Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

PORTLAND PRESS

Review Article

Reading patterns of proteome damage by glycation, oxidation and nitration: quantitation by stable isotopic dilution analysis LC-MS/MS

Naila Rabbani[1,2,3] and **Paul J. Thornalley**[4]

[1]Department of Basic Medical Science, College of Medicine, QU Health, Qatar University, P.O. Box 2713, Doha, Qatar; [2]AGE-Omics and Systems Biology Research Group, Warwick Medical School, Clinical Sciences Research Laboratories, University Hospital, Coventry CV2 2DX, U.K.; [3]Proteomics research technology platform, University of Warwick, Coventry CV4 7AL, U.K. ; [4]Diabetes Research Center, Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Qatar Foundation, P.O. Box 34110, Doha, Qatar

Correspondence: Naila Rabbani (n.rabbani@qu.edu.qa)

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) provides a high sensitivity, high specificity multiplexed method for concurrent detection of adducts formed by protein glycation, oxidation and nitration, also called AGEomics. Combined with stable isotopic dilution analysis, it provides for robust quantitation of protein glycation, oxidation and nitration adduct analytes. It is the reference method for such measurements. LC-MS/MS has been used to measure glycated, oxidized and nitrated amino acids – also called glycation, oxidation and nitration free adducts, with a concurrent quantitation of the amino acid metabolome in physiological fluids. Similar adduct residues in proteins may be quantitated with prior exhaustive enzymatic hydrolysis. It has also been applied to quantitation of other post-translation modifications, such as citrullination and formation of N_{ε}-(γ-glutamyl)lysine crosslink by transglutaminases. Application to cellular and extracellular proteins gives estimates of the steady-state levels of protein modification by glycation, oxidation and nitration, and measurement of the accumulation of glycation, oxidation and nitration adducts in cell culture medium and urinary excretion gives an indication of flux of adduct formation. Measurement of glycation, oxidation and nitration free adducts in plasma and urine provides for estimates of renal clearance of free adducts. Diagnostic potential in clinical studies has been enhanced by the combination of estimates of multiple adducts in optimized diagnostic algorithms by machine learning. Recent applications have been in early-stage detection of metabolic, vascular and renal disease, and arthritis, metabolic control and risk of developing vascular complication in diabetes, and a blood test for autism.

Introduction – historical challenges in the detection and quantitation of protein AGEomics (glycation, oxidation and nitration)

Proteins in physiological systems undergo continual spontaneous modification by glycating agents such as glucose, methylglyoxal (MG) and other saccharide derivatives forming glycation adducts. The physiological environment is often oxidizing, leading to the oxidative damage of proteins with the formation of protein oxidation adducts. There is also formation of reactive nitrogen species, such as peroxynitrite and nitryl chloride, leading to the formation of protein nitration adducts. Protein glycation, oxidation and nitration (GON) produces glycated, oxidized and nitrated amino acid residues in proteins that are typically in the range 0.1–0.0001% of the amino acid residue modified, depending on the type of adduct formed and protein substrate on which it is produced. After proteolysis, glycated, oxidized and nitrated amino acids

Received: 17 November 2019
Revised: 08 January 2020
Accepted: 13 January 2020

Version of Record published:
17 February 2020

Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

are formed – also called GON free adducts. These are found in plasma, urine and other physiological fluids in the range 0.1 nM to 10 μM. This low level and high dynamic range of modified amino acids posed a challenging analytical problem that was only solved conclusively after the year 2000 by application of stable isotopic dilution analysis liquid chromatography-tandem mass spectrometry (LC-MS/MS) in this area of biochemical research [1].

Prior to 2000, measurement of protein GON was challenging – excepting the measurement of glucose glycation adducts of hemoglobin and albumin which are of relatively high abundance and for which specific methods had been developed for the application of indicators of glycemic control in diabetes mellitus [2]. For other glycation adduct assay applications, fructosamine adducts formed in the early-stage reaction of protein glycation by glucose and advanced glycation endproducts (AGEs) formed in later-stage reactions, measurement was mainly by immunoassay using antibodies often of uncertain epitope specificity and affinity. Robust quantitation was difficult because there was often no quantitation of adducts in calibration standards and extent of glycation of standard antigens was much higher than of glycated protein analytes in physiological systems. Some steps in the immunoassay procedure were often done at high pH that may stimulate the formation of some glycation adducts and degrade others; and blocking proteins often used to decrease non-specific binding of the antibody contained glycation adducts. So, there were problems with specificity, sensitivity, analyte formation during pre-analytic processing, calibration and high background response – as previously reviewed [1]. Similar difficulties applied to detection of oxidized and nitrated proteins, with indirect methods used, such as binding of chromophoric or fluorophoric amines to carbonyl moieties [3] – reviewed in [4]. Other procedures involved measurement of non-tryptophan fluorescence as a measure of AGE fluorescence or oxidized fluorophores [5]. Exceptions where the detection of the AGE and intense fluorophore, pentosidine, by high-performance liquid chromatography with fluorimetric detection [6] and stable isotopic dilution analysis gas chromatography-liquid chromatography (GC-MS) of selected protein glycation and oxidation adducts [7]. The latter technique required chemical derivatization of analytes to confer the required volatility for detection that added complexity and vulnerability to change of analyte content during pre-analytic processing. These limitations of available analytical methods were restricting scientific advances – particularly in biomedical applications. Consensus statements on how best to deploy the methods available and guidance on the qualified interpretations of outcomes were made [8]. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) offered a markedly improved analytical approach.

Our involvement in analysis of protein GON began in 2001 when we purchased our first LC-MS/MS system. We had been using the high-performance liquid chromatography (HPLC)-based amino acid analysis method with pre-column derivatization under relatively mild chemical conditions with 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate (AQC) – pH 8.8 and 55°C for 10 min [9,10] – and used the same protocol with detection of the AQC adducts by tandem mass spectrometry. To our disappointment, we found even this mild derivatization produced migration of the MG moiety in the AGE, hydroimidazolone MG-H1, from arg to stable isotopic standard [$^{15}N_2$]arg, thereby compromising the use of stable isotopic dilution analysis. We thereafter sought to analyze GON adducts directly without pre-column derivatization. One of the few chromatographic stationary phases that retained underivatized amino acids without the added complexity of using an ion pair reagent was Hypercarb™ graphitic chromatography columns. With Hypercarb™ chromatography columns, we could readily retain highly hydrophilic amino acid analytes on column whilst non-volatile salts in samples was diverted to waste prior to mass spectrometric data collection. We used two columns in series with column switching to easily elute analytes with a diverse range of hydrophobicity, cleaned the Hypercarb™ columns between each injection by elution with 50% tetrahydrofuran and established an effective re-equilibration protocol (elution with initial mobile phase for 15 min at twice the flow rate used in data collection, 0.4 ml/min). This produced reproducible chromatography, stable analyte ionization and long working life of the Hypercarb™ columns. This is the basis of the quantitation of protein GON adducts by LC-MS/MS that we have used in subsequent studies [11]. Our methods have been adopted, reproduced and used by other research teams.

Application of stable isotopic dilution analysis LC-MS/MS for quantitation of protein GON has advanced research profoundly. We now have reliable estimates of steady-state levels of a comprehensive range of GON adduct residues in cellular and extracellular proteins of physiological systems. We have robust estimates of the physiological concentration range of GON free adducts in plasma, cerebrospinal fluid, synovial fluid, urine, dialysis fluids, foods, beverages and plants [11–18]. We have also steady-state levels of protein GON adducts in protein of cultured cells and *in situ* rate of protein modification estimated from the flux of release of protein, glycation, oxidation and nitration free adducts into the culture medium [19,20]. This has enabled appropriate design of experiments to study the effect of protein and cell function by challenge with proteins glycated, oxidized and nitrated to physiological extents and in a concentration range appropriate for physiological relevance. This was very difficult to judge prior to the quantitative estimates for individual protein GON adducts provided by LC-MS/MS. Our studies on analysis of GON free adducts in cell culture

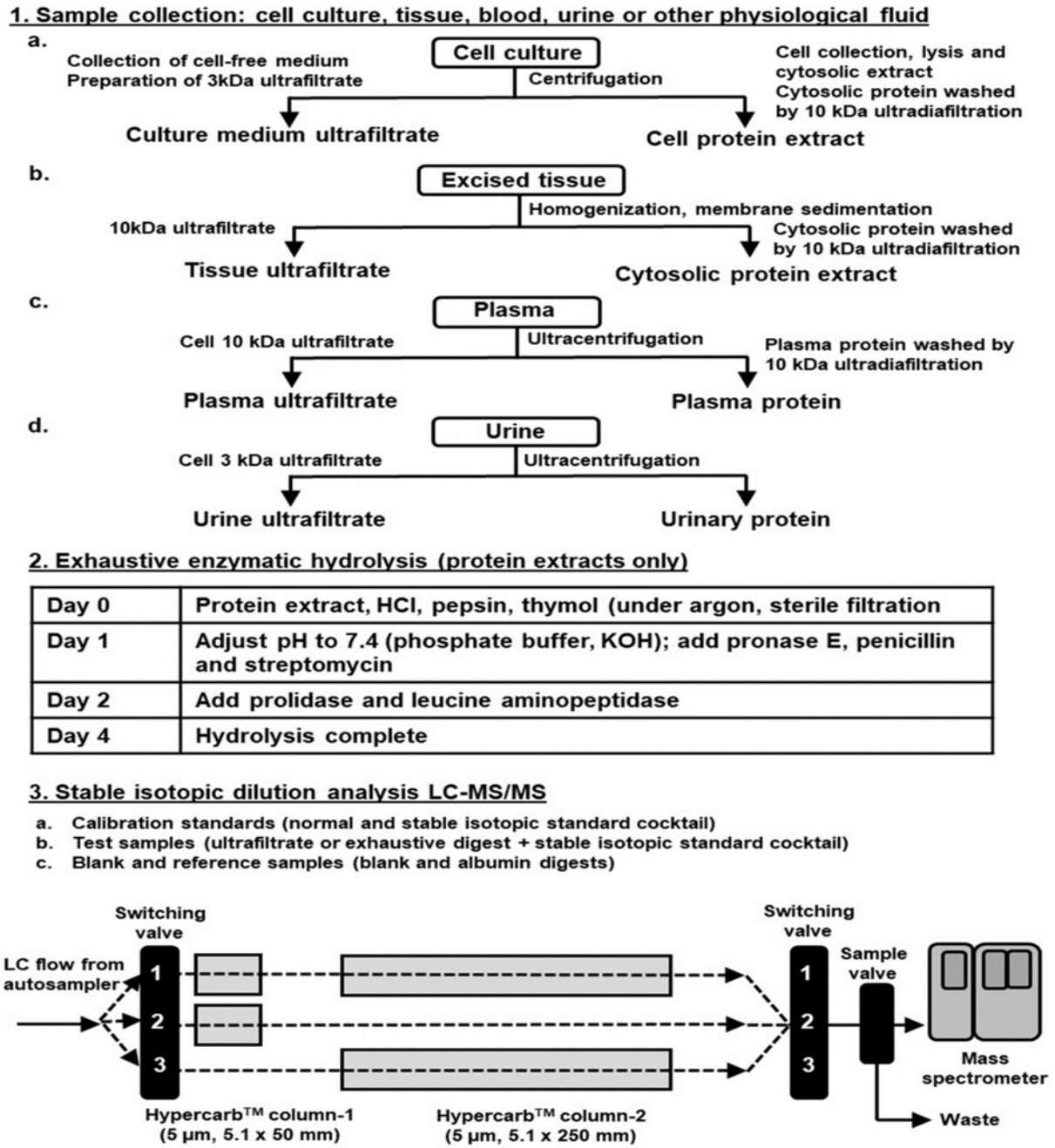


Figure 1. Assay of protein glycation, oxidation and nitration adducts by liquid chromatography-tandem mass spectrometry

(**1**, **a–d**) Sample collection and processing for GON adduct residues and free adducts. (**2**) Exhaustive enzymatic hydrolysis schedule. (**3**) Stable isotopic dilution analysis of GON adducts – illustrating column switching for both or either of 2 Hypercarb™ columns in line (elution sequence: 1, 2 and 3).

systems, laboratory animal experiments and clinical studies of human subjects revealed how there is a continuous release of GON adducts from proteins by cellular proteolysis, transport of GON free adducts from cells into tissue fluids and plasma, and filtration from plasma with their excretion in urine [13,16,20,21]. We were also recently able to estimate the amounts of glycation free adducts absorbed from food [22].

Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

Assay of protein glycation, oxidation and nitration adduct residues of proteins and free adducts in physiological fluids

The basis of the LC-MS/MS detection of protein GON adducts is: (i) addition of a cocktail of stable isotope substituted standard analytes to the sample matrix, (ii) reversed phase ultrahigh performance liquid chromatography (UPLC) of analytes in sample matrix on Hypercarb™ graphitic columns with aqueous 0.1% trifluoracetic acid mobile phase and custom acetonitrile gradient elution (Figure 1); and (iii) detection of adducts, natural and stable isotopic abundance, by positive electrospray ionization with triple quadrupole mass spectrometric detection operating in multiple reaction monitoring (MRM) mode [11,23]. The detection is specific for chromatographic retention time and m/z values of the molecular ion and major fragment ion producing by collision with argon atoms in the collision cell. Under the experimental conditions used, the analytes are detected with single positive charge, $z = 1$. With use of the latest range of high sensitivity tandem mass spectrometers, limits of detection of amino acid analytes range from low 100 attomol to low fmol, depending on analyte. There is also typically the option to collect data on 200–500 MRM mass transitions, covering detection of many analytes and related stable isotopic substituted standards in each analytical run. Typically, 2 or more MRM transitions are optimized for each analyte for surety of detection where the fragmentation and mass transition most specific for the analyte is used for analyte quantitation. The characteristics of the quantitation MRMs for protein GON adducts are given in Table 1. The detailed protocol of the method used in our laboratory has been published previously and advantages compared with other methods [1,11,23]. The method is also applicable for the assay of advanced lipoxidation products (ALEs) [24,25]. AGEs derived from glyoxal are considered to be ALEs: glyoxal-derived adducts are G-H1, CML, CMA and GOLD – assay characteristics for these are given herein. Other ALEs are S-carboxymethyl-cysteine and stable adducts derived from malondialdehyde and 4-hydroxy-nonenal. Assay of these adducts by the current method remains to be validated.

A future challenge is to make LC-MS/MS analysis of protein GON adducts more widely available in research laboratories. This may be achieved by advances in automation of LC-MS/MS and improved commercial availability of analytical standards – including stable isotopic standards.

Measurements of protein glycation, oxidation and nitration adduct residues and free adducts by LC-MS/MS – pre-analytic processing

GON adducts in proteins were initially considered to be post-translational modifications that accumulated throughout life. Whilst this occurs for chemically stable modifications on long-lived proteins, it is now appreciated that FL and some AGEs (e.g. hydroimidazolones) have a relatively short half-life *in vivo* and they and other GON modifications are formed on cellular proteins or short-lived extracellular proteins [9,11]. Proteolysis of glycated, oxidized and nitrated proteins released glycated, oxidized and nitrated amino acids – GON free adducts [11]. Protein GON free adducts are released into plasma from cells, readily filtered in renal glomeruli (their molecular masses being $<$500 Da) and – unless repaired or further metabolized – they are excreted in the urine. There are two major protocols of pre-analytic processing in the LC-MS/MS-based assay: one for assay of GON free adducts and another for assay of GON adduct residues of protein.

For protein GON free adduct measurement in plasma and urine, ultrafiltrate is prepared by microspin ultrafiltration over 10 and 3 kDa cut-off ultrafilters, respectively. The ultrafiltrate is then mixed with isotopic standard mixture and analysed by LC-MS/MS. We have also used the 10 kDa cut-off filter for ultrafiltrates of synovial fluid, cerebrospinal fluid and homogenates of tissues and foodstuffs. This is the simplest pre-analytic processing method. It provides for quick and simple analysis of plasma and urine for clinical diagnostic applications and may be readily automated for the clinical chemistry laboratory. The free adduct estimates are determined in concentration units in the physiological fluid tested. Estimates in plasma and urine can be used to calculate renal clearance of GON free adducts [13,16].

For assay of protein GON adduct residues, pre-analytic processing is more complex - involving washing of the protein sample by diafiltration (cycles of concentration and dilution with water over a 10 kDa microspin filter) and exhaustive enzymatic hydrolysis under antioxidant conditions (under argon and addition of thymol antioxidant). Prior to enzymatic hydrolysis, washed protein extract is delipidated by extraction with water-saturated diethylether. Enzymatic digestion is performed under argon with automated addition of sterile filtered reagents to minimize bacterial contamination. Antibiotics are also added to suppress bacterial growth. Metal ion chelators such as diethylenetriaminepenta-acetic acid (DETAPAC) cannot be used as they inhibit proteolytic digestion. In this procedure, the following proteases are added sequentially: pepsin, pronase E and together, prolidase and aminopeptidase.

Table 1 Mass spectrometric multiple reaction monitoring detection of amino acids, glycated, oxidized and nitrated amino acids and other protein modification related analytes

Analyte group	Analyte	R_t (min)	Molecular ion (Da)	Fragment ion (Da)	CE (eV)	Neutral fragment loss (es)	Isotopic standard
Amino acids							
	Ala	5.2	90.1	44.1	8	H_2CO_2	[$^2H_3$]Ala
	Arg	29.2	175.2	70.3	15	H_2CO_2, $NH_2C(=NH)NH_2$	[$^{15}N_2$]Arg
	Asn	7.2	133.2	74.1	14	CH_3CONH_2	[$^{13}C_4$]Asn
	Asp	7.5	134.1	88.0	10	H_2CO_2	[$^{13}C_2$,$^{15}N_1$]Asp
	Cys	6.6	122.0	59.0	18	H_2CO_2 + NH_3	[$^{13}C_3$,$^{15}N_1$]Cys
	Cystine	32.6	241.1	120.1	20	$C_3H_7NO_2S$ (cys)	[3,3,3′,3′- $^2H_4$]Cystine
	Gln	9.9	147.2	84.1	16	H_2CO_2 + NH_3	[$^{13}C_5$]Gln
	Glu	28.4	148.1	102.1	10	H_2CO_2	[$^{13}C_5$]Glu
	Gly	4.8	76.2	30.1	6	H_2CO_2	[$^{13}C_2$ $^{15}N_1$]Gly
	His	32.6	156.1	93.0	22	H_2CO_2 + NH_3	[$^{13}C_6$]His
	Hyp	5.6	132.0	86.1	12	H_2CO_2	4,5-[$^{13}C_2$]Hyp
	Ile	31.5	132.3	86.2	10	H_2CO_2	[$^{13}C_6$]Ile
	Leu	27.6	132.3	86.2	10	H_2CO_2	[$^2H_3$]Leu
	Lys	5.5	147.1	84.3	15	H_2CO_2, NH_3	[$^{13}C_6$]Lys
	Met	29.5	150.0	104.2	11	H_2CO_2	[$^2H_3$]Met
	Phe	17.2	166.1	103.1	26	H_2CO_2, NH_3	[*ring*-$^2H_5$]Phe
	Pro	6.6	116.1	70.1	12	H_2CO_2	[$^{13}C_5$]Pro
	Ser	5.2	106.0	42.0	18	H_2CO_2, H_2O	[$^{13}C_3$]Ser
	Thr	5.9	120.1	56.1	14	H_2CO_2, H_2O	[$^{13}C_4$]Thr
	Tyr	18.3	182.1	136.2	13	H_2CO_2	[$^2H_4$]Tyr
	Trp	23.5	205.0	159.1	15	H_2CO_2	[$^{15}N_2$]Trp
	Val	8.6	117.8	72.0	19	H_2CO_2	[$^2H_8$]Val
Early-stage glycation							
Fructosamines	FL	28.5	291.0	84.3	31	H_2CO_2, fructosylamine	[$^2H_4$]FL
AGEs							
Hydroimidazolones†	G-H1	12.4	215.0	100.2	14	$NH_2CH(CO_2H)CH_2CH=CH_2$	[$^{15}N_2$]G-H1
	MG-H1	11.6 & 12.5	229.2	114.3	14	$NH_2CH(CO_2H)CH_2CH=CH_2$	[$^{15}N_2$]MG-H1
Hydroimidazolones†	3DG-H	11.2, 12.6 & 13.5	319.1	114.8	20	$NH_2CH(CO_2H)CH_2CH=CH_2$	[$^{15}N_2$]3DG-H
Monolysyl AGEs	CEL	28.8	219.2	130.1	13	$NH_2CH(CH_3)CO_2H$	[$^{13}C_6$]CEL
	CML	28.5	204.9	130.1	12	$NH_2CH_2CO_2H$	[$^{13}C_6$]CML
	Pyrraline	17.9	255.2	84.3	23	2-CHO-5-HOCH$_2$-pyrrole, H_2CO_2	[$^{13}C_6$, $^2N_2$]Pyrraline
Fluorescent AGEs	Argpyrimidine	17.7	255.3	140.3	17	$NH_2CH(CO_2H)CH_2CH=CH_2$	[$^{15}N_2$]argpyrimidine
	Pentosidine‡	21.1	379.3	250.4	22	$NH_2CH(CO_2H)CH_2CH_2CH=CH_2$	[$^{13}C_6$]pentosidine
Non-fluorescent crosslinks	GOLD	11.7	327.1	198.3	21	$NH_2CH(CO_2H)CH_2CH_2CH=CH_2$	[$^2H_8$]GOLD
	MOLD	14.0	341.2	212.3	21	$NH_2CH(CO_2H)CH_2CH_2CH=CH_2$	[$^2H_8$]MOLD
	Glucosepane	16.5	429.2	382.1	38	C_2H_5O	[$^{13}C_6$]Glucosepane
Other	CMA	12.1	233.0	70.1	27	H_2CO_2, $NH_2C(=NH)NHCH_2CO_2H$	[$^{13}C_2$]CMA
	Orn	5.2	133.1	70.1	9	H_2CO_2, NH_3	[$^2H_6$]Orn
Oxidative damage							
Met oxidation	MetSO	8.7	166.1	102.2	14	CH_3-SOH	[$^2H_3$]MetSO
Tyr oxidation	Dityrosine	19.9	361.2	315.3	15	H_2CO_2	[$^2H_6$]DT
Tyr oxidation	NFK	21.5	235.8	191.2	18	H_2CO_2	[$^{15}N_2$]NFK
Lys oxidation	AASA	10.7	128.0	82.0	15	H_2CO_2	AAA (R_t = 29.4 min)
Arg and Pro oxidation	GSA	32.2	114.0	68.0	15	H_2CO_2	AAA
Nitration damage							
Tyrosine nitration	3-NT	23.2	227.1	181.2	13	H_2CO_2	[$^2H_3$]3-NT
Other							
Transglutaminase crosslinking	N_ε-(γ-Glutamyl)lysine (GEEK)	9.3	276.1	147.1	12	$NH_2CH(CO_2H)CH_2CH=C=O$	Nε-(γ-[$^{13}C_5$]Glutamyl)lysine
Citrullination	Citrulline	13.5	176.1	70.1	21	H_2CO_2, $NH_2C(=O)NH_2$	[5-^{13}C,4,4,5,5-$^2H_4$]Citrulline

†For hydroimidazolones, R_t values for the two epimers if MG-H1 are given and of the three structural isomers of 3DG-H, 3DG-H1, 3DG-H2 and 3DG-H3, are all detected. From [1]. ‡ Pentosidine is detected to higher sensitivity by in-line fluorimetry, excitation wavelength 320 nm, emission wavelength 365 nm. GEEK and citrulline were detected with custom elution gradients to resolve from isobaric interferences: for GEEK, from glu-lys and lys-glu dipeptides [68]; and for citrulline, from arg which partly degrades to citrulline in the electrospray ionization source [50]. Data from [50,67,69].

Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

Where lipoproteins and collagen is the substrate protein, the pepsin step is omitted for lipoproteins [26,27] and replaced by initial incubation for 24 h with collagenase for collagen [28]. For hemoglobin, digests were performed under carbon monoxide instead of argon to make the released heme inert with respect to catalysis of oxidation reactions [12]. Automation was achieved using a relatively low-cost, stand-alone sample autoprocessor, such as produced by CTC Analytics (Zwingen, Switzerland). The use of enzymatic hydrolysis of proteins enables measurement of acid or base hydrolysis labile analytes. The hydrolysis process is exhaustive and runs over 4 days. The protocol avoids oxidative degradation of protein adduct residues and subsequent overestimation of oxidation adducts. Assay of GON adducts in the hydrolysate is normalized to corresponding amino acid (mmol analyte per mol unmodified amino acid) or protein (pmol analyte per mg protein).

Although the use of exhaustive enzymatic hydrolysis has advantages, it introduces a potential source of error through contribution to analyte estimates by autohydrolysis of proteolytic enzymes. A robust basis for correction of the contribution to protein GON adducts by autohydrolysis of proteases was required. Hydrolysis in protein free blanks overestimates this contribution since autohydrolysis of proteases is faster when no protein substrate is present. The correction method uses triplicate digests of zero protein blanks and a known amount of human serum albumin (HSA) and quantitation of an amino acid that is not modified – valine. In HSA digests, the amount of valine detected is that from HSA + autohydrolysis of proteases ($V_{\mathrm{HSA}} + V_{\mathrm{Proteases\ (Protein)}}$). In the zero protein blanks, the valine detected is only from autohydrolysis ($V_{\mathrm{Proteases\ (Blank)}}$). The valine liberated from proteases in the presence of HSA is decreased by the suppression of autohydrolysis by presence of HSA. The factor $V_{\mathrm{Proteases\ (Protein)}}/V_{\mathrm{Proteases\ (Blank)}}$, typically approximately 0.7, is applied to the amount of protein GON adduct quantified in the blank to produce corrections applied to digests of sample protein. That is, the presence of protein substrate slows autohydrolysis of proteases by approximately 30%. This correction factor is determined for each sample batch for enzymatic hydrolysis samples.

The efficiency of exhaustive enzymatic digestion of proteins is >95% when based on recovery of amino acids from proteins of known molecular mass and amino acid composition. It is essentially quantitative when applied to proteins with low extent of glycation as found *in vivo* [10]. This is evidenced by quantitative recovery of amino acids and recovery of all protein modifications produced by a particular modifying agent when total adduct formation is also estimated independently by use of a radiolabeled modifying agent [10,29]. For example, summation of all MG-derived AGEs by the LC-MS/MS techniques compares well with total MG-derived adduct formation estimated by use of [^{14}C]labelled MG for analysis of proteins minimally modified by MG [28,29]. Exhaustive enzymatic digestion gives poor recoveries when proteins are highly glycated *in vitro* – such as prepared by a commonly used protocol for preparation of AGE-modified albumin which had approximately 40 equivalents of modification (as assessed by decrease in unmodified lysine and arginine residues) and a mass increment of albumin of approximately 7 kDa [30,31]. Such highly modified proteins are not found *in vivo* [32] – although may be present in some foodstuffs. Therefore, generally the exhaustive enzymatic digestion technique is appropriate for use in analysis of protein extracts of physiological samples but caution and further investigation of analyte recovery is required in foodstuffs when high extent of protein modification is suspected.

A further critical feature for reliability of analyte estimation is stability of analytes in storage and during pre-analytic processing. Stabilities of GON analytes vary. In general stability is decreased by high pH and high temperature; these conditions should be avoided for storage and processing of analytical standards and collection, storage and processing of test samples. Stocks of standard and internal standard analytes are stored as pure solids or as concentrated neutral and acidic solutions at −80°C. This provides stability of analytes for >10 years. Working solutions of standard analytes are diluted in initial mobile phase, 0.1% trifluoracetic acid, and kept on ice during preparation of batches of test samples and calibration standards for subsequent LC-MS/MS analyses. Aliquots of standard analyte solutions are mixed to prepare standard analyte cocktails. Aliquots of stable isotope-substituted standard cocktail are mixed with aliquots of standard analyte cocktail and sample extracts to prepare calibration and test samples for analysis, respectively. Calibration and test samples are placed at 4°C in the autosampler of the LC-MS/MS until analysis. This provides for good chemical stability of analytes in calibration standard and test samples during batch analysis of up to 50 test samples, 14 calibration standards, reference samples and blanks over 3 days. For collection of samples, it has to be kept in mind that cultured cells, physiological tissues, fluids and blood samples contain GON protein adduct residues and free adducts in the presence of excess unmodified protein and amino acid precursors and low steady-state levels of glycating agents, ROS and reactive nitrogen species. If the samples are not fractionated and chilled rapidly after collection, there is a risk of further GON analyte formation *ex vivo*. Samples should be processed immediately and not later than 1 h after collection for blood samples – except for certain analytes (see below). Sample fractionation is implemented where required: for example, sedimentation by centrifugation of cells in cultured cell suspensions or blood samples. When processing is complete, samples are snap frozen and stored at −80°C until analysis. For preparation of serum, blood is incubated for 30 min at ambient temperature to allow clotting, centrifuged (2000***g***, 10 min)

and serum removed immediately. A critical time-limiting feature for sample processing is rate of sample oxidation; for example, as judged by loss of reduced glutathione (GSH) and other thiols that leads to subsequent increase in protein oxidation adducts. Processing of blood samples within 1 h of sample collection maintains >70% GSH and thereby avoids artefactual increase in protein oxidation adducts in sample collection [33,34]. Repeated freeze thawing is also avoided as this increases estimates of protein oxidation adduct, methionine sulfoxide (MetSO) [35]. Shipment on dry ice should be used to transfer samples between laboratories. For surety of long-term storage of clinical or other samples, it is preferable to assay GON adducts in a representative pilot scale number of control samples from storage and compare the analyte estimates to reference estimates before commencing a large study with multiple batch sample analysis.

Validation of LC-MS/MS assay of protein glycation, oxidation, nitration adducts for application in epidemiological studies – Nurses' Health Study and the Health Professionals Follow-up Study

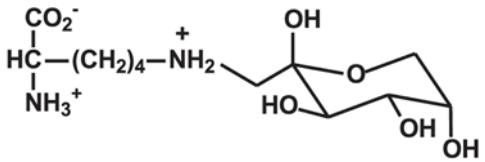
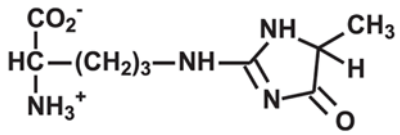
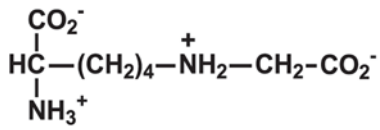
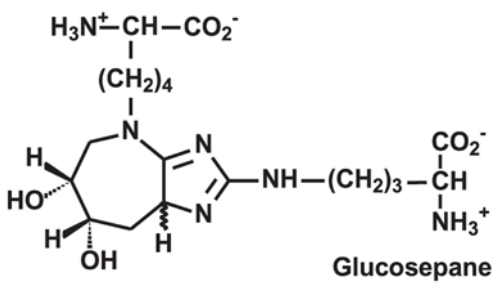
We conducted a reliability study of the LC-MS/MS assays for 12 adducts in stored repository blood samples from the Nurses' Health Study (NHS) and the Health Professionals Follow-up Study (HPFS). These adducts were: FL, CML, MG-H1, CMA, pentosidine, glucosepane, MetSO, AASA, GSA, dityrosine, NFK and 3-NT. In this investigation, we performed three studies to determine the suitability of our LC-MS/MS platform for measuring the 12 adducts in the archived human plasma from healthy volunteers and from participants of the NHS and the HPFS. Details of the sample collection and processing protocols prior to storage in the biorepository have been described elsewhere [36]. Briefly, the plasma samples were collected from participants in the NHS (n = 32,826; 1989–1990) and the HPFS (n = 18,159; 1993–1995) [36,37]. Procedures in the NHS and the HPFS were identical, except that the blood samples were collected in tubes with sodium heparin for the NHS and with EDTA for the HPFS. Samples were shipped with an icepack in a Styrofoam container to the biorepository via overnight courier. On arrival, blood samples were processed (centrifuged and divided into aliquots of plasma), and then stored in the vapor phase of liquid nitrogen freezers at less than −130°C. The three studies were designed to evaluate sources of measurement error of particular concern in epidemiologic studies, including: (i) blinded duplicate samples study – samples processed after a 24 h delay, (ii) sample with delayed processing – where whole blood samples were processed after 0, 24 or 48 h delay and then frozen at −130°C, and (iii) within-person reproducibility study over time – assessing whether a single analyte measurement reasonably reflected a measurement obtained 1–2 years later (testing the physiological stability of the analyte of this extended time period). These studies were performed to assess the applicability of the LC-MS/MS assay in future epidemiologic studies of age-related diseases in the NHS and HPFS, where most samples were shipped and processed with a delay and stored for long periods. The reliability study showed that 9 of the 12 analytes passed the blinded duplicate study and MG-H1 and glucosepane also passed the sample with delayed processing and the within-person reproducibility study. These data suggest that the LC-MS/MS platform for measuring MG-H1 and glucosepane in the archived human plasma from the NHS and the HPFS qualifies it for use in future large-scale epidemiological studies. It may also be useful for trials. All analytes passed for glycation and oxidation free adducts for the blinded duplicate samples study; 3-NT free adduct failed (unpublished observations).

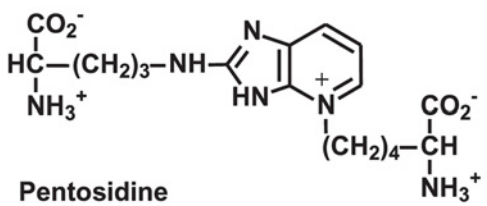
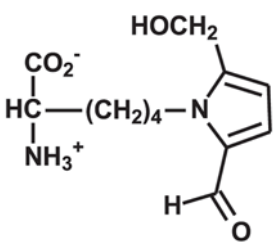
Protein AGEomics and other modifications as reporters of metabolic and physiological processes

Protein modifications may be reporters of metabolic and physiological processes – Table 2. The most well-established evidence of this is the clinical use of glycated hemoglobin and plasma glycated albumin as markers of glycemic control in diabetes [2,38]. Levels of the major glycation adduct of methylglyoxal, hydroimidazolone MG-H1, are linked to fasting and postprandial glucose exposure, insulin resistance and cardiovascular disease [12,22,39,40]. The AGE, CML, is mainly formed by oxidative degradation of FL: the CML/FL ratio in proteins is therefore a marker of oxidative stress [41]. Glucosepane is the major quantitative crosslink formed in protein glycation, formed from FL residues [42]. It is a marker of glycemic control [43] and the major glycation adduct released from joint proteins during early-stage development of osteoarthritis [41]. Pentosidine is pentose sugar-derived AGE and intense fluorophore. Levels of pentosidine are considered to reflect pentosephosphate pathway activity [44]. Pyrraline free adduct in plasma and urine is a glucose-derived AGE formed at high temperatures of culinary processing, originating only from food. Urinary pyrraline free adduct excretion is linked to amount of food consumed and intestinal permeability [45,46].

Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

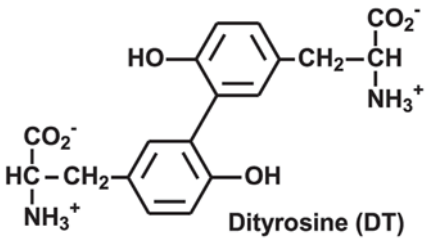
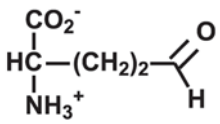
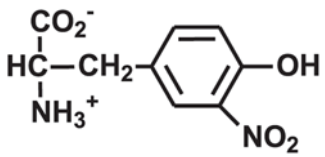
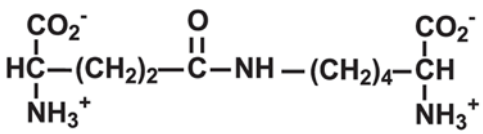
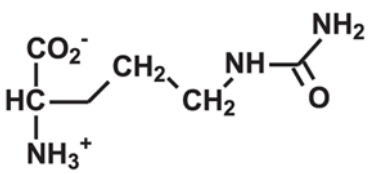
Table 2 Protein AGEomics and other modifications as reporters of metabolic and physiological processes

Protein modification	Analyte	Comment
Glycation	 N_ε-(1-Deoxy-D-fructos-1-yl)lysine (FL)	Early-stage glycation adduct [1]. Formed from glucose non-enzymatically and reporting on exposure to increased glucose concentration. Repaired intracellularly by fructosamine 3-phosphokinase [70]. Free adduct absorbed after digestion of food proteins [71].
	 MG-derived hydroimidazolone (MG-H1)	A major quantitative arginine-derived AGE formed from methylglyoxal. Linked to increased fasting and postprandial glucose exposure, insulin resistance and cardiovascular disease [12,22,39,40]. Free adduct absorbed after digestion of food proteins [22].
	 N_ε-Carboxymethyl-lysine (CML)	A major quantitative lysine-derived AGE – particularly in food. Formed by the oxidative degradation of FL and from other sources. Free adduct absorbed after digestion of food proteins [72]. The ratio of CML to FL is an indicator of oxidative stress [73].
	 Glucosepane	Major quantitative crosslink formed in protein glycation [42]. It is a marker of glycemic control [43] and the major glycation adduct released from joint proteins during early-stage development of osteoarthritis [41].

Metabolite class	Urinary metabolite	Comment
Glycation	 Pentosidine	Low-level pentose sugar-derived glycation crosslink and intense fluorophore. Considered to reflect pentosephosphate pathway activity [44].
	 Pyrraline	Glucose-derived AGE formed at high temperatures of culinary processing; originating only from food [45,46]. Urinary-free adduct excretion is linked to amount of food consumed and intestinal permeability.

Continued over

Table 2 Protein AGEomics and other modifications as reporters of metabolic and physiological processes (Continued)

Metabolite class	Urinary metabolite	Comment
Oxidation	 Dityrosine (DT)	Oxidative crosslink formed spontaneously in oxidative stress and enzymatically by dual oxidase (DUOX) [1,47]. Linked to host immunity.
	 Glutamic semialdehyde (GSA)	Major 'protein carbonyl' formed by the oxidative de-guanidylation of arginine and oxidative ring-opening of proline [74].
Nitration	 3-Nitrotyrosine (3-NT)	Major product of proteins endogenously nitrated by peroxynitrite and nitryl chloride [1,75]. May be a reporter of oxidative damage and also of bioavailability of nitric oxide.
Transglutaminase-linked crosslink	 Nε-(γ-glutamyl)lysine (GEEK)	Major protein crosslink formed enzymatically by transglutaminases [76].
Citrullination		Formed in proteins by arginine deiminases. Formed in early stage osteoarthritis and rheumatoid arthritis; immunogenic only in rheumatoid arthritis [15,50]

Molecular structures showing ionization under physiological conditions.

For markers of oxidative damage, dityrosine is formed spontaneously in oxidative stress and enzymatically by dual oxidase (DUOX) [1,47]. DUOX has an important role in gut mucosal immunity, host–microbe homeostasis and signaling for neutrophil recruitment into allergic pulmonary airways [48,49]. For protein nitration, 3-nitrotyrosine levels may reflect oxidative stress and also bioavailability of nitric oxide. For other modifications, protein citrullination is catalyzed by arginine deiminases with increased expression in inflammatory processes driving early-stage arthritis. Protein citrullination is increased in both early stage osteoarthritis and rheumatoid arthritis and is immunogenic only in rheumatoid arthritis [15,50]. Levels of citrullinated protein and auto-antibodies to citrullinated protein are thereby useful in diagnosis and classifying type of early stage arthritis [15,50].

Use of machine learning in combination of protein AGEomics for clinical diagnosis

Protein GON occurs endogenously in healthy human subjects. It may also increase and sometimes decrease in abnormal health conditions and disease. For example, decreased glycation of plasma protein in obesity and early-stage arthritis linked to increased transcapillary escape rate of albumin and increased residence time of albumin in tissue

Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

fluid [15,51]. Often change of a single protein GON adduct may not be specific to a health condition or disease. Moreover, it may be advantageous to combine biomarkers of protein GON with other biomarkers to improve diagnostic performance. For example, combination of plasma concentrations of protein GON free adducts with plasma concentration of bone resorption marker hydroxyproline and early-stage rheumatoid arthritis marker, anti-cyclic citrullinated peptide (CCP) antibody status in the detection and classification of early-stage arthritis [15,50]. In this case, combination of different adducts with optimized weighting may be used to distinguish or classify cases and controls. This is produced by computer-based artificial intelligence machine learning and involves the development of algorithms for optimum data classification.

In a typical study using machine learning, an algorithm is optimized or trained for classifying case and control samples on a sample analyte estimate dataset – the training dataset. The algorithm performance is then validated or tested on an independent dataset or test set. The algorithm may be internally validated by leave-one-out routines where repeatedly one datum is randomly omitted from the algorithm training and 2-fold or similar cross-validation where repeatedly half of the dataset is selected at random to train the algorithm and the remaining half is used for algorithm testing. Such routines are typically run 50–100 times for algorithm optimization. Where the outcome classifications are set by the investigator – such as with or without a particular disorder or disease (as applies herein), this is called supervised machine learning. Common algorithms often used are: Generalized Linear Model (GLM); GLM with elastic net (GLMNET) – allows for non-normal data distributions and non-linear relationship between the classification response and the predicting features; Random Forests – a nonlinear, decision tree-based method; and Support Vector Machines – which uses hyperplanes and maximum margins for optimum classification [52,53]. The strengths of using the machine learning approach in producing the diagnostic algorithm are that it provides a data-driven discovery of complex combinations of independent analytes and other variables or features and their relative influence on the dependent variable of interest (for example, diagnosis of disease) by logical, step-wise analysis without preconception and bias. The limitations are: access to application specific data – variables recorded and dataset size; and, as a post hoc analysis, all data collection must be completed before implementation. It also requires expertise for awareness of variables available for inclusion in the analysis – which also applies to other types of data analysis.

To develop a secure diagnostic algorithm, it is important to have reliable and robust analytical data for algorithm training and testing. The particular advantages of features in algorithms of GON adduct residues of proteins is that these analytes provide a report of the *in situ* rate of protein modification in the physiological compartment analyzed and over the lifespan of the major protein studied. For example, for plasma samples this is mainly the vascular compartment for the lifespan of albumin. The particular advantages of features in algorithms of GON free adducts is that these analytes are not re-incorporated into proteins and thereby changes in rates of formation are more easily resolved. For example, the changes in GON free adducts found in early-stage arthritis where no significant changes in related unmodified amino acids are found. Protein GON analytes are also often found to be risk predictors and/or risk factors of disease where GON processes are involved in the pathogenesis. We have applied this approach to develop diagnostic algorithms for early-stage diagnosis and classification by type of arthritis, diagnosis of autism spectrum disorder (ASD) and diagnosis of early-stage decline in metabolic health, vascular health and renal health – Table 3.

Diagnostic algorithm for autism spectrum disorder (ASD)

In a recent study we applied machine learning selection and optimum combination of protein GON adducts in plasma protein and free adducts in plasma and urine for the diagnosis of autism spectrum disorder (ASD) or autism. Autism is a developmental disorder of childhood that is considered to affect 12 million children worldwide with a global prevalence of 0.62% [54]. The prevalence is much higher in the U.S.A. at 2.47% [55] and 1.15% in Europe [56]. There are long delays of up to 4 years for referral of children with suspected autism to experts in childhood development for diagnosis by behavioral tests and observational studies. If autism is diagnosed early, intervention with counselling and cognitive restructuring can produce remission [57]. In a study of 69 children with and without autism, we found higher plasma protein content of CML, CMA and DT and lower 3DG-H adduct residues in plasma protein of children with autism, compared with children with normal development (Table 3). A diagnostic algorithm combining these analytes gave a test with accuracy 88%, sensitivity 0.92, specificity 0.84 and thereby diagnostic odds ratio = 68. This requires further validation. A simple blood test may meet the currently unmet need for improved availability of autism diagnosis.

In this report [58], the study design was cross-sectional and so inferences on causal relationships between the features in the diagnostic algorithms and autism cannot be made. Nevertheless, the features – particularly the algorithm of highest accuracy – provide suggestions to explore further in experimental mechanistic and interventional studies

Table 3 Protein glycation, oxidation and nitration adducts as features in algorithms for clinical diagnosis

Health condition or disease	Algorithm classifications (no)	Sensitivity	Specificity	Random outcome
Autism spectrum disorder (ASD)	Normal development versus ASD (2)	0.92	0.84	0.50
Early-stage arthritis – diagnosis and classification	Healthy control, eOA, eRA and non-RA (4)	Control: 0.41 eOA: 0.73 eRA: 0.57 non-RA: 0.25	Control: 0.75 eOA: 0.87 eRA: 0.91 non-RA: 0.76	0.25
Early-stage arthritis – diagnosis and classification	Stage 1: healthy control versus arthritis - any type (2) Stage 2: eOA, eRA and non-RA (3)	0.92 eOA: 0.92 eRA: 0.80 non-RA: 0.70	0.91 eOA: 0.90 eRA: 0.78 non-RA: 0.65	0.50 0.33
Early-stage impaired metabolic health	Heathy control versus prediabetes and/or insulin resistance (2)	0.80	0.87	0.50
Early-stage impaired vascular health Early-stage impaired renal health	Heathy control versus asymptomatic CVD (2) Heathy control versus stage 3 – 4 chronic kidney disease (2)	0.69 0.88	0.75 0.92	0.50 0.50

as to the basis of the association: diagnostic marker, diagnostic predictor or mechanistic risk factor. We also noted that the diagnostic algorithm of highest accuracy was based on features of AGE and dityrosine adduct residues of plasma protein – Algorithm-1 in the study. Protein glycation and oxidation adducts from dietary protein do not contribute to levels of glycation and oxidation adduct residues of protein; the latter reflecting rather rates of endogenous glycation and oxidation of protein in mainly the vascular compartment [51]. The dominance of plasma protein AGE and oxidation adducts in Algorithm-1 and the modest improvement by addition of plasma protein free adducts – which have contributions from glycated and oxidized proteins of dietary origin [22,41] – may indicate that there is limited influence of dietary glycated and oxidized proteins on the diagnosis of ASD by this blood test.

Diagnostic algorithms for early detection and typing of arthritis

Musculoskeletal disease is the most common cause of chronic disability worldwide, of which osteoarthritis (OA) accounts for 1.1% and rheumatoid arthritis (RA) 0.3% of total global disability adjusted life years [59]. If OA is detected early and subjects identified for whom the disease progresses to severe symptoms relatively rapidly, lifestyle and other interventions may be made and targeted to individuals at risk to prevent disability. If RA is detected at the early stages, treatment with disease modifying drugs may produce complete remission from disease [60]. The challenge is to detect and classify arthritis at the early stages. Rheumatoid arthritis has long been considered a disease associated with increased protein oxidation and nitration [61,62]. In a study of early-stage OA (eOA) of the knee, detected by arthroscopy before any damage to joint could be discerned by radiography, and early stage inflammatory arthritis – early stage RA (eRA) and other inflammatory arthritic disease which is often self-resolving (non-RA), we found higher citrullinated protein (CP) in plasma of eOA and eRA, higher Hyp of eOA and non-RA and anti-CCP antibody positivity of eRA, with respect to normal healthy controls. We developed a 4-classification algorithm to detect and distinguish eOA, eRA and non-RA from healthy controls [50]. We subsequently developed a further 2-step diagnosis and classification based on plasma GON free adducts, hyp and anti-CCP antibody positivity. The first step involved an algorithm classifying between healthy controls and early-stage arthritis (any type) and involved optimum features: protein GON free adducts and hyp of plasma or serum (Table 3). This gave sensitivity 0.92, specificity 0.91 and diagnostic odds ratio = 116 – a highly powerful test for early-stage arthritis. The second step was a 3-classification algorithm to diagnose the type of arthritis: eOA, eRA and non-RA. Interestingly, hyp was a feature required only to distinguish between healthy controls and arthritis (any type); and anti-CCP antibody status was a feature required only to identify the type of arthritis. Subsequent studies of the time course of changes of protein GON free adducts in plasma of a guinea pig model of osteoarthritis showed a progressive increase of these analytes as arthritis develops and cartilage is eroded, suggesting that this effect is due to increased proteolysis of proteins of the joint as the severity of arthritis increases [41]. Glycated, oxidized and nitrated amino acids are likely better reporters of early stage arthritis development than unmodified amino acids because they are not re-incorporated into protein and hence accumulate.

Essays in Biochemistry (2020) **64** 169–183
https://doi.org/10.1042/EBC20190047

Two step diagnostic algorithms for detection of early-stage decline of metabolic, vascular and renal health

In other studies, we addressed the possibility that urinary fluxes of protein GON adducts may form the basis of a health screen for early decline in metabolic health (insulin resistance and prediabetes), early decline in vascular metabolic health (assessed by carotid artery intima-media thickness, CIMT) and early decline in renal health (stage 3 – 4 chronic kidney disease). We also included assay of branched chain amino acids (BCAAs) – leu, ile and val, as disturbances in BCAA metabolism has been associated with impaired metabolic and renal health: increased plasma levels of BCAA were found previously in overweight and obese subjects and correlated positively with insulin resistance [63,64]; and decreased BCAA in chronic renal insufficiency was linked to both change in tubular transport and decreased inter-organ exchange by peripheral tissues [65,66]. We found optimum diagnostic algorithms for early decline in metabolic and vascular health contained only features of age, BMI and urinary BCAA and for early decline in renal health, age, BMI and urinary val and FL free adduct (Table 3). This approach suggested that measurement of urinary branched-chain amino acids and FL provide small, moderate and conclusive evidence of early-stage cardiovascular, metabolic and renal disease, respectively. Although this was disappointing from the expectation that biomarkers of protein GON may be linked to early decline in metabolic, vascular and renal health, we did find several interesting correlations. In subjects with impaired vascular health, there were positive correlations of CIMT, glycated hemoglobin A1C and glucosepane with age ($r = 0.86$, $r = 0.51$ and $r = 0.37$, respectively) and negative correlation of eGFR with age ($r = -0.51$). Subjects with impaired vascular health, therefore, have age-related increase in CIMT thickness along with age-related early-stage decline in glucose tolerance and renal function and increased glucose-mediated protein crosslinking. In subjects with impaired renal health, there were also positive correlations of CIMT ($r = 0.79$) and glucosepane with age ($r = 0.46$), suggesting that subjects with impaired renal function have increased CIMT and increased glucose-mediated protein crosslinking with age [67].

Concluding remarks

Detection and quantitation of protein AGEomics by stable isotopic dilution analysis LC-MS/MS has provided a gold standard reference analytical technique to assist studies of glycation and oxidative stress research. Access to its use is restricted by limited commercial availability of analytical standards. Data and insights gained in the last, nearly 20 years of its use have greatly advanced understanding, experimental design and study outcomes – particularly in biomedical research. With a fully automated LC-MS/MS platform, it could provide a valuable addition to clinical chemistry laboratories where an increasing number of diagnostic algorithms are revealing the importance of assessment of protein GON well beyond the established application of glycated hemoglobin A1C in diabetes and prediabetes. It has also been validated for use in epidemiological studies.

Summary

- Stable isotopic dilution analysis LC-MS/MS is the gold standard reference technique for quantitation of protein glycation and oxidative damage – a technique also called AGEomics.
- Studies of protein glycation, oxidation and nitration provide important mechanistic insights into the physiological consequences of hyperglycaemia, dicarbonyl stress and oxidative stress.
- LC-MS/MS provides a valuable analytical tool in diagnostic techniques for physiological and clinical applications, particularly through inclusion of protein damage analytes in diagnostic algorithms.
- Greater access to AGEomics in the future may be provided through improved commercial availability of analytical standards and wider availability of fully automated LC-MS/MS platforms.
- AGEomics has been validated for use in epidemiological studies and may feature increasingly in future advances in physiological and biomedical investigations.

Competing Interests

The authors are joint owners of a patent of the blood test for autism described above.

Author Contribution

N.R. and P.J.T. wrote and approved the manuscript.

Acknowledgements

We thank our current and former research team colleagues and collaborators for their help in developing LC-MS/MS analysis in AGEomics and Systems Biology research.

Abbreviations

3DG-H, N_{δ}-[5-(2,3,4-trihydroxybutyl)-5-hydro-4-imidazolon-2-yl]ornithine and related structural isomers; 3-NT, 3-nitrotyrosine; A1C, glycated hemoglobin HbA1c; AASA, 2-aminoadipic semialdehyde; AGE, advanced glycation end product; ASD, autism spectrum disorder; AQC, 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate; BCAA, branched chain amino acid; BMI, body mass index; CCP, anti-cyclic citrullinated peptide; CEL, N_{ε}-(1-carboxyethyl)lysine; CMA, N_{ω}-carboxymethyl-arginine; CML, N_{ε}-carboxymethyl-lysine; CP, citrullinated protein; DUOX, dual oxidase; eGFR, estimated glomerular filtration rate; FL, N_{ε}-(1-deoxyfructosyl)lysine; GC-MS, gas chromatography-liquid chromatography; GEEK, N_{ε}-(γ-glutamyl)lysine; G-H1, N_{δ}-(5-hydro-4-imidazolon-2-yl)ornithine; GLM, generalized linear model; GLMNET, GLM with elastic net; GOLD, glyoxal-derived lysine dimer, 1,3-di(N_{ε}-lysino)imidazolium salt; GON, glycation, oxidation and nitration; GSA, glutamic semialdehyde; GSH, reduced glutathione; HSA, human serum albumin; HPLC, high performance liquid chromatography; LC-MS/MS, liquid chromatography-tandem mass spectrometry; MetSO, methionine sulfoxide; MG, methylglyoxal; MG-H1, N_{δ}-(5-hydro-5-methyl-4-imidazolon-2-yl)-ornithine; MOLD, MOLD, methylglyoxal-derived lysine dimer, 1,3-di(N_{ε}-lysino)-4-methyl-imidazolium salt; MRM, multiple reaction monitoring; NFK, N-formylkynurenine; non-RA, inflammatory arthritic disease which is often self-resolving; OA, osteoarthritis; RA, rheumatoid arthritis; ROS, reactive oxygen species; UPLC, ultrahigh performance liquid chromatography .

References

1 Thornalley, P.J. and Rabbani, N. (2014) Detection of oxidized and glycated proteins in clinical samples using mass spectrometry - A user's perspective. *Biochim. Biophys. Acta* **1840**, 818–829
2 Sacks, D.B. (2012) Measurement of Hemoglobin A1c: A new twist on the path to harmony. *Diabetes Care* **35**, 2674–2680, https://doi.org/10.2337/dc12-1348
3 Buss, H., Chan, T.P., Sluis, K.B., Domigan, N.M. and Winterbourn, C.C. (1997) Protein carbonyl measurement by a sensitive ELISA method. *Free Radic. Biol. Med.* **23**, 361–366, https://doi.org/10.1016/S0891-5849(97)00104-4
4 Hawkins, C.L. and Davies, M.J. (2019) Detection, identification and quantification of oxidative protein modifications. *J. Biol. Chem.* **294**, 19683–19708, https://doi.org/10.1074/jbc.REV119.006217
5 Xue, M., Adaikalakoteswari, A., Rabbani, N. and Thornalley, P.J. (2009) Protein damage in the ageing process: advances in quantitation and the importance of enzymatic defences. *SEB Exp Biol Ser.* **62**, 227–265
6 Sell, D.R. and Monnier, V.M. (1989) Structure elucidation of a senescence crosslink from human extracellular matrix. Implication of pentoses in the aging process. *J. Biol. Chem.* **264**, 21597–21602
7 Ahmed, M.U., Thorpe, S.R. and Baynes, J.W. (1986) Identification of Nε-carboxymethyl-lysine as a degradation product of fructoselysine in glycated protein. *J. Biol. Chem.* **261**, 4889–4894
8 Murphy, M.P., Holmgren, A., Larsson, N.G., Halliwell, B., Chang, C.J., Kalyanaraman, B. et al. (2011) Unraveling the Biological Roles of Reactive Oxygen Species. *Cell Metab.* **13**, 361–366, https://doi.org/10.1016/j.cmet.2011.03.010
9 Ahmed, N., Argirov, O.K., Minhas, H.S., Cordeiro, C.A. and Thornalley, P.J. (2002) Assay of advanced glycation endproducts (AGEs): surveying AGEs by chromatographic assay with derivatisation by aminoquinolyl-N-hydroxysuccimidyl-carbamate and application to Nε-carboxymethyl-lysine- and Nε-(1-carboxyethyl)lysine-modified albumin. *Biochem. J.* **364**, 1–14, https://doi.org/10.1042/bj3640001
10 Ahmed, N. and Thornalley, P.J. (2002) Chromatographic assay of glycation adducts in human serum albumin glycated in vitro by derivatisation with aminoquinolyl-N-hydroxysuccimidyl-carbamate and intrinsic fluorescence. *Biochem. J.* **364**, 15–24, https://doi.org/10.1042/bj3640015
11 Thornalley, P.J., Battah, S., Ahmed, N., Karachalias, N., Agalou, S., Babaei-Jadidi, R. et al. (2003) Quantitative screening of advanced glycation endproducts in cellular and extracellular proteins by tandem mass spectrometry. *Biochem. J.* **375**, 581–592, https://doi.org/10.1042/bj20030763
12 Ahmed, N., Babaei-Jadidi, R., Howell, S.K., Beisswenger, P.J. and Thornalley, P.J. (2005) Degradation products of proteins damaged by glycation, oxidation and nitration in clinical type 1 diabetes. *Diabetologia* **48**, 1590–1603, https://doi.org/10.1007/s00125-005-1810-7
13 Agalou, S., Ahmed, N., Babaei-Jadidi, R., Dawnay, A. and Thornalley, P.J. (2005) Profound mishandling of protein glycation degradation products in uremia and dialysis. *J. Am. Soc. Nephrol.* **16**, 1471–1485, https://doi.org/10.1681/ASN.2004080635
14 Ahmed, N., Ahmed, U., Thornalley, P.J., Hager, K., Fleischer, G.A. and Munch, G. (2004) Protein glycation, oxidation and nitration marker residues and free adducts of cerebrospinal fluid in Alzheimer's disease and link to cognitive impairment. *J. Neurochem.* **92**, 255–263
15 Ahmed, U., Anwar, A., Savage, R.S., Thornalley, P.J. and Rabbani, N. (2016) Protein oxidation, nitration and glycation biomarkers for early-stage diagnosis of osteoarthritis of the knee and typing and progression of arthritic disease. *Arthritis Res. Ther.* **18**, 250
16 Karachalias, N., Babaei-Jadidi, R., Rabbani, N. and Thornalley, P.J. (2010) Increased protein damage in renal glomeruli, retina, nerve, plasma and urine and its prevention by thiamine and benfotiamine therapy in a rat model of diabetes. *Diabetologia* **53**, 1506–1516, https://doi.org/10.1007/s00125-010-1722-z
17 Ahmed, N., Mirshekar-Syahkal, B., Kennish, L., Karachalias, N., Babaei-Jadidi, R. and Thornalley, P.J. (2005) Assay of advanced glycation endproducts in selected beverages and food by liquid chromatography with tandem mass spectrometric detection. *Mol. Nutrit. Food Res.* **49**, 691–699
18 Bechtold, U., Rabbani, N., Mullineaux, P.M. and Thornalley, P.J. (2009) Quantitative measurement of specific biomarkers for protein oxidation, nitration and glycation in Arabidopsis leaves. *Plant J.* **59**, 661–671, https://doi.org/10.1111/j.1365-313X.2009.03898.x

19 Stratmann, B., Engelbrecht, B., Espelage, B.C., Klusmeier, N., Tiemann, J., Gawlowski, T. et al. (2016) Glyoxalase 1-knockdown in human aortic endothelial cells – effect on the proteome and endothelial function estimates. *Sci. Rep.* **6**, 37737, https://doi.org/10.1038/srep37737
20 Irshad, Z., Xue, M., Ashour, A., Larkin, J.R., Thornalley, P.J. and Rabbani, N. (2019) Activation of the unfolded protein response in high glucose treated endothelial cells is mediated by methylglyoxal. *Sci. Rep.* **9**, 7889, https://doi.org/10.1038/s41598-019-44358-1
21 Rabbani, N., Sebekova, K., Sebekova, Jr, K., Heidland, A. and Thornalley, P.J. (2007) Protein glycation, oxidation and nitration free adduct accumulation after bilateral nephrectomy and ureteral ligation. *Kidney Internat.* **72**, 1113–1121, https://doi.org/10.1038/sj.ki.5002513
22 Xue, M., Weickert, M.O., Qureshi, S., Ngianga-Bakwin, K., Anwar, A., Waldron, M. et al. (2016) Improved glycemic control and vascular function in overweight and obese subjects by glyoxalase 1 inducer formulation. *Diabetes* **65**, 2282–2294, https://doi.org/10.2337/db16-0153
23 Rabbani, N., Shaheen, F., Anwar, A., Masania, J. and Thornalley, P.J. (2014) Assay of methylglyoxal-derived protein and nucleotide AGEs. *Biochem. Soc. Trans.* **42**, 511–517, https://doi.org/10.1042/BST20140019
24 Pamplona, R. (2011) Advanced lipoxidation end-products. *Chem. Biol. Interact.* **192**, 14–20, https://doi.org/10.1016/j.cbi.2011.01.007
25 Baynes, J.W. and Thorpe, S.R. (2000) Glycoxidation and lipoxidation in atherogenesis. *Free Radical Biol. Med.* **28**, 1708–1716
26 Rabbani, N., Godfrey, L., Xue, M., Shaheen, F., Geoffrion, M., Milne, R. et al. (2011) Conversion of low density lipoprotein to the pro-atherogenic form by methylglyoxal with increased arterial proteoglycan binding and aortal retention. *Diabetes* **60**, 1973–1980, https://doi.org/10.2337/db11-0085
27 Godfrey, L., Yamada-Fowler, N., Smith, J.A., Thornalley, P.J. and Rabbani, N. (2014) Arginine-directed glycation and decreased HDL plasma concentration and functionality. *Nutr. Diabetes* **4**, e134, https://doi.org/10.1038/nutd.2014.31
28 Dobler, D., Ahmed, N., Song, L.J., Eboigbodin, K.E. and Thornalley, P.J. (2006) Increased dicarbonyl metabolism in endothelial cells in hyperglycemia induces anoikis and impairs angiogenesis by RGD and GFOGER motif modification. *Diabetes* **55**, 1961–1969, https://doi.org/10.2337/db05-1634
29 Ahmed, N., Dobler, D., Dean, M. and Thornalley, P.J. (2005) Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. *J. Biol. Chem.* **280**, 5724–5732, https://doi.org/10.1074/jbc.M410973200
30 Westwood, M.E. and Thornalley, P.J. (1995) Molecular characteristics of methylglyoxal-modified bovine and human serum albumins. Comparison with glucose-derived advanced glycation endproduct-modified serum albumins. *J. Prot. Chem.* **14**, 359–372, https://doi.org/10.1007/BF01886793
31 Thornalley, P.J., Argirova, M., Ahmed, N., Mann, V.M., Argirov, O.K. and Dawnay, A. (2000) Mass spectrometric monitoring of albumin in uraemia. *Kidney Int.* **58**, 2228–2234, https://doi.org/10.1111/j.1523-1755.2000.00398.x
32 Rabbani, N. and Thornalley, P.J. (2018) Advanced glycation end products in the pathogenesis of chronic kidney disease. *Kidney Int.* **93**, 803–813, https://doi.org/10.1016/j.kint.2017.11.034
33 McLellan, A.C., Thornalley, P.J., Benn, J. and Sonksen, P.H. (1994) The glyoxalase system in clinical diabetes mellitus and correlation with diabetic complications. *Clin. Sci.* **87**, 21–29, https://doi.org/10.1042/cs0870021
34 Kleinman, W.A. and Richie, Jr, J.P. (2000) Status of glutathione and other thiols and disulfides in human plasma. *Biochem. Pharmacol.* **60**, 19–29, https://doi.org/10.1016/S0006-2952(00)00293-8
35 Breier, M., Wahl, S., Prehn, C., Fugmann, M., Ferrari, U., Weise, M. et al. (2014) Targeted Metabolomics Identifies Reliable and Stable Metabolites in Human Serum and Plasma Samples. *PLoS One* **9**, e89728, https://doi.org/10.1371/journal.pone.0089728
36 Hankinson, S.E., Willett, W.C., Manson, J.E., Colditz, G.A., Hunter, D.J., Spiegelman, D. et al. (1998) Plasma sex steroid hormone levels and risk of breast cancer in postmenopausal women. *J. Natl. Cancer Inst.* **90**, 1292–1299, https://doi.org/10.1093/jnci/90.17.1292
37 Chu, N.F., Spiegelman, D., Yu, J., Rifai, N., Hotamisligil, G.S. and Rimm, E.B. (2001) Plasma leptin concentrations and four-year weight gain among US men. *Int. J. Obes. Relat. Metab. Disord.* **25**, 346–353, https://doi.org/10.1038/sj.ijo.0801549
38 Inaba, M., Okuno, S., Kumeda, Y., Yamada, S., Imanishi, Y., Tabata, T. et al. (2007) Glycated Albumin Is a Better Glycemic Indicator than Glycated Hemoglobin Values in Hemodialysis Patients with Diabetes: Effect of Anemia and Erythropoietin Injection. *J. Am. Soc. Nephrol.* **18**, 896–903, https://doi.org/10.1681/ASN.2006070772
39 Ahmed, N., Babaei-Jadidi, R., Howell, S.K., Thornalley, P.J. and Beisswenger, P.J. (2005) Glycated and oxidized protein degradation products are indicators of fasting and postprandial hyperglycemia in diabetes. *Diabetes Care* **28**, 2465–2471, https://doi.org/10.2337/diacare.28.10.2465
40 Kilhovd, B.K., Juutilainen, A., Lehto, S., Ronnemaa, T., Torjesen, P.A., Hanssen, K.F. et al. (2009) Increased serum levels of methylglyoxal-derived hydroimidazolone-AGE are associated with increased cardiovascular disease mortality in nondiabetic women. *Atherosclerosis* **205**, 590–594, https://doi.org/10.1016/j.atherosclerosis.2008.12.041
41 Legrand, C., Ahmed, U., Anwar, A., Rajpoot, K., Pasha, S., Lambert, C. et al. (2018) Glycation marker glucosepane increases with the progression of osteoarthritis and correlates with morphological and functional changes of cartilage in vivo. *Arthritis Res. Ther.* **20**, 131
42 Biemel, K.M., Friedl, D.A. and Lederer, M.O. (2002) Identification and quantification of major Maillard cross-links in human serum albumin and lens protein - Evidence for glucosepane as the dominant compound. *J. Biol. Chem.* **277**, 24907–24915, https://doi.org/10.1074/jbc.M202681200
43 Chiu, C.-J., Rabbani, N., Rowan, S., Chang, M.-L., Sawyer, S., Hu, F.B. et al. (2018) Studies of advanced glycation end products and oxidation biomarkers for type 2 diabetes. *Biofactors* **44**, 281–288, https://doi.org/10.1002/biof.1423
44 Wang, F., Zhao, Y., Niu, Y., Wang, C., Wang, M., Li, Y. et al. (2012) Activated glucose-6-phosphate dehydrogenase is associated with insulin resistance by upregulating pentose and pentosidine in diet-induced obesity of rats. *Horm. Metab. Res.* **44**, 938–942
45 Foerster, A. and Henle, T. (2003) Glycation in food and metabolic transit of dietary AGEs (advanced glycation end-products): studies on the urinary excretion of pyrraline. *Biochem. Soc. Trans.* **31**, 1383–1385, https://doi.org/10.1042/bst0311383
46 Hohmann, C., Liehr, K., Henning, C., Fiedler, R., Girndt, M., Gebert, M. et al. (2017) Detection of Free Advanced Glycation End Products in Vivo during Hemodialysis. *J. Agric. Food Chem.* **65**, 930–937, https://doi.org/10.1021/acs.jafc.6b05013
47 Edens, W.A., Sharling, L., Cheng, G.J., Shapira, R., Kinkade, J.M., Lee, T. et al. (2001) Tyrosine cross-linking of extracellular matrix is catalyzed by Duox, a multidomain oxidase/peroxidase with homology to the phagocyte oxidase subunit gp91 phox. *J. Cell Biol.* **154**, 879–891
48 Bae, Y.S., Choi, M.K. and Lee, W.-J. (2010) Dual oxidase in mucosal immunity and host-microbe homeostasis. *Trends Immunol.* **31**, 278–287, https://doi.org/10.1016/j.it.2010.05.003

49 Chang, S., Linderholm, A., Franzi, L., Kenyon, N., Grasberger, H. and Harper, R. (2013) Dual Oxidase Regulates Neutrophil Recruitment in Allergic Airways. *Free Radical Biol. Med.* **65**, 38–46
50 Ahmed, U., Anwar, A., Savage, R.S., Costa, M.L., Mackay, N., Filer, A. et al. (2015) Biomarkers of early stage osteoarthritis, rheumatoid arthritis and musculoskeletal health. *Sci. Rep.* **5**, 9259, 9251-9257, https://doi.org/10.1038/srep09259
51 Masania, J., Malczewska-Malec, M., Razny, U., Goralska, J., Zdzienicka, A., Kiec-Wilk, B. et al. (2016) Dicarbonyl stress in clinical obesity. *Glycoconj. J.* **33**, 581–589, https://doi.org/10.1007/s10719-016-9692-0
52 Sidey-Gibbons, J.A.M. and Sidey-Gibbons, C.J. (2019) Machine learning in medicine: a practical introduction. *BMC Med. Res. Method.* **19**, 64, https://doi.org/10.1186/s12874-019-0681-4
53 Breiman, L. (2001) Random Forests. *Machine Learn.* **45**, 5–32, https://doi.org/10.1023/A:1010933404324
54 Elsabbagh, M., Divan, G., Koh, Y.-J., Kim, Y.S., Kauchali, S., Marcín, C. et al. (2012) Global prevalence of autism and other pervasive developmental disorders. *Autism Res.* **5**, 160–179, https://doi.org/10.1002/aur.239
55 Xu, G., Strathearn, L., Liu, B. and Bao, W. (2018) Prevalence of Autism Spectrum Disorder Among US Children and Adolescents, 2014-2016. *JAMA* **319**, 81–82, https://doi.org/10.1001/jama.2017.17812
56 Narzisi, A., Posada, M., Barbieri, F., Chericoni, N., Ciuffolini, D., Pinzino, M. et al. (2018) Prevalence of Autism Spectrum Disorder in a large Italian catchment area: a school-based population study within the ASDEU project. *Epidemiol. Psychiatr. Sci.* **29**, 1–10, https://doi.org/10.1017/S2045796018000483
57 Zwaigenbaum, L. and Penner, M. (2018) Autism spectrum disorder: advances in diagnosis and evaluation. *BMJ* **361**, k1674
58 Anwar, A., Abruzzo, P.M., Pasha, S., Rajpoot, K., Bolotta, A., Ghezzo, A. et al. (2018) Advanced glycation endproducts, dityrosine and arginine transporter dysfunction in autism - a source of biomarkers for clinical diagnosis. *Molecular Autism* **9**, 3, https://doi.org/10.1186/s13229-017-0183-3
59 Connelly, L.B., Woolf, A. and Brooks, P. (2006) Cost-Effectiveness of Interventions for Musculoskeletal Conditions. In *Disease Control Priorities in Developing Countries* (Jamieson, D.T., Breman, J.G., Measham, A.R., Alleyne, G., Claeson, M., Evans, D.B., Jha, P., Mills, A. and Musgrove, P., eds), pp. 963–980, World Bank, Washington DC
60 Isaacs, J.D. (2010) The changing face of rheumatoid arthritis: sustained remission for all? *Nat. Rev. Immunol.* **10**, 605–611, https://doi.org/10.1038/nri2804
61 McCord, J.M. (1974) Free Radicals and Inflammation: Protection of Synovial Fluid by Superoxide Dismutase. *Science* **185**, 529–531, https://doi.org/10.1126/science.185.4150.529
62 Kaur, H. and Halliwell, B. (1994) Evidence for nitric oxide-mediated oxidative damage in chronic inflammation Nitrotyrosine in serum and synovial fluid from rheumatoid patients. *FEBS Lett.* **350**, 9–12, https://doi.org/10.1016/0014-5793(94)00722-5
63 Newgard, C.B., An, J., Bain, J.R., Muehlbauer, M.J., Stevens, R.D., Lien, L.F. et al. (2009) A Branched-Chain Amino Acid-Related Metabolic Signature that Differentiates Obese and Lean Humans and Contributes to Insulin Resistance. *Cell Metab.* **9**, 311–326, https://doi.org/10.1016/j.cmet.2009.02.002
64 Cheng, S., Rhee, E.P., Larson, M.G., Lewis, G.D., McCabe, E.L., Shen, D. et al. (2012) Metabolite Profiling Identifies Pathways Associated With Metabolic Risk in Humans. *Circulation* **125**, 2222–2231, https://doi.org/10.1161/CIRCULATIONAHA.111.067827
65 Cano, N.J.M., Fouque, D. and Leverve, X.M. (2006) Application of Branched-Chain Amino Acids in Human Pathological States: Renal Failure. *J. Nutr.* **136**, 299S–307S, https://doi.org/10.1093/jn/136.1.299S
66 Garibotto, G., Paoletti, E., Fiorini, F., Russo, R., Robaudo, C., Deferrari, G. et al. (1993) Peripheral metabolism of branched-chain keto acids in patients with chronic-renal-failure. *Miner. Electrolyte Metab.* **19**, 25–31
67 Masania, J., Faustmann, G., Anwar, A., Hafner-Giessauf, H., Rajpoot, R., Grabher, J. et al. (2019) Urinary metabolomic markers of protein glycation, oxidation and nitration in early-stage decline in metabolic, vascular and renal health. *Oxid. Med. Cell. Longev.*, Article ID: 4851323, https://doi.org/10.1155/2019/4851323
68 Schafer, C., Schott, M., Brandl, F., Neidhart, S. and Carle, R. (2005) Identification and Quantification of ε-(γ-Glutamyl)lysine in Digests of Enzymatically Cross-Linked Leguminous Proteins by High-Performance Liquid Chromatography−Electrospray Ionization Mass Spectrometry (HPLC-ESI-MS). *J. Agric. Food Chem.* **53**, 2830–2837
69 Thornalley, P.J. and Rabbani, N. (2013) Detection of oxidized and glycated proteins in clinical samples using mass spectrometry - A user's perspective. *Biochim. Biophys. Acta*, in press
70 Delpierre, G., Vertommen, D., Communi, D., Rider, M.H. and Van Schaftingen, E. (2004) Identification of fructosamine residues deglycated by fructosamine-3-kinase in human hemoglobin. *J. Biol. Chem.* **279**, 27613–27620
71 Erbersdobler, H.F. and Faist, V. (2001) Metabolic transit of Amadori products. *Nahrung-Food* **45**, 177–181, https://doi.org/10.1002/1521-3803(20010601)45:3%3c177::AID-FOOD177%3e3.0.CO;2-A
72 Liardon, R., de Weck-Gaudard, D., Philipossian, G. and Finot, P.A. (1987) Identification of Nε-carboxymethyllysine: a new Maillard reaction product, in rat urine. *J. Agric. Food Chem.* **35**, 427–431, https://doi.org/10.1021/jf00075a035
73 Knecht, K.J., Dunn, J.A., McFarland, K.F., McCance, D.R., Lyons, T.J., Thorpe, S.R. et al. (1991) Effect of diabetes and aging on carboxymethyllysine levels in human urine. *Diabetes* **40**, 190–196, https://doi.org/10.2337/diab.40.2.190
74 Requena, J.R., Chao, C.C., Levine, R.L. and Stadtman, E.R. (2001) Glutamic and aminoadipic semialdehydes are the main carbonyl products of metal-catalyzed oxidation of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 68–74
75 Tsikas, D. (2012) Analytical methods for 3-nitrotyrosine quantification in biological samples: the unique role of tandem mass spectrometry. *Amino Acids* **42**, 45–63, https://doi.org/10.1007/s00726-010-0604-5
76 Eckert, R.L., Kaartinen, M.T., Nurminskaya, M., Belkin, A.M., Colak, G., Johnson, G.V.W. et al. (2014) Transglutaminase Regulation of Cell Function. *Physiol. Rev.* **94**, 383–417, https://doi.org/10.1152/physrev.00019.2013