 (A) and pH pool 7 (B). The 2D maps were generated from MALDI–TOF-MS spectra of pH pool RP-fractions and contain the molecular mass in the *x*-dimension and the fractions in the *y*-dimension. Each asterisk represents a molecular mass. The masses with a spectrum intensity >500 are shown resulting in approximately 1500 masses per map. Without restriction of the spectrum intensity each map contains approximately 3500 masses. Maps demonstrate repeated detection of the same masses especially in the first and last fractions of each chromatography.

predominantly peptides with masses <20 000 and that MALDI–TOF-MS predominantly detects smaller peptides. Due to the molecular cut-off of the hemofilters (20 000), plasma proteins are excluded to such an extent that the total protein content is only 50–70 mg/l compared to 70 g/l in plasma. As shown by Schepky et al. [22] plasma and HF concentrations of small hormones such as angiotensin II, vasopressin, gastrin, endothelin and insulin are in the same range whereas only 0.02% of blood albumin is filtered through the hemofilter. This filtration is useful for the enrichment of smaller peptides during purification procedures and enables us to isolate peptides from 10 000 l of plasma equivalent. Analysis of the molecular mass distribution of HF peptides by size-exclusion gel chromatog-

raphy reveals that approximately 45% of the peptides in HF are in the range >15 000 and contain 15% albumin [22]. Analysis of fractions from pH pool 6 and pH pool 7 by gel electrophoresis reveals diverse bands in the size range between M_r 15 000 and 66 000 (data not shown). This suggests that HF contains a diversity of peptides and proteins >15 000 but due to the hemofilterś molecular cut-off of 20 000 in significantly lower concentrations than in blood plasma.

Another explanation for the detection of predominantly smaller peptides (see Table 1) is that in MALDI–TOF-MS the intensity of a signal is most dominant at about M_r 2000 to 3000. For molecules >3000 there is a continuous decrease in the intensity of the MALDI–TOF-MS signal, e.g., the MALDI–TOF-MS S/N ratio for 1 pmol of GLP-1 (M_r 3298) ranges from 30 to 60, whereas the S/N ratio for 1 pmol of albumin (M_r 66 000) ranges from 3 to 4. Low intensity of the M_r signal for proteins and peptides with M_r>10 000 reflects peak spreading due to the increasing number of isotopes and salt adducts and a decreased sensitivity of the detector.

In our institute the mass database is a useful tool with which to map each hemofiltrate peptide bank for identification of known peptides of interest (e.g., HCC-1 [13], guanylin [11], and β-defensin [9]). Subsequently, these peptides are purified for further biological testing. Furthermore, the mass spectrometry of peptide fractions is a method by which to identify peptides with special molecular characteristics (e.g., cysteine-rich peptides, amidation) by chemical modification of these groups with subsequent mass spectrometric analysis.

Contrary to our MALDI–TOF-MS mass database, 2D gel plasma maps [23,24] predominantly detect proteins of M_r>10 000. The smallest proteins found in this 2D gel map are kininogen light chain and apolipoprotein A-II with an M_r~10 000. This shows that our mass data base contains information com-

Table 1
Regulatory peptides and protein fragments from HF[a]

Peptide hormones	Angiotensin 1, guanylin-22-115, uroguanylin-89-112, cardiodilatin/atrial natriuretic peptide (CDD/ANP 99-126), β-defensin 1 (h-BD1), neutrophil defensin 1, neutrophil defensin 3, kininogen (LMW chain)
Cytokines, growth factors, growth inhibitors:	HCC-1, IGF-1, IGF-2, osteoinductive factor, platelet derived growth factor (PDGF), osteopontin, CTAP III, pigment endothelium derived factor, angiogenin I, collagen XVIII
Complement factors:	Complement factor C3, complement factor C4A (Anaphylatoxin) complement factor C9, complement factor D (CFAD)
Enzymes, enzyme inhibitors:	Lysozyme, carboxypeptidase N, pancreatic trypsin inhibitor, cystatin C, plasminogen, α-2-antiplasmin, inter-α-trypsin inhibitor complex component II (ITI2), α-1-antitrypsin, hexokinase type II, ribonuclease
Transport proteins:	Transthyretin, serotransferrin, retinol binding protein (RBP), transforming growth factor-binding protein (TGF-BP), insulin-like growth factor-binding protein (IBP3)
Plasma proteins:	Albumin, fibrinogen A (RGD Peptides), fibrinogen B, α-1-microglobulin, β-2-microglobulin, zinc-α-2-glycoprotein (ZAG), α-2-HS-glycoprotein (fetuin), serum amyloid A protein (SAA), haptoglobin, profilin, vitronectin, desmocollin, thymosin β4, apolipoprotein C-III, uteroglobin, ubiquitin, gelsolin, somatomedin B, hemopexin

[a] The fragments belong to approximately 75 precursor peptides and proteins. The known 60 peptides and proteins are shown. For peptide hormones and growth factors biologically active sequences as well as other prohormone fragments were found.

plementary to a 2D gel plasma map for predominant detection of peptides M_r<10 000.

3.2. Peptide sequence data base

Our sequence database was produced in the last five years. Up to now, the sequence data base (Fig. 3) contains over 340 entries from about 75 different protein and peptide precursors. Fifty-five percent of the entries are fragments from plasma proteins (fibrinogen A 13%, albumin 10%, β2-microglobulin 8.5%, cystatin C 7%, and fibrinogen B 6%). Thirty-three percent belong to protein families such as complement factors, enzymes, enzyme inhibitors and transport protein fragments. Four percent of the entries are contributable to peptide hormones, 3% to growth factors and cytokines.

Five percent are new sequences, most of which do not show homology to known peptide and protein families. Due to the progress in genome sequencing, increasing numbers of these novel peptides are found in the Expressed Sequence Tags database (dbEST).

Identified peptides are in an M_r from 500 to 30 000. The largest isolated protein is albumin-309-585, with an M_r of 31 493. The sequence data base shows that proteins and peptides with a M_r<8500 represent predominantly protein fragments and processed peptides. Twenty-five percent of the identified peptides with an M_r>8500 are precursor peptides and proteins.

Concerning possible non-specific degradation of the peptides in HF during collection and transport, there are a number of different findings. (1) Most of the identified hormones (e.g., CDD/ANP, IGF-1, IGF-2) show 100% identity with postulated circulating hormone sequences. The processing sites of these hormones are identical to those identified in earlier investigations [25,26]. (2) Biologically active factors such as guanylin [11], HCC-1 [13], or uteroglobin [27] are isolated from HF as precursor peptides. (3) Alignments of over 150 sequenced fragments of albumin, fibrinogen A, fibrinogen B, cystatin C, β2-microglobulin, and complement factor C3 with their precursor proteins reveal a specific tryptic or

Structure of the Peptide Data Base

detected mass	calculated mass	NT AA	sequence	CT AA	substance	pH-pool	fraction	function	reference/accession/ DB-source
1573	1572	Q	VVRAR.//.AGVNY	F	Cystatin C 75-88			proteinase inhibitor; Site 81..85: secondary area of contact	ACCESSION 118183 SWISS-PROT
2314	2315	L	SKHVE.//.NDLQF	F	Zinc-α2-Glycoprotein 35-55	1	18	may play a role in the expression of the immune response	ACCESSION 141596 SWISS-PROT
2428	2428	signal	DAHKS.//.ENFKA	L	Albumin 25-45	1	18	Site 27: copper (His)-binding	ACCESSION 2144898 PIR: locus ABHUS
2672	2672	K	SYKMA.//.KRGHA	K	Fibrinogen Aα 600-624	8		yielding monomers that polymerize into fibrin, cofactor in platelet aggregation	ACCESSION 1706799 SWISS-PROT
3731	3708	F	AVGEY.//.AGVNY	F	Cystatin C 56-88			proteinase inhibitor; Site 81..85: secondary area of contact	ACCESSION 118183 SWISS-PROT
3925	3926	S	DHYNC.//.AKCCK	end	ß-Defensin 1 (new)	7		antibiotic	Bensch et al., FEBS Lett. 368, 331-335 (1995)
5423	5425	Y	FRDPC.//.HGSVL	G	Meprin Aα 670-719 (new)			EGF homology	
7664	7664				new				
8567	8564	signal	MQIFV.//.RLRGG	end	Ubiquitin 1-76			involved in ATP-dependent selective degradation of cellular proteins, maintenance of chromatin structure, regulation of gene expression, stress response, ribosome biogenesis	ACCESSION 13667 SWISS-PROT
8673	8672	signal	TKTES.//.DMKEN	end	Cytokin HCC-1 (new)	7	16	?(stem cell proliferation)	Schulz-Knappe et al., J. Exp. Med., 183, 295-299 (1996)
10344	10337	signal	VTVQD.//.ACTGC	end	Guanylin-22-115 (new)	1	23	intestinal chloride secretion	Kuhn et al., FEBS Lett. 318, 205-209 (1993)
12933	12917				new				
15856	15830		EICPS.//.SSLCN		Uteroglobin	7		inhibitor of phospholipase A2, binding of progesterone	ACCESSION 112672 SWISS-PROT

Fig. 3. Structure of the peptide sequence database. The sequence database is a collection of sequenced peptides from HF. The five C- and N-terminal amino acids of a peptide sequence and the N- and C-terminally adjacent amino acids of a peptide sequence are shown. The peptides are arranged according to their molecular mass (average mass). The measured mass is given. If the peptide is isolated from the HF peptide bank, the fraction and the pH pool is shown. The (putative) function of the isolated peptides are given in the last column as indicated by the Swiss-Prot database or other literature (NT AA=N-terminal amino acid; CT AA=C-terminal amino acid). References, accession numbers and data base entries of the precursor proteins and peptides are given in the last column.

chymotryptic endoproteolysis with subsequent exoproteolytic digestion. Twenty percent of the fragment cleavages are C-terminal to Lys or Arg, suggesting tryptic digestion. Forty-four percent of the fragment cleavages are C-terminal to Phe, Tyr or Leu suggesting a chymotryptic digestion. Sixteen, 6, 4 and 4% of the fragment cleavages are at a distance of 1, 2, 3 and 4 amino acids, respectively, to tryptic or chymotryptic cleavage sites, suggesting exoproteinase digestion after tryptic or chymotryptic digestion (Table 2). (4) Predicted signal peptide processing sites from cDNA sequences are compared to the processing sites determined by sequencing the N-terminus of circulating peptides using SignalP V1.1 WorldWideWeb Server. Twelve percent of the sequenced peptides were either N-terminal fragments of the propeptides or non-processed peptides. Comparison revealed 100% identity of the predicted processing site with the processing site identified by sequencing the peptides (Fig. 4). These findings strongly suggest that HF peptides do not undergo non-specific degradation and reflect the peptide composition in human blood.

Alignment of identified fragments with their protein precursors revealed that processing of plasma proteins is not unidirectional but there are different processing pathways in vivo. Furthermore, this analysis suggests different proteolytic rates for different degradation products of plasma proteins, e.g., processing of N-terminal fibrinogen B does not only result in activation of fibrinogen B by cleavage of the fibrinopeptide B at the cleavage site between AS 14 and AS 15, but 8 different fragments spanning this cleavage site are also found. Identification of exclusively N-terminal fibrinogen B fragments spanning this cleavage site raises the question of a physiological function of these fragments. Only N-terminal fragments containing the thrombin cleavage site and polymerization site are found (Fig. 5).

Table 2
Processing sites of the isolated peptide fragments[a]

Tryptic digestion (C-terminal to Lys and Arg)	20%
Chymotryptic digestion (C-terminal to Phe, Tyr, Leu)	44%
Cleavage site at a distance of 1–4 AS from a tryptic or chymotryptic cleavage site	30%

[a] They were identified from over 150 fragments of albumin, fibrinogen A, fibrinogen B, cystatin C, β2-microglobulin and complement factor C3.

Peptide	predicted signal processing; C-terminal signal sequence/ N-terminal prohormone sequence	sequenced N-terminus
guanylin	AGG/VTV	VTV
HCC-1	ALG/TKT	TKT
lysozym	VLG/KVF	KVF
ß2-microglobulin	LEA/IQR	IQR
trypsin-inhibitor	TGA/DSL	DSL
uteroglobin	ASA/EIC	EIC
vitronectin	ALA/DQE	DQE

Fig. 4. Comparison of predicted signal peptide processing sites to the processing sites determined by sequencing of the N-terminus of seven peptides isolated from HF. Prediction of the signal cleavage site was performed using SignalP V1.1 WorldWideWeb Server. Twelve percent of the database entries are N-terminal fragments of propeptides or non-processed propeptides, demonstrating 100% identity with the predicted signal cleavage site.

Assessment of post-translational modifications was achieved by comparison of the theoretical average mass and the detected mass. This suggests that approximately 20% of the peptides are post-translationally modified. Due to a mass deviation of 2‰, salt adducts and detection of different isotopes of a molecule specification of the post-translational modification is difficult by the use of MALDI–MS. Up to now glycosylation of 1% of the isolated peptides was identified and elucidated (e.g., HCC-1 [28], C-terminal domain of IGFBP-5 [29]). The disulfide bridges of different proteins and peptides were identified (e.g., kazal-type inhibitor domain, C-terminal domain of IGFBP-4, endostatin [15]).

As the knowledge of protein sequences grows, it is becoming apparent that many proteins are constructed from relatively few molecular units, which occurs repeatedly. These modules often correspond to single exons. Modules can act as growth factors, in receptor-growth factor interactions, or tightly control cascades of enzyme-catalysed proteolysis. Interestingly, we have identified different modular peptides such as vitronectin fragments (somatomedin-B domain) [14], meprin Aα-fragments (EGF-domain), insulin-like growth factor-binding

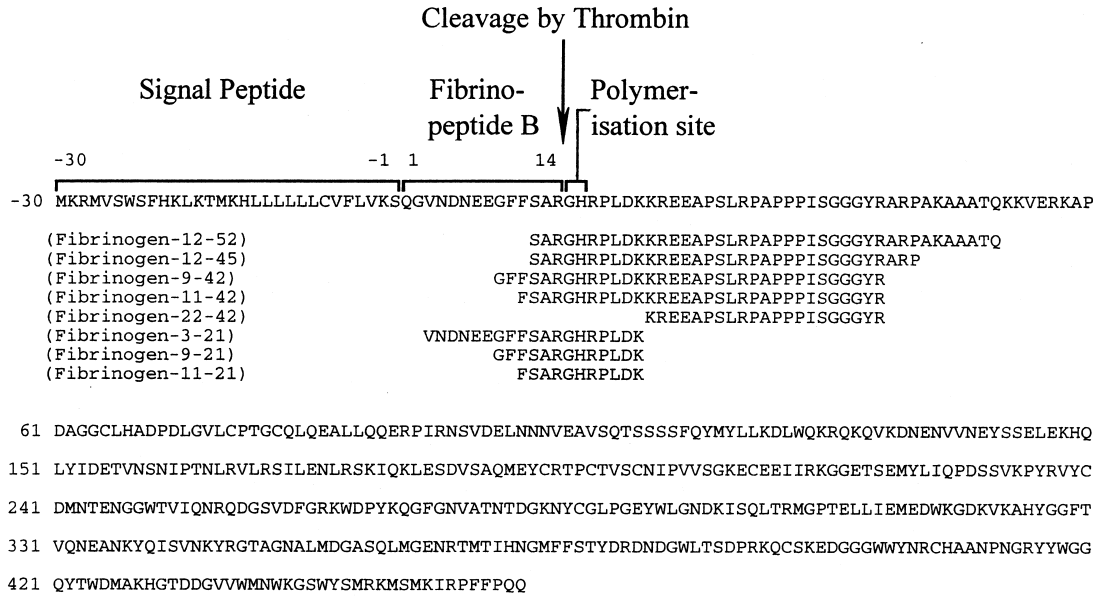

Fig. 5. Alignment of isolated fibrinogen B fragments with the precursor protein. Signal sequence is amino acid −30 to −1. Fibrinogen B is activated by cleavage between amino acid 44 and 45 releasing fibrinopeptide B. Some identified fibrinogen fragments are carrying the cleavage sites, suggesting different processing pathways for fibrinogen B.

proteins (thyroglobulin domain), and a kazal-type inhibitor domain. The identified vitronectin module (residues 1 to 44–50) is an effective competitor of the interaction of plasmin activator inhibitor (PAI)-1 with intact vitronectin or extracellular matrix. This suggests the regulation of PAI-1 in part by proteolytic processing of vitronectin modules [30]. The newly identified meprin Aα-fragments show a 94% homology to human meprin and this is a soluble EGF-like protein, suggesting that the isolated peptide has growth modulatory function as demonstrated for other EGF domains.

Concerning the biological relevance of HF peptide hormones, growth factors and cytokines were isolated in the biologically active form (e.g., angiotensin I, CDD/ANP [31], insulin like growth factor I and II). Using HF, new regulatory peptides such as guanylin [11], HCC-1 [13] or β-defensin-1 [9] could be identified (Table 1). Isolation of proteolytic products of fibrinogen and vitronectin containing cell attachment sites (RGD sequences) that carry antithrombotic activity or antagonistic activity to entire vitronectin demonstrate that degradation products of plasma proteins present in the circulation may have important biological activities [30,32].

4. Conclusions

Our database of human circulating peptides demonstrates that hemofiltrate reveals a source to characterize the composition of the peptide fraction in human blood. The mass database is a useful tool to identify chromatographically and mass spectrometrically characterized peptides in each newly prepared hemofiltrate peptide bank. As the mass deviation of the MALDI–TOF-MS technique improves, the attachment of molecular masses to their molecules becomes more reliable. Isolation of peptides from HF appears to be of general help to identify biologically active peptides and proteins to understand the processing and metabolism of peptides and plasma proteins and to identify their post-translational modifications. The mass database is a good complement to a 2D gel plasma map by predominant detection of peptides $<10\,000$.

References

[1] M. Zhang, Proc. Natl. Acad. Sci. USA, 94 (1997) 565-568
[2] E. Uberbacher, R. Mural, Proc. Natl. Acad. Sci. USA 88 (1991) 11261–11265.
[3] V. Solovyev, A. Salamov, C. Lawrence, Nucl. Acids Res. 22 (1994) 5156–5163.
[4] M. Nehls, D. Pfeifer, T. Boehm, Oncogene 9 (1994) 2169–2175.
[5] M. Nehls, D. Pfeifer, G. Micklem, C. Schmoor, T. Boehm, Curr. Biol. 4 (1994) 883–989.
[6] P. Schulz-Knappe, M. Schrader, L. Ständker, R. Richter, R. Hess, M. Jürgens, W.-G. Forssmann, J. Chromatogr. A 776 (1997) 125–132.
[7] V. Mutt, Gut Hormones, in: S.R. Bloom (Ed.), C. Livingstone, Edinburgh, London, New York, 1978, pp. 21–27.
[8] P. Juhasz, C. Costello, K. Biemann, J. Am. Soc. Mass Spectrom. 4 (1993) 399–409.
[9] K. Bensch, M. Raida, J. Mägert, P. Schulz-Knappe, W.G. Forssmann, FEBS Lett. 368 (1995) 331–335.
[10] R. Hess, M. Kuhn, P. Schulz-Knappe, M. Raida, M. Fuchs, K. Adermann, V. Kaever, Y. Cetin, W.-G. Forssmann, FEBS Lett. 374 (1995) 34–38.
[11] M. Kuhn, M. Raida, K. Adermann, P. Schulz-Knappe, R. Gerzer, J. Heim, W.G. Forssmann, FEBS Lett. 318 (1993) 205–209.
[12] A. Schepky, P. Schulz-Knappe, W.-G. Forssmann, J. Chromatogr. A 691 (1995) 255–261.
[13] P. Schulz-Knappe, H. Mägert, B. Dewald, M. Meyer, Y. Cetin, M. Kubbies, J. Tomeczkowski, K. Kirchhoff, M. Raida, K. Adermann, A. Kist, M. Reineke, R. Sillard, A. Pardigol, M. Uguccioni, M. Baggiolini, W.G. Forssmann, J. Exp. Med. 183 (1996) 295–299.
[14] L. Ständker, G. Heine, E. Hartmann, R. Hess, P. Schulz-Knappe, W.G. Forssmann, Biol. Chem. Hoppe-Seyler 376 (1995) 156.
[15] L. Ständker, M. Schrader, S. Kanse, M. Jurgens, W.G. Forssmann, K.T. Preissner, FEBS Lett. 420 (1997) 129–133.
[16] M. Jürgens, M. Schrader, M. Raida, W.G. Forssmann, P. Schulz-Knappe, J. Biol. Tech. http://www.medstv.unimelb.edu.au/JBT/Articles/JBT0006/JBT0006.html, submitted for publication
[17] A. Bairoch, B. Boeckmann, Nucleic Acids Res. 19 (Suppl.) (1991) 2247–2249.
[18] M. Boguski, T. Lowe, C. Tolstoshev, Nat. Genet. 4 (1993) 332–333.
[19] S. Altschul, T. Madden, A. Schaffer, J. Zhang, Z. Zhang, W. Miller, D. Lipman, Nucleic Acids Res. 25 (1997) 3389–3402.
[20] S. Karlin, S. Altschul, Proc. Natl. Acad. Sci. USA 87 (1990) 2264–2268.
[21] H. Nielsen, J. Engelbrecht, S. Brunak, G. von Heijne, Protein Eng. 10 (1997) 1–6.
[22] A. Schepky, K. Bensch, P. Schulz-Knappe, W.G. Forssmann, Biomed. Chromatogr. 8 (1994) 90–94.
[23] C. Hoogland, J. Sanchez, L. Tonella, A. Bairoch, D. Hochstrasser, R. Appel, Nucleic Acids Res. 26 (1998) 332–333.
[24] J. Sanchez, R. Appel, O. Golaz, C. Pasquali, F. Ravier, A. Bairoch, D. Hochstrasser, Electrophoresis 16 (1995) 1131–1151.
[25] R. Humbel, Hormonal Proteins and Peptides, in: C. Li (Ed.), Academic Press, New York, 1984, p. 57.
[26] C. Glembotski, C. Irons, A. Sprenkle, C. Sei, Can. J. Physiol. Pharmacol. 69 (1991) 1525–1536.
[27] A. Aoki, H. Pasolli, M. Raida, M. Meyer, P. Schulz-Knappe, H. Mostafavi, A. Schepky, R. Znottka, J. Elia, D. Hock, H. Beier, W.-G. Forssmann, Mol. Hum. Reprod. 2 (1996) 489–497.
[28] R. Richter, M. Schrader, M. Hagmann, W.-G. Forssmann, P. Schulz-Knappe, in preparation.
[29] L. Standker, P. Wobst, S. Mark, W.-G. Forssmann, FEBS Lett. 441 (1998) 281–286.
[30] L. Ständker, A. Enger, P. Schulz-Knappe, K. Wohn, M. Germer, M. Raida, W.G. Forssmann, K. Preissner, Eur. J. Biochem. 241 (1996) 557–563.
[31] K. Forssmann, D. Hock, F. Herbst, P. Schulz-Knappe, J. Talartschik, F. Scheler, W.G. Forssmann, Klin. Wochenschrift 64 (1986) 1276–1280.
[32] L. Ständker, R. Sillard, K. Bensch, A. Ruf, M. Raida, P. Schulz-Knappe, A. Schepky, H. Patscheke, W.G. Forssmann, Biochem. Biophys. Res. Commun. 215 (1995) 896–902.