**Diagnostic accuracy of protein glycation sites in long-term controlled patients with type 2 diabetes mellitus and their prognostic potential for early diagnosis**

Sandro Spiller1,2, Yichao Li3, Matthias Blüher4, Lonnie Welch3, Ralf Hoffmann1,2\*

1Institute of Bioanalytical Chemistry, Faculty of Chemistry and Mineralogy, Universität Leipzig

2Center for Biotechnology and Biomedicine (BBZ), Universität Leipzig

3School of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio

4Department of Medicine, Endocrinology and Nephrology, Universität Leipzig

\*Corresponding author

Prof. Dr. Ralf Hoffmann, Biotechnologisch-Biomedizinisches Zentrum (BBZ), Deutscher Platz 5, 04103 Leipzig, Germany

E-Mail: bioanaly@rz.uni-leipzig.de

**Keywords**

Biomarker, fasting plasma glucose (FPG), glycated hemoglobin (HbA1c), glycation sites, multiple reaction monitoring (MRM), plasma proteins, type 2 diabetes mellitus (T2DM)

**Abstract**

*Background*

More than 90% of the estimated 415 million diabetes mellitus patients worldwide aged 20 to 79 suffer from type 2 diabetes mellitus (T2DM), which includes a projected 193 million people who remain undiagnosed. Current screening tests, i.e., glycated hemoglobin (HbA1c) and elevated fasting plasma glucose (FPG), identify less than 50% of undiagnosed T2DM patients and provide no information how the disease will develop in prediabetic patients who are generally considered to advance to T2DM. Thus, additional diagnostic and prognostic biomarkers are required for personalized treatment.

*Methods*

Twenty-nine glycation sites previously identified in ten plasma proteins were quantified after tryptic digestion of plasma samples and enrichment by affinity chromatography at the peptide level using tandem mass spectrometry (multiple reaction monitoring) and isotope-labelled peptides as internal standard. The glycation degrees were determined in 48 patients with a duration of T2DM exceeding ten years, 48 non-diabetic individuals matched for gender, BMI, and age (range: 20-70 years), and 20 prediabetic men (age range: 25-60 years).

*Results*

In long-term controlled diabetic patients were 27 glycated peptides detected at significantly higher levels providing moderate diagnostic accuracies (ACCs) from 61 to 79% allowing a subgrouping of patients in three distinct clusters. Moreover, a feature set of three glycated peptides and four established clinical parameters provided an ACC of 95%. The same number of clusters was identified in prediabetic males, which could be identified with an ACC of 95% using a set of eight glycation sites (mostly from serum albumin). Interestingly, all patients present in one cluster showed progression of prediabetic state or advanced towards diabetes in the following five years.

*Conclusions*

Glycation sites of plasma proteins might be a promising prognostic tool to subgroup prediabetic patients to estimate their risk for the development of T2DM.

**Introduction**

Diabetes mellitus (DM) is a group of diseases characterized by hyperglycemia resulting from absolute insulin deficiency (type 1 DM), or insulin resistance together with a relative insulin secretion defect (type 2 DM). In 2013, DM affected approximately 382 million people worldwide, accounted for more than 1.3 million deaths making it the 8th leading cause of death and reduced life expectancy globally according to a recent report of the World Health Organization (WHO) (1). Around 90 to 95% of the currently estimated 415 million DM patients aged 20 to 79 suffer from type 2 DM (T2DM), which includes an estimated 193 million people who remain undiagnosed (2). Already in the state of prediabetes, the risk for macrovascular complications is increased (3). In this context, an early diagnosis of T2DM is urgently needed to increase the efficacy of established therapeutic strategies and to prevent or at least delay the development of complications, including macro- and microvascular diseases and diabetic neuropathy.

Recent data indicate that every third adult living in developed countries has prediabetes (4; 5), which is an intermediate state of hyperglycemia defined as impaired fasting glucose (IFG) and/or impaired glucose tolerance (IGT). The latter is detected by an oral glucose tolerance test (OGTT). Both IFG and IGT are risk factors for type 2 diabetes, and risk is even greater when IFG and IGT occur together. Clinically, a prediabetic state is typically diagnosed (WHO criteria, (6)) by either fasting plasma glucose (FPG) levels between 6.1 and 7.0 mmol/L (110 and 125 mg/dL; IFG) or plasma glucose levels between 7.8 and 11.1 mmol/L (140 and 200 mg/dL) after a 2-h OGTT indicating IGT. The American Diabetes Association (ADA) uses the same IGT definition, but a reduced lower cut-off for IFG (5.6 mmol/L) and additionally glycated hemoglobin (HbA1c) levels between 5.7 and 6.4% as criteria (7). Prediabetes is connected to a high risk for developing diabetes and associated complications (8) indicating that early detection of prediabetic states followed by immediate lifestyle interventions including weight reduction, calorie-reduced diets and increased physical activity will allow preventing or slowing down the transition from prediabetes to T2DM (9).

Current diagnostic criteria for the diagnosis of T2DM include plasma glucose levels above the borderline concentrations used for prediabetes, i.e., FPG ≥ 7.0 mmol/L (125 mg/dL) and OGTT ≥ 11.1 mmol/L (200 mg/dL), and HbA1c levels ≥ 6.5% (48 mmol/mol) (10; 11). Worryingly, recent epidemiological studies indicate that HbA1c and FPG tests with the currently applied diagnostic criteria only identify ~30-50% of previously undiagnosed T2DM patients (12-16). In addition, HbA1c poorly correlates with IFG and IGT (17-19). Even the diagnosis of prediabetes based on IFG and IGT is still questionable due to the inability of these tests predicting the development of the disease and the rather poor reproducibility in adults and children (20-22).

Besides HbA1c, a well-established diagnostic marker and glucose control monitoring parameter determined in red blood cells, glycated serum proteins have been recognized as markers of hyperglycemia (23-27). Due to the rapid turnover of serum proteins compared to erythrocytes (life span ~4 months), glycated proteins reflect mean concentrations of glucose in blood over a shorter period (days to weeks) than HbA1c (28; 29). Thus, they might be useful biomarkers for monitoring short-term fluctuations of glucose plasma concentrations. Fructosamine and glycated albumin strongly correlate with HbA1c and FPG (30-34), but the lack of clinical cut-off points and standardized assays are reasons for low clinical acceptance (29; 35; 36). This might be attributed to the currently applied methods determining overall protein glycation degrees, whereas recent data indicate that specific glycation sites in plasma proteins might be more favorable biomarkers for early diagnoses of T2DM (37-40).

Recently, we reported 27 glycation sites of nine plasma proteins as markers of manifest T2DM (41). All 27 glycation sites were present at significantly higher levels in samples from patients with T2DM compared to age- and body mass index (BMI)-matched control individuals. When combined with established diagnostic criteria, i.e., HbA1c and FPG, the diagnostic accuracies improved significantly. Interestingly, glycation site Lys141 of haptoglobin (HP K141) provided the best sensitivities of ~94% and ~78% when combined with HbA1c and FPG levels, respectively, and a specificity of ~98%.

Here we extended our recent study by exploring the diagnostic value of 29 glycation sites originating from ten plasma proteins for prediabetes and long-term controlled T2DM patients. The glycation sites were quantified in 48 samples obtained from long-term controlled T2DM patients, 20 individuals with prediabetes, and 48 normoglycemic individuals.

**Research design and methods**

*Study participants*

Forty-eight patients with a duration of T2DM for > 10 years and 48 non-diabetic individuals matched for gender, BMI, and age (range: 20-70 years) as well as 20 prediabetic men (age range: 25-60 years) were included into this study (Supplementary Materials, Tables S1-S3). Anthropometric and laboratory chemistry parameters were measured for all plasma samples or calculated as previously described (42; 43). The study was approved by the Ethics Committee of Universität Leipzig (approval no: 159-12-21052012) and performed in accordance to the declaration of Helsinki. All subjects gave written informed consent before taking part in this study. T2DM and prediabetes were diagnosed according to the criteria of ADA (44). T2DM patient samples were grouped by HbA1c levels, i.e., < 6.5% (male n=7, female n=11) ≥ 6.5% (male n=17, female n=13). Patients with HbA1c < 6.5% (48 mmol/mol) were diagnosed on the basis of repeated FPG (> 7.0 mmol/L) or OGTT (> 11.1 mmol/L) assessments (44). Some individuals of the T2DM group received anti-hyperglycemic medication (metformin, DPP-4 inhibitors). Prediabetic individuals (n=20, BMI ≥ 30 kg/m2, age at baseline: 30-61 years) were identified according to ADA criteria (44). After a mean observation period of ~4.1 years, blood samples were taken again from prediabetic subjects for a follow-up examination to measure the same clinical parameters as previously. All subjects had a BMI ≥ 25.0 kg/m2 and were therefore included into a multimodal intervention program consisting of regular dietary advice and 1-2 times per week physical exercise. Noteworthy, the adherence rate to this program was <50%. EDTA blood samples were collected after a twelve-hour fasting period between 8 am and 9 am, centrifuged (500 x g, 5 min) and an aliquot was used to determine routine laboratory parameters within one hour. Cell debris was removed from the remaining aliquot by filtration (Rotilabo® syringe filter) and stored at -80ºC for analyzing the glycation sites. Plasma insulin and proinsulin were measured with an enzyme immunometric assay (IMMULITE automated analyzer, Diagnostic Products Corporation, Los Angeles, CA, USA). Serum high-sensitive CRP (C-reactive protein) was measured by immunonephelometry (Dade-Behring, Milan, Italy). HbA1c, plasma glucose, serum total high-density lipoprotein (HDL)-, low-density lipoprotein (LDL)-cholesterol, triglycerides, and free fatty acids were measured as previously described (42). Homeostasis model assessment as an index of insulin resistance (HOMA-IR) was calculated by multiplying the FPG (mmol/L) with fasting plasma insulin (FPI, mU/L) divided by 22.5 (45). Updated HOMA (HOMA2) of insulin sensitivity (HOMA2 %S) and an estimate of beta cell function (HOMA2 %B) were calculated using FPG and FPI (HOMA calculator Version 2.2.3 at http://www.dtu.ox.ac.uk/homacalculator) (46).

*Peptide quantification*

Twenty-nine glycation sites were quantified at the peptide level (Supplement, Table S4) by electrospray ionization mass spectrometry (ESI-MS) on a QTRAP4000 (AB Sciex, Darmstadt, Germany) coupled on-line to reversed-phase high-performance liquid chromatography (RP-HPLC) using timed multiple reaction monitoring (MRM) (41). Briefly, plasma was ultrafiltrated (5 kDa cut-off), digested with trypsin (37°C, 18 h, 5% w/w), spiked with a concentration-balanced mixture of 13C,15N-labelled glycated peptides as internal standard, enriched for glycated peptides by boronic acid affinity chromatography (BAC), and desalted by solid phase extraction (SPE) (38; 47; 48). Peptides were separated on a C18-column (AdvanceBio Peptide Mapping column, pore size 12 nm, length 15 cm, internal diameter 2.1 mm, particle size 2.7 µm, Agilent Technologies, Böblingen, Germany) coupled on-line to the QTRAP4000. Eluents A and B were water and acetonitrile, respectively, containing both formic acid (0.1%, v/v). The column was equilibrated with 3% eluent B, the sample injected, and peptides were eluted by linear gradients starting 3 min after sample injection to 10% eluent B within 1 min, to 20% eluent B within 10 min, and finally to 95% eluent B in 7 min. The flow rate was 0.3 mL/min and the column temperature was set to 60 °C. Quantification relied on timed MRM using specific transitions of each targeted and isotope-labelled peptide by integrating individual peaks in extracted ion chromatograms (XICs) using Analyst 1.6 software (AB Sciex) (41).

*Statistics and Bioinformatics*

Datasets were evaluated by statistical tests, i.e., Kolmogorow-Smirnow, Mann-Whitney, and t-test, and Spearman rank correlation coefficients using Prism 6 (GraphPad software; La Jolla, USA). Receiver operating characteristic (ROC) analyses and screening for variable combinations relied on the Excel-add-in Multibase 2015 (Numerical Dynamics) and Prism 6 software, respectively.

Diabetes and control samples were classified by a decision tree algorithm using HbA1c, FPG, C-peptide, FPI, HOMA-IR, and HOMA2 %S (Supplementary Tables S5 to S10) in combination with each glycated peptide. The decision tree algorithm was implemented using Scikit-Learn (49). Accuracies were evaluated using nested 10-fold cross validation (50). The best feature set for classification was identified by a random forest-recursive feature elimination (RF-RFE) method (51) that was applied for all glycated peptides and clinical parameters, such as HbA1c, FPG, and BMI. Feature normalization and missing value imputation relied on the WEKA toolkit (52). Accuracies were evaluated using nested 10-fold cross validation (50). The k-means algorithm in Scikit-Learn (49) was applied to find subclasses in diabetic samples. The clustering stability score (53) and elbow criterion (54) were used to find the optimal number of subclass. Predictor screening test based on decision tree algorithm (to optimally bin the predictors) and Fisher exact test [citation]. Positive (+) and negative (-) likelihood ratios (LR) were calculated as previously reported (55; 56).

**Results**

*Long-term controlled diabetic patients*

Twenty-seven out of the 29 analyzed glycated peptides were detected at significantly higher levels in digested sera from T2DM patients than in the control samples (P < 0.05, Figure 1), although both groups were not separated by a cut-off. When all T2DM and control samples were subdivided by an HbA1c threshold of 6.5%, a similar distribution was obtained indicating that the glycation degrees of hemoglobin and the tested serum proteins correlated. Spearman’s rank correlation coefficients (*rS*), which were calculated for all glycated peptide levels and diagnostic parameters, ranged for most combinations of two glycated peptides from 0.37 to 0.98 (P < 0.001, data not shown). Moderate to strong correlations were achieved for all glycation sites with FPG (28 sites, 0.35 < *rS* <0.70, P < 0.001), proinsulin (20 sites, 0.37 < *rS* < 0.54, P < 0.001), HOMA-IR (23 sites, 0.35 < *rS* < 0.62, P < 0.001), and HbA1c values for most sites (0.42 < *rS* < 0.76) (Supplement, Table S11). Nine of the 14 glycation sites studied here for HSA and one site of fibrinogen beta chain (FGB K163) moderately correlated with free fatty acids (FFAs, 0.36 < *rS* < 0.46, P < 0.001), whereas correlations between peptide glycation and BMI were weak (-0.16 < *rS*< 0.25, 0.03 < P < 0.98).

Moderate to strong correlations were observed for HbA1c values with HOMA-IR (*rS* = 0.73, P < 0.001) and FPG (*rS* = 0.79, P < 0.001) as well as alanine aminotransferase (ALAT), gamma glutamyl transferase (GGT), C-peptide, fasting plasma insulin (FPI), proinsulin, triglycerides (TAGs), and FFA (0.36 < *rS*  < 0.68, P < 0.001). FPG moderately correlated (0.38 < *rS*  < 0.68, P < 0.001) with aspartate aminotransferase (ASAT), ALAT, GGT, FPI, proinsulin, TAGs, HOMA-IR, and FFAs.

A ROC curve analysis relative to HbA1c (cut-off 6.5%) and FPG (cut-off 7.0 mmol/L) provided for all glycated peptides maximal sensitivities (SNs) and specificities (SPs) up to 75% and 88%, respectively, and areas under curves (AUCs) up to 82% for certain cut-off concentrations (Supplement, Table S12), which is in accordance with our recent data (57). In comparison, sensitivities, specificities, and areas under curves of HbA1c were 63%, 98%, and 89% (95% confidence interval [0.82-0.96]) and 54%, 98%, and 81% (95% confidence interval [0.72-0.91]) for FPG using the same cut-off values of 6.5% (LR+ = 29.8, LR- = 0.38) and 7.0 mmol/L (LR+ = 25.8, LR- = 0.48), respectively. The analysis determined the best cut-off values maximizing sensitivities and specificities as 6.0% (LR+ = 19.5, LR- = 0.20) for HbA1c (SN = 81% and SP = 96%) and 5.72 mmol/L (LR+ = 4.8, LR- = 0.25) for FPG (SN = 79% and SP = 83%). Interestingly, four further diagnostic parameters also showed sufficient values (AUC ≥ 80%) for the evaluation metrics: AUCs were 81% for C-peptide (cut-off 1.51 nmol/L; SN = 75%, SP = 77%), 84% for FPI (cut-off 115.0 pmol/L; SN = 81%, SP = 85%), 88% for HOMA-IR (cut-off 4.68, SN = 81%, SP = 94%), and 86% for HOMA2 %S (cut-off 36.6%, SN = 77%, SP = 98%).

As the diagnostic accuracies (ACCs) of all glycated peptides (61 to 79%) were insufficient, the data set was screened for variable combinations of each glycated peptide with HbA1c, FPG, C-peptide, FPI, HOMA-IR, and HOMA2 %S for optimal SN, SP, and ACC considering different cut-off points. However, only the combination with C-peptide showed a notably better ACC of 88% (Supplement, Tables S5-10).

Thus, the RF-RFE method was applied to find a set of diagnostic parameters and glycated peptides for maximizing the classification of T2DM patients and controls. It revealed a set of seven features, i.e., C-peptide, FAAs, FPG, FPI, HbA1C, HOMA-IR, and glycated Lys141 of haptoglobin (HP K141, peptide 26), providing a SN of 94%, a SP of 96%, and an ACC of 95% (Figure 2). A cluster analysis performed for all 48 T2DM plasma samples using a k-means algorithm considering all 29 glycated peptides and 43 clinical parameters identified three clusters as optimal considering the elbow criterion (Figure 3).

*Prediabetic patients*

In addition to the diagnostic value of glycation sites, their potential as prognostic biomarkers were also investigated using samples of 20 males diagnosed with prediabetes. Peptide 19 (FGB K295, Supplement, Table S4) was removed from the dataset as it was always present at concentrations below its LODs leaving 28 glycated peptides considered in the following statistical analyses. The correlation coefficients considering two glycated peptide levels typically ranged from 0.37 < *rS* < 0.99 (P < 0.05; data not shown). In combination with the other diagnostic parameters, correlations were usually weak (*rS* > 0.36, *rS* < -0.36), except for FFAs, FPI, and HOMA-IR that showed moderate correlation coefficients for selected peptides, i.e., between FFAs and 15 glycations sites (0.36 < *rS* < 0.57 (0.01 < P < 0.12), glycations sites Lys41, Lys75, and Lys99 of protein Ig kappa chain C region and FPI (*rS* = 0.63, 0.70, and 0.39, respectively; P < 0.05), and HOMA-IR (*rS* = 0.59, 0.67, and 0.40, respectively, ; P < 0.05). Moderate correlations were also observed for HbA1c with total cholesterol (*rS* = -0.46, P =0.04), HDL-cholesterol (*rS* = -0.56, P = 0.01), and OGTT (*rS* = 0.37, P = 0.1). Additionally, the levels of HDL-cholesterol and C-peptide moderately correlated (*rS* = -0.37, P = 0.05). A cluster analysis of prediabetic plasma samples considering all glycation sites and a cluster stability test provided an optimal cluster number of three (Supplement, Figure S1).

The recently reported cut-offs to classify newly diagnosed diabetic patients and control subjects (57) allowed subgrouping prediabetic men (Supplement, Table S13) by counting how often each subject was above the cut-off values of all glycated peptide (Supplement, Table S14). Intriguingly, three clusters could be distinguished again, i.e., 18-24 counts in (cluster 1 (“highly-remarkable”), 6-12 counts in cluster 2 (“remarkable”), and up to 3 counts in cluster 3 (“unremarkable”). The clusters and the respective members were in agreement with the above-mentioned cluster analysis (Supplement, Table S14 and Figure S1). Considering these three clusters, a RF-RFE method was applied to find a set of glycated peptides for maximizing the classification. A set of nine peptides representing eight glycation sites of HSA (Lys262, Lys378, Lys73, Lys525, Lys574, Lys359, Lys174, Lys64) and one of serotransferrin (Lys683) was identified providing an ACC of 95% (Supplement, Figure S2). Noteworthy, Lys262 followed by Lys378, Lys73, and K525 of HSA contributed most to the classification verified by random forest feature importance. (Supplement, Table S15).

The predictive values of all glycation sites were evaluated by reexamining the individuals of clusters 1 to 3 after three to five years (Supplement, Tables S2 and S3). Considering the diagnostic criteria of HbA1c (> 6.5%) and FPG (> 7.0 mmol/L), eight persons converted from prediabetes to T2DM (ctT2DM), seven remained prediabetic with a clear trend towards diabetes (DPD, HbA1c fold change: 1.10-1.21, FPG fold change: 0.93-1.59), and the glycemic status of five individuals remained stable or improved (PD, Figure 4A and B) within the 4 years observation period. Importantly, nine glycation sites, i.e., Lys93, Lys181, Lys262, Lys525, and Lys545 of HSA, Lys99 of Ig kappa chain C region, Lys1003 of alpha-2-macroglobulin, Lys50 of Ig lambda-1 chain C regions, and Lys120 of apolipoprotein A-I precursor were predominantly higher glycated in nine to twelve prediabetic patients. The nine glycation sites showed higher glycation degrees in six out of seven DPD plasma samples (86%), but only in four out of eight ctT2DM samples (50%).

Noteworthy, the status of all persons classified “remarkable” advanced towards diabetes with four already diagnosed with T2DM. Among the six patients classified as “highly remarkable” only one advanced to T2DM, while half of the “unremarkable” group developed T2DM. In addition, highest HOMA-IR values (> 10) and the strongest HOMA-IR changes (> 5.2 to 16.3) were observed for persons classified “remarkable” (Supplement, Table S2). In general, HOMA2 %B (n=14; 3-92% loss) and HOMA2 %S values (n=13, 10-89% loss) decreased in most persons, which were diagnosed with hyperinsulinemia or diabetes (Figure 4C and D). Noteworthy, recommendations for healthy diet and increased physical activity were followed by the patients with a low (<50%) adherence rate (BMI fold change: 0.97 ± 0.05) and therefore, this intervention could not prevent progression of prediabetic state or development towards T2DM.

**Discussion**

Recently, we quantified glycation sites in plasma proteins that might be valuable diagnostic markers to complement currently established diagnostic criteria based on HbA1c and FPG (57), as both criteria failed to detect T2DM, especially in early phases. In this study, persons newly diagnosed with T2DM using different criteria (HbA1c, FPG, OGTT, or random plasma glucose) showed characteristic glycation patterns in plasma proteins that allowed their differentiation from matched healthy controls. The highest sensitivity to diagnose chronic hyperglycemia could be achieved by a combination of the glycation levels of four sites in plasma proteins, i.e., K93, K262, and K414 in HSA and K141 in haptoglobin, in combination with other routine parameters including HbA1c. This feature set provided a favorable diagnostic accuracy of around 98% compared to only 76 and 70% when only HbA1c and FPG, respectively, were used. These promising results motivated us to extend our previous studies by including individuals with varying degrees of disturbances of glucose metabolism and insulin sensitivity who have been monitored for ~4 years. Here, we investigated the diagnostic and prognostic potential of plasma protein glycation sites longitudinally. Thus, we quantified 29 glycation sites originating from ten proteins in plasma samples from long-term controlled diabetic patients (n=48) and matched non-diabetic subjects (control, n=48). In addition, we included analyses of plasma samples obtained from 20 individuals who have initially been classified as prediabetic and re-evaluated after 3-5 years.

A ROC analysis of the data set clearly indicated that the glycation levels of the 29 investigated sites typically provide better SNs (up to 79%) than both HbA1c (SN = 63%, AUC = 89%) and FPG (SN = 54%, AUC = 89%) and only slightly lower AUCs (up to 82%) for classifying T2DM patients and controls. Noteworthy, the three diagnostic parameters FPI (cut-off 115.0 pmol/L; AUC = 84%), HOMA-IR (cut-off 4.68, AUC = 88%), and HOMA2 %S (cut-off 4.68, AUC = 86%), which are typically used to characterize insulin resistance, provided clinically relevant AUCs. Additionally, data reiterate that lowering the cut-off values of HbA1c (from 6.5 to 6.0%) and FPG (from 7.0 to 5.72 mmol/L), as recommended by the WHO, improved the SNs from 63 to 81% and from 54 to 79%, respectively, supporting recent reports (12-16).

In contrast to our previous study on newly diagnosed T2DM patients (57), glycation degrees of long-term controlled patients showed moderate to strong correlations with HbA1c and other parameters of glucose metabolism and insulin sensitivity including FPG and HOMA-IR. This indicates that plasma proteins follow similar glycation kinetics as hemoglobin after manifestation of the disease and that their glycation degree generally increase as insulin sensitivity deteriorates. The previously reported negative correlation of BMI with glycation sites (57) was not observed in the current cohort suggesting that the assumption of higher capacity of adipose tissue in patients with obesity leads to higher uptake of excess glucose is not valid after manifestation of T2DM.

Statistical evaluation of the data revealed a feature set matrix using one glycation site of haptoglobin (K141) and twelve routine parameters typically used to characterize T2DM (FPG, HbA1c, FPI, C-peptide) and IR (HOMA-IR, FFAs) (58) providing a high accuracy of 95% for the cohort, which confirms our previous results (57). This result emphasizes the relevance of HP K141 acting as biomarker for the disease, as HP K141 provided also the highest accuracies in combination with HbA1c and FPG in the previous cohort (57).

Based on cut-off values for the 29 glycation sites determined previously in patients with newly diagnosed T2DM (57), prediabetic subjects could be subdivided into three clusters. Importantly, these three clusters may reflect the individual risk to progress from prediabetes to T2DM, to remain prediabetic or to improve hyperglycemia. Reexamination of these individuals after three to five years, indicated that individuals of one cluster predominantly advanced to T2DM or deterioration of prediabetic status. According to the RF-RFE method, glycation sites of HSA and serotransferrin contributed most to the classification of the three clusters at baseline underlining their relevance for identifying prediabetic subgroups. Since subgroups might respond differently to T2DM treatment strategies or show individual risks for disease progression and developing diabetic complications, the glycation sites might provide a prognostic tool to overcome current diagnostic limitations. However, this study was limited by the small size of discovery cohorts and the limited transferability into clinical routine providing ‘only’ first hints and need to be tested in larger cohorts using at least the set of glycation sites applied here, maybe even further sites from different plasma proteins. Moreover, the value of our glycation site cluster to predict T2DM should be tested in the context of large, prospective epidemiologic studies on the long-term incident risk to develop T2DM. In addition, we propose to investigate whether the combination of HbA1c with glycation levels of plasma proteins (especially HSA, serotransferrin, and haptoglobin) and their dynamics may reflect the individual risk to develop long-term complications of diabetes including diabetic retinopathy, nephropathy and neuropathy.

**Conclusions**

The data obtained here for small, well defined cohorts of long-term diabetic and prediabetic patients confirms the diagnostic potential and for the first time indicates the prognostic value of glycation sites of plasma proteins, which provide similar or better diagnostic accuracies as routinely applied clinical parameters. Interestingly, the combination of glycation sites and established clinical parameters provided the best accuracy (95%). Moreover, the studied glycation sites can subgroup prediabetic patients allowing an estimation of the individual risk of patients to develop T2DM in the following years, which identify persons subject to early therapeutic inventions beyond dietary changes and exercises. In all cases, certain glycation sites of serum albumin, serotransferrin, and haptoglobin provided the best diagnostic and prognostic measures.

**Declarations**

*Ethics approval and consent to participate*

The study was approved by the Ethics Committee of Universität Leipzig (approval no: 159-12-21052012), and performed in accordance to the declaration of Helsinki. Written informed consent was obtained from all participants in this study.

*Consent for publication*

Not applicable

*Availability of supporting data*

The datasets during the current study are available from the corresponding author on reasonable request.

*Competing interests*

The authors declare that they have no competing interests.

*Funding*

Financial support from Deutsche Forschungsgemeinschaft (HO-2222/7-1 to RH and SFB 1052: Obesity Mechanisms, B01 to MB) is gratefully acknowledged.

*Authors’ contributions*

S.S. analyses, data interpretation, statistical analysis, contributed to discussion and manuscript writing; Y.L. and L.W. statistical analysis, data interpretation, contributed to discussion, reviewed/edited manuscript. M.B. sample collection and study design, clinical testing, data interpretation, contributed to discussion and manuscript writing. R.H. study design, contributed to discussion, wrote manuscript.

*Acknowledgements*

We thank Dr. Uta Greifenhagen for support during sample preparation, Dr. Andrej Frolov for helpful discussions about analytical setup, and Dr. Daniel Knappe, Dr. David Singer, and Tina Goldbach for their support in synthesizing the peptides.

*Authors’ information*

1Institute of Bioanalytical Chemistry, Faculty of Chemistry and Mineralogy, Universität Leipzig, Leipzig, Germany. 2Center for Biotechnology and Biomedicine (BBZ), Universität Leipzig, Leipzig, Germany. 3School of Electrical Engineering and Computer Science, Ohio University, Athens, Ohio, USA. 4University Hospital Leipzig, Department for Internal Medicine, Clinic for Endocrinology and Nephrology, Universität Leipzig, Leipzig, Germany.

**Figure 1.** Contents of 12 glycated peptides in tryptic plasma digests obtained from long-term controlled T2DM patients and non-diabetic controls. Samples were split in two groups using an HbA1C level of 6.5% as cut-off. Each dot represents the peptide level of the corresponding peptide in one plasma sample. Dotted line indicate the limit of quantification (LOQ) of the peptide. Peptide sequences, glycation sites, and the corresponding protein are provided as Supplement (Table S2). Statistical significance was tested by a Mann-Whitney U-test (\*\*\* denotes *p* < 0.0001 and \* denotes *p* < 0.05).

**Figure 2.** Principle component plot of 48 T2DM patients (triangle) and non-diabetic controls (circle) using seven diagnostic parameters: HbA1C, fasting plasma glucose (FPG), free fatty acids (FAA), C-peptide levels, HOMA-IR, fasting plasma insulin (FPI), and glycated peptide 26. PC1 and PC2 denote first and second principle components, respectively.

**Figure 3.** Principle component plot separating 48 T2DM patients in three clusters based on all available patient information, i.e., diagnostic parameters and glycated peptide levels. Missing values were imputed using Weka. The ellipses are drawn at the 95% confidence interval using the *level* parameter in ggplot. PC1 and PC2 denote first and second principle components, respectively.

**Figure 4.** Categorization of twenty individuals with prediabetes (ADA criteria) by diagnostic parameters at baseline (A and C) and follow-up studies (B and D) using FPG versus HbA1c levels (A and B) or HOMA2 %B versus insulin sensitivity HOMA2 %S values (C and D). Thresholds of FPG and HbA1c levels (dashed lines) relied on the current diagnostic criteria of WHO and ADA for normal glucose tolerance (NGT), impaired fasting glucose (IFG), prediabetes (PD), and type 2 diabetes mellitus (T2DM). Updated homeostasis model assessment (HOMA2) separated individuals in normal, diabetic (100% values), insulinopenic or hyperinsulinemic state. The numbers indicate the prediabetes samples (Supplement, Table S14 and S15).

**References**

1. Collaborators GMaCoD: Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015;385:117-171

2. (IDF) IDF: *Diabetes Atlas 7th edn*. Brussels, Belgium, 2015

3. Grundy SM: Pre-diabetes, metabolic syndrome, and cardiovascular risk. J Am Coll Cardiol 2012;59:635-643

4. Bullard KM, Saydah SH, Imperatore G, Cowie CC, Gregg EW, Geiss LS, Cheng YJ, Rolka DB, Williams DE, Caspersen CJ: Secular changes in U.S. Prediabetes prevalence defined by hemoglobin A1c and fasting plasma glucose: National Health and Nutrition Examination Surveys, 1999-2010. Diabetes Care 2013;36:2286-2293

5. Mainous AG, Tanner RJ, Baker R, Zayas CE, Harle CA: Prevalence of prediabetes in England from 2003 to 2011: population-based, cross-sectional study. BMJ Open 2014;4

6. Organization WH: Global report in diabetes. 2016;

7. Association AD: Classification and Diagnosis of Diabetes. Sec. 2. . Diabetes Care 2017;40:S11-S24

8. Tabák AG, Herder C, Rathmann W, Brunner EJ, Kivimäki M: Prediabetes: A high-risk state for developing diabetes. Lancet 2012;379:2279-2290

9. Dunkley AJ, Bodicoat DH, Greaves CJ, Russell C, Yates T, Davies MJ, Khunti K: Diabetes prevention in the real world: effectiveness of pragmatic lifestyle interventions for the prevention of type 2 diabetes and of the impact of adherence to guideline recommendations: a systematic review and meta-analysis. Diabetes Care 2014;37:922-933

10. World Health Organization IDF: Definition and diagnosis of diabetes mellitus and

intermediate hyperglycaemia: report of a WHO/IDF consultation. Geneva: World Health

Organization; 2006. 2006;

11. World Health Organization (WHO), Abbreviated report of a WHO consultation.

Use of glycated hemoglobin (HbA1c) in the diagnosis if diabetes mellitus. 2011.

12. Buell C, Kermah D, Davidson MB: Utility of A1C for Diabetes Screening in the 1999–2004 NHANES Population. Diabetes Care 2007;30:2233-2235

13. Kramer CK, Araneta MRG, Barrett-Connor E: A1C and Diabetes Diagnosis: The Rancho Bernardo Study. Diabetes Care 2010;33:101-103

14. Carson AP, Reynolds K, Fonseca VA, Muntner P: Comparison of A1C and fasting glucose criteria to diagnose diabetes among U.S. adults. Diabetes Care 2010;33:95-97

15. Lorenzo C, Wagenknecht LE, Hanley AJ, Rewers MJ, Karter AJ, Haffner SM: A1C between 5.7 and 6.4% as a marker for identifying pre-diabetes, insulin sensitivity and secretion, and cardiovascular risk factors: the Insulin Resistance Atherosclerosis Study (IRAS). Diabetes Care 2010;33:2104-2109

16. Zhou X, Pang Z, Gao W, Wang S, Zhang L, Ning F, Qiao Q: Performance of an A1C and fasting capillary blood glucose test for screening newly diagnosed diabetes and pre-diabetes defined by an oral glucose tolerance test in Qingdao, China. Diabetes Care 2010;33:545-550

17. Gosmanov AR, Wan J: Low positive predictive value of hemoglobin A1c for diagnosis of prediabetes in clinical practice. The American journal of the medical sciences 2014;348:191-194

18. Guo F, Moellering DR, Garvey WT: Use of HbA1c for diagnoses of diabetes and prediabetes: comparison with diagnoses based on fasting and 2-hr glucose values and effects of gender, race, and age. Metabolic syndrome and related disorders 2014;12:258-268

19. van 't Riet E, Alssema M, Rijkelijkhuizen JM, Kostense PJ, Nijpels G, Dekker JM: Relationship between A1C and glucose levels in the general Dutch population: the new Hoorn study. Diabetes Care 2010;33:61-66

20. Genuth S, Kahn R: A Step Backward—or Is it Forward? Diabetes Care 2008;31:1093

21. Balion CM, Raina PS, Gerstein HC, Santaguida PL, Morrison KM, Booker L, Hunt DL: Reproducibility of impaired glucose tolerance (IGT) and impaired fasting glucose (IFG) classification: a systematic review. Clinical chemistry and laboratory medicine 2007;45:1180-1185

22. Libman IM, Barinas-Mitchell E, Bartucci A, Robertson R, Arslanian S: Reproducibility of the oral glucose tolerance test in overweight children. J Clin Endocrinol Metab 2008;93:4231-4237

23. Armbruster DA: Fructosamine: structure, analysis, and clinical usefulness. Clinical Chemistry 1987;33:2153-2163

24. Hill RP, Hindle EJ, Howey JE, Lemon M, Lloyd DR: Recommendations for adopting standard conditions and analytical procedures in the measurement of serum fructosamine concentration. Ann Clin Biochem 1990;27 ( Pt 5):413-424

25. Furusyo N, Hayashi J: Glycated albumin and diabetes mellitus. Biochim Biophys Acta 2013;1830:5509-5514

26. Anguizola J, Matsuda R, Barnaby OS, Hoy KS, Wa C, DeBolt E, Koke M, Hage DS: Review: Glycation of human serum albumin. Clin Chim Acta 2013;425:64-76

27. Arasteh A, Farahi S, Habibi-Rezaei M, Moosavi-Movahedi AA: Glycated albumin: an overview of the In Vitro models of an In Vivo potential disease marker. Journal of Diabetes and Metabolic Disorders 2014;13:49-49

28. Goldstein DE, Little RR, Lorenz RA, Malone JI, Nathan D, Peterson CM, Sacks DB: Tests of glycemia in diabetes. Diabetes Care 2004;27:1761-1773

29. Welsh KJ, Kirkman MS, Sacks DB: Role of Glycated Proteins in the Diagnosis and Management of Diabetes: Research Gaps and Future Directions. Diabetes Care 2016;39:1299

30. Ai M, Otokozawa S, Schaefer EJ, Asztalos BF, Nakajima K, Shrader P, Kathiresan S, Meigs JB, Williams G, Nathan DM: Glycated albumin and direct low density lipoprotein cholesterol levels in type 2 diabetes mellitus. Clin Chim Acta 2009;406:71-74

31. Juraschek SP, Steffes MW, Selvin E: Associations of alternative markers of glycemia with hemoglobin A(1c) and fasting glucose. Clin Chem 2012;58:1648-1655

32. Nathan DM, McGee P, Steffes MW, Lachin JM: Relationship of glycated albumin to blood glucose and HbA1c values and to retinopathy, nephropathy, and cardiovascular outcomes in the DCCT/EDIC study. Diabetes 2014;63:282-290

33. Selvin E, Rawlings AM, Grams M, Klein R, Sharrett AR, Steffes M, Coresh J: Fructosamine and glycated albumin for risk stratification and prediction of incident diabetes and microvascular complications: a prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) study. The lancet Diabetes & endocrinology 2014;2:279-288

34. Beck R, Steffes M, Xing D, Ruedy K, Mauras N, Wilson DM, Kollman C: The interrelationships of glycemic control measures: HbA1c, glycated albumin, fructosamine, 1,5-anhydroglucitrol, and continuous glucose monitoring. Pediatric diabetes 2011;12:690-695

35. Danese E, Montagnana M, Nouvenne A, Lippi G: Advantages and pitfalls of fructosamine and glycated albumin in the diagnosis and treatment of diabetes. J Diabetes Sci Technol 2015;9:169-176

36. Parrinello CM, Selvin E: Beyond HbA1c and glucose: the role of nontraditional glycemic markers in diabetes diagnosis, prognosis, and management. Current diabetes reports 2014;14:548

37. Rondeau P, Bourdon E: The glycation of albumin: Structural and functional impacts. Biochimie 2011;93:645-658

38. Frolov A, Hoffmann R: Identification and relative quantification of specific glycation sites in human serum albumin. Anal Bioanal Chem 2010;397:2349-2356

39. Zhang Q, Monroe ME, Schepmoes AA, Clauss TRW, Gritsenko MA, Meng D, Petyuk VA, Smith RD, Metz TO: 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

40. 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

41. Spiller S, Frolov A, Hoffmann R: Quantification of specific glycation sites in human serum albumin as prospective type 2 diabetes mellitus biomarkers. Protein and peptide letters 2017;24:(Epub ahead of print)

42. Kloting N, Fasshauer M, Dietrich A, Kovacs P, Schon MR, Kern M, Stumvoll M, Bluher M: Insulin-sensitive obesity. American journal of physiology Endocrinology and metabolism 2010;299:E506-515

43. Kannt A, Pfenninger A, Teichert L, Tonjes A, Dietrich A, Schon MR, Kloting N, Bluher M: Association of nicotinamide-N-methyltransferase mRNA expression in human adipose tissue and the plasma concentration of its product, 1-methylnicotinamide, with insulin resistance. Diabetologia 2015;58:799-808

44. American Diabetes A: Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 2013;36:67-74

45. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC: Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985;28:412-419

46. Wallace TM, Levy JC, Matthews DR: Use and Abuse of HOMA Modeling. Diabetes Care 2004;27:1487-1495

47. Anderson L, Hunter CL: Quantitative mass spectrometric multiple reaction monitoring assays for major plasma proteins. Mol Cell Proteomics 2006;5:573-588

48. Frolov A, Hoffmann R: Analysis of Amadori Peptides Enriched by Boronic Acid Affinity Chromatography. Annals of the New York Academy of Sciences 2008;1126:253-256

49. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V: Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research 2011;12:2825-2830

50. Chen L, Xuan J, Wang C, Shih Ie M, Wang Y, Zhang Z, Hoffman E, Clarke R: Knowledge-guided multi-scale independent component analysis for biomarker identification. BMC Bioinformatics 2008;9:416

51. Granitto PM, Furlanello C, Biasioli F, Gasperi F: Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. Chemometrics and Intelligent Laboratory Systems 2006;83:83-90

52. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH: The WEKA data mining software: an update. ACM SIGKDD explorations newsletter 2009;11:10-18

53. von Luxburg U: Clustering Stability: An Overview. Foundations and Trends® in Machine Learning 2009;2:235-274

54. Ketchen DJ, Shook CL: The application of cluster analysis in strategic management research: an analysis and critique Strategic Management Journal 1996;17:441-458

55. Florkowski CM: Sensitivity, Specificity, Receiver-Operating Characteristic (ROC) Curves and Likelihood Ratios: Communicating the Performance of Diagnostic Tests. The Clinical Biochemist Reviews 2008;29:S83-S87

56. Hajian-Tilaki K: Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. Caspian Journal of Internal Medicine 2013;4:627-635

57. 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 HbA(1c) and fasting plasma glucose. Clin Proteom 2017;14:10

58. Meigs JB: Multiple Biomarker Prediction of Type 2 Diabetes. Diabetes Care 2009;32:1346-1348