



Article

# Multiple Glycation Sites in Blood Plasma Proteins as an Integrated Biomarker of Type 2 Diabetes Mellitus

Alena Soboleva <sup>1,2,†</sup> , Gregory Mavropulo-Stolyarenko <sup>1,†</sup>, Tatiana Karonova <sup>3,4</sup>,  
Domenika Thieme <sup>5</sup> , Wolfgang Hoehenwarter <sup>5</sup> , Christian Ihling <sup>6</sup>, Vasily Stefanov <sup>1</sup>,  
Tatiana Grishina <sup>1</sup> and Andrej Frolov <sup>1,2,\*</sup>

<sup>1</sup> Department of Biochemistry, St. Petersburg State University, 199034 Saint Petersburg, Russia; st021585@student.spbu.ru (A.S.); gm2124@mail.ru (G.M.-S.); vastef@mail.ru (V.S.); tgrishina@mail.ru (T.G.)

<sup>2</sup> Department of Bioorganic Chemistry, Leibniz Institute of Plant Biochemistry, D-06120 Halle (Saale), Germany

<sup>3</sup> Almazov National Medical Research Centre, 197341 Saint Petersburg, Russia; karonova@mail.ru

<sup>4</sup> Department of Faculty Therapy, The First Pavlov St. Petersburg State Medical University, 197022 Saint Petersburg, Russia

<sup>5</sup> Proteome Analytics Research Group, Leibniz Institute of Plant Biochemistry, D-06120 Halle (Saale), Germany; Domenika.Thieme@ipb-halle.de (D.T.); Wolfgang.Hoehenwarter@ipb-halle.de (W.H.)

<sup>6</sup> Institute of Pharmacy, Martin Luther University of Halle-Wittenberg, D-06120 Halle (Saale), Germany; christian.ihling@pharmazie.uni-halle.de

\* Correspondence: afrolov@ipb-halle.de; Tel.: +49-0-345-5582-1350

† These authors contributed equally to this work.

Received: 15 March 2019; Accepted: 7 May 2019; Published: 10 May 2019



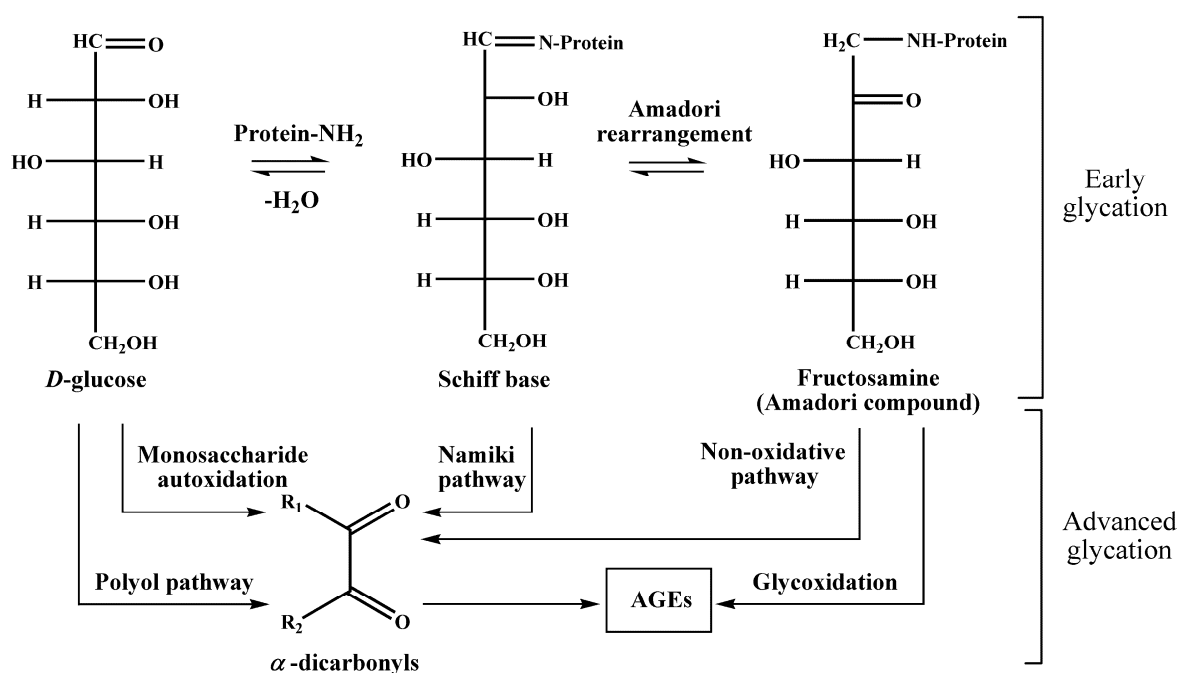
**Abstract:** Type 2 diabetes mellitus (T2DM) is one of the most widely spread metabolic diseases. Because of its asymptomatic onset and slow development, early diagnosis and adequate glycaemic control are the prerequisites for successful T2DM therapy. In this context, individual amino acid residues might be sensitive indicators of alterations in blood glycation levels. Moreover, due to a large variation in the half-life times of plasma proteins, a generalized biomarker, based on multiple glycation sites, might provide comprehensive control of the glycemic status across any desired time span. Therefore, here, we address the patterns of glycation sites in highly-abundant blood plasma proteins of T2DM patients and corresponding age- and gender-matched controls by comprehensive liquid chromatography-mass spectrometry (LC-MS). The analysis revealed 42 lysyl residues, significantly upregulated under hyperglycemic conditions. Thereby, for 32 glycation sites, biomarker behavior was demonstrated here for the first time. The differentially glycosylated lysines represented nine plasma proteins with half-lives from 2 to 21 days, giving access to an integrated biomarker based on multiple protein-specific Amadori peptides. The validation of this biomarker relied on linear discriminant analysis (LDA) with random sub-sampling of the training set and leave-one-out cross-validation (LOOCV), which resulted in an accuracy, specificity, and sensitivity of 92%, 100%, and 85%, respectively.

**Keywords:** Amadori compounds; biomarkers; glycation; glycation sites; label-free quantification; linear discriminant analysis; mass spectrometry; plasma proteins; type 2 diabetes mellitus

## 1. Introduction

Diabetes is an ubiquitously spread disease with a worldwide occurrence exceeding 387 million in 2014, and expected to reach 592 million by the year 2035 [1]. Among the total number of cases, more than 90% are represented by the type 2 diabetes mellitus (T2DM), characterized with impaired insulin action and/or insulin secretion [2]. As the onset of the early metabolic alterations (insulin tolerance and

hyperglycaemia) is asymptomatic, vascular diabetic complications are often revealed at the moment of T2DM diagnosis [3]. Hence, early discovery of T2DM, or even preceding changes in glucose metabolism, would essentially increase therapy efficiency and reduce the costs required for the treatment of diabetic complications [4]. Obviously, even minor changes in blood glucose concentrations might affect the rates of blood protein glycation [5], i.e., interaction of glucose with free amino groups of proteins, yielding fructosylamines, also known as Amadori compounds [6] (Figure 1). Their further degradation results in the formation of a highly heterogenic group of advanced glycation end products (AGEs) [7], known for their pro-inflammatory impact in the development of vascular diabetic complications [8,9]. However, although AGEs are well-established biomarkers of diabetic complications [10], Amadori compounds might suit better for glycaemic control and early diagnostics of T2DM [11].



**Figure 1.** Early and advanced glycation in human blood plasma. The major pathways of advanced glycation end product (AGE) formation: monosaccharide autoxidation (Wolff-pathway), autoxidation of aldimins (Namiki-pathway), polyol pathway, autoxidation of Amadori products (Hodge-pathway), and non-oxidative pathway.

Fasting plasma glucose (FPG) is the preferred diagnostic parameter, but an increase of blood glucose is also detected in patients suffering from other diseases besides diabetes mellitus [12]. Currently, the blood content of hemoglobin isoform, HbA<sub>1c</sub>, glycated by the N-terminus of its  $\alpha$ -chain (>6.5% of the total hemoglobin fraction), is one of the principle diagnostic criteria of this disease [13] and an efficient tool of long-term (60–90 days) glycaemic control [14]. However, HbA<sub>1c</sub> does not deliver any information about short-term alterations in plasma glucose concentrations accompanying the onset of metabolic syndrome [15]. In contrast, short-living plasma proteins provide a good possibility to decrease the time dimensions of glycaemic control. Thus, human serum albumin (HSA), the major blood plasma protein with a half-life of 21 days, can be used as a marker of T2DM [16]. Its global glycation rates can be quantitatively assigned by an array of enzymatic [17], colorimetric [18], immunochemical [19], electrophoretic [20], and chromatographic [21] methods. Importantly, the levels of HSA glycation vary from 1% and 10% in healthy individuals to 20% to 90% in patients with diabetes [22]. However, as the glycation rates at individual lysyl residues in the HSA molecule differ essentially [23], the sensitivity of the potential glycation sites to short-term changes of blood glucose levels might also be different.

In this context, the monitoring of glycation rates at specific glycation sites might be advantageous in comparison to the quantification of global glycation levels. Therefore during the last decade,

mass spectrometry was intensively employed in the establishment of such techniques [24–26]. In all cases, analytics relied on the bottom-up proteomic approach, based on nano-scaled liquid chromatography-mass spectrometry (nanoLC-MS) and tandem mass spectrometry (MS/MS). Thus, we showed that only some HSA lysyl residues are differentially glycosylated in plasma of T2DM patients, whereas glycosylation rates at other potential modification sites seemed to not be affected [27]. Hence, tryptic peptides, representing differentially glycosylated sites, might be considered as prospective T2DM biomarkers. For some of them, it was additionally confirmed by absolute quantification using internal standardization with dabsylated bi-labeled Amadori-modified peptides [28] or  $^{13}\text{C}$ ,  $^{15}\text{N}$  synthetic analogs of specifically glycosylated peptides [29]. As was shown recently, one of the plasma glycosylation sites (namely K<sub>141</sub> in haptoglobin) might provide an additional diagnostic tool in combination with well-established T2DM markers, such as fasting plasma glucose (FPG) and HbA<sub>1c</sub>. The main advantage of combining two markers—HbA<sub>1c</sub> and glycosylated haptoglobin, K<sub>141</sub>—is the simultaneous consideration of two proteins with different half-life times, i.e., 3 to 4 and 2 to 4 days, respectively. It makes this biomarker sensitive to long- and short-term fluctuations of blood glucose concentrations. The set of glycosylated K<sub>141</sub> of haptoglobin and HbA<sub>1c</sub> provided a sensitivity of 94%, a specificity of 98%, and an accuracy of 96% to identify T2DM [30].

In this context, it is logical to assume that a biomarker strategy, based on multiple specific glycosylation sites in plasma proteins, could essentially increase the efficiency of glycemic control and disease prediction. Indeed, the involvement of several glycosylated proteins with different half-life times ( $\tau_{1/2}$ ) allows several time segments in the glycemic control to be addressed without additional analyses. Moreover, this approach might decrease the impact of individual glycosylation sites in the overall dispersion of the data, when larger cohorts are considered. Therefore, here, we present a mass spectrometry-based biomarker approach relying on multiple glycosylation sites. We demonstrate its applicability to different time points in the span of glycemic control. Based on the results of linear discriminant analysis (LDA) performed for the cohorts of T2DM patients and individuals without diabetes, we characterize a set of Amadori-modified peptides with a diagnosis accuracy of up to 92%.

## 2. Results

### 2.1. Normalization by Plasma Protein Contents and Tryptic Digestion

According to the applied workflow (Figure 2), the normalization of analyte abundances relied on the determination of plasma protein contents and quality controls. The plasma protein concentrations, determined by the Bradford assay, varied from 36.2 to 82.4 mg/mL (Table S1), and were  $58.8 \pm 9.85$  and  $61.9 \pm 13.5$  mg/mL in the T2DM and normoglycemic groups, respectively. These values were used for the normalization of protein amounts taken for tryptic digestion and were considered in the label-free quantification procedure (Figure 2). Hence, as an adequate normalization is a pre-requisite for the satisfactory precision of label-free quantification [31], the plasma protein contents, determined by the Bradford assay, were cross-verified by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (protein load of 5  $\mu\text{g}$  per line) (Figure S1), and total line densities were recorded (Table S1). The whole-line densitometric analysis revealed the average density of  $19,483 \pm 1289$  arbitrary units (AU, relative standard deviation (RSD) = 6.6%), thereby providing sufficient precision for reliable quantification of individual glycosylated peptides by the signal intensities of corresponding quasi-molecular ions. The following tryptic digestion was considered to be quantitative, as the human serum albumine (HSA) band could not be detected in the SDS-PAGE experiments with tryptic digests (Figure S2), indicating a digestion efficiency better than 95% [27].

### 2.2. Annotation of Prospectively Glycosylated Peptides

Due to the high resolution and mass accuracy of the measurements (typically more than 12,000 and less than 3 ppm, respectively), individual glycosylated peptides could be reliably annotated by their retention times ( $t_R$ ) and exact masses ( $m/z$ ) of multiple-charged quasi-molecular ions. This step relied on

the list of glycated tryptic peptides, obtained from the plasma proteins of T2DM patients, based on the comprehensive works of Zhang et al. [26,32] and representing, to the best of our knowledge, the most complete set of reliably identified *in vivo* plasma glycation sites (more than 350). Therefore, for each glycated peptide identified by Zhang et al., all possible charge states were calculated with consideration of their amino acid composition. On the basis of this data, characteristic extracted ion chromatograms (XICs) were constructed in the mass ranges of  $m/z \pm 0.02$  for all matched signals with  $z \geq 2$ . For this, the time-of-flight (TOF) scans obtained with pooled diabetic plasma protein digests ( $n = 3$ ) were used. Most of the glycated peptides were annotated with sub-ppm mass accuracies and in all cases represented the minimal mass deviations from the theoretically predicted values in the ranges used for building the corresponding XICs (Table 1). This strategy revealed a total of 51 Amadori-modified peptides present in the pooled plasma protein digests in each sample, representing 10 major plasma proteins. As these features could be reliably detected, i.e., demonstrated a signal to noise ratio  $\geq 3$  in the corresponding XICs, their relative abundances in the T2DM and without diabetes cohorts were addressed by the label free quantification approach.

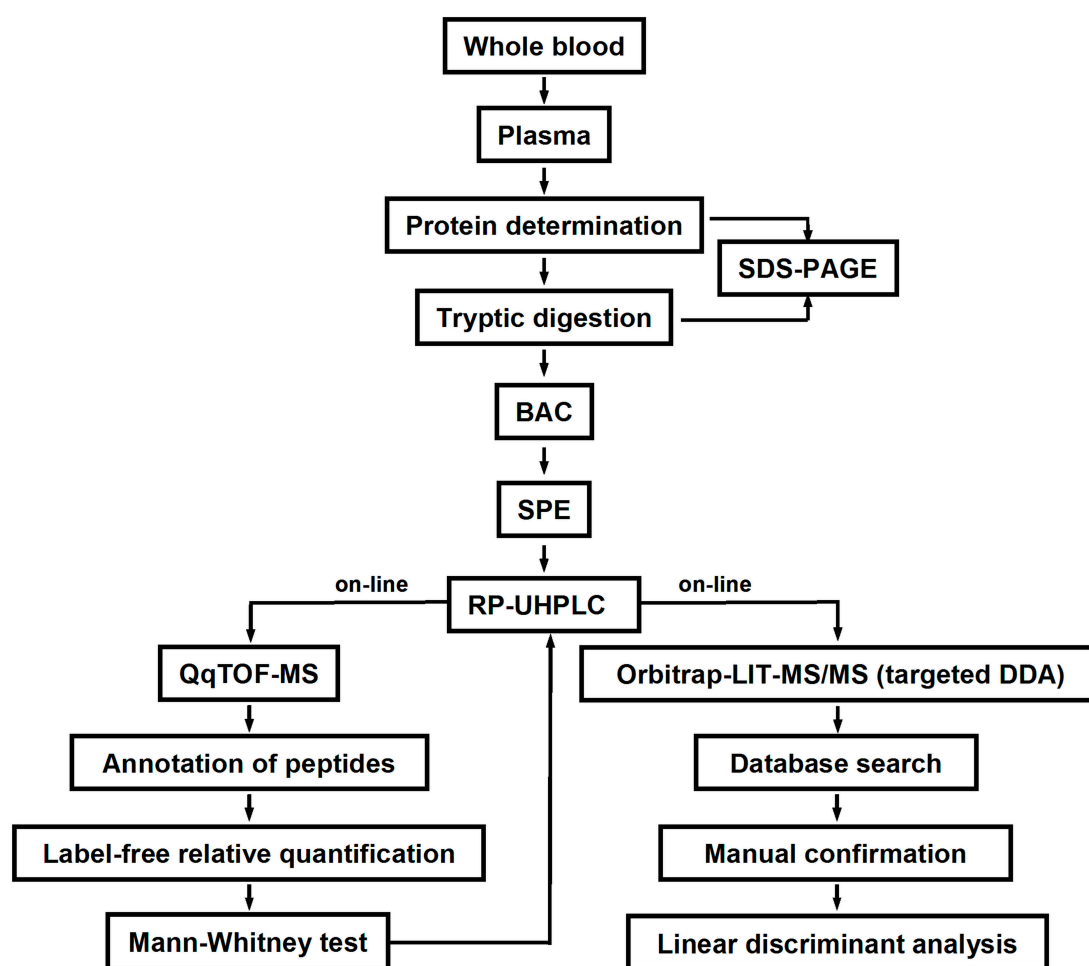


Figure 2. Overview of the experimental workflow.

**Table 1.** Individual glycated peptides differentially abundant in tryptic digests obtained from the T2DM patients and individuals without diabetes.

#	Sequence	Glycation Site	$t_R$ (min)	$m/z_{obs}$	$z$	Error (ppm)	U-Statistic (z)	$p$ Value (z)	Ac	Sp	Sn	LDA Variable
Human serum albumin												
1	VTK*C*C*TESLVNR	K <sub>475</sub>	14.5	543.5926	3	0.2	45, (3)	<0.0001, (3)	81.6	77.7	85	V24
2	RYK*AAFTEC*C*QAADK	K <sub>162</sub>	22.5	495.9759/660.9649	4/3	0.4/0.2	12, (4)	<0.0001, (4)	97.4	100	95	V14
3	LK*EC*C*EKPLEK	K <sub>276</sub>	24.4	427.9688/570.2889/854.9281	4/3/2	0.9/0.5/1.0	19, (3)	<0.0001, (3)	86.8	83.3	90	V20
4	C*ASLQK*FGER	K <sub>205</sub>	25.4	453.2191/679.3240	3/2	0.2/1.5	14, (3)	<0.0001, (3)	92.1	94.4	90	V22
5	LAK*TYETTLEK	K <sub>351</sub>	25.5	486.9246/729.8828	3/2	0.2/0.0	32, (3)	<0.0001, (3)	84.2	77.7	90	V15
6	YK*AAFTEC*C*QAADK	K <sub>162</sub>	26.2	608.9315/912.8924	3/2	0.0/0.8	21, (3)	<0.0001, (3)	86.8	100	75	V16
7	ETYGEMADC*C*AK*QEPER	K <sub>93</sub>	28.0	745.9667	3	0.0	43, (3)	<0.0001, (3)	84.2	94.4	75	V27
8	YIC*ENQDSISSK*LK	K <sub>274</sub>	29.9	616.2968	3	0.2	44, (3)	<0.0001, (3)	86.8	75	100	V25
9	FK*DLGEENFK	K <sub>12</sub>	31.0	463.5581/694.8330	3/2	0.2/0.1	44, (3)	<0.0001, (3)	81.6	77.7	85	V17
10	TYETTLEK*C*C*AAADPHEC*YAK	K <sub>359</sub>	31.7	670.7864/894.0452	4/3	0.1/0.7	45, (4)	<0.0001, (4)	79.0	66.6	90	V28
11	K*YLIEIAR	K <sub>137</sub>	31.9	406.5524/609.3246	3/2	0.0/0.3	22, (3)	<0.0001, (3)	86.8	88.8	85	V18
12	K*QTALVELVK	K <sub>525</sub>	33.1	430.9227/645.8801	3/2	0.5/0.3	46, (3)	<0.0001, (3)	84.2	88.8	80	V19
13	K*VPQVSTPTLVEVSR	K <sub>414</sub>	35.5	601.3355/901.4988	3/2	0.0/0.7	18, (3)	<0.0001, (3)	89.5	94.4	85	V11
14	ADLAK*YIC*ENQDSISSK	K <sub>262</sub>	36.0	701.9975	3	0.7	49, (3)	0.0001, (3)	86.8	100	75	V1
15	LVTDLTK*VHTEC*C*HGDLEEC*ADDR	K <sub>240</sub>	36.1	604.6759/755.5923/1007.1200	5/4/3	1.2/0.4/0.3	43, (5)	<0.0001, (5)	84.2	70	100	V23
16	TC*VADESAENC*DK*SLHTLFGDK	K <sub>64</sub>	37.4	665.5443/887.0572	4/3	1.4/2.1	34, (4)	<0.0001, (4)	84.2	66.6	100	V2
17	K*LVAASQAALGL	K <sub>574</sub>	37.5	652.3774	2	0.2	56, (2)	0.0003, (2)	79.0	66.6	90	V12
18	AAFTEC*C*QAADK*AAAC*LLPK	K <sub>174</sub>	38.5	763.0193	3	0.7	30, (3)	<0.0001, (3)	84.2	66.6	100	V3
19	LVNEVTEFAK*TC*VADESAENC*DK	K <sub>51</sub>	38.7	698.5634/931.0817	4/3	0.4/0.2	17, (4)	<0.0001, (4)	84.2	94.4	75	V4
20	AAAC*LLPK*LDELRL	K <sub>181</sub>	40.9	520.9491/780.9191	3/2	0.8/0.1	40, (3)	<0.0001, (3)	86.8	88.8	85	V21
21	SLHTLFGDK*LC*TVATLR	K <sub>73</sub>	42.5	524.2790/698.7018	4/3	1.0/0.3	41, (3)	<0.0001, (3)	81.6	83.3	80	V13
22	EFNAETFTFHADIC*TLSEK*ER	K <sub>519</sub>	42.9	677.5613/903.0777	4/3	0.7/0.8	67, (4)	0.0010, (4)	73.7	55.5	90	V26
23	TC*VADESAENC*DK*SLHTLFGDKLC*TVATLR	K <sub>64</sub>	43.1	894.1641	4	5.0	51, (4)	0.0001, (4)	81.6	72.2	90	V29
24	EQLK*AVM*DDFAAFVEK	K <sub>545</sub>	43.9	673.6598	3	0.3	68, (3)	0.0011, (3)	76.3	88.8	65	V7
25	VFDEFK*PLVEEPQNLIK	K <sub>378</sub>	45.8	736.3884	3	0.4	80, (3)	0.0036, (3)	76.3	72.2	80	V5
26	AEFAEVSK*LVTDLTK	K <sub>233</sub>	48.5	604.9884	3	0.8	79, (3)	0.0033, (3)	79.0	88.8	70	V6
27	RHPYFYAPELFFAK*	K <sub>159</sub>	50.1	516.0182	4	0.4	86, (4)	0.0063, (4)	73.7	44.4	100	V30

Table 1. Cont.

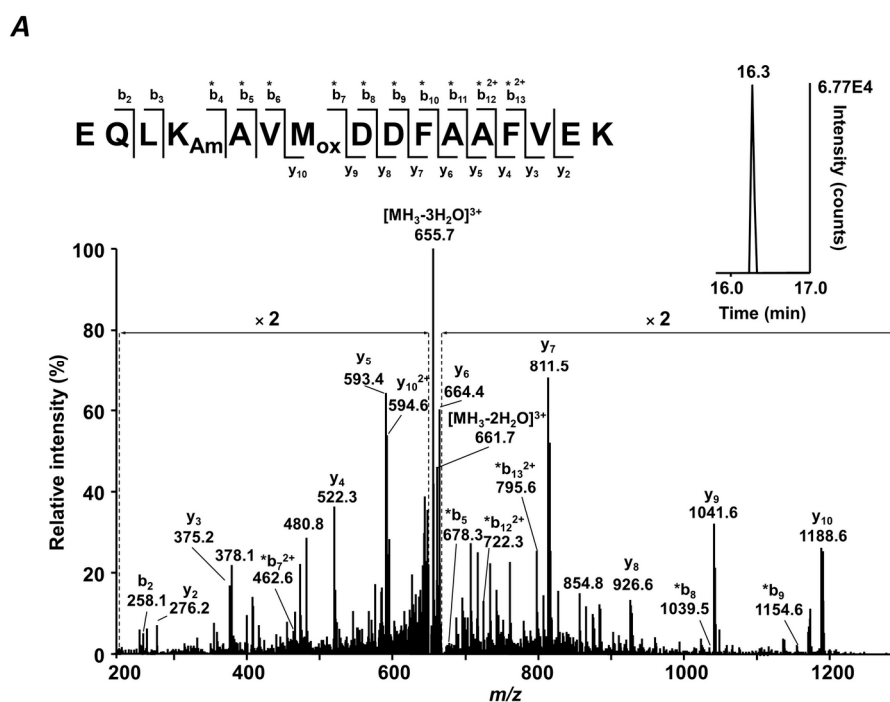
#	Sequence	Glycation Site	$t_R$ (min)	$m/z_{obs}$	$z$	Error (ppm)	U-Statistic ( $z$ )	$p$ Value ( $z$ )	Ac	Sp	Sn	LDA Variable
Alpha-2-macroglobulin												
28	LVDGK*GVPIPNK	K <sub>375</sub>	28.5	466.9332	3	0.2	33, (3)	<0.0001, (3)	86.8	94.4	80	V32
29	ALLAYAFALAGNQDK*R	K <sub>1162</sub>	46.2	628.6674	3	0.0	77, (3)	0.0027, (3)	76.3	66.6	85	V10
Apolipoprotein A1												
30	LAELYHAK*ATEHLSTLSEK	K <sub>219</sub>	28.5	548.2794	4	0.5	80, (4)	0.0036, (4)	76.3	94.4	60	V39
Ceruloplasmin precursor												
31	VTFHMK*GAYPLSIEPIGVR	K <sub>468</sub>	39.9	565.8053	4	0.0	57, (4)	<0.0001, (4)	84.2	70	100	V40
Complement C4-A												
32	RC*SVFYGAPSK*SR	K <sub>1575</sub>	39.0	838.9029	2	5.0	27, (2)	<0.0001, (2)	84.2	72.2	95	V42
FLJ00385 protein												
33	VSNK*ALPAPIEK	K <sub>319</sub>	26.6	476.9368/714.9006	3/2	0.0/1.0	30, (3)	<0.0001, (3)	86.8	75	100	V38
Igkappachain C region												
34	VQWK*VDNALQSGNSQESVTEQDSK	K <sub>41</sub>	34.3	947.1123	3	0.6	51, (3)	0.0002, (3)	81.6	94.4	70	V8
35	DSTYLSSTLTLK*ADYEK	K <sub>75</sub>	38.7	757.6970	3	0.5	35, (3)	<0.0001, (3)	86.8	88.8	85	V9
Serotransferrin												
36	GDVAFVK*HQTPQNTGGK	K <sub>553</sub>	25.7	512.0137	4	0.6	37, (4)	<0.0001, (4)	86.8	94.4	80	V31
37	DDTVC*LAK*LHDR	K <sub>659</sub>	29.0	401.9458/535.5914	4/3	0.5/0.2	25, (4)	<0.0001, (4)	84.2	94.4	75	V36
38	KPVDEYK*DC*HLAQVPSHTVVAR	K <sub>258</sub>	30.2	543.0768/678.5909/904.4496	5/4/3	2.4/2.2/4.8	34, (5)	<0.0001, (5)	86.8	88.8	85	V37
39	SK*EFQLFSSPHGK	K <sub>299</sub>	34.3	414.2098/551.9418	4/3	2.1/1.4	42, (3)	<0.0001, (3)	81.6	77.7	85	V34
40	DLLFK*DSAHGFLK	K <sub>315</sub>	40.3	413.9693/551.6227	4/3	0.2/0.5	47, (4)	0.0001, (4)	81.6	83.3	80	V35
41	DGAGDVAFAVK*HSTIFENLANK	K <sub>225</sub>	41.6	599.5509/799.0649	4/3	0.8/0.5	53, (3)	0.0002, (3)	84.2	94.4	75	V33
Vitamin D-binding protein precursor												
42	TSALSAK*SC*ESNSPFVHPGTAEC*C*TK	K <sub>94</sub>	32.3	772.0957	4	1.3	39, (4)	<0.0001, (4)	86.8	83.3	90	V41

Individual tryptic digests obtained from T2DM patients ( $n = 20$ ) and corresponding controls ( $n = 18$ ) were analyzed by capillary RP-HPLC-ESI-QqTOF-MS. The relative quantification relied on a label-free strategy using a digest prepared from T2DM pooled plasma as a quality control (injected after each 8th sample). Statistical analysis relied on the Mann-Whitney test. C\*, M\*, and K\* denote carbamidomethylated cysteine, methionine sulfoxide, and fructosamine lysine residues, respectively. Ac, Sp, and Sn denote the mass accuracy, accuracy, specificity, and sensitivity, respectively (calculated in MS Excel with RealStatisticsaddon, <http://www.real-statistics.com>).

### 2.3. Label-free Quantification

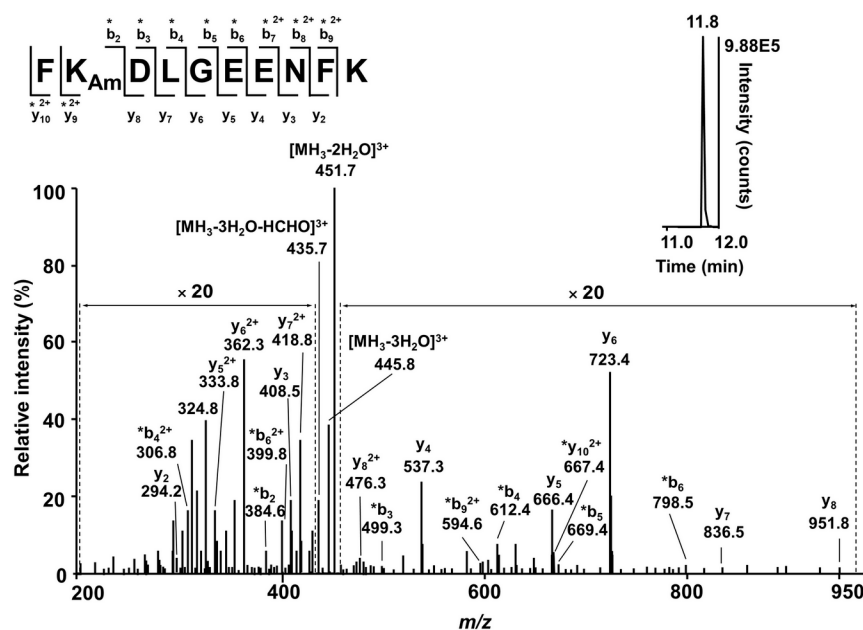
A paired Mann-Whitney test performed for the intensities of individual peptide signals (expressed as integral peak areas of corresponding XICs at defined retention times ( $t_{RS}$ )) revealed 42 differentially ( $p \leq 0.01$ ) glycosylated peptides representing nine plasma proteins (Figure 3 and Figure S3). Among this number, for 10 peptides, the biomarker properties were already proposed earlier [27,30], whereas for 32 species, a statistically significant T2DM-related abundance increase was observed here for the first time. Thereby, the levels of glycation at 25 HSA lysyl residues (represented by 27 tryptic peptides and comprising 43% of the total 58 present in the protein sequence) were significantly increased in the plasma of T2DM patients in comparison to the healthy controls. In contrast, only six modification sites were up-regulated in serotransferrin (six tryptic peptides, 10.3% of the totally available lysines), whereas the other proteins were represented with only two ( $\alpha$ -2-macroglobulin and Ig kappa chain C region) or one site, demonstrating enhanced glycation in T2DM plasma (Table 1). Most of the annotated prospective biomarkers could distinguish T2DM patients with sensitivities of 85% to 100% and accuracies exceeding 80% (Table 1). Thus, each of the differentially glycosylated proteins was represented with at least one marker peptide with biomarker sensitivities typically higher than 85% (with one exception for the apolipoprotein A-derived peptide, LAEYHAKATEHLSTLSEK, modified at lysine 219, Table 1). Interestingly, the  $\tau_{1/2}$  values of these proteins completely covered a time span of three weeks (Table 2).

Integration of selected peptide signals in the quality control (QC) samples revealed high intra- and inter-day precision of analyte retention times ( $t_{RS}$ ) and abundances. Thus, the RSD for intra- and inter-day precision of  $t_{RS}$  typically did not exceed 0.54% and 0.94%, whereas for peak areas, it was not higher than 11.0% and 16.46%, respectively (Table S2).





## B



**Figure 3.** Tandem mass spectra acquired at  $m/z$  673.66 (A) and 463.56 (B) corresponding to the  $[M + 3H]^{3+}$  ions of the peptides, EQLK<sub>Am</sub>AVM<sub>Ox</sub>DDFAAFVEK and FK<sub>Am</sub>DLGEENFK (both human serum albumin). The spectra were acquired with the LTQ-Orbitrap Velos Pro mass spectrometer operated in the positive ion mode as a part of targeted RP-HPLC-Orbitrap-LIT-MS/MS DDA workflow. The inserts represent extracted ion chromatograms ( $m/z \pm 0.02$ ) acquired for the  $m/z$  673.66 (A) and 463.56 (B). K<sub>Am</sub> and M<sub>Ox</sub> denote glycosylated lysyl and oxidized methionyl residues, respectively.

**Table 2.** The half-life times of plasma proteins, containing specific glycation sites differentially abundant in T2DM patients and normoglycemic individuals.

Glycated Protein	Glycation Site	Representative marker Peptide (#)	$\tau_{1/2}$ (Days)	Ref.
$\alpha$ -2-macroglobulin	K <sub>1162</sub>	ALLAYAFALAGNQDK*R (29)	1–2	[33]
Complement C4-A	K <sub>1575</sub>	RC*SVFYGAPSK*SR (32)	2–3	[34]
Vitamin D-binding protein precursor	K <sub>94</sub>	TSALSAK*SC*ESNSPPFVHPGTAEC*C*TK (42)	2.5–3	[35]
Ceruloplasmin	K <sub>468</sub>	VTFHNK*GAYPLSIEPIGVR (31)	5.5	[36]
Apolipoprotein A1	K <sub>219</sub>	LAEYHAK*ATEHLSTLSEK (30)	5–6	[37]
Serotransferrin	K <sub>299</sub>	SK*EFQLFSSPHGK (39)	8–10	[38]
Igkappa chain C region	K <sub>75</sub>	DSTYLSSTLTLSEK*ADYEK (35)	19–21	[39]
Human serum albumin	K <sub>159</sub>	RHPYFYAPELLFFAK* (27)	19–21	[39]
FLJ00385 protein	K <sub>319</sub>	VSNK*ALPAPIEK (33)	no data	-

#### 2.4. Sequence Assessment of Differentially Abundant Glycated Peptides

All features demonstrating significantly higher abundances in patients with diabetes were identified by tandem mass spectrometry (MS/MS, Table S3). The MS/MS spectra of 18 glycated peptides could be acquired in data-dependent acquisition (DDA) experiments with tryptic digests obtained from pooled T2DM plasma ( $n = 3$ ). The fragmentation of the other 24 species was achieved in targeted LC-MS/MS experiments. The sequences of the annotated peptides were unambiguously confirmed by a search against Uniprot human database (downloaded 7 June 2016) and verified by manual interpretation of the MS/MS spectra as exemplified in Figure 3 (for the spectral information see Figure S4). The sequences of six peptides (3, 4, 32, 34, 37, 40) could not be identified by the Mascot search engine and were assigned by manual interpretation (the complete spectral information is presented in Figure S5).

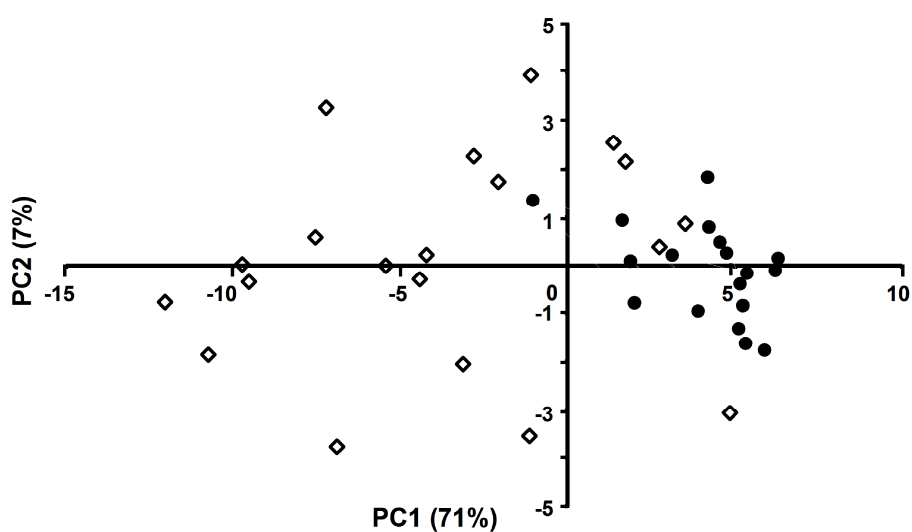
For all annotated peptides, manual interpretation revealed a good coverage of the peptide sequences by intense b- and y-fragment ions (Figure 3), whereas their glycosylated status, in the majority of



cases, could be confirmed by the patterns of water and formaldehyde losses from the quasi-molecular ions, characteristic for glucose-derived Amadori compounds [40,41]. Additionally, the presence of a sugar moiety at the corresponding lysyl residues was confirmed by pyrylium Amadori-related fragment ions.

### 2.5. Linear Discriminant Analysis

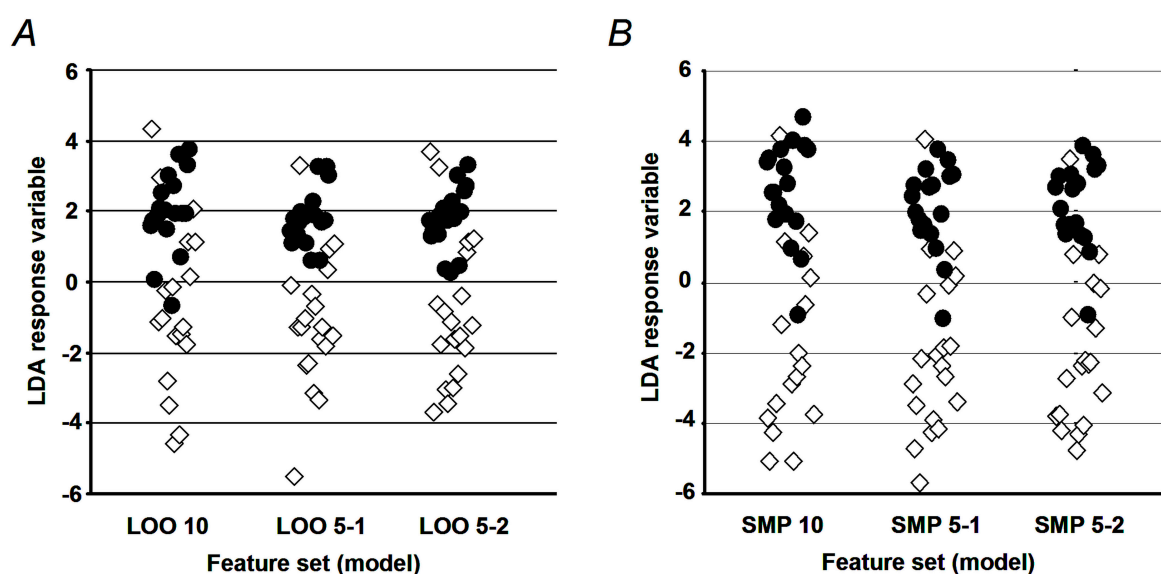
First, the data were subjected to principle component analysis (PCA), and the relative influence of the extracted individual principle components (PCs) was addressed, leaving out only the first nine components, which are responsible for 95% of the sample variance (see the corresponding variable loadings listed in Table S4). As can be seen from Figure 4, even the first two principal components provide decent class separation.



**Figure 4.** Results of the principle component analysis (PCA) performed for T2DM patient (opened diamonds) and normoglycemic (closed circles) groups. The analysis was based on all 42 differentially glycosylated peptides and indicated clear separation between the groups.

Alternatively, another procedure for the reduction of dimensionality, i.e., a variance inflation factor (VIF)-based factor filtering (following the workflow for generation of the LDA model presented in the Materials and Methods part), was applied to the original data. Therefore, diverse feature subsets could be generated depending on the order of the feature testing. In order to characterize possible non-collinear feature combinations, this algorithm was run 1000 times for the VIF cutoff values of 5 and 10. On the basis of the obtained results (Tables S5 and S6), the most abundant combinations of factors were selected for further training with LDA.

The results of the filtering procedure, based on the VIFs, are summarized in Tables S5 and S6, where individual variables are labeled as in Table 1. For the further processing, one of the most highly represented sets of variables with VIF = 10 and two sets, represented at 70% of all optimization runs with VIF = 5, were selected. Validation of the training and prediction accuracy was performed for the selected sets of variables in leave-one-out cross validation (LOOCV) and the sub-sampling LDA parameterizations modes (Figure 5). The resulting values for the accuracy, specificity, and sensitivity are summarized in Table 3.



**Figure 5.** Values of the linear discriminant analysis (LDA) response variable plotted for T2DM (opened diamonds) and control (closed circles) individuals, calculated using selected predictive variable sets (as summarized in Table 3). (A) Estimated via leave-one-out cross-validation; (B) estimated from original samples, with LDA trained on generated samples.

**Table 3.** Summary of the predictor performance validation under distinct training and performance estimation conditions.

Model Type	Training \ Prediction Mode						Variable Set
	LOO Cross Validated			Sub-Sample Trained			
	Ac	Sp	Sn	Ac	Sp	Sn	
VIF10	81.58	88.89	75	84.21	94.44	75	V6 V7 V9 V10 V12 V28 V29 V30 V32 V33 V34 V39 V40 V41 V42
VIF5_1	92.11	100	85	86.84	94.44	80	V6 V10 V12 V28 V29 V30 V33 V34 V39 V41 V42
VIF5_2	89.47	100	80	89.47	94.44	85	V7 V10 V12 V28 V29 V30 V32 V33 V34 V39 V41 V42
PCA	84.2	100	70	-	-	-	PC1 PC2 PC3 PC4 PC5 PC6 PC7 PC8 PC9

Ac, accuracy; Sp, specificity; Sn, sensitivity.

### 3. Discussion

#### 3.1. Experimental Setup

Protein glycation is one of the most universal markers of diabetes mellitus. It is recognized as an indirect, but adequate marker of blood glucose contents over a certain period of time [42]. Indeed, as blood sugar concentrations fluctuate in a wide range, depending on circadian rhythm, food intake, medication, and other factors [43], the determination of actual blood glucose levels usually does not deliver clinically valuable information. In this context, HbA<sub>1c</sub> is commonly used both in diagnostics of T2DM and long-term therapy control of the disease [42], whereas determination of short-term markers, like glycated albumin and total plasma fructosamine fraction [44,45], are still much less common in clinical practice. Thus, persisting hyperglycemia is typically confirmed only over the last three months prior to determination, whereas no or little information about the dynamics of glycation (and, hence, blood glucose levels) is available.

Therefore, the monitoring of several glycated proteins, varying in their half-life times, would be a promising alternative to the traditional approach for glycemic control. However, as the glycation rates at individual protein lysyl residues vary essentially [23], the total glycation levels of each protein would be less sensitive to sugar fluctuations than modification levels at individual residues. To address this concept in statistically representative cohorts, we analyzed the relative quantities of corresponding

glycated species in the group ( $n = 20$ ) of T2DM patients matched by age and gender to healthy controls. This cohort size was a good starting point for this pilot study, but due to a high number of potential marker peptides, we employed an  $\alpha$ -level of 0.01 (i.e., lower, than usual) and implemented a family-wise error rate control procedure [46], thus ensuring that the samples sizes used in our study were sufficient to prove all related inferences. As for the PCA\LDA, dimensionality reduction techniques are well suited for the high parameters of samples ratio, and according to Sharma et al. [47] and references therein, our case (approximately 40 dimensions to 40 samples) is several orders of magnitude below typical thresholds, above which small sample size problems should be specifically addressed.

To access individual glycation sites, we applied an experimental setup relying on the bottom-up proteomic approach and LC-MS-based label-free quantification (Figure 2). As we have shown previously, the abundances of such Amadori peptides (i.e., the areas of corresponding peaks in appropriate XICs) can be used to quantify corresponding glycation sites in bovine serum albumin [25]. Moreover, testing of this approach in small cohorts by nano-scaled reversed-phase high performance liquid chromatography-electrospray ionization-tandem mass-spectrometry (nanoRP-HPLC-ESI-MS/MS) DDA experiments revealed the differential abundance of some glycation sites in T2DM patients and normoglycemic individuals, although the poor precision of the quantification restricted the reliability of the results' interpretation [27].

Besides employing larger cohorts, this limitation can be generally overcome by using a reliable single-step chromatography system (i.e., skipping trapping procedure) as a part of a targeted or untargeted quantitative platform. Although the first approach, recently proposed by Spiller and co-workers [30], provides sufficient precision [29], it targets only a limited number of potential marker sites. Moreover, the selection of marker peptides, addressed in the multiple reaction monitoring (MRM) triple quadrupole (QqQ) method, proposed by the authors was based on our pilot study, which was performed in small cohorts consisting of only five probands [27]. This experimental setup could certainly result in missing many glycation sites with biomarker properties.

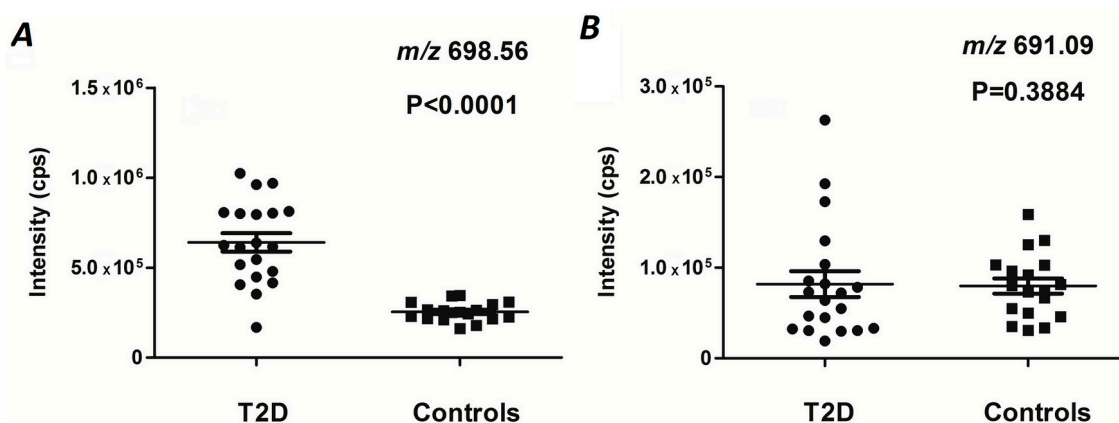
Therefore, here, we decided on an unbiased profiling approach based on high resolution mass spectrometry (HR-MS) coupled with on-line to microbore reversed-phase high performance liquid chromatography (RP-HPLC) with subsequent annotation of previously identified glycated peptides by the  $m/z$  of corresponding quasi-molecular ions. For this, we relied on the best available database of in vivo plasma glycation sites, built by Metz and co-workers on the basis of multi-dimensional in-depth plasma proteomics analysis [26]. Although in the absence of enrichment and pre-fractionation steps our straightforward approach (comprising boronic acid affinity chromatography (BAC) and RP-HPLC steps) yielded only 51 of the most abundant peptides, representing 46 modification sites among 2205 identified by nano-LC-MS/MS-based proteomics [26], their quantification was accurate and extremely precise. Indeed, our method provided a 14- and 7-fold improvement of intra- and inter-day precision for the peak area in comparison to the values, obtained by Frolov et al. (2014) [27], respectively (Table S2). Thereby, inter-day precision values were even slightly better than those generated with QqQ-MS instrumentation [29] with an essentially higher number of considered peptides. Without any doubt, it positively affected the significance of differences between T2DM and control groups (Table 1, Figure 3 and Figure S3), with a simultaneous reduction of the analysis time and costs in comparison to the conventional LC-based proteomic approach.

In this context, the analytical workflow (Figure 2) was reduced to the combination of three principle steps—enrichment of glycated peptides with BAC, desalting of the samples by SPE, and subsequent reversed phase chromatography (RPC) separation [24,48]. In our recent work we proved that the the solid phase extraction (SPE) step is absolutely mandatory for the acceptable recovery of Amadori peptides from the reversed phase [29]. Therefore, it was ultimately included in the proposed workflow.

### 3.2. Biomarker Potential of Glycated Peptides

Although the size of the cohorts used in our previous study was insufficient for reliable conclusions, five individuals per comparison group was a good starting point to establish new biomarkers based

on specific glycation sites in plasma proteins [27]. The increase of the cohort size in combination with an untargeted LC-MS approach resulted in the identification of a higher number of significantly up-regulated individual glycation sites with higher confidence (Table 1) in comparison to the published data, obtained with both spectral counting [49] and integration of specific signals in individual XICs [27]. Thus, our setup resulted in a higher biomarker potential of the plasma protein glycation sites considered in this context earlier [27–29] (Figure 6), and yielded new prospective Amadori peptides, whose biomarker behavior has not been reported yet.



**Figure 6.** Peak areas of the signals at 38.7 min (A), and 48.9 min (B), integrated in the extracted ion chromatograms (XICs) of  $m/z$   $698.56 \pm 0.02$  (A) and  $691.09 \pm 0.02$  (B), corresponding to the  $[M + 4H]^{4+}$  ions of glycated peptides LVNEVTEFAK<sub>Am</sub>TCCAMVADESAENCCAMDK (human serum albumin, A) and QNCCAMELFEQLGEYK<sub>Am</sub>FQNALLVR (human serum albumin, B), respectively. The data were acquired in the RP-HPLC-QqTOF-MS experiments performed with boronic acid affinity chromatography (BAC) enriched tryptic digests obtained from individual T2DM and control plasma samples. Statistical significance was estimated by the Mann-Whitney *U*-test, and *p* values below 0.05 were considered to be significant. K<sub>Am</sub> and C<sub>CAM</sub> denote glycated lysyl and carbamidomethylatedcysteinylnyl residues, respectively.

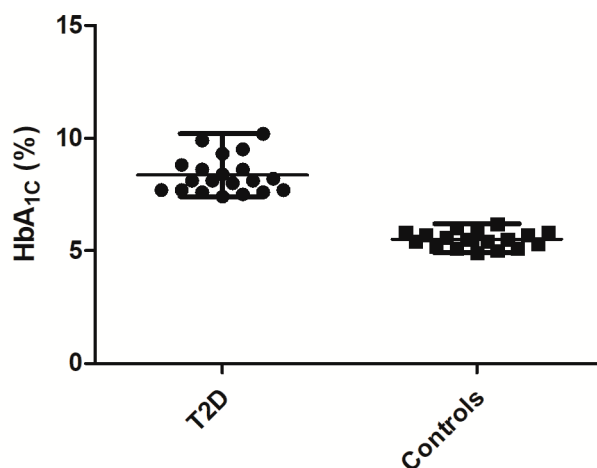
As all these peptides belonged to proteins with varying half-life times, we succeeded in addressing the time scale of glycemic control (Table 2). Indeed, relatively long-living proteins (HSA and the immunoglobulin (Ig) kappa chain C region with  $\tau_{1/2}$  of up to three weeks [39]) are exposed to plasma glucose for times much longer than those relevant for  $\alpha$ -2-macroglobulin and complement C4-A protein ( $\tau_{1/2}$  of 1 to 3 days [33,34]). Hence, the glycation sites representing the latter proteins would deliver valuable information about blood sugar levels over the days directly preceding the analysis. Such data would directly show if short-term fluctuations of plasma glucose levels do occur. Our results indicate a significantly lower degree of such fluctuations in the plasma of healthy individuals in comparison to T2DM patients (Table 1, Figure 6 and Figure S4). Thus, in the future, the corresponding tryptic peptides might be addressed as possible markers of impaired glucose tolerance.

### 3.3. Specific Set of Glycation Sites as an Integrated Biomarker of T2DM

As the proteins listed in Table 2 cover a large range of half-lives, consideration of the interference between representative glycation sites might provide a generalized T2DM marker, addressing a dynamic aspect of glycemic control. In this context, we tried to identify a set of multiple glycated sites in several marker proteins with various  $\tau_{1/2}$ . This set is expected to provide sensitivity sufficient for the prediction of T2DM and efficient glycemic control, i.e., provide separation of groups different from existing ones, e.g., glycated hemoglobin (Figure 7). However, a biomarker, based on multiple glycated proteins would give a higher degree in the flexibility of the glycemic control depth.

A basic T2DM marker, relying on two glycation sites, was proposed recently. Using a decision tree algorithm, Spiller and co-workers successfully combined quantitative information on a selected

haptoglobin glycation site (K<sub>141</sub>) with a well-established T2DM marker, such as HbA<sub>1c</sub> [30]. Unfortunately, due to its “step-by-step” decision tree design, this methodology is generally not applicable to the assessment of multiple factor contributions. Besides, when large cohorts are used, uncertainties, related to overfitting and multiple testing, need to be considered [49], i.e., a proper inspection of all resulting decision trees becomes critical for the accuracy of disease prediction. Indeed, as this calculation strategy does not assess factors simultaneously, the risk that the chosen set of classification steps and cut-off procedures “learns” some specific features of the training set is relatively high.



**Figure 7.** The level of HbA<sub>1c</sub> in plasma samples obtained from patients with T2DM and healthy persons. Statistical significance was estimated by the Mann-Whitney *U*-test,  $p < 0.0001$ .

Therefore, to avoid a potential bias associated with the application of the decision tree design to multiple features, we decided on LDA as the main classification algorithm. The corresponding output variable represented a simple weighed sum of the analyzed features, which could successfully meet the requirements for the establishment of a time dimension in glycaemic control. Thereby, the procedures for selection of the model parameters aimed at the removal of highly-correlated sites (duplicates or linear combinations) in the feature space (VIF-based filtering) or for transformation of the feature space itself (PCA). We believe that the application of several complementary feature space transformation and randomization techniques (i.e., PCA, VIF-based filtering, random sub-sampling of the training set, and leave-one-out cross-validation) allowed us to exclude any experimental bias and ensure a high reliability of the prediction model set.

Generally, all model generation and validation procedures resulted in similar metric values (Table 3). It can be noted that the models generated with a VIF cutoff of 5 demonstrated a clearly better performance (sensitivity, accuracy, and specificity of 80%, 90%, and 100%, respectively), in comparison to the models with a VIF cutoff of 10, containing higher numbers of features. Performance of the PCA-based model was somewhere in the middle, possibly due to the averaging of the specifics of individual features upon calculation of the principle component (PC) variable values. It is important to stress that the diagnostic results presented in Table 3 are conservative, i.e., bias-free [49]. Also, as the sub-sampling training sets discard the information about feature cross-correlations, they might be less informative than ones used for leave-one-out cross-validation, thus providing less accurate predictions.

Also, in the future, for the collection of blood samples, it is necessary to keep in mind not only therapy, but also the diet of patients. Thus, the influence of dietary polyphenols on blood glucose at different levels may also help control and prevent diabetes complication via a decrease of hyperglycemia and an improvement of acute insulin secretion and insulin sensitivity [50].

## 4. Materials and Methods

### 4.1. Reagents

Unless stated otherwise, materials were obtained from the following manufacturers. AMRESCO LLC (Fountain Parkway, Solon, OH, USA): ammonium persulfate (ACS grade), glycine (biotechnology grade), *N,N'*-methylene-bis-acrylamide (ultra-pure grade), tris(hydroxymethyl)aminomethane (tris, ultra-pure grade); Aptec Diagnostics nv (Saint-Petersburg, Russia): Albumin kit; Becton Dickinson (Moscow, Russia): BD Vacutainer® Plus Plastic K2EDTA tubes; Carl Roth GmbH & Co (Karlsruhe, Germany): acetonitrile ( $\geq 99.95\%$ , LC-MS grade), formic acid ( $\geq 98\%$ , p.a., ACS), ethanol ( $\geq 99.8\%$ ), methanol (HPLC Ultra Gradient Grade), sodium dodecyl sulfate (SDS) ( $>99\%$ ), tris-(2-carboxyethyl)-phosphine hydrochloride (TCEP,  $\geq 98\%$ ); Component-Reactiv (Moscow, Russia): phosphoric acid (analytical quality); PanReac AppliChem (Darmstadt, Germany): acrylamide (2K Standard Grade), glycerol (ACS grade); QIAGEN GmbH (Hilden, Germany): 1 mL polypropylene columns; Reachim (Moscow, Russia): Hydrochloric acid (analytical quality); Roche Diagnostics (Saint-Petersburg, Russia): Precinorm U plus Control, Precipath U plus Control; SERVA Electrophoresis GmbH (Heidelberg, Germany): Coomassie Brilliant Blue G-250, 2-mercaptoethanol (research grade), trypsin NB (sequencing grade, modified from porcine pancreas); Thermo Fisher Scientific (Waltham, MA, USA): PageRuler™ Unstained Protein Ladder #26614, PageRuler™ Prestained Protein Ladder #26616; Vekton (Saint-Petersburg, Russia): ammonia hydroxide (ACS grade); Waters GmbH (Eschborn, Germany): Oasis HLB cartridges (10 mg, 30  $\mu\text{m}$  particle size). All other chemicals were purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany). Water was purified in house (resistance  $>16 \text{ m}\Omega/\text{cm}$ ) on a water conditioning and purification system "UVOI-MF-1-NA(18)-N" (Mediana-Filtr, Moscow, Russia).

### 4.2. Setup of Experimental Cohorts

The T2DM patient ( $n = 20$ ) and normoglycemic control ( $n = 18$ ) groups comprised non-smoking female volunteers aged 45 to 75 years ( $63.4 \pm 7.9$  and  $60.7 \pm 4.7$ , respectively), who were not receiving hormone replacement therapy and had no clinically manifested diabetes complications (Tables 4 and 5). The control individuals' HbA<sub>1C</sub> levels did not exceed 6.5%, and did not have diagnosed diabetes and anti-hyperglycemic therapy in their medical history. All participants provided written informed consent. The study was approved 02-03-2015 by the Local Ethical Committee of the Federal Almazov North-West Medical Research Centre, Saint-Petersburg, Russian Federation, and was performed in agreement with the Declaration of Helsinki.

**Table 4.** Physical and biochemical parameters of the T2DM patients, as well as corresponding diabetes-related and non-related therapy.

Participants	Height (cm)	Weight (kg)	Body Mass Index	Age (Years)	Disease Duration (Years)	Therapy		HbA <sub>1C</sub> (%)	Content of Albumin ( $\mu\text{g}/\mu\text{L}$ ) in Blood Serum	Complications of Type 2 Diabetes Mellitus
						Diabetes-Specific Therapy	Diabetes-Nonspecific Therapy			
1	157	60.3	24.5	55	13	Gliclazide, Metformin	Verapamil, Enalapril, Hydrochlorothiazide, Atorvastatin	8.6	44.4	Polyneuropathy, nephropathy
2	165	97	35.6	50	14	Metformin, Vildagliptin	Valsartan, Verapamil, Rosuvastatin, Fenofibrate	9.5	35.2	Neuropathy, nephropathy, retinopathy
3	162	89.5	34.1	64	7	Metformin, Vildagliptin	Atorvastatin, Perindopril, Indapamide	8.0	44	Neuropathy
4	175	98.5	32.2	56	8	Metformin, Gliclazide	Losartan, Levothyroxine, Acetylsalicylic acid, Magnesiumhydroxide	8.6	48	Neuropathy
5	155	80.0	33.3	68	10	Glimepiride, Metformin, Vildagliptin	Perindopril, Bisoprolol, Amlodipine, Atorvastatin, Acetylsalicylic acid	8.8	34.6	Neuropathy, retinopathy, nephropathy
6	164	105	39.0	65	14	Metformin, Vildagliptin, Glimepiride	Acetylsalicylic acid, Losartan, Bisoprolol, Amlodipine, Moxonidine, Furosemide, Atorvastatin, Fenofibrate, Doxazosin	7.5	49	Neuropathy, retinopathy, nephropathy
7	163	72	27.0	57	1	Liraglutide	Rosuvastatin	8.1	48	-
8	166	76	27.6	72	3	Metformin	Amlodipine, Acetylsalicylic acid	7.4	49.3	Neuropathy
9	157	70	28.4	65	5	Metformin	-	7.6	41.4	Neuropathy, retinopathy
10	158	105.0	42	72	10	Metformin, Glibenclamide	Metoprolol, Digoxin, Atorvastatin, Famotidine, Dabigatran, Torasemide, Azilsartan	8.1	47	Neuropathy
11	166	69	25.0	65	10	Metformin, Liraglutide	Moxonidine, Rosuvastatin	7.7	49.4	Neuropathy
12	156	72	29.5	58	5	Metformin, Glimepiride, Vildagliptin	Losartan	8.4	46.7	Neuropathy



Table 4. Cont.

Participants	Height (cm)	Weight (kg)	Body Mass Index	Age (Years)	Disease Duration (Years)	Therapy		HbA <sub>1c</sub> (%)	Content of Albumin ( $\mu\text{g}/\mu\text{L}$ ) in Blood Serum	Complications of Type 2 Diabetes Mellitus
						Diabetes-Specific Therapy	Diabetes-Nonspecific Therapy			
13	156	67	27.5	72	9	Metformin	Levothyroxine, Perindopril, Indapamide, Metoprolol, Acetylsalicylic acid, Magnesiumhydroxide, Fenofibrate, Rosuvastatin	7.7	48.4	Neuropathy
14	160	76	29.7	57	10	Metformin	Hydrochlorothiazide, Eprosartan, Acetylsalicylic acid, Magnesiumhydroxide	8.1	48.3	Neuropathy
15	167	89	31.9	67	10	Metformin, Sitagliptin	Amlodipine, Losartan, Acetylsalicylic acid, Magnesium hydroxide, Thioctic acid	7.7	41.4	Neuropathy, retinopathy
16	148	72	32.8	70	7	Metformin, Gliclazide	Atorvastatin, Amlodipine, Acetylsalicylic acid, Magnesiumhydroxide, Allopurinol	9.9	29.4	Neuropathy, retinopathy
17	172	105	35.5	64	3	Metformin, Glimepiride, Vildagliptin	Losartan, Hydrochlorothiazide, Atorvastatin, Acetylsalicylic acid	7.6	53.4	Neuropathy, retinopathy
18	160	76	29.7	70	12	Metformin, Liraglutide	Valsartan, Verapamil, Indapamide, Thioctic acid	8.2	50.2	Neuropathy, retinopathy
19	159	144	45	46	18	Metformin, Vildagliptin, Gliclazide	Losartan, Amlodipine, Moxonidine, Indapamide, Acetylsalicylic acid	10.2	34.4	Neuropathy, retinopathy
20	164	72.1	26.8	74	10	Metformin, Glibenclamide	Amlodipine, Bisoprolol, Atorvastatin, Acetylsalicylic acid	9.3	49	Neuropathy

**Table 5.** Physical and biochemical parameters of individuals without diabetes, as well as corresponding therapy, not related to diabetes.

Participants	Height (cm)	Weight (kg)	Body Mass Index	Age (Years)	Therapy		HbA <sub>1C</sub> (%)	Content of Albumin in Blood (µg/µL)	Complications of Type 2 Diabetes Mellitus
					Diabetes-Specific Therapy	Diabetes-Nonspecific Therapy			
1	162	78	29.7	54	-	Enalapril	5.8	47.3	-
2	156	82	33.7	67	-	-	6.2	46	-
3	168	84	29.8	67	-	-	5.7	47.8	-
4	171	77	26.6	57	-	Acetylsalicylic acid	5.4	49.2	-
5	159	83	32.8	57	-	Amlodipine, Atorvastatin	5.5	53.4	-
6	158	75	30.0	65	-	Losartan, Indapamide, Metoprolol, Acetylsalicylic acid, Atorvastatin	6.0	46.4	-
7	158	64.1	25.7	64	-	-	4.9	45.9	-
8	163	83	31.2	61	-	Losartan, Acetylsalicylic acid, Calcium carbonate, Colecalciferol, Alfacalcidol	5.1	39.8	-
9	158	64	25.6	64	-	Enalapril	5.2	50	-
10	173	67	22.4	58	-	-	5.5	40.1	-
11	163	72	27.1	61	-	-	5.7	48.4	-
12	165	91.8	33.7	55	-	Amlodipine, Valsartan, Hydrochlorothiazide	5.3	49	-
13	154	59	24.9	57	-	Atorvastatin, Acetylsalicylic acid, Magnesiumhydroxide, Indapamide, Calciumcarbonate, Colecalciferol	5.4	49.9	-
14	166	70	25.4	56	-	Bisoprolol, Red vine leafextract	5.9	43.2	-
15	162	85	32.4	63	-	Losartan, Hydrochlorothiazide	5.8	49	-
16	164	78	29.0	55	-	Losartan, Hydrochlorothiazide	5.0	50.2	-
17	172	95	32.1	64	-	Perindopril, Indapamide, Amlidipine	5.1	39	-
18	155	72.5	29.8	68	-	Rosuvastatin, Perindopril, Indapamide, Levothyroxine, Calcium carbonate, Colecalciferol, Ibandronic acid	5.6	48.2	-

#### 4.3. Blood Sampling and Plasma Isolation

The blood samples (approximately 10 mL each) were collected in polypropylene tubes coated with ethylenediaminetetraacetic acid (Becton Dickinson, Franklin Lakes, NJ, USA). Plasma was separated by centrifugation (1200× *g*, 15 min, 4 °C) and transferred to 1.5 mL polypropylene tubes. The total plasma protein contents were determined by Bradford assay in a 96-well microtiter plate format as described by Greifenhagen and co-workers [51]. The precision of protein determination was verified by SDS-PAGE according to an established protocol [11]. Average densities across individual lanes (expressed in arbitrary units) were determined by a ChemiDoc XRS imaging system controlled by Quantity One<sup>®</sup> 1-D analysis software (Bio-Rad Laboratories Ltd., Moscow, Russia). Thereby, for inter-gel normalization, the first and the last plasma protein samples loaded on each gel were replicated in the previous and following gels, respectively (Figure S1). For the calculation of RSDs, the densities of individual lines were normalized to the gel average value. Individual plasma samples were split into aliquots of 20 µL, and stored at −80 °C. Alternatively, 30 µL of each T2DM sample were combined to obtain a representative pool of diabetic material. The plasma contents of human serum albumin were determined calorimetrically at 628 nm after a color reaction with bromocresol purple (Clinical Chemistry Analyzer CA90, Furuno Electric Co. LTD, Nishinomiya, Japan) using an Albumin kit (Aptec Diagnostics nv) and the control serum kits, Precinorm U and Precipath U (Roche Diagnostics). The plasma HbA<sub>1c</sub> level was measured using the commercial kit, BioRad (Hercules, CA, USA), on a hemoglobin analyzer BioRad D-10 (Bio-Rad Laboratories Inc, Hercules, CA, USA).

#### 4.4. Tryptic Digestion

The plasma proteins were digested according to Frolov and co-workers [27] with slight modifications. Briefly, aliquots of plasma containing 150 µg of protein were diluted with 100 mmol/L ammonium bicarbonate buffer (pH 8.0), complemented with 10 µL of SDS (0.5% in water, *w/v*) and 10 µL of 50 mmol/L TCEP in 100 mmol/L ammonium bicarbonate buffer, and incubated for 30 min at 37 °C under continuous shaking (450 rpm). Afterwards, the samples were cooled to room temperature (RT), and 11 µL of 100 mmol/L iodoacetamide in 100 mmol/L ammonium bicarbonate buffer were added, and alkylation of free thiols was performed during 15 min in darkness at RT. After completion of the incubation, the proteins were sequentially digested at 37 °C with trypsin (25 µg/mL in 100 mmol/L ammonium bicarbonate buffer) taken in the 1:20 and 1:40 (*w/w*) enzyme-protein ratio for 5 and 12 h, respectively, under continuous shaking (450 rpm).

The completeness of the digest was verified by SDS-PAGE as described by Schmidt and co-workers [52] with modifications. Briefly, aliquots of digested samples containing 5 µg of protein were diluted with sample buffer (65.8 mmol/L Tris-HCl, pH 6.8, 20% (*v/v*) glycerol, 2% SDS, 10% (*v/v*) β-mercaptoethanol, 0.05% (*v/v*) bromophenol blue) at least 2-fold and heated at 95 °C for 5 min. Afterwards, the samples were separated on a polyacrylamide gel (T = 12.00%, C = 2.65%), and stained with colloidal Coomassie Brilliant Blue G 250 dye. The digests were frozen and stored at −80 °C before further analysis.

#### 4.5. Boronic Acid Affinity Chromatography

Enrichment of glycosylated peptides was performed according to Soboleva and co-workers [28] with slight modifications. In detail, the pH of tryptic digests was adjusted to 8.0 with 25% (*v/v*) ammonia hydroxide using indicator paper (Lachema, Brno, Czech Republic), before 400 µL of ice cold (4 °C) loading buffer (250 mmol/L ammonium acetate, 50 mmol/L magnesium acetate, pH 8.1) were added. The samples were loaded on 1 mL polypropylene gravity flow columns packed with m-aminophenylboronic acid (mAPBA) agarose, and unbound peptides were washed out with 12 mL of ice-cold (4 °C) loading buffer. Afterwards, glycosylated peptides were sequentially eluted with 0.1 and 0.2 mol/L warm (37 °C) acetic acid (8 and 2 mL, respectively). The eluates were combined, and loaded on

Oasis HLB SPE cartridges, installed on the VacElut 12 Manifold (Agilent Technologies, Moscow, Russia), pre-conditioned with 1 mL of methanol, and pre-equilibrated with 2 mL of 0.1% (*v/v*) aqueous (aq.) formic acid. After a wash with 2 mL of 0.1% (*v/v*) formic acid, peptides were eluted in a step gradient of 40%, 60%, and 80% acetonitrile (0.33 mL each), as described by Spiller et al. [29]. The SPE-eluates were combined, dried under vacuum by a CentriVap Vacuum Concentrator (Labconco, Kansas City, MO, USA), and stored at  $-20\text{ }^{\circ}\text{C}$  for further analysis.

#### 4.6. LC-MS Analysis

The dried eluates were reconstituted in 40  $\mu\text{L}$  of 3% (*v/v*) acetonitrile in aq. 0.1% (*v/v*) formic acid, and 8  $\mu\text{L}$  of the obtained solutions were loaded on a ZORBAX SB column (C18, ID 0.3 mm, length 150 mm, particle size 3.5  $\mu\text{m}$ , Agilent Technologies, Moscow, Russia) using an Agilent 1200 Compact liquid chromatograph equipped with an Agilent 1200 Infinity autosampler and Agilent 1260 Infinity capillary pump (Agilent Technologies, Moscow, Russia). The eluents, A and B, were 4% and 90% acetonitrile, respectively, both containing 0.1% (*v/v*) formic acid. After a 5-min isocratic step (0% eluent B), glycosylated peptides were eluted at the flow rate of 5  $\mu\text{L}/\text{min}$  at 25  $^{\circ}\text{C}$  in sequential linear gradients to 45% and to 100% eluent B in 30 and 2 min, respectively. The column effluents were introduced on-line in an Agilent 6538 Ultra High Definition Accurate-Mass Q-TOF quadrupole-time of flight (QqTOF) mass spectrometer via a dual electrospray ionization (ESI) source (Agilent Technologies, Moscow, Russia). The instrument was operated in the positive ion mode under the settings summarized in Table S7 and controlled by MassHunter Workstation software (Agilent Technologies, Moscow, Russia). Analyte annotation and label free quantification relied on TOF-MS scans acquired in the mass range of 400–2000 *m/z*. Thereby, a pooled enriched tryptic digest, obtained from T2DM patients, was used as an external QC, injected after each eighth sample. Prospective Amadori-modified peptides were annotated by  $t_{\text{RS}}$  and exact *m/z* values (mass accuracy better than 3 ppm). Relative abundances of individual prospectively glycosylated peptides were calculated by integration of characteristic extracted ion chromatograms (XICs,  $m/z \pm 0.02$ ) built for the annotated *m/z* values at specific  $t_{\text{RS}}$  (quantitative analysis tool of the MassHunter Workstation software).

#### 4.7. MS/MS Analysis

The sequences and glycosylation status of differentially abundant glycosylated peptides were confirmed by MS/MS using a combination of DDA and targeted MS/MS experiments using an LTQ-Orbitrap Velos Pro mass spectrometer. For this, a digest of pooled T2DM plasma (1.5  $\mu\text{g}$ ) was loaded on the Acclaim®PepMap100 pre-column (C18-phase, ID 75  $\mu\text{m}$ , length 2 cm, particle size 3  $\mu\text{m}$ ), and separated on the EASY-Spray ES803 C18 column (500  $\times$  0.075 mm, 2  $\mu\text{m}$  particle size, 40  $^{\circ}\text{C}$ ) using an EASY-nLC 1000 nano liquid chromatography system controlled by the Xcalibur 2.1.0 software. Eluents A and B were water and acetonitrile, respectively, both containing 0.1% (*v/v*) formic acid. The analytes were eluted at the flow rate of 0.3  $\mu\text{L}/\text{min}$  as follows: 5% B (0–15 min), 5% to 40% B (15–195 min), 40% to 80% B (195–197 min). The column effluent was transferred on-line to MS (operated in positive ion mode and controlled by Xcalibur 2.1.0 software) via an EASY-spray ion source. DDA analyses relied on survey Orbitrap-MS scans followed by dependent linear ion trap (LIT) ones (collision induced dissociation (CID) fragmentation, He, 35% normalized collision energy) acquired for the 20 most intense signals with charge states of two or more. For targeted MS/MS, 10 LIT MS2 sub-experiments per liquid chromatography-tandem mass-spectrometry (LC-MS/MS) run were acquired for the two most intense quasi-molecular ions of each peptide, not fragmented in DDA experiments. The MS settings are summarized in Table S8. Tandem mass spectra were searched against a FASTA file containing sequences of the proteins annotated by QqTOF-MS using the Mascot search engine within Proteome Discoverer 1.4 software (Thermo-Fisher Scientific, Bremen, Germany) with the following settings: Peptide tolerance—7 ppm, MS/MS match tolerance—0.8 Da, 3 missed cleavage sites per peptide. Glycosylated peptides were annotated by their  $t_{\text{RS}}$ , *m/z* values, and isotopic patterns. The results were

filtered with consideration of peptide confidence (medium), rank (one), false positive rate (0.05), and post-translational modifications—carbamidomethylation (C), oxidation (M), and glycation (K).

#### 4.8. Statistical Analysis

The statistical significance of differences in the relative abundances of specific glycosylated peptides detected in the plasma protein digests obtained from T2DM patients and healthy individuals was determined by the Mann-Whitney *U* test [53]. Holm-Bonferroni correction was applied to adjust significance in terms of multiple hypothesis tests [46]. Since the intensities observed for the signals representing individual charge states of the same peptides demonstrated high correlation, these values were averaged for further processing. To distinguish the T2DM patients from the age-matched controls without diabetes, a linear discriminant analysis (LDA) approach was applied as described by Venables and Ripley [54]. All LDA-related procedures were performed in R (MASS package); all other calculations were performed in MS Excel with RealStatisticsaddon (<http://www.real-statistics.com>).

#### 4.9. Generation of the LDA Model

Because the initial correlation analysis of the feature space showed a high degree of multicollinearity between factors (see Supplementary Figure S6 for a factor cross-correlation plot), we employed two strategies for feature space optimization prior to LDA calculation: (i) PCA-based transformation of the feature space, and (ii) generation of an orthogonal set of original features by the removal of factors with high variance inflation factor (VIF) values [55]. The latter can be calculated for any variable in the set (1) and is a numeric measure of to which extent the variable in question can be predicted by the linear combination of other variables, with the most commonly used critical VIF values being 5 and 10, corresponding to the multiple correlation of 90% and 95% between the tested variables and the rest variables in the set.

$$VIF_i = \frac{1}{1 - R_i^2}, \quad (1)$$

where:  $VIF_i$ —variance inflation factor for variable  $i$ ;  $R_i^2$ —coefficient of determination, i.e., the proportion of the variance in the variable,  $I$ , which is predictable based on another independent variable

The application of PCA was straightforward and only the PCs responsible for 95% of the original sample variance were selected. The corresponding values were estimated for each sample, and the resulting reduced feature set was subjected to further processing with LDA. The VIF-based filtering relied on an iterative procedure, employing a random subset of two features (a test sub-set), and defined in the beginning of each calculation run. Afterwards, further features were randomly selected from the original set and sequentially included in the test subset in each new analysis run. Then, individual features with the VIF values above a selected cutoff (set as described above [43]) were iteratively removed from the test subset in the order of decreasing values. This procedure was further applied in this way to test all initial features.

#### 4.10. Validation of the LDA Model

To assess the model performance, two complementary approaches were employed. In terms of the first one, accuracy, sensitivity, and specificity were determined for the established set of diagnostic peptides. To avoid the overfitting problem [56], parameterization of the LDA model was performed *in silico* with a set of 400 calculated sample patterns (200 for each cohort), obtained by the sub-sampling feature distributions of original samples as described by Politis and co-authors [57]. Thereby, all selected training samples shared less than 25% of the factor values, obtained with any of the original (verification) vectors. The second approach relied on the so-called leave-one-out cross-validation procedure [58]. It employed a set of iterations, where at each run one sample from the original sample set was taken for verification, whereas the other samples were used for parameterization of the LDA model. The performance characteristics (i.e., accuracy, sensitivity, and specificity) were calculated from the set

of individual iteration outcomes as described elsewhere [59,60]. For the PCA-generated feature subset, only the leave-one-out cross-validation procedure was performed.

## 5. Conclusions

T2DM is one of the most widely spread metabolic disorders. Typically, the first stages of the disease are slow and are not accompanied by any clinically manifested symptoms. Because of this reason, T2DM is most often discovered at the step of complications, which makes therapy less efficient and more expensive. Unfortunately, HbA<sub>1c</sub>—a recognized T2DM marker—delivers information about changes in glycaemic status over three months, and, hence, is insensitive to short-term glucose excursions preceding the disease. Thus, control of glycemic status over various, especially short, periods of time might increase the rates of early T2DM discovery. In this context, our approach might bring the desired “time dimension” in glycemic control. Indeed, the integrated biomarker, proposed here, not only covers three weeks before blood sampling, but also indicates a continuous character of the glycation process throughout this period. Secondly, although our integrated biomarker relies on multiple glycation sites, all required information is acquired in one experiment, which is advantageous in comparison to the approaches employing several tests. Finally, the proposed marker has potential for further optimization. Thus, it can be “tuned” for shorter times by excluding relatively long-living proteins. On the the other hand, implementation of immunoaffinity depletion might essentially increase the pattern of marker peptides, and, hence, the selection of marker proteins and the reliability of the marker.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/1422-0067/20/9/2329/s1>.

**Author Contributions:** A.S. performed experiments, analyzed mass-spectrometry (MS) data, made statistical analysis. A.F. designed experiment, supervised the project, performed MS experiments and wrote the manuscript. T.G. contributed advice and reagents/materials. T.K. selected patients for experiment and provided plasma samples. D.T. and W.H. made MS analysis and peptide search. C.I. performed MS analysis. G.M.-S. and V.S. made statistical analysis.

**Funding:** The reported study was funded by Russian Foundation for Basic Research according to the research project № 18-34-00927 and travel grant from Saint-Petersburg State University (№ 28223644).

**Acknowledgments:** We thank Anna Bystrova (Federal Almazov North-West Medical Research Centre) for blood collection and analysis of HbA<sub>1c</sub> and serum albumin. We thank Elena Lukashева, Inna Zorina, Maria Vikhnina, Anna Didio, Veronika Chantseva, Mikhail Bankin (Department of Biochemistry and Department of Plant Physiology and Biochemistry, Faculty of Biology, St. Petersburg State University) for help in Bradford assay. We thank also Cell Technology core facilities of the St. Petersburg University for technical support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

AGEs	advanced glycation end products
AU	arbitrary units
BAC	boronic acid affinity chromatography
DDA	data-dependent acquisition
ESI	electrospray ionization
FPG	fasting plasma glucose
Ig	immunoglobulin
HbA <sub>1c</sub>	glycated hemoglobin
HPLC	high performance liquid chromatography
HR-MS	high resolution mass spectrometry
HSA	human serum albumin
LC-MS	liquid chromatography-mass spectrometry
LC-MS/MS	liquid chromatography-tandem mass-spectrometry
LDA	linear discriminant analysis
LOOCV	leave-one-out cross validation
LIT	linear ion trap

mAPBA	m-aminophenylboronic acid
MRM	multiple reaction monitoring
MS	Mass-spectrometry
MS/MS	tandem mass spectrometry
nanoLC-MS	nano-scaled liquid chromatography-mass spectrometry
nanoRP-HPLC-ESI/MS	nano-scaled reversed-phase high performance liquid chromatography-electrospray ionization- tandem mass-spectrometry
PCA	principle component analysis
PCs	principle components
QC	quality control
QqQ	triple quadrupole
QqTOF	quadrupole-time of flight
RPC	reversed phase chromatography
RP-HPLC	reversed-phase high performance liquid chromatography
RSD	relative standard deviation
RT	room temperature
SDS	sodium dodecyl sulfate
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SPE	solid phase extraction
$\tau_{1/2}$	half-life times
T2DM	type 2 diabetes mellitus
TCEP	tris-(2-carboxyethyl)-phosphine hydrochloride
TOF	time-of-flight
$t_{R_s}$	retention times
VIF	variance inflation factor
XICs	extracted ion chromatograms

## References

1. Berbudi, A.; Ajendra, J.; Wardani, A.P.F.; Hoerauf, A.; Hübner, M.P. Parasitic helminths and their beneficial impact on type 1 and type 2 diabetes. *Diabetes Metab. Res. Rev.* **2016**, *32*, 238–250. [[CrossRef](#)] [[PubMed](#)]
2. Association, A.D. 2. Classification and diagnosis of diabetes. *Diabetes Care* **2015**, *38*, S8–S16. [[CrossRef](#)] [[PubMed](#)]
3. Colagiuri, S.; Davies, D. The value of early detection of type 2 diabetes. *Curr. Opin. Endocrinol. Diabetes Obes.* **2009**, *16*, 95–99. [[CrossRef](#)] [[PubMed](#)]
4. Nyenwe, E.A.; Jerkins, T.W.; Umpierrez, G.E.; Kitabchi, A.E. Management of type 2 diabetes: Evolving strategies for the treatment of patients with type 2 diabetes. *Metabolism* **2011**, *60*, 1–23. [[CrossRef](#)]
5. Singh, V.P.; Bali, A.; Singh, N.; Jaggi, A.S. Advanced glycation end products and diabetic complications. *Korean J. Physiol. Pharmacol.* **2014**, *18*, 1–14. [[CrossRef](#)] [[PubMed](#)]
6. Hodge, J.E. The Amadori rearrangement. *Adv. Carbohydr. Chem.* **1955**, *10*, 169–205.
7. Ahmed, N.; Thornalley, P.J. Quantitative screening of protein biomarkers of early glycation, advanced glycation, oxidation and nitrosation in cellular and extracellular proteins by tandem mass spectrometry multiple reaction monitoring. *Biochem. Soc. Trans.* **2003**, *31*, 1417–1422. [[CrossRef](#)] [[PubMed](#)]
8. Janda, K.; Krzanowski, M.; Gajda, M.; Dumnicka, P.; Jasek, E.; Fedak, D.; Pietrzycka, A.; Kuźniewski, M.; Litwin, J.A.; Sułowicz, W. Vascular effects of advanced glycation end-products: Content of immunohistochemically detected AGEs in radial artery samples as a predictor for arterial calcification and cardiovascular risk in asymptomatic patients with chronic kidney disease. *Dis. Markers* **2015**, *2015*, 153978.
9. Wautier, M.P.; Wautier, J.L. Advanced glycation end products and retinal vascular lesions in diabetes mellitus. *Austin J. Endocrinol. Diabetes* **2015**, *2*, 1034.
10. Rabbani, N.; Thornalley, P.J. Hidden complexities in the measurement of fructosyl-Lysine and advanced glycation end products for risk prediction of vascular complications of diabetes. *Diabetes* **2015**, *64*, 9–11. [[CrossRef](#)] [[PubMed](#)]



11. Greifenhagen, U.; Frolov, A.; Blüher, M.; Hoffmann, R. Site-specific analysis of advanced glycation end products in plasma proteins of type 2 diabetes mellitus patients. *Anal. Bioanal. Chem.* **2016**, *408*, 5557–5566. [[CrossRef](#)] [[PubMed](#)]
12. Nielson, C.; Lange, T. Blood glucose and heart failure in nondiabetic patients. *Diabetes Care* **2005**, *28*, 607–611. [[CrossRef](#)] [[PubMed](#)]
13. Cohen, R.M.; Haggerty, S.; Herman, W.H. HbA1c for the diagnosis of diabetes and prediabetes: is it time for a mid-course correction? *J. Clin. Endocrinol. Metab.* **2010**, *95*, 5203–5206. [[CrossRef](#)]
14. Bozkaya, G.; Ozgu, E.; Karaca, B. The association between estimated average glucose levels and fasting plasma glucose levels. *Clin. Sao Paulo Braz.* **2010**, *65*, 1077–1080. [[CrossRef](#)]
15. Škrha, J.; Šoupal, J.; Škrha, J.; Prázný, M. Glucose variability, HbA1c and microvascular complications. *Rev. Endocr. Metab. Disord.* **2016**, *17*, 103–110. [[CrossRef](#)] [[PubMed](#)]
16. Roohk, H.V.; Zaidi, A.R. A review of glycated albumin as an intermediate glycation index for controlling diabetes. *J. Diabetes Sci. Technol.* **2008**, *2*, 1114–1121. [[CrossRef](#)] [[PubMed](#)]
17. Kouzuma, T.; Usami, T.; Yamakoshi, M.; Takahashi, M.; Imamura, S. An enzymatic method for the measurement of glycated albumin in biological samples. *Clin. Chim. Acta Int. J. Clin. Chem.* **2002**, *324*, 61–71. [[CrossRef](#)]
18. Mashiba, S.; Uchida, K.; Okuda, S.; Tomita, S. Measurement of glycated albumin by the nitroblue tetrazolium colorimetric method. *Clin. Chim. Acta Int. J. Clin. Chem.* **1992**, *212*, 3–15. [[CrossRef](#)]
19. Ikeda, K.; Sakamoto, Y.; Kawasaki, Y.; Miyake, T.; Tanaka, K.; Urata, T.; Katayama, Y.; Ueda, S.; Horiuchi, S. Determination of glycated albumin by enzyme-linked boronate immunoassay (ELBIA). *Clin. Chem.* **1998**, *44*, 256–263. [[PubMed](#)]
20. Morais, M.P.P.; Mackay, J.D.; Bhamra, S.K.; Buchanan, J.G.; James, T.D.; Fossey, J.S.; van den Elsen, J.M.H. Analysis of protein glycation using phenylboronate acrylamide gel electrophoresis. *Proteomics* **2010**, *10*, 48–58. [[CrossRef](#)]
21. Shima, K.; Ito, N.; Abe, F.; Hirota, M.; Yano, M.; Yamamoto, Y.; Uchida, T.; Noguchi, K. High-performance liquid chromatographic assay of serum glycated albumin. *Diabetologia* **1988**, *31*, 627–631. [[CrossRef](#)]
22. Rabbani, G.; Ahn, S.N. Structure, enzymatic activities, glycation and therapeutic potential of human serum albumin: A natural cargo. *Int. J. Biol. Macromol.* **2019**, *123*, 979–990. [[CrossRef](#)]
23. Barnaby, O.S.; Cerny, R.L.; Clarke, W.; Hage, D.S. Comparison of modification sites formed on human serum albumin at various stages of glycation. *Clin. Chim. Acta Int. J. Clin. Chem.* **2011**, *412*, 277–285. [[CrossRef](#)]
24. Frolov, A.; Hoffmann, R. Analysis of amadori peptides enriched by boronic acid affinity chromatography. *Ann. N. Y. Acad. Sci.* **2008**, *1126*, 253–256. [[CrossRef](#)]
25. Frolov, A.; Hoffmann, R. Identification and relative quantification of specific glycation sites in human serum albumin. *Anal. Bioanal. Chem.* **2010**, *397*, 2349–2356. [[CrossRef](#)]
26. Zhang, Q.; Monroe, M.E.; Schepmoes, A.A.; Clauss, T.R.W.; Gritsenko, M.A.; Meng, D.; Petyuk, V.A.; Smith, R.D.; Metz, T.O. Comprehensive identification of glycated peptides and their glycation motifs in plasma and erythrocytes of control and diabetic subjects. *J. Proteome Res.* **2011**, *10*, 3076–3088. [[CrossRef](#)]
27. Frolov, A.; Blüher, M.; Hoffmann, R. Glycation sites of human plasma proteins are affected to different extents by hyperglycemic conditions in type 2 diabetes mellitus. *Anal. Bioanal. Chem.* **2014**, *406*, 5755–5763. [[CrossRef](#)]
28. Soboleva, A.; Modzel, M.; Didio, A.; Płóciennik, H.; Kijewska, M.; Grischina, T.; Karonova, T.; Bilova, T.; Stefanov, V.; Stefanowicz, P.; et al. Quantification of prospective type 2 diabetes mellitus biomarkers by stable isotope dilution with bi-labeled standard glycated peptides. *Anal. Methods* **2017**, *9*, 409–418. [[CrossRef](#)]
29. Spiller, S.; Frolov, A.; Hoffmann, R. Quantification of specific glycation sites in human serum albumin as prospective type 2 diabetes mellitus biomarkers. *Protein Pept. Lett.* **2017**. [[CrossRef](#)]
30. Spiller, S.; Li, Y.; Blüher, M.; Welch, L.; Hoffmann, R. Glycated lysine-141 in haptoglobin improves the diagnostic accuracy for type 2 diabetes mellitus in combination with glycated hemoglobin HbA1c and fasting plasma glucose. *Clin. Proteomics* **2017**, *14*, 10. [[CrossRef](#)]
31. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **2007**, *389*, 1017–1031. [[CrossRef](#)]
32. Zhang, Q.; Tang, N.; Schepmoes, A.A.; Phillips, L.S.; Smith, R.D.; Metz, T.O. Proteomic profiling of nonenzymatically glycated proteins in human plasma and erythrocyte membranes. *J. Proteome Res.* **2008**, *7*, 2025–2032. [[CrossRef](#)]

33. Demirag, B.; Sarisozen, B.; Durak, K.; Bilgen, O.F.; Ozturk, C. The effect of alpha-2 macroglobulin on the healing of ruptured anterior cruciate ligament in rabbits. *Connect. Tissue Res.* **2004**, *45*, 23–27. [[CrossRef](#)]
34. Carpenter, C.B.; Ruddy, S.; Shehadeh, I.H.; Müller-Eberhard, H.J.; Merrill, J.P.; Austen, K.F. Complement metabolism in man: Hypercatabolism of the fourth (C4) and third (C3) components in patients with renal allograft rejection and hereditary angioedema (HAE). *J. Clin. Investig.* **1969**, *48*, 1495–1505. [[CrossRef](#)]
35. Weinstein, S.J.; Stolzenberg-Solomon, R.Z.; Kopp, W.; Rager, H.; Virtamo, J.; Albanes, D. Impact of circulating vitamin D binding protein levels on the association between 25-hydroxyvitamin D and pancreatic cancer risk: A nested case-control study. *Cancer Res.* **2012**, *72*, 1190–1198. [[CrossRef](#)]
36. Olivieri, S.; Conti, A.; Iannaccone, S.; Cannistraci, C.V.; Campanella, A.; Barbariga, M.; Codazzi, F.; Pelizzoni, I.; Magnani, G.; Pesca, M.; et al. Ceruloplasmin oxidation, a feature of Parkinson's disease CSF, inhibits ferroxidase activity and promotes cellular iron retention. *J. Neurosci.* **2011**, *31*, 18568–18577. [[CrossRef](#)]
37. Nanjee, M.N.; Crouse, J.R.; King, J.M.; Hovorka, R.; Rees, S.E.; Carson, E.R.; Morgenthaler, J.-J.; Lerch, P.; Miller, N.E. Effects of intravenous infusion of lipid-free Apo A-I in humans. *Arterioscler. Thromb. Vasc. Biol.* **1996**, *16*, 1203–1214. [[CrossRef](#)]
38. Chung, M.C.-M. Structure and function of transferrin. *Biochem. Educ.* **1984**, *12*, 146–154. [[CrossRef](#)]
39. Sand, K.M.K.; Bern, M.; Nilsen, J.; Noordzij, H.T.; Sandlie, I.; Andersen, J.T. Unraveling the interaction between FcRn and albumin: Opportunities for design of albumin-based therapeutics. *Front. Immunol.* **2015**, *5*. [[CrossRef](#)] [[PubMed](#)]
40. Fedorova, M.; Frolov, A.; Hoffmann, R. Fragmentation behavior of Amadori-peptides obtained by non-enzymatic glycosylation of lysine residues with ADP-ribose in tandem mass spectrometry. *J. Mass Spectrom.* **2010**, *45*, 664–669. [[CrossRef](#)]
41. Frolov, A.; Hoffmann, P.; Hoffmann, R. Fragmentation behavior of glycated peptides derived from D-glucose, D-fructose and D-ribose in tandem mass spectrometry. *J. Mass Spectrom.* **2006**, *41*, 1459–1469. [[CrossRef](#)]
42. American Diabetes Association Diagnosis and classification of diabetes mellitus. *Diabetes Care* **2006**, *29* (Suppl. 1), S43–S48.
43. Forrestel, A.C.; Miedlich, S.U.; Yurcheshen, M.; Wittlin, S.D.; Sellix, M.T. Chronomedicine and type 2 diabetes: Shining some light on melatonin. *Diabetologia* **2017**, *60*, 808–822. [[CrossRef](#)]
44. Neelofar, K.; Ahmad, J. An overview of in vitro and in vivo glycation of albumin: A potential disease marker in diabetes mellitus. *Glycoconj. J.* **2017**, *34*, 575–584. [[CrossRef](#)]
45. Selvin, E.; Rawlings, A.M.; Lutsey, P.L.; Maruthur, N.; Pankow, J.S.; Steffes, M.; Coresh, J. Fructosamine and Glycated Albumin and the Risk of Cardiovascular Outcomes and Death. *Circulation* **2015**, *132*, 269–277. [[CrossRef](#)]
46. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
47. Sharma, A.; Paliwal, K.K. Linear discriminant analysis for the small sample size problem: An overview. *Int. J. Mach. Learn. Cybern.* **2015**, *6*, 443–454. [[CrossRef](#)]
48. Zhang, Q.; Tang, N.; Brock, J.W.C.; Mottaz, H.M.; Ames, J.M.; Baynes, J.W.; Smith, R.D.; Metz, T.O. Enrichment and analysis of nonenzymatically glycated peptides: Boronate affinity chromatography coupled with electron-transfer dissociation mass spectrometry. *J. Proteome Res.* **2007**, *6*, 2323–2330. [[CrossRef](#)]
49. Cawley, G.C.; Talbot, N.L.C. On over-fitting in model selection and subsequent selection bias in performance evaluation. *J. Mach. Learn. Res.* **2010**, *11*, 2079–2107.
50. Aryaeian, N.; Sedehi, S.K.; Arablou, T. Polyphenols and their effects on diabetes management: A review. *Med. J. Islam. Repub. Iran* **2017**, *31*, 134. [[CrossRef](#)]
51. Greifenhagen, U.; Frolov, A.; Blüher, M.; Hoffmann, R. Plasma proteins modified by advanced glycation end products (AGEs) reveal site-specific susceptibilities to glycemic control in patients with type 2 diabetes. *J. Biol. Chem.* **2016**, *291*, 9610–9616. [[CrossRef](#)] [[PubMed](#)]
52. Schmidt, R.; Böhme, D.; Singer, D.; Frolov, A. Specific tandem mass spectrometric detection of AGE-modified arginine residues in peptides. *J. Mass Spectrom.* **2015**, *50*, 613–624. [[CrossRef](#)] [[PubMed](#)]
53. Mann, H.B.; Whitney, D.R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Stat.* **1947**, *18*, 50–60. [[CrossRef](#)]
54. Venables, W.N.; Ripley, B.D. *Modern Applied Statistics with S*; Springer: New York, NY, USA, 2002; ISBN 978-0-387-95457-8.

55. Kutner, M.; Nachtsheim, C.; Neter, J.; Li, W. *Applied Linear Statistical Models*, 5th ed.; McGraw-Hill/Irwin: Boston, MA, USA, 2005; ISBN 978-0-07-310874-2.
56. Pang, Y.; Wang, S.; Yuan, Y. Learning regularized LDA by clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 2191–2201. [[CrossRef](#)] [[PubMed](#)]
57. Politis, D.N.; Romano, J.P.; Wolf, M. *Subsampling*; Springer Series in Statistics; Springer: New York, NY, USA, 1999; ISBN 978-0-387-98854-2.
58. Stone, M. Cross-validated choice and assessment of statistical predictions. *J. R. Stat. Soc. Ser. B Methodol.* **1974**, *36*, 111–147. [[CrossRef](#)]
59. Lachenbruch, P.A.; Mickey, M.R. Estimation of error rates in discriminant analysis. *Technometrics* **1968**, *10*, 1–11. [[CrossRef](#)]
60. Luntz, A.; Brailovsky, V. On estimation of characters obtained in statistical procedure of recognition. *News Acad. Sci. USSR Tech. Cybern.* **1969**, *3*, 48–52.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).