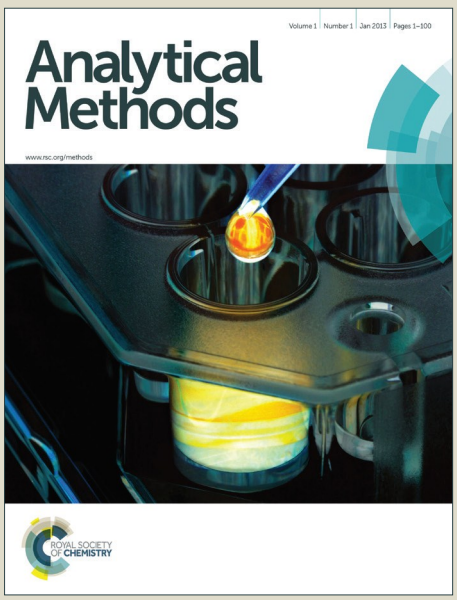
Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Journal Name

ARTICLE

Biophysical characterization of a protein for structure comparison: methods for identifying insulin structural changes

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

M. Sklepari,[a] A. Rodger,*[a] A. Reason,[b] S. Jamshidi,[a] I. Prokes,[a] C. A. Blindauer[a]

Although protein structure has been studied for many decades it remains the case that we cannot state with confidence whether two samples have the same molecular structure, particularly in solution. The increasing number of biosimilar biopharmaceutical drugs that are being tested means this is not an academic exercise. In this work we consider how four well-established techniques: dynamic light scattering (DLS), circular dichroism (CD), nuclear magnetic resonance spectroscopy (NMR), and molecular modelling can be combined to provide information about the supposedly well-understood protein insulin. A goal of this work was to establish a systematic means of detecting differences between insulin samples as a function of pH, temperature, and the presence or absence of zinc, all of which are known to change the oligomerisation state and to affect molecular structure. We used the recently developed Secondary Structure Neural Network (SSNN) circular dichroism algorithm to facilitate analysis of the CD spectra.

Introduction

Since the launch of recombinant human insulin (Eli Lilly & Co.'s Humulin) as the first biopharmaceutical drug in 1982,[1] the challenge of knowing whether a new batch of a protein drug is the same – not only at primary structure but also secondary and tertiary structure levels – as the one on which toxicology and pharmacokinetics had been studied has been recognised. This issue has become more challenging with the advent of biosimilars where one cannot even rely on a largely similar production process being used. Although we have regulations such as those of the Food and Drug Administration[2] and the European Medical Evaluation Agency (EMEA),[3,4] the reality is that we currently do not have the analytical methods to enable us to state with confidence whether or not two protein samples have the same molecular structure. Insulin products have recently lost their patent protection and several new drugs have been developed or are under development for diabetes treatment, such as insulin degludec.[5] The case where Marvel Ltd withdrew its application for a marketing authorisation for a biosimilar insulin renders even more significant the need for high standards of quality control.[6] In this paper we focus on insulin as a case study, considering what information can be extracted from four of the well-established techniques often used to characterise protein structure. We have complemented room temperature data collection with temperature dependence which enables us to probe subtle differences in structure which may not be apparent in room temperature data.

Insulin plays a key role in the regulation of blood glucose levels, exists in different forms, and has been extensively studied since the 1920s using circular dichroism (CD)[7-9] and nuclear magnetic resonance (NMR) spectroscopies,[10-16] dynamic light scattering,[17-21] X-Ray crystallography,[22-25] mass spectrometry (MS)[26,27] and other techniques. It is often described as a peptide hormone since it consists of two polypeptide chains A and B, with 21 and 30 amino acids respectively, held together by two disulfide bridges CysA7–B7 and CysA20–B19, while a third intrachain disulfide bond links cysteines A6 and A11.[28-34]

One of the main characteristics of insulin is its ability, under different conditions, to exist as:

monomers (pH 10–11) or

dimers (acidic pH, with the C-terminal portions of two adjacent B chains forming anti-parallel strands)[35] or

tetramers and hexamers (neutral pH, with the hexamer being three dimers assembled around 2 Zn^{2+} ions to bury hydrophobic surfaces).[14,19,36-39]

Monomeric units within the different oligomerisation states have similar but not identical structures, with the hexamer being the most stable and slightly more helical than any other oligomerisation state[40] (Figure 1).[41]

The variation of structure with conditions is important in the formulation of commercial insulin products, as insulin's oligomerisation state is directly connected with its biological activity.[42-46] In the present study, our focus is on the relationship between oligomerisation state, stability, and secondary structure and the simplest analytical methodologies that can be used to probe this. The behaviour of the monomer,

[a] Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK. E-mail: a.rodger@warwick.ac.uk
[b] CEO and managing director at Biopharmaspec Ltd, Suite 3.1, Lido Medical Centre, St. Saviour, Jersey JE2 7LA, UK.

† Electronic Supplementary Information (ESI) and multimedia files available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

dimer and hexamer as a function of temperature was measured by three well-established experimental techniques – dynamic light scattering (DLS), CD spectroscopy, NMR spectroscopy – and was simulated through molecular dynamics (MD) simulations to aid interpretation of the experimental data.

The main goal of this work was to establish the extent to which different techniques could be used to establish rapidly whether or not different samples had the same structure. We selected techniques based on their simplicity, relatively low time consumption, ease of sample preparation, and because the three instrumental techniques provide different and complementary information. DLS is probably the simplest way to estimate protein size. CD spectroscopy similarly is the simplest way to estimate secondary structure for solution phase proteins. NMR spectroscopy of proteins is commonly used to give atomic resolution structures; however, for our quality control purposes we chose easily accessible 1D ^{1}H NMR for its ability to probe variations in atomic environments. MD complements these techniques by giving indications of how structures can change with temperature. As with NMR we chose not to use it in a sophisticated way but simply as an indicator of the effect of heating the protein.

Experimental

Materials and methods

Materials. Insulin from bovine pancreas (I6634) and ethylenediaminetetraacetic acid disodium salt dihydrate (E5134) (Na_2EDTA) were purchased from Sigma-Aldrich. Bovine insulin differs in sequence from human insulin at A8 (human: Thr; bovine: Ala), A10 (human: Ile; bovine: Val) and B30 (human: Thr; bovine: Ala), resulting in a slightly more compact structure.[22] EDTA can be used to remove Zn-binding in the case of neutral and basic pH,[46] while at acidic pH, the His(B10) side-chains are protonated and therefore unable to coordinate metal ions.[47] Hydrochloric acid (36%; BDH) was purchased from VWR International, sodium hydroxide pellets and sodium dihydrogen orthophosphate dihydrate were purchased from Fisher Scientific, and disodium hydrogen orthophosphate (71643) was from Fluka Biochemika. Deuterium oxide was from Sigma-Aldrich (151882), and ultrapure water was used (18 MΩ.cm at 25 °C) in all cases. The 0.1 M HNO_3 for ICP-OES sample preparation was purified in-house by sub-boiling point distillation, and for calibration, zinc ICP/DCP 356743 and sulfur ICP/DCP 356603 standards (Fluka) were used.

Spectroscopic data collection. A Jasco V-660 UV/Vis or a NanoPhotometer P300 UV/Vis Spectrophotometer was used for measuring insulin concentrations using a molar absorptivity value of ε_{278} = 6080 $M^{-1}cm^{-1}$.[48]

DLS data were collected on a Malvern Zetasizer Nano-series with a laser wavelength of 633 nm and a detection angle of 173˚. A quartz cuvette (ZEN2112) QS 3.00 mm from Malvern Instruments was used for the measurements. Averages over 6 measurements at each temperature for each sample were taken and are presented in terms of intensity and number values. Number values present the number of particles of each size, thus emphasizing smaller particles, whereas intensity values scale by the volume, thus emphasizing the presence of larger particles.

CD spectra were collected on a JASCO J-1500 CD spectropolarimeter using a PTC-510 Peltier thermostated cell holder. The CD spectra (mdeg) were baseline-corrected by subtracting the corresponding buffer blank from each spectrum at all temperatures. The baseline-corrected spectra were zeroed by subtracting the averaged value from 260–250 nm (mdeg) from the whole spectrum. The data were converted into Δε per amino acid residue units, the wavelength step was changed to 1 nm and data from 240 nm to 190 nm were used as the input format required for the secondary structure analysis by SSNN[49] and SELCON.[50]

Protein (*via* S) and zinc concentrations were determined using an Inductively Coupled Plasma Optical Emission Spectrometer (ICP-OES; PerkinElmer Optima 5300 DV, model S10) by measuring S at 180.669 and 181.975 nm and Zn at 206.200 and 213.857 nm. Plasma operating conditions were: Argon (Ar) flow rate 13.0 L/min, auxiliary gas flow rate 0.2 L/min, nebulizer flow rate 0.8 L/min and RF power 1300W. For calibration, 6 Zn and S standards of 0.2, 0.5, 0.7, 1.0, 2.0 and 5.0 ppm were prepared gravimetrically in ultra-pure 0.1 M HNO_3. Samples were prepared using the same 0.1 M HNO_3, which was also used as a blank. Typically, correlation coefficients of better than 0.9998 were obtained.

NMR experiments were carried out on a Bruker AV III-600 spectrometer operating at 600.13 MHz for ^{1}H. The ^{1}H-NMR spectra were Fourier-transformed, phased and baseline-corrected in Topspin 1.3. The chemical shift of the residual HDO was corrected for its temperature-dependence and used as internal reference.[51] The line widths of selected peaks, which could be easily followed across the increasing temperature, were measured in Topspin 1.3 at half height of the peaks.

The *balance* used was a Mettler Toledo XP2U and the pH meter Mettler Toledo Seven Compact pH/Ion S220 InLab Nano Sensors.

Refractive indices were measured on a Bellingham+Stanley Abbe60/DR Refractometer. The centrifuge was a Fisher Scientific microcentrifuge 7200g. A Merck Millipore Direct-Q system was used for water purification.

Temperature dependent measurements. For all *CD temperature-dependent* experiments, the same parameters were used: wavelength range 260–180 nm, 1 s data integration time (DIT), 2 nm band width, data pitch 0.2 nm, 100 nm/min continuous scanning speed. Data were obtained every 5 °C from 20 °C to 110 °C, at a ramp rate of 2.5 °C/min with 300 s equilibration time once the desired temperature was reached. Each spectrum is the average of 12 accumulations. A 0.1 cm path length cuvette was used for 0.1 mg/mL insulin solutions.

For the *^{1}H-NMR temperature-dependent experiments*, the water peak was suppressed using double pulsed field gradient spin echo for excitation sculpting (DPFGSE).[52] The acquisition

time was either 0.7 or 1.2 seconds, a relaxation delay of 1 s, and a 90˚ pulse width of 8.05 µs were used. The final spectra are the sum of 128 scans. An equilibration time of 5 minutes was allowed after the desired temperature was reached. The chemical shift of the water resonance was determined at each temperature and the excitation frequency (centred on the water resonance) was adjusted accordingly. Spectra were acquired from 20–75 °C, with a 5 °C temperature interval.

For the *DLS experiments*, the viscosity was estimated[53] for the neutral and basic samples (Table S1†), by running measurements of solutions consisting of 1 µL of polystyrene latex spheres of known size (100 nm) in 99 µL of blank solution at each temperature, while for the acidic sample the viscosity of water was used, as latex spheres appeared to aggregate in these conditions due to increased hydrophobicity.[54] In a temperature range from 20–75 °C with a 5 °C temperature interval, 6 measurements were recorded for each temperature. 300 s were allowed as the equilibration time after the desired temperature was reached. Each measurement consisted of 6 runs, with 60 s total duration.

Sample Preparation. For the *CD experiments*, the different samples were prepared as follows.

Low pH samples: insulin was initially dissolved in HCl (0.1 M). Sodium phosphate buffer (100 mM, pH 7.0) and water were added. The pH was further adjusted with HCl (0.1 M) and the final volume was adjusted with the addition of water, resulting in an insulin solution of ~0.1 mg/mL in 10 mM sodium phosphate buffer and pH 2.8.

Neutral pH samples: insulin was dissolved in 0.1 M NaOH, sodium phosphate buffer was added and the solution was split into two parts. EDTA (1 mM) was added to one of the aliquots (to complex zinc), aiming at a concentration of 1.5 times that of the concentration of insulin in the sample. Water was added resulting in insulin solutions of ~0.1 mg/mL in 10 mM sodium phosphate buffer. The pH of both samples was pH 7.3.

Basic pH sample: two samples were prepared as for the neutral samples except that pH was adjusted with 0.1 M NaOH to pH 10.6 in 10 mM sodium phosphate buffer.

All samples were left to equilibrate for 30 mins and the pH was checked again before starting the experiment. For all samples, identical control solutions omitting the protein (blank) were prepared for baseline correction of CD spectra.

Samples for *DLS* were prepared in the same way as samples for CD spectroscopy, but with a 50-fold increase in all concentrations. The resulting pH values were: 2.5, 7.7 and 10.9 for the EDTA-free basic sample and 10.5 for the EDTA-containing basic sample. The refractive indices of the blank samples were measured. All samples and blanks were centrifuged with a centrifugal force of 6,000 rpm for 10 minutes and filtered through 0.2 µm syringe filters (Thermo Scientific) directly into a Quartz cuvette prior to the measurement.

For the *NMR experiments*, the samples were prepared as for the DLS experiments, but solutions contained 10% D_2O for the field-frequency lock.[39] The resulting pH values (determined with the pH meter calibrated using buffers in pure H_2O solutions) were: 2.5, 7.4 and 10.8 for the EDTA-free basic sample and 10.7 for the EDTA-containing basic sample.

For the *ICP-OES* sample preparation, insulin was initially dissolved in 1 M NH_4OH. 20 mM ammonium bicarbonate buffer (pH 8.0) and water were added, resulting in a final insulin solution of 1 mg/mL in 10 mM ammonium bicarbonate and 0.1 M NH_4OH. The sample was separated into 100, 200, 300 and 500 µL aliquots, and each was diluted to a final volume of 3 mL using ultra-pure 0.1 M HNO_3.

Simulated Annealing (SA). Simulated annealing molecular dynamics (SA-MD) simulations were performed using the AMBER 12.0 package. Each system was prepared by using the Generalised Born implicit solvent model.[55] The Particle Mesh Ewald (PME) method was employed in all of the simulations.[56] During each simulation, all bonds to H atoms were fixed and all other bonds were constrained to their equilibrium values by applying the SHAKE algorithm.[57] A cut-off radius of 9 Å was used. The minimisations were performed for 10000 cycles. The temperature was controlled by applying the Langevin algorithm, while the pressure was controlled by the isotropic position scaling protocol used in AMBER.[58] Three separate simulations of insulin monomer, dimer and hexamer were performed. The initial PDB (Protein Data Bank) structures for monomer, dimer and hexamer were extracted from the PDB ID code 4E7T. The pH was set to 10.8, 2.5 and 7.5 for insulin monomer, dimer and hexamer systems respectively with appropriate states of protonation, deprotonation and neutralisation of the amino acid residues present in each insulin system, as established using Accelrys discovery studio 2.5. The temperature range used was 293–383 K. At the first step after minimisation, the systems were heated to 293 K and allowed to relax slightly. A total run of 1.68 ns of MD simulated annealing with the time and temperature profile given in Table S2 was performed for all insulin systems.

For the *secondary structure determination*, the DSSP (Dictionary of protein Secondary Structure)[59] module of AMBER was used and three snapshots of the annealing trajectories at 20, 50, 65, 75 and 110 °C were analysed for each system. The results presented are the average over three snapshots at 1/3, 2/3 and last trajectory of each temperature. α-helix and 3/10-helix were combined to give the total α-helix content; β-strand was used for the total β-sheet content; turns were given by addition of turns and bends; and "other" is the sum of β-bridge, π-helix and other, according to Micsonai et al. and Sreerama et al.[60,61] The images from the simulations were generated using Swiss-PdbViewer.[62]

Results and discussion

Data from our set of techniques are reported and analysed below with an eye to using the data sets to identify relatively small structural differences between samples.

ICP-OES measurements

According to the manufacturer's specification sheet, the Zn content in the product is ≤1.0 % (on a dry basis). ICP-OES was

conducted in order to verify the amount of Zn present, and its ratio to the protein's concentration was found to be [Zn]/[P]=0.65 (0.7 % on a dry basis).

DLS Spectroscopy

Insulin assembly and aggregation is pH and concentration dependent.[19,63] Based on the X-ray structure,[28] the monomer, dimer, tetramer and hexamer diameters with one layer of water have been calculated previously as 3.1, 3.7, 4.5, and 5.7 nm, respectively.[17,63,64] In our dynamic light scattering (DLS) experiments at 1 mM protein, three main populations (Figure 2i,ii Table S3†) were observed below 40 °C. At 20 °C, an equilibrium of monomers and dimers (3.2–3.3 nm) is dominant at acidic and basic pH, whereas tetramers (4.8 nm) are predominant at neutral pH in the presence of EDTA (no Zn), and hexamers (5.3 nm) are favoured at neutral pH in the presence of Zn, possibly in equilibrium with smaller oligomers as judged by the relatively small size value. The acidic and neutral samples presented mid-range polydispersity, while the basic samples appeared to be highly polydisperse. All solutions contained very high molecular weight species. A nearly monodisperse hexamer sample was prepared (Table S4†) in conditions that strongly favoured the hexamer formation (1.4 mM insulin, 100 mM NaCl, pH 8.0).[17]

CD and Secondary Structure Analysis

The CD spectra for different insulin samples at 20 °C (Figure 2iii) all have obvious α-helical content, but differ somewhat from the characteristic α-helix spectrum[65] with 1–2 nm shifts to longer wavelength compared to the typical bands at 192 nm/208 nm/222 nm reflecting the relatively high β-sheet content of insulin, as the classic β-sheet spectrum has a positive peak between 195 nm and 202 nm and a single negative signal between 215 nm and 220 nm. The relative intensities of the peaks are consistent with the previously reported pH dependence of the insulin CD spectra.[66] The most significant variation concerns the 222 nm band, which increases in magnitude according to: basic pH < acidic pH, neutral-EDTA < neutral EDTA-free (Zn available) sample following the change from predominance of monomer, to dimer, to hexamer.[8] The largely monomeric basic samples present the lowest intensities of the 194 nm positive peak with their 223 nm peak almost disappearing in accord with an increase in random coil content (usually a negative signal at 200 nm and little, perhaps positive intensity at 222 nm)[67] consistent with the increased flexibility of the N- and C-termini of the B chain of the monomer.[47,68] In the neutral EDTA-free sample, the 222:208 nm ratio is almost 1:1 which is generally accepted to reflect increased hydrophobicity of the environment for the respective helices as would occur in the hexamer where surfaces are protected from the polar aqueous environment.[65] The 195 nm peak increases are a result of the stabilisation of β-sheets between chains in the oligomeric complexes.

The differences between the samples were more apparent when we measured the CD as a function of temperature. All samples showed a gradual reduction in α-helical character (222 and 208 nm signals, Figure 2iv), but the temperature at which this occurred and the degree of shift towards the 200 nm negative band of the random coil differed. The basic (monomeric) samples unfolded whereas the acidic and the neutral samples shifted to 202–204 nm and maintained a considerable amount of helical structure even at 75 °C. We quantified these changes using SSNN[49,69] (a neural network algorithm which uses a self-organising map for protein secondary structure prediction from CD spectra and is based on reference set 7 protein database[67] enhanced by 5 denatured and helical proteins) and SELCON3 using the same reference set 7 (Figure S1†).[50,61] Reference set 7 has been annotated by DSSP. All five samples showed similar, but not identical, trends (Figure 2v), namely gradual decrease in their helical content (of the order of 20–40%), increase in unordered content (15–25%), and a relatively abrupt increase in β-sheet content (10–20%) at different temperatures. In general, the data reflect the higher stability of the oligomers relative to the monomers. The *formation* of β-structure with temperature increase is intriguing and discussed below in the context of the molecular modelling. Finally, spectra were recorded at the end of each experiment after cooling down to 20 °C. Partial reversibility was apparent only at acidic pH (Figures S2, S3†).

^{1}H-NMR Spectroscopy

Applications of NMR spectroscopy to study protein structure and dynamics in solution typically involve an impressive arsenal of state-of-the-art multi-nuclear multi-dimensional techniques such as nuclear Overhauser effect spectroscopy (NOESY), hetero-correlated (^{1}H,^{13}C,^{15}N) experiments, and their hyphenated equivalents, ^{15}N-NMR-based relaxation measurements to study protein dynamics, and diffusion-ordered spectroscopy (DOSY) for estimating molecular size and oligomer speciation.[70] In the case of insulin especially, specific ^{1}H-NMR experiments have been used for detection of aggregation.[71,72] However, as our goal was to detect (as simply as possible) qualitative differences between two samples we chose to exploit the information-rich,[73] but often hard to interpret, 1-dimensional ^{1}H-NMR spectroscopy as a simple and relatively fast approach. The time required to record each spectrum did not exceed 5 mins and approximately 4 hours were needed in total for the temperature-dependent measurements. Key diagnostic variations were observed in line widths (quantitative values are given where possible), chemical shifts and peak dispersion as a function of solution conditions. More details are given in the SI.

Line widths (Figure 3): The *aromatic and backbone* H^N signals between 6 and 10 ppm show clear differences in line widths (Figure 3i) between samples at room temperature. At *basic pH*, most of the backbone H^N signals have disappeared due to exchange with the water protons on an intermediate timescale, and the narrow line widths of the signals for aromatic and aliphatic protons suggest that mono- and dimeric states are favoured.[74] The broadening of the aromatic peaks

above 40°C (Figure 3ii C, D) is probably an indication of the onset of aggregation (see below).

At *acidic pH*, the backbone H^N peaks are widely dispersed, indicating that the protein is in a well-folded state, and sharp enough to confirm that smaller oligomers, such as dimers, are predominant in these conditions.[39]

The broader peaks throughout the *neutral-EDTA* spectra (up to 55°C, Figure 3ii B) are a sign of shorter relaxation times, which are, amongst other parameters, the result of the presence of larger oligomers (i.e. tetramers and hexamers) with longer rotational correlation times (τ_c). Above 55 °C the decrease in line width of the aromatic peaks around 6.4–7.6 ppm (Figure 3ii B) indicates that dissociation to dimers occurs. Aggregation seems to start at 70 °C (Figure 3ii B).

The pronounced broadness of all peaks in the spectrum of the *neutral EDTA-free* sample is indicative of the dominance of the hexamer[10] (Figure 3ii A), while at higher temperatures (60–70°C) the aromatic peaks (6.4–7.6 ppm) become progressively sharper, suggesting an increase in the presence of dimeric or monomeric units.[26] The line widths for the peak at 6.71 ppm were measured over the increasing temperature to give an example of how this feature can be exploited in a quantitative manner for quality control of proteins where molecular size or oligomerisation state is relevant, and complement DLS data.

Aromatic region (6.4–7.6 ppm) (Figure 3ii): Insulin has Tyr and Phe residues, which have protons that can be monitored to report on their environment. Qualitatively, it is evident that the two *basic samples* have a high degree of similarity in this region, whilst the *two neutral samples* are significantly different. The Tyr(B26) Hδ proton signal at 6.63 ppm, characteristic for the monomeric state,[13] shows a gradual decrease and eventual disappearance at 50 °C for the two *basic samples* (Figure 3ii C, D) indicating the transition from the monomeric to an aggregated state.

Histidine Hε1 region (7.45–8.7 ppm) (Figure 4i): Histidine NMR signals are particularly sensitive to pH as their imidazole rings gain/lose a proton between pH 6–7, unless they are buried, hydrogen-bonded, or metal-bound (ESI Figure S4†). *Basic samples* (Figures 4i D, E) have two strong signals at 7.51 and 7.66 ppm at room temperature, with measured line widths 4-6 Hz, which have been assigned to the Hε1 protons of His(B5) and His(B10) respectively[12,14] and decrease in intensity when the temperature increases. At 60 °C, a broader peak at 7.61 ppm which has been assigned to a denatured form of insulin replaces these signals. At *acidic pH* (Figure 4i A), the histidine Hε1 protons appear at 8.53 ppm for His(B5) and 8.61 ppm for His(B10) (line widths 6-7 Hz), as the imidazole ring is protonated under these conditions.[10,39,75,76] At *neutral pH* (Figures 4i B, C), the histidine Hε1 peaks (~7.5-7.6 ppm) appear very broad at 20 °C (34-39 Hz for the EDTA-free and 27-31 Hz for the EDTA-containing sample), and progressively sharper as the temperature increases (e.g. at 60 °C the line width is 4 Hz). The measured line widths of the histidine peaks give a quantitative representation of the oligomerisation state in each sample (Figure 4i, Table S5).

Methyl region (0–1 ppm) (Figure 4ii): Insulin has an extremely high-field shifted signal for the Leu(B15) Hδ methyl protons in the monomer that is present at *basic pH*[77-79] (0.16 ppm, Figures 4ii D, E). This signal moves downfield for the dimer and higher assemblies[13,80] (~0.45 ppm at 20 °C) where shielding of this methyl group by Phe(B24) and Tyr(B26) (Figure 5) is less effective, though still present.[13] The *acidic pH* and *neutral–EDTA-free* samples do not show the resonance at 0.16 ppm, confirming the absence of monomers (Figures 4ii A, B), while the addition of *EDTA at neutral pH* led to a high level of heterogeneity in the methyl region, consistent with the presence of several species (Figure 4ii C). In all cases, disappearance of all high-field shifted peaks in the methyl region indicates unfolding of the protein. The high stability of the hexamer is thus reflected in the persistence of such high-field shifted signals up to a temperature of 70 °C for the most stable neutral (pH 7.4) EDTA-free sample (Figure 4ii B).[26,81] The line widths of these methyl signals are given where possible (Figure 4ii, Table S5).

Hα resonances (3.5–7.0 ppm) (Figure 4iii): For all samples, the downfield shift of peaks in the Hα proton area, from 3.5–4.5 ppm to 4.6–5.8 ppm at increased temperatures (≥ 50 °C), is thought to reveal gradual increase in the β-sheet content,[82] which is in agreement with our CD results and usually accompanies the formation of aggregates. Based on the knowledge that the chemical shifts of Hα protons are dependent on secondary structure, it has been suggested that integration of 1-D ^{1}H-NMR spectra from 4.85–5.90 ppm (largely Hα resonances from residues in β-conformation) may give estimates of the β-strand content of a protein without the need of prior resonance assignment.[83] Amyloid-like fibrils are composed of arrays of β-strands;[84] it is therefore possible that the increase in low-field shifted Hα proton resonances is an indication for the onset of fibril formation, although the sharpness of the peaks indicates that the species are not highly aggregated. Loss of NMR signal intensity for the acidic sample at 75 °C (Figures 4i-iii A), also a consequence of aggregation which may be observed by visual inspection of the sample at the end of the experiment. Differences in the location of these peaks between the *neutral* and *acidic* (at 4.6–5.0 ppm) (Figures 4iii A-C) and the *basic* samples (at 4.6–4.9 ppm and 5.6–5.8 ppm) (Figures 4iii D, E) suggest the presence of different misfolded conformations and aggregates in each case, in accord with our DLS data.

Molecular Modelling

In the molecular dynamics simulations, the monomer, dimer and hexamer were separately investigated for their dynamical behaviour as a function of increasing temperature. We used higher temperatures than in the laboratory to compensate for the shorter time scale of the simulations.

Monomer (ESI, Video S1†): The overlaid structures at 20, 50, 65, 75 and 110 °C (Figure 6i) illustrate increasing flexibility and progressive displacement of the B chain C-terminus,[85] and increasingly flexible N-termini for both A and B chains, while the α-helix content seems to be significantly reduced and the turn content increased at the end of the simulation (Table 1).

Dimer (ESI, Video S2†): Figure 6ii A shows the overlaid structures at 20 and 110 °C. Throughout the dimer simulation, the N-termini of both chains A and B appear to be flexible (Figure S10†). The dimer appears to dissociate at 110 °C, resulting in two misfolded monomers with an extended B chain C-terminus (which turns the Leu(B15) methyl group away from the aromatic residues Phe(B24), Phe(B25) and Tyr(B26)). The overlay of the 110 °C monomer structure with one of the two monomers from the dimer simulation (Figure 6ii B, and Figure S11†) shows that the final structures are similar.

Hexamer (ESI, Video S3†): Figure 6iii depicts the overlaid structures at 20, 50, 65, 75 and 110 °C. The structure gradually expands as the temperature increases, suggesting a tendency towards dissociation. At 110 °C the interactions between the antiparallel β-strands of the B-chain C-terminal portions are weakened and the B-chain N-terminus is completely extended (Figure S12†). Each one of the three dimers is separately depicted in Figure S13†. At the end of this simulation, a significant amount of helix content is still maintained (Table 1) which indicates that despite the evident onset of dissociation into monomers, the latter do not completely unfold on the short timescale of the simulations.

Conclusions

The main conclusion from this work is that empirical comparisons of data collected with the complementary techniques of DLS, CD, NMR and molecular modelling coupled with innovative use of temperature dependence can be used to identify the occurrences of differences in structure in solution even when the samples are heterogeneous. This is therefore the beginning of the development of a batch-to-batch comparison methodology for comparing the similarity of a new formulation to the original product as part of the Quality Control procedure, which will ensure the safety and efficacy of a biopharmaceutical product.[4] DLS gives a simple indication of particle size which helps understanding of line-width variations in NMR spectra. Chemical shifts in different regions of the NMR spectrum give evidence for structural changes but more detailed analysis of methyl protons, histidines, or other aromatic protons (particularly when coupled with temperature dependence) enabled us to monitor the presence or loss of intermolecular interactions as well as stabilities of samples. CD gave a picture of structural differences and the relative stabilities of different samples which could be plotted as percentage changes of different secondary structure motifs as a function of temperature. Although the simulations were only run for a very short time (ns), they provided a useful complement to the experiments when considered over a greater temperature range than the experiments, letting us visualise which parts of the molecules were likely to unfold first.

The significance of temperature as a variable is that we were able to observe quite subtle differences in structure and/or stability by seeing the different temperatures at which the structural motifs were lost and β-strand structure appeared There is no evidence of β-strand structure in the molecular modelling, so we conclude that the β-strand formation is an intermolecular event that reflects the early stages of the process that later leads to fibril formation. Thus combining temperature dependent DLS, CD (with structure analysis) and molecular modelling where possible should be explored for probing early aggregate formation. Although most of our data confirmed and consolidated existing literature data for insulin, we are not aware of any combined use of these techniques in conjunction with temperature variation to probe the early stages of fibre formation.

Acknowledgements

We thank the European Union FP7 Marie Curie Initial Training Network (Innovative Doctoral Programme) for the funding for Meropi Sklepari. Hasan Tanvir Imam's help with the ICP–OES measurements is gratefully acknowledged.

References

1. L. K. Altman, A new insulin given approval for use in U.S., http://www.nytimes.com/1982/10/30/us/a-new-insulin-given-approval-for-use-in-us.html, (accessed 3rd May 2016).
2. U.S. Food and Drug Administration, Scientific Considerations in Demonstrating Biosimilarity to a Reference Product, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM291128.pdf, (accessed 11 Nov, 2015).
3. European Medicines Agency, Guideline on non-clinical and clinical development of similar biological medicinal products containing recombinant human insulin and insulin analogues, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2015/03/WC500184161.pdf, (accessed 26 Oct, 2015).
4. European Medicines Agency, Note for guidance on specifications: Test procedures and acceptance criteria for biotechnological/biological products (CPMP/ICH/365/96) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002824.pdf, (accessed 28 Oct, 2015).
5. M. Eisenstein, *Nat. Biotechnol.*, 2011, **29**, 782-785.
6. European Medicines Agency, Marvel LifeSciences Ltd withdraws its marketing authorisation applications for Solumarv, Isomarv and Combimarv (human insulin), http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2012/11/WC500135156.pdf, (accessed 26 Oct, 2015).
7. Y. Pocker and S. B. Biswas, *Biochemistry*, 1980, **19**, 5043-5049.
8. S. N. Timasheff, M. J. Ettinger and M. G. D. Strycharz, *Biochemistry*, 1971, **10**, 824-831.
9. J. Goldman and F. H. Carpenter, *Biochemistry*, 1974, **13**, 4566-4574.
10. J. H. Bradbury, V. Ramesh and G. Dodson, *J. Mol. Biol.*, 1981, **150**, 609-613.
11. J. H. Bradbury and V. Ramesh, *Biochem. J.*, 1985, **229**, 731-737.

12. R. Palmieri, R. W. K. Lee and M. F. Dunn, *Biochemistry*, 1988, **27**, 3387-3397.
13. M. Roy, R. W. Lee, J. Brange and M. F. Dunn, *J. Biol. Chem.*, 1990, **265**, 5448-5452.
14. M. Roy, R. W. K. Lee, N. C. Kaarsholm, H. Thøgersen, J. Brange and M. F. Dunn, *Biochim. Biophys. Acta, Mol. Cell Res.*, 1990, **1053**, 63-73.
15. W. Kadima, M. Roy, R. W. Lee, N. C. Kaarsholm and M. F. Dunn, *J. Biol. Chem.*, 1992, **267**, 8963-8970.
16. W. Bocian, J. Sitkowski, E. Bednarek, A. Tarnowska, R. Kawęcki and L. Kozerski, *J. Biomol. NMR*, 2008, **40**, 55-64.
17. W. Kadima, L. Øgendal, R. Bauer, N. Kaarsholm, K. Brodersen, J. F. Hansen and P. Porting, *Biopolymers*, 1993, **33**, 1643-1657.
18. S. Hvidt, *Biophys. Chem.*, 1991, **39**, 205-213.
19. A. K. Attri, C. Fernández and A. P. Minton, *Biophys. Chem.*, 2010, **148**, 28-33.
20. H. B. Bohidar, *Biopolymers*, 1998, **45**, 1-8.
21. Y. Xu, Y. Yan, D. Seeman, L. Sun and P. L. Dubin, *Langmuir*, 2012, **28**, 579-586.
22. G. D. Smith, W. A. Pangborn and R. H. Blessing, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **61**, 1476-1482.
23. C. G. Frankær, M. V. Knudsen, K. Norén, E. Nazarenko, K. Ståhl and P. Harris, *Acta Crystallogr., Sect. D*, 2012, **68**, 1259-1271.
24. I. Margiolaki, A. E. Giannopoulou, J. P. Wright, L. Knight, M. Norrman, G. Schluckebier, A. N. Fitch and R. B. Von Dreele, *Acta Crystallogr., Sect. D*, 2013, **69**, 978-990.
25. O. Gursky, J. Badger, Y. Li and D. L. Caspar, *Biophys. J.*, **63**, 1210-1220.
26. E. J. Nettleton, P. Tito, M. Sunde, M. Bouchard, C. M. Dobson and C. V. Robinson, *Biophys. J.*, 2000, **79**, 1053-1065.
27. D. Fabris and C. Fenselau, *Anal. Chem.*, 1999, **71**, 384-387.
28. E. N. Baker, T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. J. Dodson, G. G. Dodson, D. M. Hodgkin, R. E. Hubbard, N. W. Isaacs, C. D. Reynolds and et al., *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 1988, **319**, 369-456.
29. F. Sanger, *Biochem. J.*, 1949, **44**, 126-128.
30. F. Sanger and H. Tuppy, *Biochem. J.*, 1951, **49**, 463-481.
31. F. Sanger and H. Tuppy, *Biochem. J.*, 1951, **49**, 481-490.
32. F. Sanger and E. O. P. Thompson, *Biochem. J.*, 1953, **53**, 353-366.
33. F. Sanger and E. O. P. Thompson, *Biochem. J.*, 1953, **53**, 366-374.
34. H. R. Morris and P. Pucci, *Biochem. Biophys. Res. Commun.*, 1985, **126**, 1122-1128.
35. M. L. Dieken, M. Federwisch, P. De Meyts and A. Wollmer, *Insulin & Related Proteins — Structure to Function and Pharmacology*, Springer Netherlands, 2002.
36. B. Skelbaek-Pedersen, J. Brange, L. Langkjaer, U. Damgaard, H. Ege, S. Havelund, L. G. Heding, K. H. Joergensen, J. Lykkeberg and J. Markussen, *Galenics of Insulin: The Physico-chemical and Pharmaceutical Aspects of Insulin and Insulin Preparations*, Springer Berlin Heidelberg, 2012.
37. A. H. Pekar and B. H. Frank, *Biochemistry*, 1972, **11**, 4013-4016.
38. P. D. Jeffrey and J. H. Coates, *Biochemistry*, 1966, **5**, 489-498.
39. S. Ludvigsen, M. Roy, H. Thoegersen and N. C. Kaarsholm, *Biochemistry*, 1994, **33**, 7998-8006.
40. S. G. Melberg and W. C. Johnson, *Proteins: Struct., Funct., Bioinf.*, 1990, **8**, 280-286.
41. L. Schrödinger, *The PyMOL Molecular Graphics System, Version 1.74*, 2015.
42. J. Brange, *Stability of Insulin*, Springer, 1994.
43. G. Dodson and D. Steiner, *Curr. Opin. Struct. Biol.*, 1998, **8**, 189-194.
44. J. Brange, U. Ribel, J. F. Hansen, G. Dodson, M. T. Hansen, S. Havelund, S. G. Melberg, F. Norris, K. Norris, L. Snel, A. R. Sorensen and H. O. Voigt, *Nature*, 1988, **333**, 679-682.
45. F. Y. H. Lin, D. Y. Kwok, Z. Policova, W. Zingg and A. W. Neumann, *Colloids Surf., B*, 1995, **3**, 281-286.
46. J. Gavrilova, V. Tougu and P. Palumaa, *Metallomics*, 2014, **6**, 1296-1300.
47. Z.-P. Yao, Z.-H. Zeng, H.-M. Li, Y. Zhang, Y.-M. Feng and D.-C. Wang, *Acta Crystallogr., Sect. D*, 1999, **55**, 1524-1532.
48. L. Weil, T. S. Seibles and T. T. Herskovits, *Arch. Biochem. Biophys.*, 1965, **111**, 308-320.
49. V. Hall, A. Nash and A. Rodger, *Anal. Methods*, 2014, **6**, 6721-6726.
50. N. Sreerama and R. W. Woody, *Anal. Biochem.*, 1993, **209**, 32-44.
51. D. Wishart, C. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. Markley and B. Sykes, *J. Biomol. NMR*, 1995, **6**, 135-140.
52. T. L. Hwang and A. J. Shaka, *J. Magn. Reson., Ser. A*, 1995, **112**, 275-279.
53. E. L. Gilroy, M. R. Hicks, D. J. Smith and A. Rodger, *Analyst*, 2011, **136**, 4159-4163.
54. T. Mokhtari, Doctor of Philosophy, Kansas State University, 2007.
55. V. Tsui and D. A. Case, *Biopolymers*, 2000, **56**, 275-291.
56. T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089-10092.
57. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
58. D. A. Case, T. E. Cheatham, 3rd, T. Darden, H. Gohlke, R. Luo, K. M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
59. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577-2637.
60. A. Micsonai, F. Wien, L. Kernya, Y.-H. Lee, Y. Goto, M. Réfrégiers and J. Kardos, *Proc. Natl. Acad. Sci.*, 2015, **112**, E3095-E3103.
61. N. Sreerama, S. Y. Venyaminov and R. W. Woody, *Protein Science : A Publication of the Protein Society*, 1999, **8**, 370-380.
62. N. Guex and M. C. Peitsch, *Electrophoresis*, 1997, **18**, 2714-2723.
63. A. K. Attri, C. Fernández and A. P. Minton, *Biophys. Chem.*, 2010, **148**, 23-27.
64. L. Hovgaard, H. Jacobs, N. A. Mazer and S. W. Kim, *Int. J. Pharm.*, 1996, **132**, 107-113.
65. B. Nordén, A. Rodger, T. Dafforn and R. S. o. Chemistry, *Linear Dichroism and Circular Dichroism: A Textbook on Polarized-light Spectroscopy*, Royal Society of Chemistry, 2010.
66. J. W. S. Morris, D. A. Mercola and E. R. Arquilla, *Biochim. Biophys. Acta, Protein Struct.*, 1968, **160**, 145-150.

67. L. Whitmore and B. A. Wallace, *Nucleic Acids Res.*, 2004, **32**, W668-W673.
68. Y. Zhang, J. L. Whittingham, J. P. Turkenburg, E. J. Dodson, J. Brange and G. G. Dodson, *Acta Crystallogr., Sect. D*, 2002, **58**, 186-187.
69. V. Hall, A. Nash, E. Hines and A. Rodger, *J. Comput. Chem.*, 2013, **34**, 2774-2786.
70. A. Bellomaria, R. Nepravishta, M. Marchetti and M. Paci, *Food Chemistry*, 2016, **194**, 733-739.
71. M. B. Taraban, H. C. Truong, Y. Feng, E. V. Jouravleva, M. A. Anisimov and Y. B. Yu, *J. Pharm. Sci.*, 2015, **104**, 4132-4141.
72. M. F. Lin and C. K. Larive, *Anal. Biochem.*, 1995, **229**, 214-220.
73. R. Page, W. Peti, I. A. Wilson, R. C. Stevens and K. Wüthrich, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 1901-1905.
74. V. Ramesh and J. H. Bradbury, *Arch. Biochem. Biophys.*, 1987, **258**, 112-122.
75. K. A. Higgins, D. J. Craik and J. G. Hall, *Biochem. Int.*, 1990, **22**, 627-637.
76. A. D. Kline and R. M. Justice, *Biochemistry*, 1990, **29**, 2906-2913.
77. K. Huang, B. Xu, S.-Q. Hu, Y.-C. Chu, Q.-x. Hua, Y. Qu, B. Li, S. Wang, R.-y. Wang, S. H. Nakagawa, A. M. Theede, J. Whittaker, P. De Meyts, P. G. Katsoyannis and M. A. Weiss, *J. Mol. Biol.*, 2004, **341**, 529-550.
78. K. Huang, S. J. Chan, Q.-x. Hua, Y.-C. Chu, R.-y. Wang, B. Klaproth, W. Jia, J. Whittaker, P. De Meyts, S. H. Nakagawa, D. F. Steiner, P. G. Katsoyannis and M. A. Weiss, *J. Biol. Chem.*, 2007, **282**, 35337-35349.
79. J. G. Menting, Y. Yang, S. J. Chan, N. B. Phillips, B. J. Smith, J. Whittaker, N. P. Wickramasinghe, L. J. Whittaker, V. Pandyarajan, Z.-l. Wan, S. P. Yadav, J. M. Carroll, N. Strokes, C. T. Roberts, F. Ismail-Beigi, W. Milewski, D. F. Steiner, V. S. Chauhan, C. W. Ward, M. A. Weiss and M. C. Lawrence, *Proc. Natl. Acad. Sci.*, 2014, **111**, E3395-E3404.
80. S. M. Kristensen, A. M. M. Jørgensen, J. J. Led, P. Balschmidt and F. B. Hansen, *J. Mol. Biol.*, 1991, **218**, 221-231.
81. T. Blundell, G. Dodson, D. Hodgkin and D. Mercola, in *Advances in Protein Chemistry*, eds. J. T. E. C.B. Anfinsen and M. R. Frederic, Academic Press, 1972, vol. Volume 26, pp. 279-402.
82. S. P. Mielke and V. V. Krishnan, *Progress in Nuclear Magnetic Resonance Spectroscopy*, 2009, **54**, 141-165.
83. D. S. Wishart, B. D. Sykes and F. M. Richards, *FEBS Lett.*, 1991, **293**, 72-80.
84. R. Tycko and R. B. Wickner, *Accounts of Chemical Research*, 2013, **46**, 1487-1496.
85. V. Zoete, M. Meuwly and M. Karplus, *J. Mol. Biol.*, 2004, **342**, 913-929.

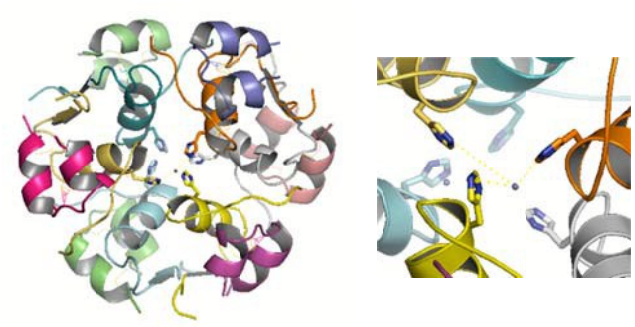


Figure 1: *Left*: Three dimers aggregate around 2 Zn^{2+} ions in hexamer formation. Each chain of the insulin hexamer is represented with a different color. *Right*: Coordination between a Zn^{2+} ion and three His(B10) side chains. For the representation the structure with PDB ID code 4E7T was used and PyMOL.[1]

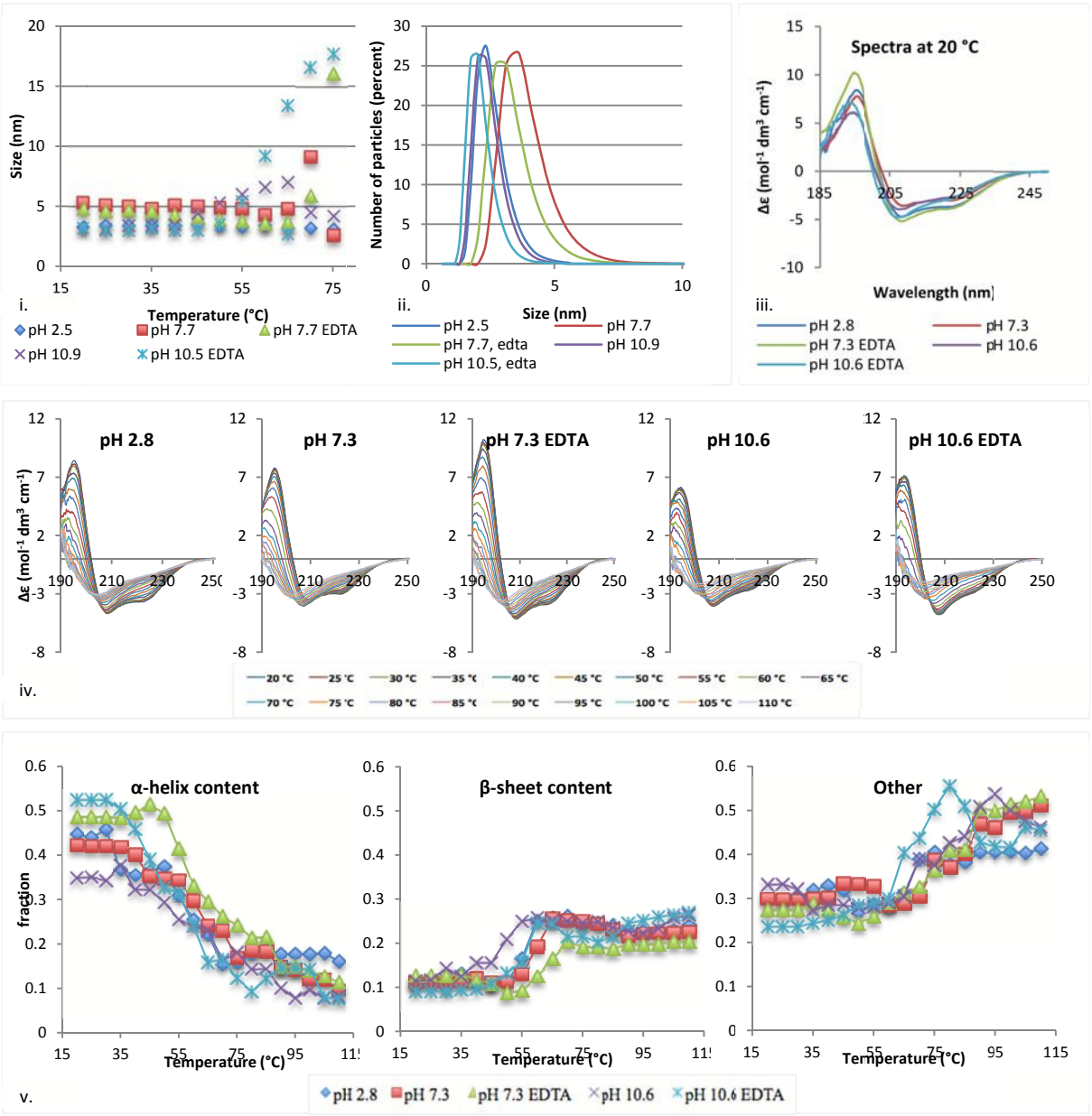


Figure 2: i, ii. Hydrodynamic diameter (nm) of Insulin (1 mM) in acidic, neutral (EDTA-free, EDTA-containing) and basic (EDTA-free, EDTA-containing) pH in 10 mM sodium phosphate buffer solutions shown by DLS. i. Dynamic light scattering temperature experiment. The size of insulin particles in terms of intensity values (nm) is shown. ii. Size distribution of the five samples at 25 °C. The number of particles (percent) of each population is shown. iii. CD spectra in Δε units for Insulin at 0.1 mg/mL, 10 mM sodium phosphate buffer at the indicated pH conditions at 20 °C. iv. Changes in the CD spectra of Insulin with increasing temperature. v. Changes in secondary structure content (fraction values) of the 5 different samples of insulin with increasing temperature. The CD spectra were analysed with SSNN.

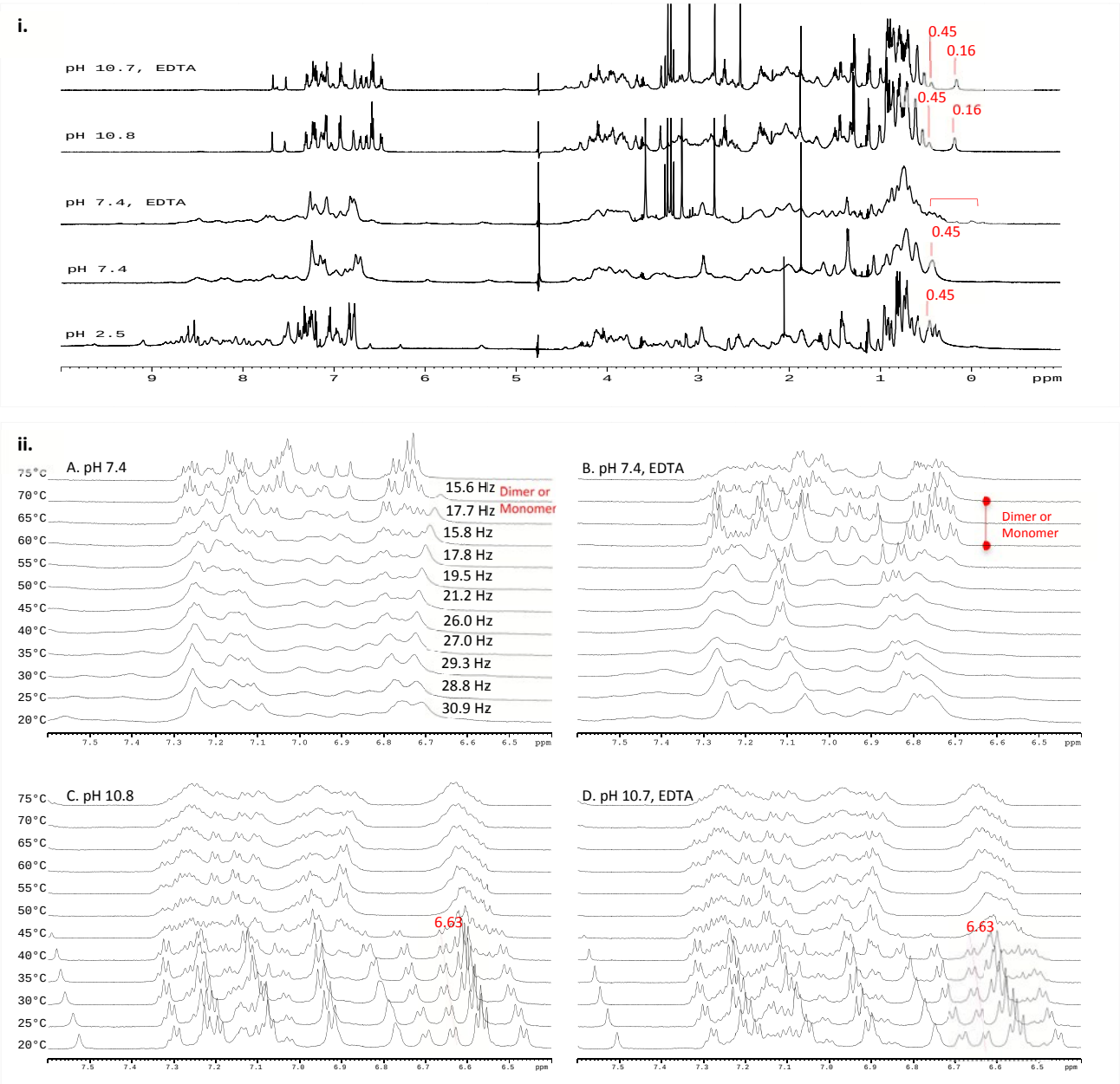


Figure 3: i. Comparison of line widths and overall dispersion in the ^{1}H-NMR spectra of Insulin (1 mM) in acidic, neutral (EDTA-free, EDTA-containing) and basic (EDTA-free, EDTA-containing) pH, 10 mM sodium phosphate buffer solutions at 25 °C. In the aliphatic region, at 0.16 and 0.45 ppm the characteristic peaks for the monomer and the dimeric unit respectively are indicated. ii. Aromatic region (6.4-7.6 ppm) of the ^{1}H-NMR spectra of insulin at neutral and basic pH at increasing temperature. The sharpness of the peaks at 70 °C for the pH 7.4 EDTA-free sample (tetramers and hexamers predominant at room temperature) (A) and at 60-65 °C for the pH 7.4 EDTA-containing sample (tetramers predominant at room temperature) (B) suggests an increase in the presence of the dimeric or monomeric units. The line widths for the peak at 6.71 ppm for the neutral EDTA-free sample (A) were measured at half height of the peak, indicating transition towards smaller oligomers with the increase of temperature. The peak at 6.63 ppm up to 45 °C for both basic samples pH 10.8 EDTA-free sample (C) and the pH 10.7 EDTA-containing sample (D) (monomers and dimers predominant at room temperature) refers to a Tyr(B26) Hδ proton signal characteristic for the monomeric state.

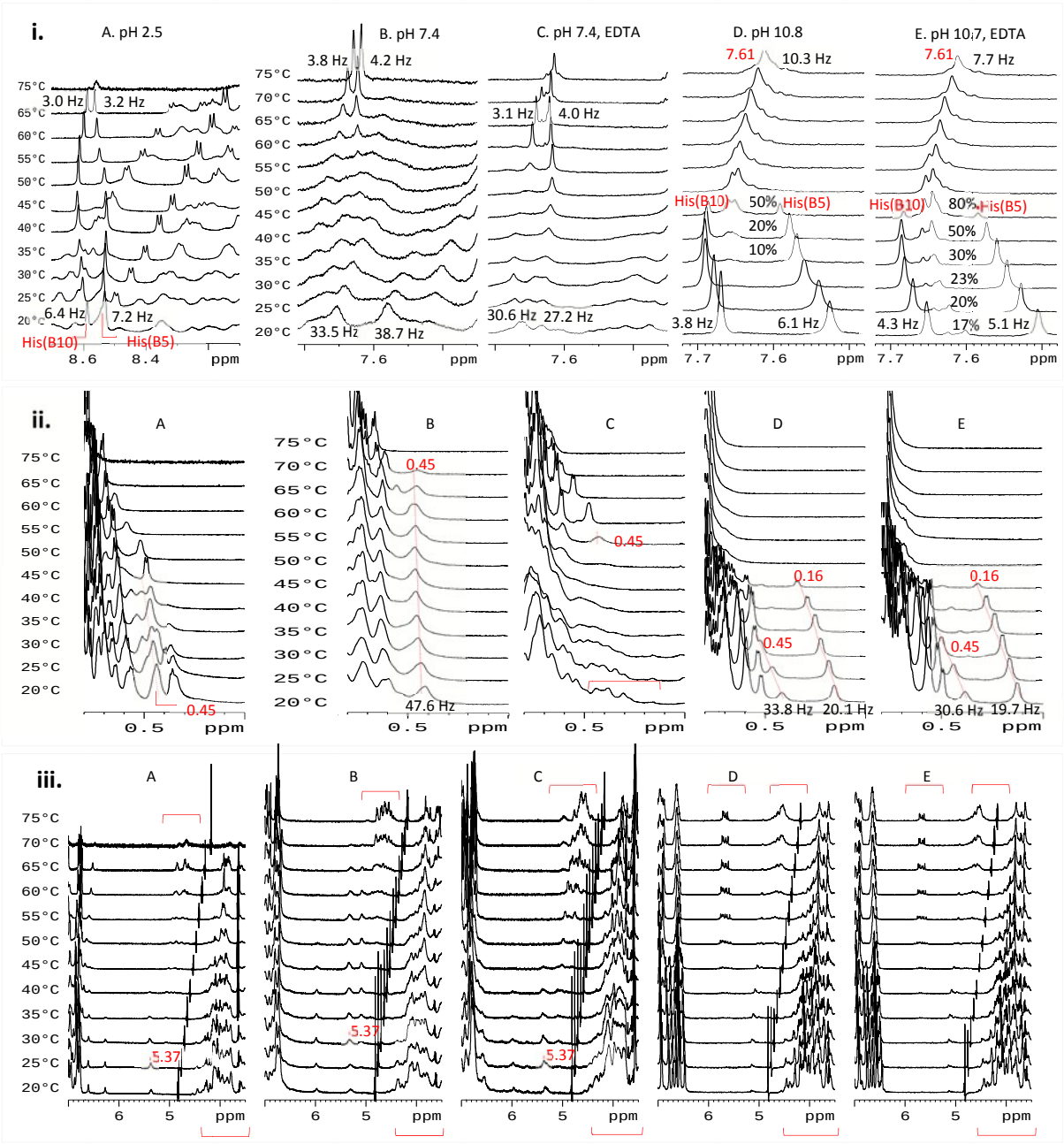


Figure 4: Selected regions from the ^{1}H-NMR spectra of insulin at acidic (A), neutral EDTA-free (B) and EDTA-containing (C),, basic EDTA-free (D) and EDTA-containing (E) pH at increasing temperature. i. Histidine Hε1 region: A. (8.1–8.7 ppm), B-E. (7.48–7.72 ppm). The His(B5) and His(B10) Hε1 proton signals are shown. For the two basic samples the appearance and increase in the population of a new species, probably corresponding to denatured forms, is indicated. The noted line widths of the histidine peaks are an indication of the size of the dominant species in each sample. ii. Methyl region (0.0–0.8 ppm). The disappearance of the peaks characteristic of the monomer (0.16 ppm) and dimeric unit (0.45 ppm) is shown, indicating unfolding of the protein. The shown line widths indicate the presence of small units at basic pH (D, E) and higher oligomers at pH 7.4 (B). iii. Hα region (3.5–7.0 ppm). The downfield shift of peaks reveals a gradual increase in the β-sheet content, indicating the formation of aggregates. The peak at 5.37 ppm probably corresponds to one of the aromatic residues on the monomer-monomer interface (ESI Figures S5–9†).

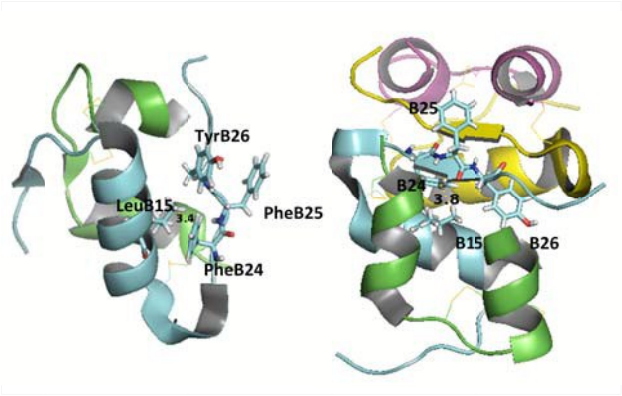


Figure 5: The distance between the Leu(B15) Hδ methyl group and the aromatic residue Phe(B24) is shown. *Left*: in the case of the monomer the distance was measured as 3.4 Å. *Right*: for the dimer the distance was measured as 3.8 Å.

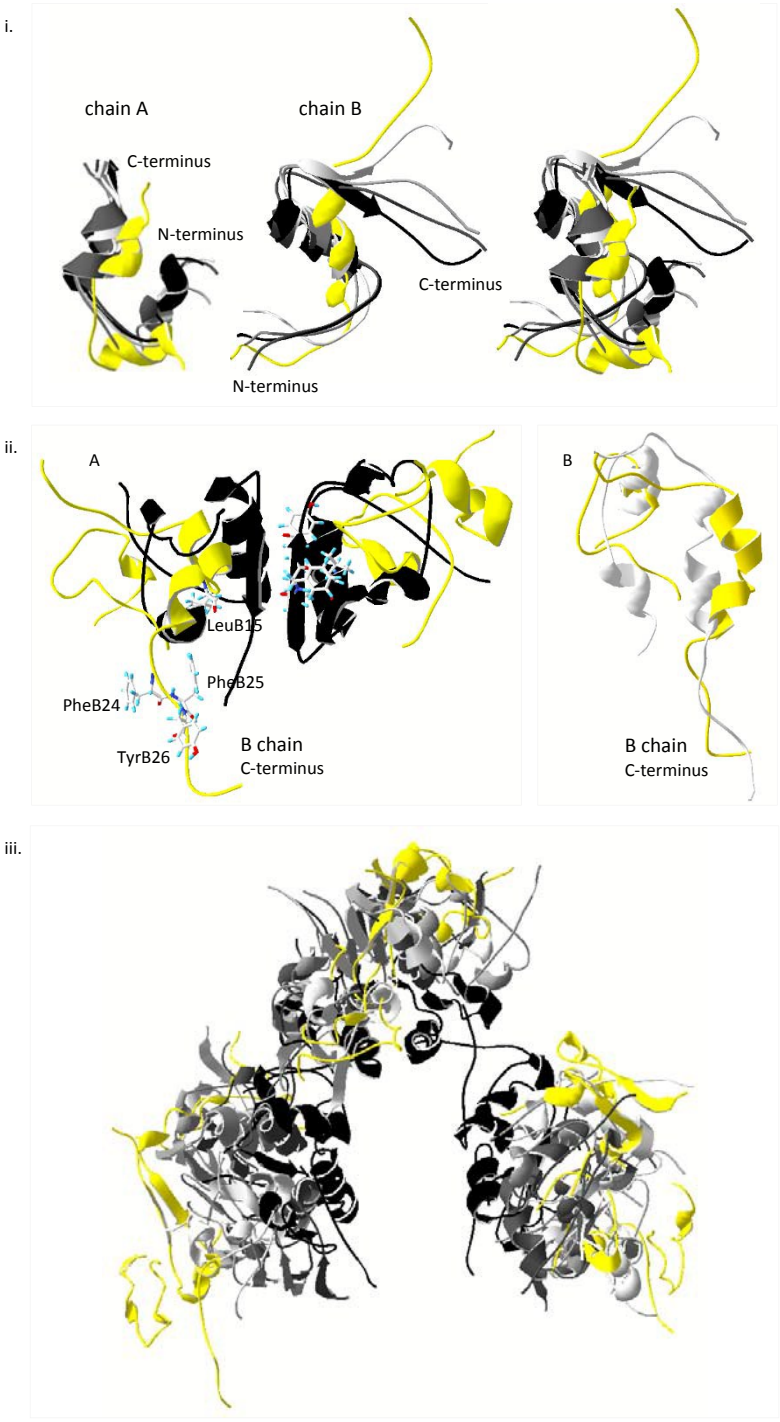


Figure 6: Overlaid structures from the SA-MD simulations at increasing temperature. From black to light grey with decreasing brightness, the frames are at 20, 50, 65 and 75 °C. The frame at 110 °C is shown in yellow. i. The monomer is depicted. Left: A chain, center: B chain, right: A and B chains. ii. Dimer. A. The frames are at 20 °C (black) and 110 °C (yellow). LeuB15 and the aromatic residues at the C-terminus of chain B are shown in both structures, indicating the increased distance at 110 °C. B. Comparison between the monomer 1 from the dimer simulation at 110 °C (yellow) and the monomer simulation at 110 °C (white). The similarity of the two unfolded structures is apparent. iii. Hexamer. The expansion of the structure and the tendency towards dissociation with increasing temperature is shown.

Table 1: Secondary structure analysis results from the simulations of Insulin monomer, dimer and hexamer using the DSSP algorithm of the AMBER package. The results were averaged over three snapshots at each one of the shown temperatures. The abbreviations stand for: (h) helix, (s) strand, (t) turn, (o) other.

		20 °C	50 °C	65 °C	75 °C	110 °C
Monomer	(h)	0.45	0.46	0.36	0.41	0.20
	(s)	0.0	0.05	0.08	0.03	0.03
	(t)	0.13	0.18	0.20	0.17	0.37
	(o)	0.42	0.31	0.36	0.39	0.40
Dimer	(h)	0.40	0.38	0.38	0.34	0.27
	(s)	0.05	0.07	0.04	0.09	0.05
	(t)	0.17	0.18	0.21	0.24	0.31
	(o)	0.38	0.37	0.37	0.33	0.37
Hexamer	(h)	0.42	0.43	0.37	0.42	0.36
	(s)	0.02	0.03	0.07	0.03	0.05
	(t)	0.16	0.17	0.21	0.18	0.22
	(o)	0.41	0.38	0.35	0.38	0.37