Molecular Immunology 166 (2024) 16–28

Contents lists available at ScienceDirect

Molecular Immunology

journal homepage: www.elsevier.com/locate/molimm

Towards detection of structurally-diverse glycated epitopes in native proteins: Single-chain antibody directed to non-A1c epitope in human haemoglobin

Miss Lucia Gaetani [a], Isabel Campos Pinto [b], Meng Li [c], Peter O'Connor [c], Susan Giorgi-Coll [d], Matthew Tyreman [d], Katherine L. Rumary [d], James A. Schouten [d], Paul Davis [d], Ann M. Dixon [c,*]

[a] Medical Research Council Doctoral Training Programme, Warwick Medical School, UK
[b] iBET, Bayer Satellite Lab, Av. República, Quinta do Marquês, Edifício iBET/ITQB, Oeiras 2780-157, Portugal
[c] Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK
[d] Global Access Diagnostics, Thurleigh, Bedfordshire MK44 2YA, UK

ARTICLE INFO

Keywords:
Protein glycation products
HbA1c
Fragment antibody
Single chain variable fragment
Epitope mapping

ABSTRACT

Over 500 million people worldwide are affected by *diabetes mellitus*, a chronic disease that leads to high blood glucose levels and causes severe side effects. The predominant biological marker for diagnosis of diabetes is glycated haemoglobin (GHb). In human blood the predominant reducing sugar, glucose, irreversibly conjugates onto accessible amine groups within Hb. Most methods for diagnosis and monitoring of diabetes selectively detect N-terminal glycation at Val-1 on the β-globin chain, but not glycation at other sites. Detection of other glycated epitopes of GHb has the potential to provide new information on the extent, duration and timing of elevated glucose, facilitating personalised diagnosis and intelligent diabetic control. In this work, a new anti-GHb Fab antibody (Fab-1) specific for haemoglobin A1c (HbA1c) with nanomolar affinity was discovered via epitope-directed immunisation and phage display. A single chain variable fragment (scFv) antibody derived from Fab-1 retained affinity and specificity for HbA1c, and affinity was enhanced tenfold upon addition of an enhanced green fluorescent protein tag. Both the scFv and Fab-1 recognised an epitope within HbA1c that was distinct from β-Val-1, and our data suggest that this epitope may include glycation at Lys-66 in the β-globin chain. To our knowledge, this is the first report of an scFv/Fab anti-glycated epitope antibody that recognises a non-A1c epitope in GHb, and confirms that fructosamine attached to different, discrete glycation sites within the same protein can be resolved from one another by immunoassay.

1. Introduction

Diabetes mellitus is a chronic disease that impacts cellular uptake of free glucose in the bloodstream, resulting in excessive blood glucose levels. This can lead to severe complications of diabetes including coronary disease, heart attack, stroke, heart failure, kidney failure, visual impairment or even blindness. Recent estimates from the International Diabetes Federation (IDF) suggest that 537 million people worldwide, or 1 in every 11 adults, are currently living with diabetes and 45% of these people are undiagnosed (Sun et al., 2022; Ogurtsova et al., 2022).

After diagnosis of diabetes, blood glucose concentration tests are usually carried out by the patient with simple test strips carrying enzymatic redox systems (glucose oxidase or glucose dehydrogenase) linked to electrode sensors that connect to an electrochemical reader device. There is a growing trend in the use of automated, in-line, wearable pumps for direct intravenous dosing of insulin which is based on the same electrochemistry linked to a pump-compatible test reader. Even with these long-established glucose tests, there is a continuing need for monitoring the overall average blood glucose concentration during the previous weeks and months, because of the possibility of wide fluctuations in glucose concentration between tests and between insulin dosing. Given the severity of the health problems that can be caused by diabetes, early diagnosis and efficient management of blood glucose levels is critical. Both early diagnosis and efficient glycaemia control can benefit from optimised detection of protein glycation patterns over wider timescales, leading to personalised medication.

* Corresponding author.
E-mail address: ann.dixon@warwick.ac.uk (A.M. Dixon).

https://doi.org/10.1016/j.molimm.2023.12.008
Received 2 October 2023; Received in revised form 15 December 2023; Accepted 18 December 2023
Available online 4 January 2024

The glycated haemoglobin known as haemoglobin A1c is an established biomarker of diabetes that provides a record of glycaemic control over a period of 2–3 months, as a consequence of haemoglobin's circulating "lifespan" of approximately 90–120 days. Glycation of Hb proceeds via a nonenzymatic condensation reaction between glucose and amino (NH_2) groups present on either the N-terminal Val residue of the Hb β-chain (β Val-1) or on amino-bearing sidechains such as Lys (Fig. 1A), leading to the formation of a Schiff base (A). An Amadori rearrangement then irreversibly converts the Schiff base (via an open-chain Amadori product, B) to a fructosamine (1-amino-1-deoxy

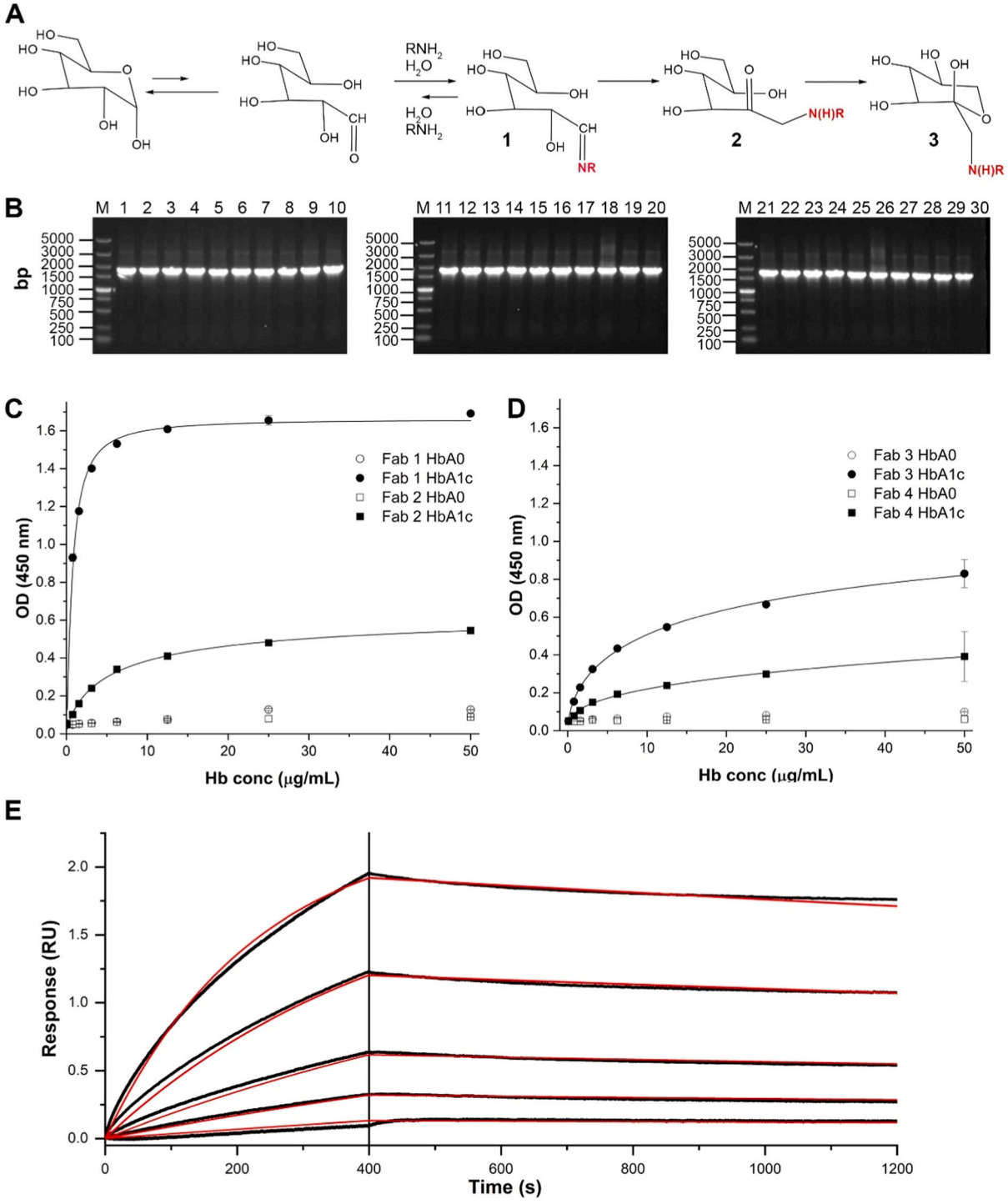


Fig. 1. Development and screening of a Fab library derived from epitope-directed immunization and phage display. (A) Hb glycation via Amadori rearrangement can lead to modification of N-terminal residues as well as residues carrying amide groups on their sidechains (e.g. Lys). **(B)** Agarose gel electrophoresis showing results of colony PCR of 30 randomly selected cloned from the Fab library. ELISA response curves for **(C)** Fabs 1–2 and **(D)** Fabs 3–4 against HbA1c and HbA0 using a sandwich ELISA, in which Hb was titrated over the concentration range 0 – 50 μg/mL (0 – 735 nM) and the Fab concentration was fixed at 1 μg/mL (20 nM). ELISA data are reported as means of three or more independent measurements, and error bars are shown for +/–1 x standard deviation. Fits of the data to the Hill equation are shown, and the resulting estimates of EC_{50} values are given in Table 1. **(E)** Biolayer interferometry sensorgrams showing binding kinetics of immobilized Fab 1 at a fixed concentration of 10 μg/mL (195 nM) titrated with HbA1c analyte (6, 12, 25, 50 and 100 nM). Raw data are shown in black, and red lines are the results of global fitting of these data to a 1:1 binding model. This global fit yielded a K_D of 3.4 ± 0.02 nM.

fructose) (C) as a pendant group on the modified protein. This covalent attachment of fructosamine is accumulative, so it follows that a measurement of its levels in the blood can give a historical record of glycaemic status during the preceding weeks and even months. By selectively measuring fructosamine attached to specific proteins such as haemoglobin and human serum albumin (HSA), it is possible to read the glycaemic status over longer time spans.

Given the nature of the Hb reaction with free glucose in the bloodstream, it is clear that several different glycation products can be formed, and a single Hb can be glycated at multiple sites. "HbA1c" is a glycated form of Hb in which glucose is bound specifically to β Val-1, and this species is thought to make up ~50% of the glucose-derivatised haemoglobin (Mortensen, 1998; Abraham, 1985). This leaves 50% of the GHb species that are glycated at alternative sites, such as the N-terminus of the α chain and the ε amino groups of lysine residues in both the α and β chains (e.g. β Lys-66, α Lys-61) (Bunn et al., 1979; Shapiro et al., 1980). It is also important to note that the presence of glycation at β Val-1 does not exclude the possible glycation at other sites. The sum of all glycated Hb species is referred to as "Total GHb". Several techniques are used to measure HbA1c in blood, including ion-exchange high-performance liquid chromatography (HPLC), immunoassays, enzymatic assays, capillary electrophoresis, and boronate affinity chromatography. However, most of these methods exclusively recognise the β Val-1 glycation of HbA1c, and do not report other glycated haemoglobin species or provide a measure of total GHb.

While the measurement of HbA1c has been extensively standardized and validated in its ability to diagnose the onset of diabetes (Saudek et al., 2006; Holman et al., 2008), recent studies suggest that the relationship between HbA1c and blood glucose level may vary among different racial groups (Herman, 2016) and people of different age. Advanced glycation end-products (AGEs) distinct from HbA1c have also been shown to have a strong correlation with diabetic patients who go on to develop cardiovascular and microvascular complications. Hence, measurements of fructosamine linked to a wider range of proteins have the potential to be a relevant source of information to diabetic patients. This is information which is currently not being extracted because readily available assays only resolve fructosamine linked to the N-terminal valine of HbA1c or, less frequently, on free amino sites within HSA. There are enzymatic assays that measure the total fructosamine in blood, but without any resolution of the proteins to which it is attached. Thus, we propose the possibility of gaining a more detailed historic record of glycaemic control spanning the three months prior to taking the sample, by developing an innovative, multiplex immunoassay capable of detecting alternative biomarkers providing the basis for truly personalised medication. The multiplex test lines would be based upon antibodies that recognise epitopes consisting of fructosamine linked to unique amino acid sequences found only at glycation sites within proteins selected as the best glycaemic status markers.

As a starting point from which to gain essential insights into the nature and discovery of complex/alternative epitopes of protein glycation (i.e. in HbA1c beyond β Val-1), we combined an epitope-directed strategy with phage display technology to discover four unique Fabs that recognise HbA1c. These Fabs were expressed and purified as soluble antibody fragments, and the antigen binding affinity and kinetics were evaluated. A monovalent scFv fusion protein was designed from the Fab sequence that showed the most favourable binding, and this scFv was expressed in a bacterial host and purified. We demonstrate that both the Fab and the scFv specifically bind HbA1c with high affinity via an epitope that is glycated but is *not* glycated β Val-1, confirming that fructosamine attached to different, discrete glycation sites within the same protein can be resolved from one another by immunoassay. Additionally, we show that the scFv binds the antigen with higher affinity than the Fab, and addition of a fusion domain (eGFP) further improves this binding.

2. Materials and methods

2.1. Animals

Adult female sheep (~70 kg) were used to generate antibodies. All animal experiments were conducted in compliance with the animal experimentation guidelines of the Home Office. Local ethical review and Home Office approval was secured by Origen Antibodies Ltd, and granted under the Animals (Scientific Procedures Act, 1986) with a Home Office project licence number PPL 70/8627.

2.2. Immunization protocol

An immune response to GHb was produced using an epitope-directed strategy. Briefly, initial immunisation of sheep with human HbA0 was used to generate a polyclonal antibody response, and HbA0-specific polyclonal antibodies were purified on an HbA0 affinity column. A second sheep was immunised with human HbA1c pre-complexed with excess purified HbA0-specific polyclonal antibodies to generate antibodies to the HbA1c-antibody complex. Animals were immunised by subcutaneous injection of antigen (1 mg/mL) in up to 8 sites (0.5 mL / site). The antigen, formulated in sterile saline, was emulsified in complete Freund's adjuvant (Sigma, UK) for the initial immunisation and incomplete Freund's adjuvant (Sigma, UK) for subsequent boosts, using recommended procedures as described by Greenfield (2022) (Greenfield, 2022). Animals were boosted with antigen (200 μg/mL) emulsified in incomplete Freund's adjuvant at 4-week intervals. Once animals had been shown to respond to antigen with antibody production (after boost 3), further boosts were carried out at 4-week intervals. Blood samples were collected by venepuncture into sterile collection bags before immunisation (pre-immunisation control) and 1 week after each boost. The volume collected was up to 10% blood volume (~ 250 mL). Blood was allowed to clot (placing in a heated water batch at 37 °C for 1 h, centrifuged (2000 rpm for 1 h) and serum collected. Serum samples was frozen at –20 °C before titration by direct Enzyme-Linked Immunosorbent Assay (ELISA).

2.3. Biopanning of Fab phage antibody library

Total RNA was extracted from sheep peripheral blood lymphocytes (PBLs) using Qiagen Kits (Qiagen, UK), according to the manufacturer's protocol, and sent to Creative Biolabs (Shirley, NY USA) for construction of a Fab library. Briefly, the total RNA was reverse transcribed into cDNA using universal primers for the heavy and light chains. The purity and concentration were determined by measuring the absorbance (A) at 260 nm and 280 nm (A260/A280) and quality was assessed on a 1% agarose gel. Heavy and light chains (λ and κ) were then amplified by polymerase chain reaction (PCR) using gene-specific primers for known variants of sheep Fab fragments, linked together, and cloned into a phagemid vector. The resulting phagemids were transformed into *E. coli* strain TG-1 to generate the Fab fragment libraries for both the λ and κ light chains. The quality of the Fab fragment library was validated by selecting 30 clones at random for colony PCR and sequencing and visualised on 1% agarose gels using GeneRuler 1 kb DNA ladder (ThermoFischer, UK). The phagemid library in *E. coli* was then infected and rescued with helper phage. Phage amplification was followed by harvesting and analysis of the phage titre. Biopanning of the library was done by incubating Fab phages with HbA1c pre-complexed with anti-HbA0 antibody. To increase surface area of the HbA1c, a biotinylated anti-HbA0 monoclonal antibody was used to capture the complex and streptavidin magnetic beads (Dynabeads™ M-280 Streptavidin,ThermoFisher, UK) were later used to capture the whole complex. Unbound phage were washed away using PBS/0.5% Tween 20, and bound phage were eluted by using 0.1 M glycine. Eluted phages were neutralized with 1 M Tris buffer and used to re-infect TG-1 *E. coli* in exponential growth phase. This biopanning cycle was carried out three times to generate an enriched library towards the HbA1c-anti-HbA0 antibody complex.

To assess enrichment of the library, direct ELISA was performed using the eluted pooled phages from each round of panning. Once enrichment was confirmed, serial dilutions of the library were plated, and 190 clones were picked and grown individually in 96-well culture plates for small scale expression. Periplasmic supernatants from each clone were used to assess binding by ELISA. Twenty positive clones were identified by monoclonal phage ELISA, and Sanger sequencing of all 20 clones identified four unique monoclonal Fabs (see Data Availability Statement for access to sequences). His_6 and Myc-tagged variants of the Fabs were expressed, purified, and analysed using SDS-PAGE and ELISA.

2.4. Anti-HbA1c scFv design

The variable heavy (V_H) and light (V_L) chains sequences of Fab 1 were identified through BLAST alignment with the immunoglobulins (IgG) gamma (IGHG3_HUMAN, IGHG4_HUMAN, IGHG2_HUMAN) and lambda (IGLC6_HUMAN, IGLC3_HUMAN, IGLC2_HUMAN, IGLC1_HUMAN) for heavy and light chains, respectively. A synthetic gene composed of the V_H and V_L chains linked by a 20 amino acid linker ($(G_4S)_5$), an N-terminal enhanced green fluorescent protein (eGFP) tag to aid stability and solubility, and a C-terminal 10 x His (His_{10}) tag for purification by immobilized metal affinity chromatography (IMAC) was synthesized at GenScript (Oxford, UK). The resulting gene was inserted into a pET22b+ plasmid lacking an N-terminal pelB sequence to direct expression of the fusion protein to the cytoplasm.

2.5. Bacterial expression and purification of the Fab 1-derived scFv

The pET22b+ plasmid containing the Fab 1-derived scFv gene was transformed into Shuffle T7 Express (New England Biolabs, Ipswich, MA) chemically competent *E. Coli* cells. To avoid leaky expression, Shuffle T7 cells were additionally transformed with the pLysS plasmid. Protein expression was induced via addition of 0.1 mM isopropyl-*β*-d-thiogalactoside (IPTG) to cells grown in Luria-Bertani (LB) medium containing ampicillin (100 μg/mL) and chloramphenicol (35 μg/mL) to an OD_{600} of 0.6. Following induction, cells were cultured at 18 °C with shaking (180 RPM) overnight. Bacterial cells were harvested via centrifugation at 3068 x g, 4 °C for 15 min, and resuspended in lysis buffer (50 mM Tris-HCl, 150 mM NaCl, 20 mM imidazole, pH 8) containing 1 mL 1X Bug Buster (EMD Millipore Corp., Bellerica, MA USA), protease inhibitor cocktail (Roche Diagnostics, GmbH, Mannheim, Germany) and 1 mg/mL of deoxyribonuclease from bovine pancreas (Sigma, USA). Cells were lysed using a cell disruptor at 20 psi. The cell lysate was centrifuged at 3068 x g and 4 °C for 20 min and the supernatant was collected for purification by IMAC, while the insoluble fraction was resuspended in 8 M urea and stored at 4 °C before being analysed by SDS-PAGE. The soluble fraction was manually loaded onto a HisTrap HP 5 mL column (GE Healthcare) with a flow rate of 1 mL/min. The target protein was purified by gradient elution in buffer A (50 mM Tris-HCl, 150 mM NaCl, pH 8) containing increasing concentrations of buffer B (50 mM Tris-HCl, 150 mM NaCl, 1 M imidazole, pH 8) (78–100% buffer B) at a flow rate of 0.5 mL/min and 0.3 MPa.

The eGFP tag was cleaved from the scFv using a two-fold molar excess of GFP-tagged TEV protease. Cleavage was carried out under oxidative conditions (without DTT) in buffer (50 mM EDTA, 10 mM Tris-HCl, 150 mM NaCl, pH 7.5). The cleavage reaction was incubated overnight at 4 °C with rotation and loaded onto GFP-trap agarose (Chromotek, Germany). Purification of the scFv from the eGFP tag and residual GFP-TEV protease was carried out according to the manufacturer's instructions. The scFv was exchanged into 50 mM sodium phosphate buffer, pH 7.5, using a PD10 desalting column, and concentrated using a centrifugal concentrator with a 3 kDa molecular weight cut-off (Pierce, UK).

2.6. Sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Protein composition and purity of fractions from bacterial expression, purification and cleavage were assessed using SDS-PAGE (visualized using Coomasie Blue R250, silver nitrate, and eGFP fluorescence ($\lambda_{ex} = 488$)). Samples for SDS-PAGE were mixed with an equal volume of 2X SDS-loading dye (4X SDS-loading dye 0.25 M Tris-HCl (pH 7.2), 8% Sodium dodecyl sulfate (SDS), 30% glycerol, 0.02% bromophenol, 0.3 M Dithiothreitol (DTT)). Novex 10–20% Tris-Glycine gels (Invitrogen, UK) were used according to the manufacturer's instructions. Band migration was related to the See blue Plus 2 (Invitrogen, UK) prestained protein marker.

2.7. Quantification of free sulfhydryl groups

Ellman's assay was used for the quantification of free sulfhydryl groups. Ellman's Reagent (4 mg, Thermo Fisher Scientific, UK) was dissolved in 1 mL of reaction buffer (0.1 M sodium phosphate, pH 8.0, containing 1 mM EDTA) and the assay was performed following the manufacturer's instructions. Sample absorbance was measured at 412 nm (A_{412}) with a Jasco V-600 Spectrophotometer.

2.8. Mass spectrometry

The identities of eGFP-scFv-His_{10} (55102.33 Da) and scFv-His_{10} (27243.95 Da) were confirmed via tryptic digest, followed by nano LC-ESI-MS/MS analysis of resulting peptides at the Proteomics Research Technology Platform (School of Life Sciences, University of Warwick). Details of the digest and liquid chromatography setup are given in Supplementary information. The LC was coupled to a hybrid timsTOF Pro (Bruker Daltonics, Germany) via a CaptiveSpray nano-electrospray ion source. The timsTOF Pro was operated in Data-Dependent Parallel Accumulation-Serial Fragmentation (PASEF) mode over a mass range of 100 - 1700 *m/z*. The raw data were processed as described in Supplementary Information.

Glycation sites within HbA1c were identified via tryptic digest of native HbA1c, followed by either (a) nano LC Fourier transform ion cyclotron resonance MS/MS (nLC-FTICR-MS/MS) or (b) direct infusion FTICR-MS. The nLC-FTICR-MS/MS experiments were performed using a Bruker nanoElute (Bruker Daltonics, Germany) coupled to a 15 Tesla Bruker SolariX FTICR-MS (Bruker Daltonics, Germany). Direct infusion analyses were run on a 15 Tesla FT-ICR-MS in both positive and negative ionization mode using a custom nanospray ion source. Further details regarding sample treatment and instrumental setup for the FTICR measurements are given in Supplementary Information.

2.9. Circular dichroism (CD)

CD spectra were collected on a Jasco J-1500 CD Spectrometer (Jasco UK, Great Dunmow, UK) equipped with Peltier temperature control. A sample of the cleaved scFv-His_{10} (3.5 μM) in 50 mM sodium phosphate buffer, pH 7.5 was placed in a quartz cuvette of 1 mm path length (Starna Scientific Ltd., UK). Spectra were recorded at 25 °C between 190 nm and 300 nm, with a bandwidth of 2 nm and a data pitch of 0.2 nm. Buffer spectra were also recorded for baseline subtraction. The machine units of mdeg were converted to units of molar circular dichroism ($\Delta\varepsilon$).

2.10. Enzyme-linked immunosorbent assay (ELISA)

The binding activity and specificity of Fabs 1–4 for HbA1c (Abcam, UK) was tested via sandwich ELISA, using unglycated haemoglobin (HbA0, Sigma, UK) as negative control. A biotinylated anti-Hb mouse monoclonal IgG (1 μg/mL in PBS) (BBI solutions) was immobilized onto a streptavidin 96-well plate (Thermo Fisher Scientific, UK) at 4 °C

overnight. The plate was washed three times with Tris-buffered saline with 0.1% Tween® 20 (TBST) and blocked with 1% bovine serum albumin (BSA; Sigma, USA) in phosphate-buffered saline (PBS) for one hour at room temperature with gentle agitation. The antigens (HbA1c and HbA0) were prepared in PBS making two-fold dilutions (0 – 50 μg/mL, or 0 – 735 nM) and captured by anti-Hb IgG onto the plate by incubation for one hour at room temperature with gentle agitation. Following a wash step with TBST, Myc-tagged Fabs 1–4 (1 μg/mL (20 nM) in 1% BSA/PBS) were incubated for one hour at room temperature with gentle agitation. Fabs that remained bound after a wash step with TBST were detected using anti-Myc IgG conjugated with horseradish peroxidase (HRP, Abcam, UK), according to the manufacturer's instructions. The plate was developed with 100 μL of 1-step Turbo TMB-ELISA substrate solution (Thermo Fisher Scientific, UK) per well until a colorimetric reaction took place. The reaction was stopped using 100 μL TMB Stop solution (Thermo Fisher Scientific, UK) per well and absorbance at 450 nm was measured with a plate reader.

Affinity and specificity of the Fab 1-derived scFv for its cognate antigen HbA1c was assessed using direct ELISA. Native HbA1c (Abcam, UK), as well as HbA1c treated with sodium periodate or denatured by high temperatures (see Supplementary Information), was immobilised in two-fold dilutions over the concentration range 0.48–250 ng/mL (7 pM – 4 nM) directly onto a high-binding 96-well plate and incubating at 4 °C overnight. As a negative control, non-glycated haemoglobin (HbA0, Abcam, UK) was also immobilised at the same concentration range of HbA1c. Recombinant GFP-His_6 was expressed and purified, and used as positive control, being directly immobilised on the plate over at a concentration range of 4–250 ng/mL (0.2 – 18 nM). Blocking and washing of the plate was carried out as above. The primary antibody (eGFP-scFV) was added at a fixed concentration of 1 μg/mL (18.15 nM) in 1% BSA/PBS and incubated for one hour at room temperature with gentle agitation. The plate was washed again with TBST and incubated with HRP-conjugated anti-GFP rabbit polyclonal IgG (300 ng/mL, Generon Ltd, UK) for one hour at room temperature with gentle agitation. The plate was developed as described above.

To assess whether the Fab 1-derived scFv was able to bind the fructosylated hexapeptide (fructosyl-VHLTPE-biotin, Sussex Research Laboratories Inc., Ontario, Canada) as a model antigen of HbA1c, a direct ELISA was performed by immobilizing the biotinylated fructosyl-hexapeptide in two-fold dilutions 150–2.3 - 150 ng/mL (2 - 175 nM) onto a pre-blocked streptavidin-coated 96-well plate (Thermo Fisher Scientific, UK), incubating at 4 °C overnight. Washing of the plate and detection of the antigen was performed with eGFP-scFV antibody as described above, and plate development was carried out with the same method.

All ELISA data in this work are reported as means of three or more independent measurements +/– the standard standard deviation.

2.11. Binding kinetics of scFv – antigen interaction

The Fab 1 – HbA1c binding kinetics were investigated using biolayer interferometry (BLI, Octet RED96e, Sartorius, Germany). Fab 1 was immobilised on Ni-NTA sensors (Sartorius, Germany) via its His_{10} tag, and hydrated in 1X kinetics buffer (PBS + 0.1% BSA, 0.02% Tween20 and Kathon) at a fixed concentration of 10 μg/mL (195 nM). Sensorgrams were collected upon submersion of the biosensors into a black 96-well plate containing the analyte (6–100 nM HbA1c, or 50 nM HbA0 as negative control (which was subtracted from each binding curve) in 1X kinetics buffer. The baseline was set using 1X kinetics buffer in 2 steps of 60 and 120 s, respectively. The kinetic analysis was performed at 25 °C with the orbital shake speed of 1000 RPM. Data were fit to a global 1:1 binding model (details in Supplementary Information). The same method was applied to study the binding kinetics between HbA1c and both eGFP-scFv-His_{10} and cleaved scFv-His_{10}, both of which were immobilised at a fixed concentration of 10 μg/mL (182 and 367 nM, respectively). The binding kinetics of the HbA1c model antigen, the biotinylated fructosyl-hexapeptide (fructosyl-VHLTPE-biotin, model antigen of the β-chain glycated at the N-terminus), with a commercial anti-HbA1c monoclonal antibody (H01369M, Meridian Life Science, UK) (64 nM) was measured by immobilisation of the peptide onto a streptavidin biosensor (Sartorius, Germany) at a fixed concentration of 64 nM in 1X kinetics buffer. Sensorgrams were collected upon submersion of the biosensors into a black 96-well plate containing the anti-HbA1c monoclonal antibody at concentrations between 1–16 nM.

2.12. Homology modelling and structure prediction

To predict the recombinant antibody structure, homology modelling studies were performed for the scFv region using SWISS-MODEL (Waterhouse et al., 2018; Bienert et al., 2017; Guex et al., 2009; Studer et al., 2020; Bertoni et al., 2017) as well as computational tools specific for antibody structure prediction such as DeepAb (Ruffolo et al., 2022) and ABodyBuilder (Leem et al., 2016). Models of the CDR loops and prediction of the confidence scores were determined by ABlooper (Abanades et al., 2022). The full length sequence of the eGFP – scFv – His_{10} fusion protein was submitted to AlphaFold2 within the ColabFold environment (Mirdita et al., 2022) and the resulting structure coloured according to the pLDDT confidence measure in the B-factor field. Highest confidence regions were coloured dark blue, lowest confidence regions were coloured orange, and intermediate confidence regions were yellow-light blue. All molecular graphics from AlphaFold2 were produced using UCSF Chimera X, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from NIH P41-GM103311 (Pettersen et al., 2004).

3. Results

3.1. Phage library of Fabs that recognise alternative epitopes of glycated haemoglobin

In order to develop new antibodies that recognize alternative epitopes of glycated haemoglobin, e.g. distinct from glycated β Val-1, an epitope-directed immunisation and panning strategy was adopted. Affinity depletion to select anti-HbA1c polyclonal antibodies (pAbs) was achieved by immunisation of sheep with human HBA1c pre-complexed with excess, purified HbA0-specific pAbs. The antibodies generated from this immunisation were expected to have enhanced selectivity to novel epitopes. Immunisation of sheep with human HbA1c pre-complexed with excess, purified HbA0-specific pAbs, as expected, resulted in antibodies against HbA1c as well as antibodies against new antigenic sites which could be composed of (i) a combination of sites within HbA1c and sites in bound HbA0-specific pAbs, (ii) alternative conformations of the primary epitope (e.g. glycated β Val-1), or (iii) alternative glycation sites within HbA1c that are distal to the primary epitope but remain accessible in the presence of the HbA0-specific pAbs.

The quality of the Fab fragment phage library (built by extraction of B-cell Ig genes of interest (see Materials and Methods)) was validated by colony PCR (Fig. 1B) and sequencing of 30 randomly selected clones. The results indicated an insertion rate of 30/30, uniqueness of 24/28, diversity of 1.7×10^9, and titres of 1.2×10^{14} CFU/mL confirming that it was a high quality, large, and diverse library. The library was panned in solution against HbA1c precomplexed with purified HbA0-specific polyclonal antibodies using streptavidin-biotin affinity technology to capture binders (see Materials and Methods). This approach ensured that any potential low frequency phage of interest was not lost before enrichment and maximised the likelihood of identifying antibodies that recognised alternative glycation sites in HbA1c. This process yielded 20 positive clones which, upon sequencing, revealed four unique monoclonal Fabs. His_6/Myc-tagged variants of all four Fabs were expressed, purified and analysed using SDS-PAGE (Fig. S1A).

3.2. Affinity and selectivity of soluble Fab fragments

Data from the sandwich enzyme-linked immunosorbent assays (ELISAs) used to evaluate the specificity of each of the four unique monoclonal Fabs resulting from epitope-directed immunization and panning are provided in Fig. 1C-D. The serial dilutions of HbA0 and HbA1c, respectively, yielded different calibration curves with Fabs 1–4, revealing their distinct specificity/selectivity and binding characteristics. The data were fitted to the Hill Equation to estimate the EC_{50} values for each of the Fabs, and the values obtained are given in Table 1. All four Fabs had at least some selectivity in this format for HbA1c vs. HbA0, and Fab 1 displayed the highest affinity for the HbA1c antigen with an apparent EC_{50} value of 0.73 μg/mL or 10 nM.

Competition ELISAs were additionally used to further evaluate the specificity of all four Fabs for HbA1c. In these assays, a 20-fold molar excess of free HbA0 was added to each Fab before adding the Fab to the ELISA plate. Any reduction of HbA1c-binding by the Fab was evidence of cross-reactivity towards HbA0, indicating selectivity for HbA1c, rather than specificity. The results of the competition assays are shown in Figs. S1B-E, where high concentrations of HbA0 in solution almost completely outcompeted the immobilised HbA1c for binding to Fabs 2 and 4. Conversely, even very high concentrations of HbA0 did not compete with HbA1c for binding to Fab 1, indicating a very specificity for HbA1c.

The binding properties of Fab 1 were examined in more detail using the Octet Biolayer interferometry (BLI) system, which detects antigen binding by means of light wave interference. Fab 1 was immobilized on Ni-NTA biosensors at a fixed concentration (10 μg/mL) and it was exposed to solutions with increasing concentrations of HbA1c (6–100 nM). The resulting sensorgrams are shown in Fig. 1E, and a global fit of these data to a 1:1 binding model (see Materials and Methods and Fig. 1E, red lines) yielded a K_D of 3.37 ± 0.016 nM, and k_{on} and k_{off} values shown in Table 2. This high affinity is in good agreement with other antibodies selected from immune repertoires via phage display (Li et al., 2000; Frenzel et al., 2016).

3.3. Design of a single-chain antibody fragment for bacterial expression

The variable heavy (V_H) and variable light (V_L) chain sequences of Fab 1 were defined by BLAST alignment (Fig. S2) of the antibody's sequence with known human immunoglobulin sequences (see Materials and Methods). Once identified, the V_H and V_L regions required a linker to connect them together. A survey of available scFv crystal structures was performed to design the length and sequence of the linker. Given the recurrence of 20 amino acid linkers amongst known scFv structures (Kaufmann et al., 2002; Leysath et al., 2009; Xiao et al., 2019; Miyanabe et al., 2018; Zdanov et al., 1994) (e.g. PDB IDs 5YD3, 6J90, and 1KTR) and successful use of the $(G_4S)_n$ amino acid repeat in scFv linkers in the literature, a $(G_4S)_4$ linker was used to connect the C-terminus of the V_H chain to the N-terminus of the V_L chain (see Fig. 2A). An enhanced green fluorescent protein (eGFP) was selected as a fusion partner as a means of increasing the scFv's solubility in the bacterial cytoplasm and offering a route to sensitive detection of protein expression. GFP / eGFP has been successfully used as a fusion protein in many scFv studies reported thus far (Cassimeris et al., 2013; Cao et al., 2008; Hink et al., 2000; Peipp et al., 2004; Rangnoi et al., 2021). Hink and co-workers explained that GFP's suitability stems from the fact that it does not influence the ability of the scFv to bind its target (Hink et al., 2000). Therefore, the presence of GFP, while improving stability and solubility, leaves the antigen-binding sites fully exposed and the scFv free to recognize its antigen. A His_{10} tag was added at the C-terminus of the fusion protein to carry out purification by immobilised metal affinity chromatography (IMAC) (see Fig. 3A). Finally, tobacco etch virus (TEV) protease and thrombin cleavage sites were added to facilitate cleavage of the eGFP and His_{10} tags, respectively.

The structures of the designed scFv and the entire fusion protein were predicted using a range of computational tools to explore the length of the linker and the relative orientations of the complementarity determining regions (CDRs), V_H, V_L, and eGFP regions with respect to one another. Structural models of the Fab 1-derived scFv were produced using both the Swiss-MODEL homology modelling server as well as the ABodyBuilder machine learning programme (see Materials and Methods), and the resulting structures are overlayed in Fig. 2B. Both methods predict very similar structures, with a high abundance of β-sheet linked by coil / loop regions. The CDRs of the scFv were modelled using ABlooper within the ABodyBuilder programme (Abanades et al., 2022) and these are shown in the sequence in supplementary Fig. S3.

The structure of the entire fusion protein, including the eGFP domain, was predicted using the AlphaFold2 machine learning programme (Tunyasuvunakool et al., 2021; Varadi et al., 2021) and the resulting top-ranking structure is shown in Fig. 2C-D. The predicted structure of the anti-HbA1c scFv is highly similar to the crystal structure of a Fab with 72.4% sequence identity (PDB ID 6QNA, shown in grey) suggesting that the V_H and V_L domains can adopt a biologically relevant fold in the context of this fusion protein. The linker between these two domains is visible in Fig. 3D and is predicted to be unstructured and of sufficient length to facilitate rearrangement of the variable domains if needed to achieve optimal orientation. The expected β-barrel structure of eGFP was also returned by AlphaFold2.

3.4. Bacterial expression, purification, and characterization of Fab 1-derived scFv

Bacterial expression of the Fab 1-derived scFv was tested in the engineered *E. coli* strains Origami (DE3) B and Shuffle T7 Express, both of which are capable of correctly folding proteins containing disulfide-bonds in the cytoplasm. Different temperatures (18–30 °C), induction times (4–16 h) and IPTG concentrations (0.01 - 1 mM) were tested to define the optimal experimental conditions yielding the highest level of

Table 1
Affinity of Fabs 1–4 and Fab 1 scFv for HbA1c in sandwich ELISA and direct ELISA assays, respectively. ELISA data shown in Figs. 1C-D and 4B were fit to the Hill equation to estimate the EC_{50} for each Fab and Fab 1 scFv.

	Fab 1	Fab 2	Fab 3	Fab 4	Fab 1 scFv
EC_{50} (μg/mL)	0.73 ± 0.06	6.18 ± 1.32	25.6 ± 9.5	644 ± 313	7.5 × 10^{-3} ± 1.3 × 10^{-3}
EC_{50} (nM)	10	90	376	9.47 × 10^{3}	0.11 ± 0.02

Table 2
Kinetic analysis and binding affinity of directly bound anti-HbA1c antibodies to HbA1c from BLI analyses.

Ligand	K_D (nM)	k_{on} (M^{-1} s^{-1})	k_{off} (s^{-1})
Fab 1	3.4 ± 0.02	4.28 × 10^{4} ± 97.1	1.44 × 10^{4} ± 6.19 × 10^{-7}
GFP-scFv-His_{10}	0.217 ± 0.00392	7.84 × 10^{4} ± 72.5	1.70 × 10^{-5} ± 3.07 × 10^{-7}
scFv-His_{10}	2.0 ± 0.01	2.31 × 10^{4} ± 41.5	4.72 × 10^{-5} ± 2.88 × 10^{-7}

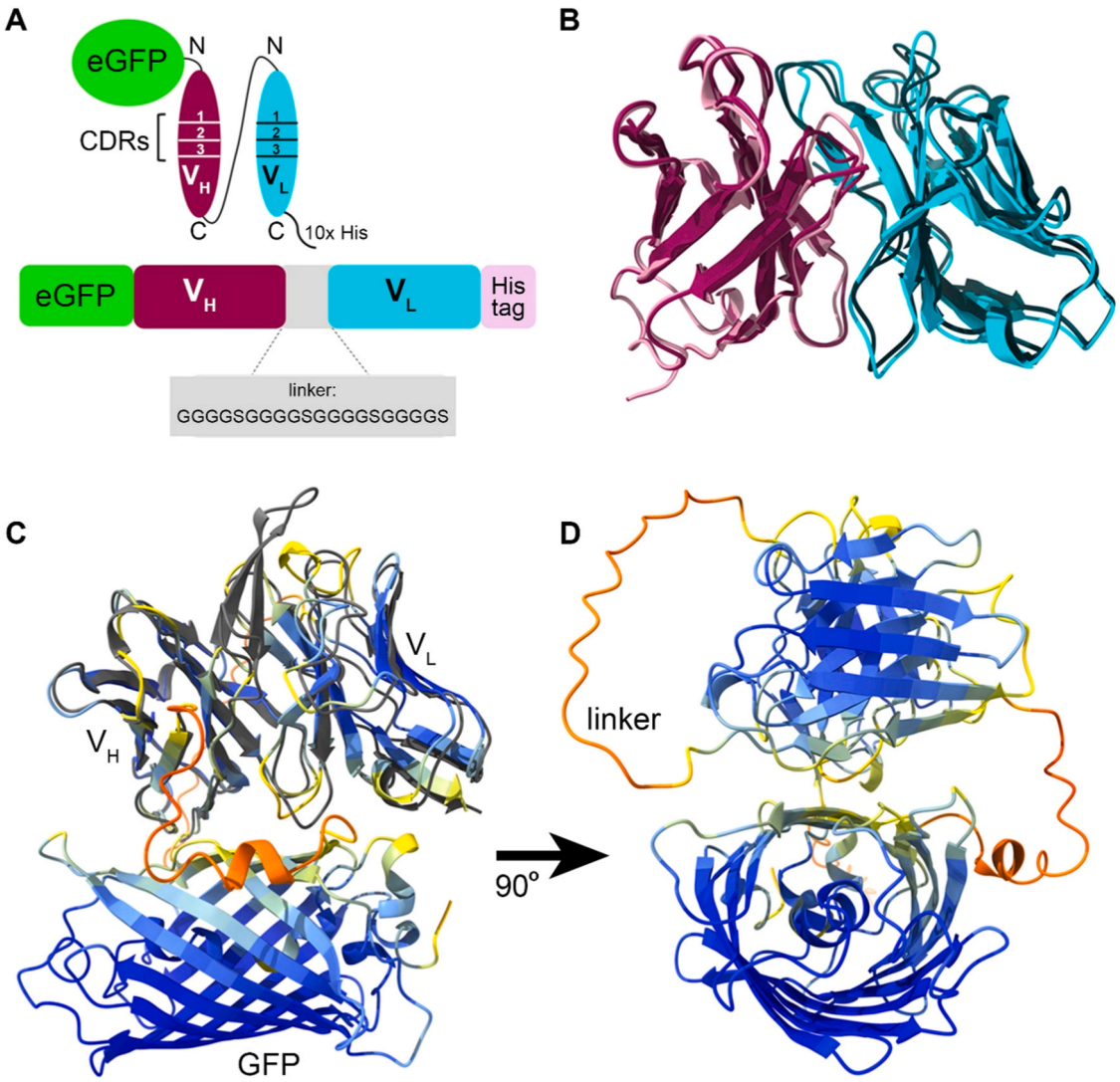


Fig. 2. Design of a single chain variable fragment (scFv) antibody fusion protein derived from Fab 1. (A) Schematic of the anti-HbA1c scFv design, containing an N-terminal enhanced green fluorescent protein (eGFP) tag to aid detection and solubility. **(B)** Overlay of structure prediction for the V_H-V_L scFv using ABodyBuilder (dark shades) and Swiss Model (Light shades). The coloring scheme used for each domain in panel A was maintained in panel B. **(C)** AlphaFold2 structure prediction for the entire fusion protein, showing the predicted arrangement of the three domains relative to one another. The V_H-V_L region of the scFv is overlayed with the crystal structure of bovine anti-RSV hybrid Fab B13HC-B4LC (PDB ID: 6QNA). **(D)** A 90° rotation of the AlphaFold2 model, showing the relative position of the (GGGGS)4 linker.

recombinant antibody in the soluble fraction. Preliminary results (not shown) led us to select Shuffle T7 Express as the best host for recombinant protein expression, as the level of target protein in the soluble fraction was greatest. Optimal expression was achieved upon induction with 0.1 mM IPTG followed by growth at 18 °C, 180 RPM, where the yield of the 55 kDa protein was seen to increase up to 16 h post induction (Fig. 3A-B, lanes 4 h-16 h). The presence of the eGFP fusion partner was confirmed by in-gel fluorescence using an UV-transilluminator (Fig. 3B). Cell fractionation was used to visualise the proportion of protein in the soluble fraction (Fig. 3A, lane SF). The soluble fraction was collected for purification via IMAC using a gradient of 0 - 1 M imidazole, and the scFv eluted at the highest imidazole concentrations (Fig. 3C). The identity of the fusion protein was confirmed using tryptic digestion followed by mass spectrometric analysis to validate the sequence of the protein, where 94% sequence coverage was achieved (Supplementary Figures 4A-B) and the sequence confirmed.

The eGFP fusion partner was cleaved from the scFv using GFP-tagged TEV protease under oxidative conditions, and the scFv was purified from both the GFP-TEV and cleaved eGFP using GFP-Trap ® resin (see Fig. 3D). Following buffer exchange and concentration, disulphide bonds formation was assessed using the Ellman's reagent (see Materials and Methods), which confirmed the absence of free sulfhydryl groups by the low absorbance at 412 nm ($A_{412} = 0.05$). The folding of the scFv was investigated using circular dichroism spectroscopy (CD), the resulting CD spectrum of the scFv is shown in Fig. 4A, and contains the characteristic peak for β-sheet secondary structure at 217 nm. A fit of the CD data using Dichroweb (Miles et al., 2022; Whitmore and Wallace, 2008) yielded a β-sheet secondary structure content of 37% sheet which is in reasonable agreement with the β-sheet content predicted from our modelling experiments (32%).

3.5. *New scFv binds HbA1c with high affinity and specificity*

Fab 1 was shown to have a high specificity for HbA1c, with affinity in the low nM range (Fig. 1C, Table 1). ELISA and biolayer interferometry (BLI) were once again used to assess the specificity and affinity of the Fab 1-derived scFv. In a direct ELISA, the recombinant eGFP-scFv fusion protein was exposed to adsorbed antigen, either HbA1c or HbA_0 (over a

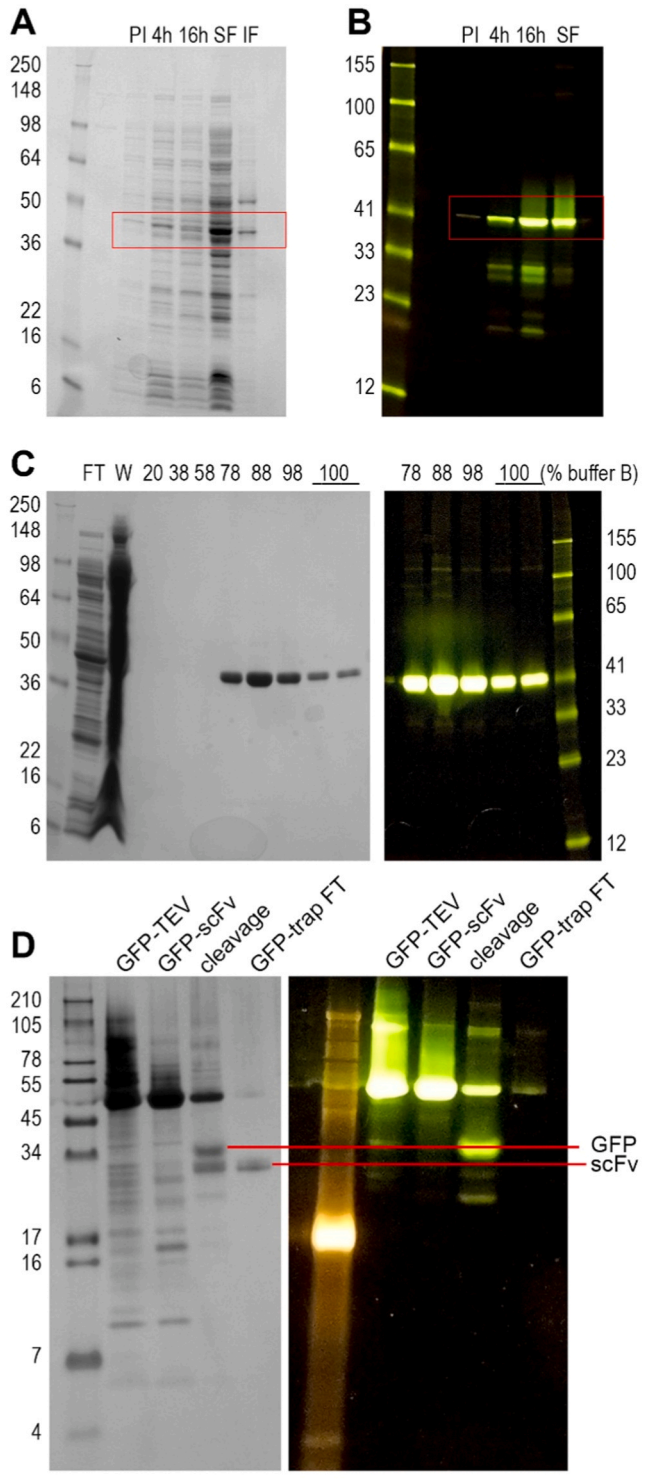


Fig. 3. Expression and purification of scFv in E. coli. (A) Coomassie-stained SDS-PAGE gel showing the expression of the scFv pre-induction sample (PI), 4 and 16 h post induction, and soluble (SF) and insoluble fractions (IF). The band corresponding to the GFP-scFv-His_{10} fusion protein is indicated by the red box. **(B)** eGFP florescence detection of the SDS-PAGE gel shown in panel (A). **(C)** Coomassie R250 and eGFP fluorescence detection of fractions collected from FPLC IMAC purification. SDS-PAGE lanes for the flow through (FT), wash (W) and elution fractions at increasing imidazole concentrations are shown. **(D)** Coomassie R250 and eGFP fluorescence detection following the cleavage of the eGFP tag from the scFv using GFP-tagged TEV protease. Cleaved GFP and GFP-TEV protease were separated from free scFv using GFP-trap resin.

concentration range of 7 pM – 4 nM), and binding was detected using an HRP-conjugated secondary antibody against GFP (Fig. 4B). The data for HbA1c were fitted to the Hill Equation to estimate the EC_{50} value (see Table 1) for the interaction between the scFv and HbA1c. Also shown is the positive control (Fig. 4B, inset) demonstrating that the HRP-conjugated anti-GFP antibody was specifically recognising its cognate antigen, GFP. These results confirm that the Fab 1-derived scFv retained high affinity for the antigen as well as maintaining high specificity for HbA1c, as no binding was observed for HbA0.

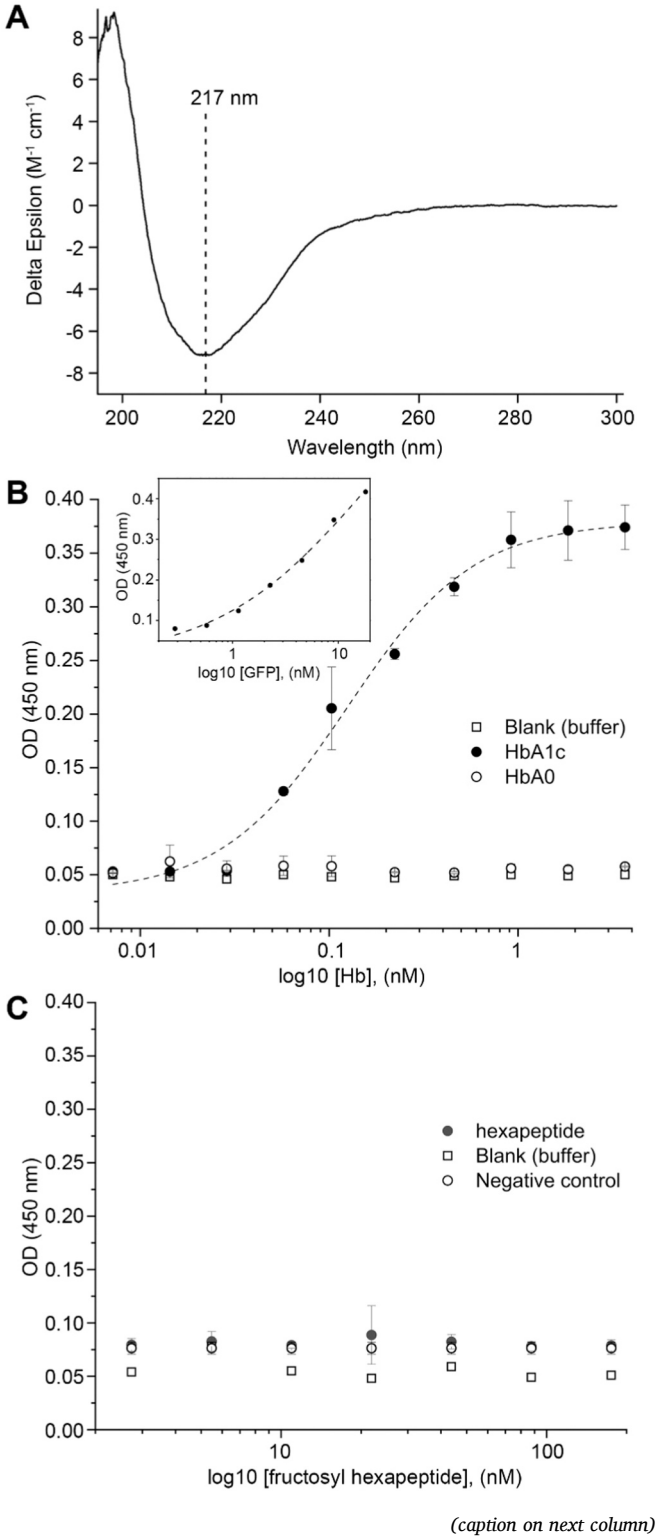
The binding affinities of the eGFP-scFv-His_{10} fusion protein, and the scFv with the eGFP-fusion tag removed, were quantified via biolayer interferometry (BLI) (see Table 2). The antibodies were immobilized at a fixed concentration (10 μg/mL) onto Ni-NTA sensor tips and exposed to solutions with increasing concentrations of HbA1c (6 - 100 nM). The resulting sensorgrams for the eGFP-scFv and the scFv in the absence of eGFP are shown in Fig. 5A-B, respectively, and the data were fitted to a global 1:1 binding model (see Materials and Methods and red lines in Fig. 5). The eGFP-scFv data yielded a K_D of 217 ± 3.9 pM, which is a 10-fold increase in affinity for HbA1c over that of the parent Fab. Interestingly, the K_D was reduced almost 10-fold to 2.0 ± 0.01 nM upon removal of the eGFP tag from the scFv, suggesting that the eGFP tag imparts a favourable influence on the ability HbA1c binding of the scFv paratope. In all measurements, the absolute specificity of the antibodies for HbA1c were confirmed using HbA_0 as negative control.

3.6. *New scFv binds HbA1c via an epitope that is distinct from the β Val-1 glycation site*

In order to probe the epitope within HbA1c recognised by the Fab 1-derived scFv, an HbA1c model antigen was tested. The model antigen was a peptide containing six amino acids at the extreme N-terminus of the β-chain of HbA1c, in which the valine residue at position 1 was conjugated with glucose. The resulting fructosylated hexapeptide is an industry standard for identifying HbA1c selective antibodies, as the majority of these antibodies recognise glycation of β Val-1. ELISA results for binding of the fructosylated hexapeptide to the eGFP-scFv are shown in Fig. 4C, where no detectable binding was observed. This contrasts with the affinity and specificity of a commercially available anti-HbA1c IgG (Meridian Life Science, UK) for the fructosylated hexapeptide (Fig. S5) which, when measured using BLI yielded a K_D of 9.1 ± 2.0 nM. These data are evidence that the scFv, derived from a Fab discovered as part of an epitope-directed/phage display approach, recognises an HbA1c epitope distinct from the N-terminal Val residue on the β-chain of Hb.

The nature of the epitope within HbA1c recognised by the Fab 1-derived scFv was first explored using ELISA. Periodate oxidation of HbA1c, resulting in cleavage of 1,2-diol bonds in pendant fructosamine groups, was carried out to chemically degrade all glycation sites in the protein (Fig. 5C) before immobilization onto the 96-well plate, and a direct ELISA was again performed. Comparison of the resulting data to that for native HbA1c (Fig. 5D) revealed that selectively eliminating the sugar moieties in HbA1c prevented recognition of the antigen by the scFv, suggesting that the scFv recognises a glycated epitope. HbA1c was also denatured by heat prior to immobilization onto the 96-well plate, thus unfolding the protein while keeping glycation sites intact (Fig. 5C). As shown in Fig. 5D, heat denaturation of HbA1c had an equally detrimental impact on antigen recognition, indicating that the binding epitope is both a conformational epitope and a glycated epitope.

High resolution mass spectrometry, specifically Fourier-transform ion cyclotron resonance MS/MS (FTICR-MS/MS) was used to identify the glycation sites in the HbA1c antigen used throughout this work, in the hopes of locating additional sites that were distinct from Val-1 on the β chain. Peptides resulting from digestion with trypsin were ionised using electrospray either via direct infusion or separated using nano-scale liquid chromatography (nano-LC) and fragmented in MS/MS experiments either using collisionally-activated dissociation (CAD) or



(caption on next column)

Fig. 4. Characterisation of structure and antigen recognition of the Fab 1 scFv. (A) Circular dichroism spectrum of purified, recombinant Fab 1 scFv following GFP cleavage. The sample was composed of 3.5 μM scFv in 50 mM $NaPO_4$ buffer (pH 7.5) and spectra were acquired at 25 °C. **(B)** eGFP-scFv against HbA1c and HbA0 using a direct ELISA assay, in which the antigen was titrated over the concentration range 0.48–250 ng/mL (7 pM – 4 nM) and the eGFP-scFv concentration was fixed at 1 μg/mL (18 nM). The fit line from fitting of the HbA1c data to the Hill equation is shown, and the resulting estimates of EC_{50} values are given in Table 1. The ELISA results confirm that the scFv retains the affinity and specificity of the parent Fab. **(Inset)** Results from a positive control, in which recognition of purified recombinant GFP-His_6 by the HRP-conjugated anti-GFP secondary antibody was measured over the concentration range 4–250 ng/mL (0.2 – 18 nM) in a direct ELISA. Fitting of these data to the Hill equation yielded an apparent EC_{50} value of 21 nM. **(C)** GFP-scFv against biotinylated fructosylated hexapeptide using a direct ELISA, in which the peptide was titrated over the concentration range 2.3 - 150 ng/mL (2 – 175 nM) and the GFP-scFv concentration was fixed at 1 μg/mL (18 nM). Comparison to the blank and negative control indicate no binding of the scFv to the hexapeptide. All ELISA data are reported as means of three or more independent measurements, and error bars are shown for +/–1 x standard deviation.

electron capture dissociation (ECD) prior to MS/MS analysis. The results from the CAD experiments indicated that the fructosamine-peptide bond is very fragile, leading to cleavage of the sugar during CAD, whilst ECD involves radical rearrangement chemistry and cleaves the backbone amide bond preferentially while ignoring glycans. Therefore, ECD is a more preferable technique to identify sites for labile modifications such as amino acid glycation (Håkansson et al., 2001; O'Connor et al., 2006; Leymarie et al., 2003). Analyses of the resulting masses yielded sequence coverage of 74% and 84% for the α-globin and β-globin subunits (see Supplementary Fig. S6), respectively. No glycation sites were detected in the α-globin peptides either because they were not present or because they were highly labile, and our data cannot distinguish these two possibilities at this stage. However, in addition to the glycation site at Val-1 (Fig. 6A, Supplementary Figures 7–9 and Supplementary Table 1), MS/MS results showed glycation of the β-globin chain at Lys-66. The mass spectrum for the peptide spanning Lys-66 – Lys-82 is shown in Fig. 6B, along with peak assignments and corresponding fragments. A full peak assignment table is given in Supplementary Table S2. Taken together, this evidence definitively shows that the HbA1c antigen used throughout this work carries at least two glycation sites in its β-globin chain, one at Val-1 and a second at Lys-66.

The location of Lys-66 in the structure of human glycated haemoglobin (Fig. 6B) lies within a 19-residue α-helix spanning Lys-59 – His-77 (shown in pink, Fig. 6B). The surface-exposed position of Lys-66 makes it unsurprising that this site undergoes glycation in vivo with the highest prevalence after Val-1 (Abraham, 1985; Bunn et al., 1979). Recognition of Lys-66 glycation by our scFv would explain its lack affinity for the HbA1c model antigen (the fructosyl hexapeptide, Fig. 4C), whilst maintaining high affinity and selectivity for full length HbA1c (Figs. 4B and 5). The loss of binding following periodate oxidation would also support the presence of a glycated Lys residue in the epitope. However, as shown in Fig. 5, binding is not solely dictated by the sugar but also requires native protein folding. Given that Lys-66 lies in a such a well-structured region of HbA1c (i.e. in the central portion of a long α-helix), in contrast to Val-1 which lies at the relatively unstructured N-terminus, residues near Lys-66 must also contribute to scFv recognition of the antigen.

4. Discussion and conclusions

Recent studies from the International Diabetes Federation (IDF) have estimated that, in 2021, almost one in two adults with diabetes were unaware and living with the disease without being diagnosed. The highest proportion of undiagnosed diabetes is in low income countries, with Africa (53.6%), Western Pacific (52.8%) and South-East Asia re-

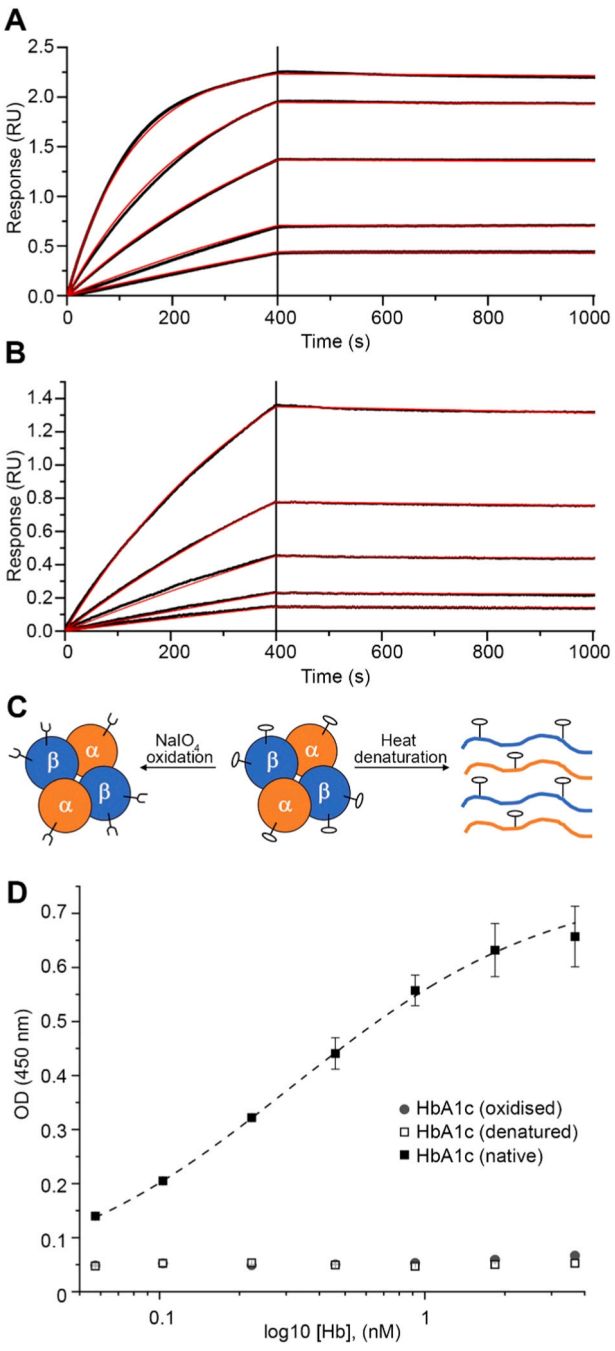


Fig. 5. Investigation of the scFv affinity and binding epitope within HbA1c. Biolayer interferometry sensorgrams showing binding kinetics of HbA1c with immobilized **(A)** eGFP-scFv and **(B)** scFv without the eGFP tag at a fixed concentration of 10 μg/mL (182 and 367 nM, respectively). In both cases, the immobilised antibody was titrated with HbA1c analyte (6–100 nM). The red lines are the results of global fitting of these data to a 1:1 binding model, yielding a K_D of 217 ± 3.9 pM for the eGFP-scFv and a K_D of 2.0 ± 0.01 nM for the scFv alone. **(C)** Schematic summarising modifications carried out on HbA1c, including sodium periodate oxidation of sugars and heat denaturation of the protein, prior to immobilisation and subsequent ELISA measurement. **(D)** Direct ELISA data following HbA1c treatment summarised in panel (C). Native, oxidised, or denatured HbA1c was titrated over the concentration range 3.9–250 ng/mL (57 pM – 4 nM) and the eGFP-scFv concentration was fixed at 1 μg/mL (18 nM). The fit line from fitting of the native HbA1c data to the Hill equation is shown, and the resulting EC_{50} value is given in Table 1. Data from oxidised and denatured HbA1c were not fit as no binding was observed.

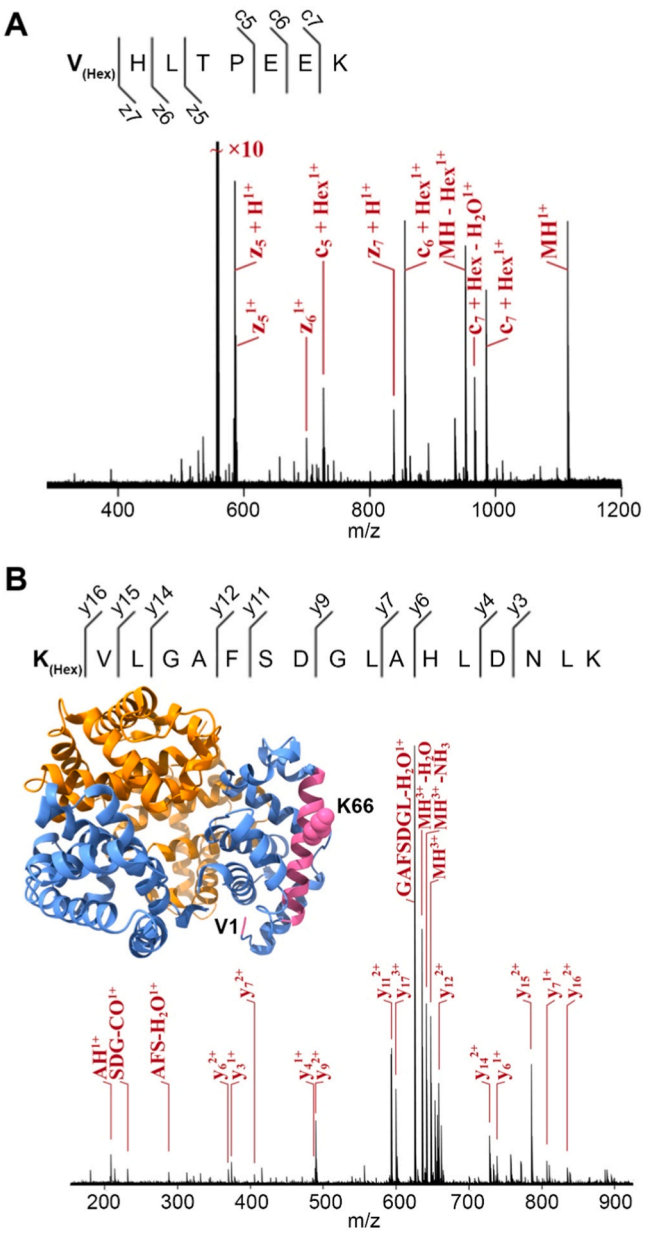


Fig. 6. Identification of two glycation sites in HbA1c using FTICR mass spectrometry. FTICR MS/MS spectra for glycated peptides derived from a tryptic digest of HbA1c. **(A)** Sequence and spectrum of a peptide corresponding to the extreme N-terminus of the β-globin chain. The spectrum was obtained via direct infusion FTICR MS/MS in positive ion mode and using electron capture dissociation (ECD) fragmentation. The peaks in the mass spectrum (see Table S1 for full assignments) confirm the presence of a hexose modification on Val-1, as expected for HbA1c. **(B)** Sequence and spectrum of a peptide corresponding to K66-K82 of the β-globin chain. The spectrum was obtained via nano-LC FTICR MS/MS in positive ion mode and using collisionally-activated dissociation (CAD). The peaks in the mass spectrum are fully assigned in Table S2, and confirm the presence of a second glycation site at Lys-66 in HbA1c. (inset) Crystal structure for human glycated haemoglobin (PDB ID 3B75), with the α-globin chains in orange and the β-globin chains in blue. The positions of Val-1, at the unstructured N-terminus, and Lys-66, in a long, surface-exposed α-helix, are shown in pink.

gions (51.3%) thought to have the highest rates (Ogurtsova et al., 2022). Therefore, accurate, reliable and affordable diagnostic tests are needed, especially in these low-medium income countries, to keep the global diabetes epidemic under control. The challenge posed by this high prevalence of undiagnosed diabetes worldwide demands even better

diagnostic tools which can deliver more information and guidance that is personalised. We propose that such improvements may become possible by unlocking the information hidden within the variety of glycated epitopes to be found in the blood. Other glycated forms of Hb which have been identified in a recent mass spectrometry study on the profile of biological biomarkers in diabetic petients (Zhang et al., 2011) point to the kind of extra information that can made accessible. As mentioned previously, haemoglobin can be glycated on different amino acids (e.g β-Lys-66, α-Lys-61) yet the current diagnostic methods for diabetes almost exclusively recognize the HbA1c epitope glycated at Val-1 on the Hb β chain (Sharma et al., 2020; Little and Roberts, 2009; Hörber et al., 2020). A new biological reagent or test that can recognise alternative glycation sites that result from diabetes is potentially of great interest.

In this work, we used an epitope-directed immunisation strategy in combination with phage display to find unique anti-HbA1c antibodies, specific for HbA1c but through epitopes within the protein that are distinct from the glycated β Val-1. While we do not yet know the full identity of the epitope that our new Fab and scFv recognise, we can confirm that neither antibody recognises glycated β Val-1 but both are specific for HbA1c, with high affinity. This is consistent with the prediction that HbA1c is glycated at more than one site, and this alternative glycation site likely forms at least part of the epitope that our new antibody recognises. Our ELISA data following periodate oxidation of HbA1c (Fig. 5D) indicate that the new binding epitope in HbA1c is glycated: selective scission of 1,2-diol bonds in pendant fructosamine moieties eliminated the ability of the scFv to recognise HbA1c. We also know that the binding epitope is, in part, conformational: denaturation of the protein fold also eliminated recognition. We therefore speculate, based on our mass spectrometry data, that the binding epitope may be composed of fructosylated K66 in the β-globin chain as well as some of the α-helical region immediately surrounding K66.

The excellent specificity and low nM affinity of the parent Fab 1 for the antigen was encouraging, but the retention of these characteristics after engineering to remove the constant domains and create an scFv (see Tables 1–2) was better than expected. Surprisingly, addition of an eGFP tag to the scFv yielded 10-fold higher affinity for the antigen without compromising specificity. The eGFP tag may increase affinity by directing favourable orientation of the antibody's CDRs towards the antigen, however more work is required to understand this impact. Regardless, this new antibody is, to our knowledge, novel in its ability to bind an alternative epitope within GHb and could be employed alongside commercial anti-HbA1c antibodies in immunoassays to detect other forms of GHb (glycated at sites other than β Val-1) and, potentially, provide more information than either a conventional HbA1c test or a total GHb measurement. A further benefit of the recombinant Fab 1–derived eGFP-scFv is that it is efficiently expressed in *E. coli*, well-folded, and highly soluble. As such it could readily be employed in the development of new diagnostic tools.

Thus far, neither immunoassays nor rapid diagnostics for HbA1c detection and quantification based on a fusion-protein scFv format have been reported in the literature, and our antibody may lay the foundations for such applications. For instance, the presence of an eGFP tag may be used to carry out a fluorophore-linked immunosorbent assay (FLISA) (Oelschlaeger et al., 2002), measuring the binding activity of the eGFP-scFv to the glycated haemoglobin and quantifying the antigen in a sample. Additionally, the subnanomolar affinity and absolute specificity of the eGFP-scFv for HbA1c make our antibody a perfect candidate for the development of a rapid diagnostic test for HbA1c blood levels in a lateral flow assay format. Indeed, the ELISA data shown in this work (e.g. Fig. 4B) indicate that the resulting scFv can readily detect HbA1c concentrations of < 0.25 μg/mL (or 0.025 mg/dL). This value is four orders of magnitude lower than the HbA1c concentration in a healthy adult, and five orders of magnitude lower than the HbA1c concentration in a diabetic adult (see values in Table 3) (Loh et al., 2015; Ang et al., 2016; Sherwani et al., 2016), suggesting that the scFv is an excellent candidate for HbA1c detection in blood samples. This may offer a cost-effective and reliable tool for the diabetes detection and monitoring especially in the low-income countries, allowing people to self-monitor their health conditions.

Table 3
Comparison of HbA1c concentrations in normal, pre-diabetic and diabetic adults. The value of 13.8 g/dL for total haemoglobin concentration ($[Hb]_{Total}$, average of male and female values) was obtained from Loh and co-workers (Loh et al., 2015).

Status	HbA1c (%)[a]	HbA1c (mmol/mol)[b]	HbA1c (mg/dL)[c]
Normal	4–6	20–42	552–828
Pre-diabetic	7	53	966
Diabetic	≥8	≥64	≥1104

[a] HbA1c (%) = (moles HbA1c / total moles Hb) × 100
[b] HbA1c (mmol/mol) = (HbA1c (%) × 10.93) – 23.5
[c] HbA1c (mg/dL) = $[Hb]_{Total}$ × (HbA1c (%) / 100)

Of wider significance, the increased understanding gained in these studies of the nature of glycated epitopes within highly structured protein domains will help to enable the generation of a broad repertoire of antibodies capable of detecting many different glycated epitopes in a variety of different proteins. Each of these may contribute uniquely to the personalised diagnostic/prognostic information that may eventually be harvested from the detection of fructosamine in a rich diversity of molecular settings.

Accession numbers

The mRNA sequences generated and described in the current study are available in the National Centre for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) database [GenBank accession numbers: **OR536599**, **OR540584**, **OR540660**, **OR540664**, **OR540673**, **OR540679**, **OR540684**, **OR540699**].

CRediT authorship contribution statement

Dixon Ann M.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Gaetani Lucia:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Davis Paul:** Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Rumary Katherine L.:** Data curation, Investigation, Methodology. **Schouten James A.:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Tyreman Matthew:** Data curation, Formal analysis, Investigation, Methodology. **O'Connor Peter:** Data curation, Investigation, Methodology, Resources, Supervision, Writing – review & editing. **Giorgi-Coll Susan:** Investigation, Methodology. **Pinto Isabel Campos:** Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing. **Li Meng:** Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing.

Data Availability

Data availability has been described in the Data Availability Statement in the Manuscript document (following the References section).

Acknowledgements

This work was supported by an industry-linked iCASE PhD studentship awarded to L. Gaetani linked to the Doctoral Training Programme in Interdisciplinary Biomedical Research, University of Warwick co-funded by the Medical Research Council and Global Access Diagnostics (grant number MR/R015910/1). The authors would like to acknowledge Dr Jerry Macaleer (Lumira Dx) for helpful discussions and guidance. The authors would also like to acknowledge Dr. Cleidiane Zampronio and Dr. Andrew Bottrill at the Proteomics Facility RTP, School of Life Sciences, University of Warwick for protein sequence analysis and data interpretation, Prof. Richard Napier and Dr. Panward Prasongpholchai for help and advice throughout the project, and Prof. T. D.H. Bugg for use of a plate reader.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.molimm.2023.12.008.

References

Abanades, B., Georges, G., Bujotzek, A., Deane, C.M., 2022. ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation. Bioinformatics 38, 1877–1880. https://doi.org/10.1093/bioinformatics/btac016.

Abraham, E.C., 1985. Glycosylated Hemoglobins: Methods of Analysis and Clinical Applications. M. Dekker,.

Ang, S.H., Thevarajah, T.M., Woi, P.M., Alias, Y.B., Khor, S.M., 2016. A lateral flow immunosensor for direct, sensitive, and highly selective detection of hemoglobin A1c in whole blood. J. Chromatogr. B Anal. Technol. Biomed. Life Sci. 1015-1016, 157–165. https://doi.org/10.1016/j.jchromb.2016.01.059.

Bertoni, M., Kiefer, F., Biasini, M., Bordoli, L., Schwede, T., 2017. Modeling protein quaternary structure of homo- and hetero-oligomers beyond binary interactions by homology. Sci. Rep. UK 7, 10480. https://doi.org/10.1038/s41598-017-09654-8.

Bienert, S., et al., 2017. The SWISS-MODEL repository-new features and functionality. D313-d319 Nucleic Acids Res. 45. https://doi.org/10.1093/nar/gkw1132. D313-d319.

Bunn, H.F., et al., 1979. Structural heterogeneity of human hemoglobin A due to nonenzymatic glycosylation. J. Biol. Chem. 254, 3892–3898.

Cao, M., et al., 2008. Construction and characterization of an enhanced GFP-tagged anti-BAFF scFv antibody. Appl. Microbiol. Biotechnol. 79, 423–431. https://doi.org/10.1007/s00253-008-1447-9.

Cassimeris, L., Guglielmi, L., Denis, V., Larroque, C., Martineau, P., 2013. Specific in vivo labeling of tyrosinated α-tubulin and measurement of microtubule dynamics using a GFP tagged, cytoplasmically expressed recombinant antibody. PLoS. One 8, e59812. https://doi.org/10.1371/journal.pone.0059812.

Frenzel, A., Schirrmann, T., Hust, M., 2016. Phage display-derived human antibodies in clinical development and therapy. mAbs 8, 1177–1194. https://doi.org/10.1080/19420862.2016.1212149.

Greenfield, E.A., 2022. Immunizing animals. Pdb.top100180 Cold. Spring Harb. Protoc. 2022. https://doi.org/10.1101/pdb.top100180. Pdb.top100180.

Guex, N., Peitsch, M.C., Schwede, T., 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. Electrophoresis 30 (Suppl 1), S162–S173. https://doi.org/10.1002/elps.200900140.

Håkansson, K., Cooper, H.J., Emmett, M.R., Costello, C.E., Marshall, A.G., Nilsson, C.L., 2001. Electron capture dissociation and infrared multiphoton dissociation MS/MS of an N-glycosylated tryptic peptic to yield complementary sequence information. Anal. Chem. 73, 4530–4536. https://doi.org/10.1021/ac0103470.

Herman, W.H., 2016. Are there clinical implications of racial differences in HbA1c? Yes, to not consider can do great harm! Diabetes Care 39, 1458–1461. https://doi.org/10.2337/dc15-2686.

Hink, M.A., et al., 2000. Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein. J. Biol. Chem. 275, 17556–17560. https://doi.org/10.1074/jbc.M001348200.

Holman, R.R., Paul, S.K., Bethel, M.A., Matthews, D.R., Neil, H.A., 2008. 10-year follow-up of intensive glucose control in type 2 diabetes. N. Engl. J. Med. 359, 1577–1589. https://doi.org/10.1056/NEJMoa0806470.

Hörber, S., Achenbach, P., Schleicher, E., Peter, A., 2020. Harmonization of immunoassays for biomarkers in diabetes mellitus. Biotechnol. Adv. 39, 107359 https://doi.org/10.1016/j.biotechadv.2019.02.015.

Kaufmann, M., et al., 2002. Crystal structure of the anti-His tag antibody 3D5 single-chain fragment complexed to its antigen. J. Mol. Biol. 318, 135–147. https://doi.org/10.1016/s0022-2836(02)00038-4.

Leem, J., Dunbar, J., Georges, G., Shi, J., Deane, C.M., 2016. ABodyBuilder: automated antibody structure prediction with data-driven accuracy estimation. MAbs 8, 1259–1268. https://doi.org/10.1080/19420862.2016.1205773.

Leymarie, N., Costello, C.E., O'Connor, P.B., 2003. Electron capture dissociation initiates a free radical reaction cascade. J. Am. Chem. Soc. 125, 8949–8958. https://doi.org/10.1021/ja028831n.

Leysath, C.E., et al., 2009. Crystal structure of the engineered neutralizing antibody M18 complexed to domain 4 of the anthrax protective antigen. J. Mol. Biol. 387, 680–693. https://doi.org/10.1016/j.jmb.2009.02.003.

Li, Y., Kilpatrick, J., Whitelam, G.C., 2000. Sheep monoclonal antibody fragments generated using a phage display system. J. Immunol. Methods 236, 133–146. https://doi.org/10.1016/s0022-1759(99)00227-6.

Little, R.R., Roberts, W.L., 2009. A review of variant hemoglobins interfering with hemoglobin A1c measurement. J. Diabetes Sci. Technol. 3, 446–451. https://doi.org/10.1177/193229680900300307.

Loh, T.P., Sethi, S.K., Wong, M.S., Tai, E.S., Kao, S.L., 2015. Relationship between measured average glucose by continuous glucose monitor and HbA1c measured by three different routine laboratory methods. Clin. Biochem. 48, 514–518. https://doi.org/10.1016/j.clinbiochem.2015.02.012.

Miles, A.J., Ramalli, S.G., Wallace, B.A., 2022. DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. Protein Sci. 31, 37–46. https://doi.org/10.1002/pro.4153.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682. https://doi.org/10.1038/s41592-022-01488-1.

Miyanabe, K., et al., 2018. Intramolecular H-bonds govern the recognition of a flexible peptide by an antibody. J. Biochem. 164, 65–76. https://doi.org/10.1093/jb/mvy032.

Mortensen, R.F. a H.B., 1998. Glycated hemoglobins. J. Chromatogr. 429, 279–292.

O'Connor, P.B., Lin, C., Cournoyer, J.J., Pittman, J.L., Belyayev, M., Budnik, B.A., 2006. Long-lived electron capture dissociation product ions experience radical migration via hydrogen abstraction. J. Am. Soc. Mass. Spectrom. 17, 576–585. https://doi.org/10.1016/j.jasms.2005.12.015.

Oelschlaeger, P., Srikant-Iyer, S., Lange, S., Schmitt, J., Schmid, R.D., 2002. Fluorophor-linked immunosorbent assay: a time- and cost-saving method for the characterization of antibody fragments using a fusion protein of a single-chain antibody fragment and enhanced green fluorescent protein. Anal. Biochem. 309, 27–34. https://doi.org/10.1016/s0003-2697(02)00290-7.

Ogurtsova, K., et al., 2022. IDF diabetes Atlas: global estimates of undiagnosed diabetes in adults for 2021. Diabetes Res. Clin. Pr. 183 https://doi.org/10.1016/j.diabres.2021.109118.

Peipp, M., et al., 2004. Efficient eukaryotic expression of fluorescent scFv fusion proteins directed against CD antigens for FACS applications. J. Immunol. Methods 285, 265–280. https://doi.org/10.1016/j.jim.2003.12.001.

Pettersen, E.F., et al., 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084.

Rangnoi, K., Rüker, F., Wozniak-Knopp, G., Cvak, B., O'Kennedy, R., Yamabhai, M., 2021. Binding characteristic of various antibody formats against aflatoxins. ACS Omega 6, 25258–25268. https://doi.org/10.1021/acsomega.1c03044.

Ruffolo, J.A., Sulam, J., Gray, J.J., 2022. Antibody structure prediction using interpretable deep learning. Patterns (N. Y) 3, 100406. https://doi.org/10.1016/j.patter.2021.100406.

Saudek, C.D., Derr, R.L., Kalyani, R.R., 2006. Assessing glycemia in diabetes using self-monitoring blood glucose and hemoglobin A1c. Jama 295, 1688–1697. https://doi.org/10.1001/jama.295.14.1688.

Shapiro, R., McManus, M.J., Zalut, C., Bunn, H.F., 1980. Sites of nonenzymatic glycosylation of human hemoglobin A. J. Biol. Chem. 255, 3120–3127.

Sharma, P., Panchal, A., Yadav, N., Narang, J., 2020. Analytical techniques for the detection of glycated haemoglobin underlining the sensors. Int. J. Biol. Macromol. 155, 685–696. https://doi.org/10.1016/j.ijbiomac.2020.03.205.

Sherwani, S.I., Khan, H.A., Ekhzaimy, A., Masood, A., Sakharkar, M.K., 2016. Significance of HbA1c test in diagnosis and prognosis of diabetic patients. Biomark. Insights 11, 95–104. https://doi.org/10.4137/bmi.S38440.

Studer, G., Rempfer, C., Waterhouse, A.M., Gumienny, R., Haas, J., Schwede, T., 2020. QMEANDisCo-distance constraints applied on model quality estimation. Bioinformatics 36, 1765–1771. https://doi.org/10.1093/bioinformatics/btz828.

Sun, H., et al., 2022. IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045 (https://doi.org:). Diabetes Res. Clin. Pr. 183, 109119. https://doi.org/10.1016/j.diabres.2021.109119.

Tunyasuvunakool, K., et al., 2021. Highly accurate protein structure prediction for the human proteome. Nature 596, 590–596. https://doi.org/10.1038/s41586-021-03828-1.

Varadi, M., et al., 2021. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444. https://doi.org/10.1093/nar/gkab1061.

Waterhouse, A., et al., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. W296-w303 Nucleic Acids Res. 46. https://doi.org/10.1093/nar/gky427. W296-w303.

Whitmore, L., Wallace, B.A., 2008. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89, 392–400. https://doi.org/10.1002/bip.20853.

Xiao, H., et al., 2019. Light chain modulates heavy chain conformation to change protection profile of monoclonal antibodies against influenza A viruses. Cell. Discov. 5, 21 https://doi.org/10.1038/s41421-019-0086-x.

Zdanov, A., et al., 1994. Structure of a single-chain antibody variable domain (Fv) fragment complexed with a carbohydrate antigen at 1.7-A resolution. Proc. Natl. Acad. Sci. U. S. A. 91, 6423–6427. https://doi.org/10.1073/pnas.91.14.6423.

Zhang, Q., et al., 2011. Comprehensive identification of glycated peptides and their glycation motifs in plasma and erythrocytes of control and diabetic subjects. J. Proteome Res. 10, 3076–3088. https://doi.org/10.1021/pr200040j.

GLOSSARY

AGEs: Advanced glycation end-products
BLI: Biolayer interferometry
BSA: Bovine serum albumin
CAD: Collisionally-activated dissociation
CD: Circular dichroism
CDRs: Complementarity determining regions
DTT: Dithiothreitol
ECD: Electron capture dissociation
eGFP: Enhanced green fluorescent protein
ELISA: Enzyme-linked immunosorbent assays
Fab-1: Anti-GHb Fab antibody
Fab: Fragment antigen-binding
FTICR-MS/MS: Fourier-transform ion cyclotron resonance MS/MS
GHb: Glycated haemoglobin
HSA: Human serum albumin
Hb: Haemoglobin
HbA0: non-glycated Haemoglobin
HbA1c: Haemoglobin A1c
HPLC: High-performance liquid chromatography
HRP: Horseradish peroxidase
IDF: International Diabetes Federation
IgG: Immunoglobulin G
IMAC: Immobilised metal affinity chromatography
LC-ESI-MS/MS: Liquid Chromatography-Electrospray Ionization-Mass Spectrometry
Ni-NTA: Nickel-nitrilotriacetic acid
PASEF: Parallel Accumulation-Serial Fragmentation
PBLs: Peripheral blood lymphocytes
scFv: Single chain variable fragment
TBST: Tris-buffered saline with 0.1% Tween® 20
TEV: Tobacco etch virus
TMB: 3,3′,5,5′-Tetramethylbenzidine