

HDX and in-silico docking reveal that excipients stabilise G-CSF via a combination of preferential exclusion and specific hotspot interactions

Victoria E. Wood¹, Kate Groves², Adam Cryar², Milena Quaglia², Paul Matejtschuk³,

Paul A. Dalby^{1}*

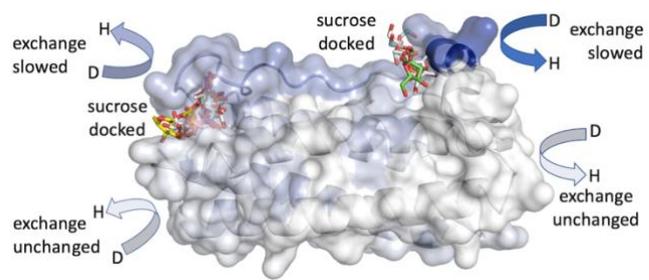
¹Department of Biochemical Engineering, University College London, Gower Street, London, WC1E 6BT

²National Measurement Laboratory at LGC Ltd, Queens Road, Teddington, TW11

0LY

³National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire, EN6 3QG

*Correspondence email: p.dalby@ucl.ac.uk



Graphical Abstract

ABSTRACT

Assuring the stability of therapeutic proteins is a major challenge in the biopharmaceutical industry, and a better molecular understanding of the mechanisms through which formulations influence their stability, is an ongoing priority. While the preferential exclusion effects of excipients are well known, the additional presence and impact of specific protein-excipient interactions has proven more elusive to identify and characterise. We have taken a combined approach of *in-silico* molecular docking, and hydrogen deuterium exchange mass spectrometry (HDX-MS), to characterise the interactions between Granulocyte Colony stimulating Factor (G-CSF), and some common excipients. These interactions were related to their influence on the thermal-melting temperatures (T_m), for the non-reversible unfolding of G-CSF in liquid formulations. The residue-level interaction sites predicted *in silico*, correlated well with those identified experimentally, and highlighted the potential impact of specific excipient interactions on the T_m of G-CSF.

Keywords: Aggregation, formulation, stability, excipient, prediction

INTRODUCTION

Improving the stability of therapeutic proteins remains a major challenge in the biopharmaceutical industry, especially as development timelines are becoming further

compressed, requiring a stable formulation at increasingly earlier stages in clinical development. While several degradation pathways must be controlled, aggregation is of primary concern as it can lead to a loss of potency, blockages in needles or blood vessels, and the potential for adverse immunogenic effects in patients [1]. Aggregate formation can occur throughout the manufacturing process [2-4], but is particularly problematic during storage, requiring formulation to minimise further formation, and enhance the shelf-life of the final dosage form [5].

The formulation design process has improved in recent years through the availability of Design of Experiment (DoE) approaches [6, 7], and high-throughput screens that measure the impact of formulations on thermal-unfolding transition mid-points (T_m) and aggregation onset temperatures (T_{agg}) [8-10]. Formulations are also assessed through intensive degradation studies with aggregation kinetics accelerated by elevated temperatures, shear flow, agitation, or through repeated freezing and thawing [11-13]. However, the initial selection of potential formulation factors of pH, ionic strength, buffers, and excipients, is often still based upon relatively sparse and inconsistent data on their effects observed with different proteins, which is further confounded by complex interactions between the different factors, as revealed by DoE [6,7].

A major goal in formulation is to improve molecular-level understanding of formulation behaviours, ultimately enabling the prediction of formulation excipients most likely to improve the stability of a given protein. Protein stability is generally composed of three components: chemical, colloidal and conformational stability. Formulation excipients can influence all three through a variety of mechanisms, which

in turn can influence protein aggregation kinetics [14-16]. For the two-state reversible unfolding of proteins, osmolyte [Os] concentration-dependent stabilisation often has a linear relationship with the free-energy of unfolding, quantifiable as $\delta \ln(K)/\delta[\text{Os}]$, which also correlates with the air-solvent surface tension [17, 18]. Extensive biophysical analyses, notably with densimetry [19-23], and recently with vapour pressure osmometry [24, 25] have generally observed that most osmolytic amino acids, sugars and polyols, are preferentially excluded from protein surfaces [26]. Such molecules are hydrophilic and should be able to interact with polar residues on the protein surface. However, if these interactions are weaker than those between protein and water molecules, the protein becomes preferentially hydrated, and the excipients mainly excluded into the bulk solution [27].

Several theories are used to explain how preferential exclusion leads to conformational stabilisation of the native protein. Firstly, enrichment of water on the protein surface may alter the free-energy associated with desolvation upon folding. Secondly, increased air-solvent surface tension, due to excipient-solvent interactions, potentially leads to an increased free-energy of cavity formation required for unfolding [20]. Finally, in scaled particle theory (SPT), a hard-sphere model of macromolecular crowding is assumed, in which a sterically crowded solution disfavours formation of the unfolded state, as it occupies more volume than the native state [17, 18].

In general, the effect of osmolytes on stability correlates well with changes in air-solvent surface tension [20], which in turn correlates with the change in the solvent accessible surface of proteins upon unfolding [20], and also with related

thermodynamic terms including the heat capacity change upon unfolding [28]. Converse relationships also exist for chemical denaturants [19, 28]. As most proteins increase their solvent accessible surface area and heat capacity, upon unfolding [28], sugars and similar osmolytes will usually increase their conformational stability in a concentration-dependent manner. However, their impact on reversible protein-protein interactions is less predictable [18], and can be either destabilising or stabilising [29-32]. For reversible homo-dimerisation of native α -chymotrypsin, stabilisation by the osmolytes glucose, sucrose, or raffinose, could be explained in part by a statistical mechanical model based on hard volume exclusion effects, but also required an additional mechanistic component working in opposition, such as a preferential chemical interaction on the order of $K < 10 \text{ M}^{-1}$, to fully explain the experimental observations [18].

The convolution of reversible unfolding, and protein-protein complex formation, with irreversible protein aggregation, as measured either kinetically under native conditions, or as apparent transition mid-points ($T_{m,app}$) from non-reversible thermal-unfolding, further complicates the effects of osmolytes and related excipients. In addition, the $T_{m,app}$ from thermal unfolding, often does not even correlate well to low-temperature aggregation kinetics, which can also follow non-Arrhenius behaviour [9, 33, 34]. Thus, osmolytes may have a more complex combination of mechanisms for the stabilisation of proteins against aggregation. Conformational stabilisation through preferential exclusion is likely to play a major role, but specific interactions with the protein surface that influence the solvent accessibility of aggregation-prone regions (APRs), or modify

surface charge, hydrophobicity and solvation, may also be important [35, 36]. This raises the question as to whether a small number of important specific interactions are hidden under a background of preferential exclusion. Indeed, the resolution of density measurement methods by which preferential exclusion was determined, was too low to rule out the simultaneous presence of a small number of specific interactions, leading to conclusions that preferential exclusion of sucrose was “*at least partially*” observed [20]. Preferential GdmCl interactions were previously resolved down to only 50 mols GdmCl per mol protein by densimetry [19], and so any specific interactions potentially formed at lower concentrations would also not be observed before preferential exclusion begins to dominate.

Alternative methods for identifying preferential interactions have had limited success. For example, circular dichroism did not observe any impact of sucrose on the native conformations of α -chymotrypsin or chymotrypsinogen [20]. Sucrose also inhibited global hydrogen deuterium exchange (HDX) as measured using IR spectroscopy, and by cysteine reactivity, for recombinant interleukin 1 receptor antagonist (rhIL-1ra) [37], consistent with a reduction of conformational flexibility and compaction of the native state through preferential exclusion. For non-glycosylated G-CSF, a 19.6 kDa four-helix bundle cytokine used to treat neutropenia [38], sucrose inhibits aggregation, and this was previously proposed to be due to preferential exclusion favouring the most compact native ensemble [39]. Peptide-level HDX-mass spectrometry (MS) in phosphate-buffered saline (PBS) at pH 7, was also consistent with preferential exclusion as it identified no regional interactions with 1 M sucrose, while

exchange protection across many G-CSF peptides indicated non-specific stabilization [40]. A heteronuclear single quantum coherence (HSQC)-NMR study of G-CSF titrated with either polysorbate 80 or sorbitol, revealed no impact on the G-CSF structure at pH 4.2 as measured by chemical shifts [41].

Peptide-level HDX-MS analysis of an IgG1 mAb with arginine and sucrose, also revealed only a global trend of protection in most peptides, correlating with conformational and storage stability [42]. Similar results were observed with an IgG4 mAb in 0.4 M trehalose at pH 6.5 – 7.4 [43]. However, trehalose was found to significantly rigidify a CH2 peptide in IgG1, previously shown to be an aggregation hotspot region [42]. This indicated that sugars may not be entirely excluded preferentially, but can form some interactions at selective sites on the protein surface, with sufficiently high affinity to displace the water.

Progress in computational methods to understand and predict formulations is emerging through a number of routes, including statistical and neural network approaches [44], molecular surface calculations [45], and molecular dynamics simulations [36, 35]. We also demonstrated recently, that molecular docking predicted specific interactions of excipients at hotspots on protein surfaces, and a potential correlation with experimental conformational stabilities [46]. Here, we explore this methodology, in combination with peptide-level HDX-MS, for a range of excipients added to non-glycosylated G-CSF.

Marketed as Filgrastim, G-CSF is formulated in 10 mM sodium acetate, 5% sorbitol, and 0.004% polysorbate-80, at pH 4.0 [47] and is much less stable at physiological pH,

where it rapidly aggregates [48, 49, 39]. G-CSF aggregation is thought to occur through a structurally-perturbed monomer (M^*), which can either revert to monomer (M), combine to form an aggregate dimer (M_2), or combine with other aggregates into larger species (M_{x+1}) [50]. The aggregation kinetics of G-CSF formulations correlated well with conformational stability under some conditions [8], but less well under conditions that strongly favoured (>99.9%) the native-state [10]. The previous peptide-level HDX-MS studies of G-CSF formulations [40] were performed only at pH 7 with excess (1 M) sucrose, under known aggregation-prone conditions of 37 °C [50], and then also at 4 °C to examine protection under slower exchange conditions. The intrinsic rate of exchange is three orders of magnitude faster at pH 7 than at pH 4, while G-CSF is also considerably more stable at pH 4. The study under destabilising conditions for G-CSF at pH 7, combined with a high concentration of sucrose, may have made it difficult to observe any exchange protection at specific peptides resulting from preferential interactions with the sucrose. Therefore, we aimed to identify specific interactions under more stable conditions of pH 4.25, normally used in G-CSF formulations, and where the intrinsic exchange is also slower. The influence of pre-equilibration time revealed a combination of preferential exclusion and specific interaction effects for sucrose, mannitol and phenylalanine.

MATERIALS AND METHODS

Chemicals were supplied by the following manufacturers: acetic acid, arginine.HCl, deuterium oxide (D₂O) 99.9%, glutamic acid, histidine, phenylalanine, sodium acetate, and sucrose (Sigma-Aldrich Co. (now Merck), Gillingham, UK); mannitol, NaCl, sorbitol and trehalose (Fisher Scientific Inc., Loughborough, UK); tris(2-carboxyethyl)phosphine (TCEP) (Thermo Fisher, Hemel Hempstead, UK); phosphate buffered saline (PBS) (Severn Biotech Ltd, Kidderminster, UK).

Expression and purification of G-CSF

G-CSF was expressed as inclusion bodies (IBs) in *E. coli* BL21 DE3 harbouring a modified pET21A plasmid (Novagen, WI, USA) described previously [51, 10]. A glycerol stock was cultured in 10 mL Terrific Broth with 0.1 mg/mL ampicillin (TB/Amp) in a 50 mL falcon tube, incubated overnight at 37 °C, 250 rpm, then seeded into sterile 2 L baffled flasks containing 500 mL TB/Amp and incubated at 37 °C, 250 rpm for 3 h. This was sterile transferred into a 7.5 L bioreactor (New Brunswick, NJ, USA) containing 5 L of TB held at 37 °C. A magnesium sulphate/ampicillin stock was added to a final concentration of 1 mM. Dissolved oxygen was controlled at 30% (v/v) via agitation up to 600 rpm followed by maintenance via gas cylinder, and a pH of 7 maintained with phosphoric acid and ammonium hydroxide. Expression was induced at mid-exponential and at stationary phase growth (OD₆₀₀= 10 and 35, respectively), with 1 M isopropyl β-d-1-thiogalactopyranoside (IPTG) for a 1 mM final concentration. Cells were harvested 3.5 h post-induction by centrifugation at 7080 x g, 20 mins, and 4 °C (Avanti J-20 XPI; Beckman Coulter Inc., Fullerton, CA, USA). Pellets were washed in 10 mM PBS (Severn Biotech Ltd, Kidderminster, UK),

centrifuged into small 4 g pellets in 50 mL falcon tubes at 7,728 x g, for 30 mins, 4 °C, then stored at -20 °C. Pellets were defrosted for 30 min at room temperature (RT), re-suspended in 10 mM PBS at 1% (w/v), lysed by a single pass through an APV LAB40 high pressure homogeniser at 1000 Bar, then stored on ice. Sodium deoxycholate was added to 1 mg/mL and the lysate rolled for 15 min at RT, before adding 20 µL of Benzonase® nuclease (25 U/mL; Merck Millipore, Billerica, Massachusetts, USA), and rolled for a further 15 mins. Inclusion bodies (IBs) were pelleted by centrifugation at 17,700 x g, 30 min, 4 °C (Avanti J20 XPI), and washed twice by resuspension in wash buffer at 1:40 (w/v) ratio at RT using a hand-held food blender and centrifugation 17,700 x g, 30 min, 4 °C. The first wash buffer was 50 mM Tris pH 8.0, 5 mM ethylenediaminetetraacetic acid (EDTA) and 2% Triton X-100 (g/v), and the second was 50 mM Tris pH 8.0, 5 mM EDTA and 1 M NaCl.

IB pellets were re-suspended in 10 mL of 4 M urea and adjusted to pH 12.0 using NaOH, then rolled for 30 min at RT. The solution was added dropwise to a stirred solution of 1 M Arginine.HCl, pH 8.25, to a final 20x dilution, then rolled for > 12 h at RT. Refolding was quenched by adjusting to pH 4.25 with glacial acetic acid, rolling for a further two hours, then clarified by centrifugation at 17,700 x g, 20 min, 4 °C, and the supernatant spin-concentrated to 10 mL using Amicon Ultra-15 10 kDa cut off filters (Merck Millipore, Billerica, Massachusetts, USA) at 1,389 x g and 4 °C, before final sterile filtration and storage at 4 °C. G-CSF was purified by size-exclusion chromatography (SEC) through a HiLoad® 26/60 Superdex® 200 prep-grade 320 mL column (GE Healthcare Life Sciences, Germany), and eluted isocratically in 50 mM

sodium acetate pH 4.25 at 3 mL/min as previously [51], into a fraction collector containing tubes with Milli-Q H₂O at 6 °C, which diluted the buffer to 10 mM. Fractions with > 95% purity and > 0.1 mg/mL G-CSF as determined by non-reducing SDS-PAGE and UV absorbance, were pooled and concentrated to 0.6 mg/mL and 1.0 mg/mL using Amicon Ultra-15 centrifugal filters at 1890 x g and 4 °C. Analytical SEC determined the G-CSF samples to be 98% monomeric. LC-MS analysis confirmed the presence of 96.4% monomer (MW of M+ 18,799 Da and M+acetonitrile(ACN)+2H 18,820 Da), and 3.6% dimer (MW M+ 37,598 Da) Further details of intact LC-MS methods are shown in the supplementary information.

Thermal transition mid-points (T_m) and aggregation onset temperatures (T_{agg})

Concentrated G-CSF preparations were mixed 1:1 with sterile-filtered excipient solutions prepared at 2X their final concentration in 10 mM sodium acetate pH 4.25, to obtain final formulated G-CSF at 0.3 mg/mL and 0.5 mg/mL. Intrinsic protein fluorescence (266 nm excitation, 280-450 nm emission scan) and static light scattering (SLS) at 266 nm, were measured simultaneously with a UNit (Unchained Laboratories, Pleasanton CA) in 1 °C steps from 20 °C to 80 °C, after 30 s equilibration at each temperature. Microcuvettes were loaded with 9 μ L samples in triplicate, using 0.5 mg/mL G-CSF. T_{agg} was determined from SLS counts at 266 nm using the instrument software, and defined as the temperature at which 10% of the maximum signal increase was reached. Fluorescence intensity at 340 nm, and also the barycentric mean (BCM) fluorescence intensity, versus temperature were each fitted to

a two-state transition model using equation 1 [52, 53] in OriginPro 2017 (Origin Lab Corp., Northampton, MA, USA):

$$I_T = \frac{I_N + aT + (I_D + bT) \exp\left[\left(\frac{\Delta H_{vh}}{R}\right)\left(\frac{1}{T_m} - \frac{1}{T}\right)\right]}{1 + \exp\left[\left(\frac{\Delta H_{vh}}{R}\right)\left(\frac{1}{T_m} - \frac{1}{T}\right)\right]} \quad \text{Eq. 1}$$

where I_T is the observed signal, I_N and I_D the native and denatured baseline intercepts, a and b the native and denatured baseline slopes, T the temperature, ΔH_{vh} the van't Hoff enthalpy, R the gas constant ($1.987 \text{ cal mol}^{-1} \text{ K}^{-1}$) and T_m the thermal transition midpoint. The van't Hoff entropy was calculated using equation 2, and the mole-fraction, f_T , of unfolded protein at any temperature T , was calculated from equation 3.

$$\Delta S_{vh} = \frac{\Delta H_{vh}}{T_m} \quad \text{Eq. 2}$$

$$f_T = \frac{\exp\left[\left(\frac{\Delta H_{vh}}{R}\right)\left(\frac{1}{T_m} - \frac{1}{T}\right)\right]}{1 + \exp\left[\left(\frac{\Delta H_{vh}}{R}\right)\left(\frac{1}{T_m} - \frac{1}{T}\right)\right]} \quad \text{Eq. 3}$$

Hydrogen deuterium exchange - mass spectrometry

Sample preparation

To study the impact of pH on the peptide-level HDX-MS, G-CSF at 0.6 mg/mL was dialysed into 50 mM sodium acetate, pH 4.25, and 10 mM PBS, pH 7.4 (Severn Biotech Ltd, Kidderminster, UK), using 10 kDa cut-off Slide-A-Lyzer Dialysis cassettes (Fisher Scientific, Leicestershire). To study the impact of excipients on the peptide-level HDX-MS, G-CSF at 0.6 mg/mL was not dialysed, but remained in 10 mM sodium acetate, pH 4.25.

Deuterium-exchanged excipients, i.e. deuterated at protonation sites labile to exchange, were obtained by dissolving the solids in 99.9% D₂O, to 10 % (w/v) (549 mM) mannitol, 50 mM arginine, 10 % (w/v) (292 mM) sucrose, and 50 mM phenylalanine, 1 mL aliquoted into 2 mL Schott glass vials (VCDIN2D, Adelphi Tubes Ltd, Haywards Heath, UK), and then capped with igloo halobutyl rubber stoppers, leaving a gap for vapour passage. Vials were freeze-dried for three cycles as described by [37], using a VirTis AdVantage freeze-dryer (Biopharma, Winchester, UK) and the parameters shown in Table S1, supplementary information. Between cycles the samples were rehydrated with 1 mL 99.9% D₂O, except after the final freeze-drying cycle, where the vials were backfilled with nitrogen to 1 bar, stoppered using the hydraulic stoppering of the freeze-dryer shelves, then snap-frozen in liquid nitrogen for storage at -80 °C.

An internal reference peptide (IRP), of amino acidic sequence PPPI, was custom synthesised (Thermo Fisher Scientific Inc., Wilmington, USA) and used to screen for differences in intrinsic HDX rates resulting from different formulation conditions [40]. The peptide was solubilised at 1 mg/mL in 50:50 (v:v) MeOH:ACN, stored in 10 µL aliquots, and frozen at -20 °C. Aliquots were defrosted as appropriate, diluted in sample buffer and spiked into protein samples prior to deuteration, where specified, at 1:100 for a final concentration of 3 µM.

Method

For each run, 15 µL of 0.3 mg/mL (15 µM) protein sample was diluted 1:10 with appropriate buffer solutions, prepared in either H₂O or D₂O, for peptide mapping/time

zero and exchange experiments respectively. For the pH comparison, the buffers were 10 mM sodium acetate, pH 4.25, and 10 mM PBS, pH 7.4. For the excipient study, the buffers were 10 mM sodium acetate, pH 4.25, with and without excipients. Deuterium-exchanged excipient vials were defrosted for 10 mins at room temperature, and reconstituted in 10 mM sodium acetate in 99.9% D₂O, pH 4.25. Resulting buffers were pH-adjusted using 99% DCl, and assuming $pD \text{ corr} = pD \text{ read} + 0.4$ [54].

For the pH comparison, samples were incubated at room temperature (21 ± 2 °C) for 30 s, 2.6 min, 5 min, 15 min, 1 h, 4 h and 8 h, in triplicate. For the excipient study, samples were incubated at room temperature (21 ± 2 °C) for 2.6 min, 1 h and 8 h, in triplicate. The exchange lengths included were chosen to cover a range of exchange kinetics.

Following the appropriate exchange time, exchange was quenched by a 1:1 dilution with 50 μ L of 4 M guanidine hydrochloride, 600 mM TCEP, in 100 mM sodium acetate buffer (pH 2.5) at 4 °C. Those conditions were selected to maximise peptide map coverage and redundancy [55]. Following a 30 s quench delay, 95 μ L of quenched sample was injected onto a refrigerated nanoACQUITY UPLC System with HDX technology (Waters, Milford, Massachusetts, USA) for on-line pepsin digestion and chromatographic separation. Sample handling and mixing steps were performed using a first-generation LEAP PAL system set up for HDX analysis (LEAP Technologies, Morrisville, North Carolina, USA). Multiple sources of deuteration buffer solutions were defined for the LEAP PAL system so that the HDX-MS experiments of a control G-CSF sample could be run in parallel to more than one formulated G-CSF. The

location of the relevant deuteration source was indicated in HDx Director Version 1.0.3.9 software through the application of a customised script (Leap Technologies).

On-line digestion was performed using a 5 μm , 2.1 x 30 mm Enzymate BEH pepsin column (Waters, Milford, Massachusetts, USA) at 25 °C. Flow rate was 100 $\mu\text{L}/\text{min}$ (mobile phase 0.05% v/v formic acid). Proteolytic peptides were trapped using an ACE C18 guard cartridge (5 μm , 2.1 mm i.d.), desalted for 3.75 min and sequentially chromatographically separated on an ACE Excel Super C18, (2 μm , 2.1 x 100 mm i.d, both Hichrom, Reading, U.K), both trap and column were held at 0 °C. Chromatographic separation was carried out at 100 $\mu\text{L}/\text{min}$ by application of a 7 min linear gradient from 92% A / 8% B to 65% A / 35% B. Mobile phases consisted of aqueous, 0.1% v/v formic acid (A) and ACN, 0.1% v/v formic (B). MS experiments were performed on a Synapt G2Si ESI-Q-TOF-MS instrument (Waters, Milford, Massachusetts, USA), as described previously [56]. Further details of the MS conditions are provided in the supplementary information.

Data analysis

ProteinLynx Global Server software v 3.02 (PLGS, Waters, Milford, Massachusetts, USA) was used to generate peak lists by inputting MS^E data. The PLGS outputs were imported into DynamX v3.0 [Waters, Milford, Massachusetts, USA] to generate peptide maps. To ensure only peptides observed reproducibly across several experiments were included, a file threshold (peptides observed in x files out of y $n=x/y$) of $n=4/5$ was set for generation of the HDX-MS peptide map. Full details of PLGS and DynamX processing conditions are provided in the supplementary information.

Deuterium exchange measurements were analysed using DynamX and all data were manually validated and curated following the approach described previously [56]. All comparisons were made relatively to a GCSF control state run in parallel, so no corrections were made for back exchange.

The IRP was analysed alongside G-CSF peptides using a retention time of 6.31 min, amino acid sequence “PPPI” and a maximum uptake of 1.0 Da. The differential between control and experimental relative uptake was calculated to obtain correction values for GCSF differential data:

$$\Delta D_t = (M_{exp,t} - M_{control,t}) - (P_{exp,t} - P_{control,t}) \quad \text{Eq. 4}$$

where $M_{exp,t}$ and $M_{control,t}$ represent the mean of triplicate measurements of the deuterium uptake at time t for the experimental and control samples, respectively, and $P_{exp,t}$ and $P_{control,t}$ represent the intrinsic HDX rate determined for the spiked IRP, present in the samples, also at time t .

The labelling times of G-CSF at pH 4.25 were corrected using equation 5 [57] and as shown in Table S2 (supporting information):

$$k_{ch} \sim k_{OH}[OH^-] = A \exp\left(\frac{E_a}{RT}\right)[OH^-] \quad \text{Eq. 5}$$

where k_{ch} is the calculated intrinsic rate, A is the frequency factor, E_a is the activation energy of the base-catalysed amide hydrogen exchange reaction at 17 kcal/mol [58], and $k_{OH}[OH^-]$ is the base-catalysed reaction, which increases at higher pH (> 2.5). Values for the base-catalysed reaction were obtained from polyalanine HDX studies performed by using $^1\text{H-NMR}$ [58].

iGEMDOCK docking

The G-CSF structure was extracted from PDBID:2D9Q [59] to remove the two G-CSF receptor structures, then protonated at pH 4 using the online server www.pbd2pqr.org and applying it to the structure in PyMOL (Schrodinger, LLC, New York, NY, USA), using the Adaptive Poisson-Boltzmann Solver (APBS) plugin tool. 3D excipient molecules were drawn in Maestro 10.3 (Schrodinger, LLC, New York, NY, USA) as described previously [46], at pH 4, using pKa data from www.chemicalize.org. A short minimisation protocol was performed using MacroModel 10.0 (Schrodinger, LLC, New York, NY, USA) and the final structures saved as MOL files. The flexible docking software Generic Evolutionary Method for Molecular DOCKing (iGEMDOCK) [60] was used to scan for excipient interaction regions across the G-CSF 3D structure. Accurate docking was selected as the default setting (population size 800, generations 80, and number of solutions 10). The 10 best-pose solution files were visualised using Discovery Studio 4.0 (Biovia, San Diego, CA, USA). Clustering of poses was analysed using the “set interaction and atom composition” drop down with 4 interaction clusters and 4 atom composition clusters.

RESULTS AND DISCUSSION

Impact of formulation equilibration time on T_m and T_{agg}

The T_m -values for thermal denaturation of 0.5 mg/mL G-CSF in sodium acetate pH 4.25, were found to be only marginally lower in 10 mM acetate buffer, compared to in 50 mM and 100 mM buffer (Figure S1, supplementary information). Further T_m -values, along with the associated T_{agg} , were determined after various pre-equilibration

times upon mixing G-CSF to 0.5 mg/mL, with buffer alone (10 mM sodium acetate pH 4.25), or with buffered excipient, under the same temperature and excipient concentration conditions to be used in the HDX-MS labelling experiments discussed below. Using 1:1 mixing of excipient stock into G-CSF samples, solubility limited the final mannitol and sucrose concentration to 8% (w/v) (234 mM). A high temperature ramp rate of 2 °C/min, scanned from 30 to 90 °C, with a measurement interval of 5 °C, ensured that the thermal transitions were complete within 10-15 mins, to minimise the impact of assay time on the incubation time study. Samples were pre-equilibrated at 21±2 °C for 5 mins, 1 hr and 3 hrs prior to measurement, but staggered so that all samples were analysed simultaneously.

For all samples, with or without the excipient, the intrinsic fluorescence and SLS showed a distinct difference between the 5 min incubated sample and the two longer incubation time points (Figure S2, supplementary information). In all cases, the 5 min sample had higher values of barycentric mean fluorescence relative to the 1 hr and 3 hr samples. Similarly, the SLS profiles at 5 min produced lower scattering intensities. This highlighted an equilibration effect upon mixing, whereby at least 1-3 hours was required to reach equilibrium. An incubation time of 1 hour was taken forward into the more accurate analysis of the impact of excipient concentration on T_m and T_{agg} .

Effect of excipients on G-CSF stability as measured by T_m and T_{agg}

Short term stability measurements such as T_m have been previously shown to correlate with real-time shelf-life results for G-CSF, over a modest range of buffer and excipient conditions [8]. More recently, this correlation was found to decrease as the

range of excipients and protein variants was widened, particularly when including excipients such as Tween 80 which also altered the colloidal stability of G-CSF [10]. This type of correlation with T_m is expected to be lost entirely for low-temperature shelf-life conditions, where the native-state population predominates [9]. Nevertheless, T_m and T_{agg} studies are useful for first-pass screening of many formulations, which aim to eliminate the least stable options, or when formulating for storage in the absence of a cold chain where ambient temperatures may extend up to 40 °C.

It was previously found that mannitol, trehalose, sorbitol and sucrose led to increases in the T_m of G-CSF, particularly for trehalose and sorbitol [61]. To create a self-consistent dataset under identical buffer conditions, we evaluated the impact of a range of excipients on T_m and T_{agg} . Excipients were selected based on previous use in G-CSF formulation studies, while spanning a range of excipient classes. The basic amino-acid arginine appears in patented G-CSF formulations [62], and is often found to suppress aggregation and solution viscosity [63, 35]. Mannitol can be used as a tonicity agent [64], and is an isomer of sorbitol which is included in the commercial formulation of G-CSF (Neupogen®) at a concentration of 50 mM [65]. Both mannitol and sorbitol were selected for study to determine any differences between the two. Hydrophobic amino-acid phenylalanine appears in a G-CSF aqueous formulation patent [66]. Sucrose is a widely used sugar in biopharmaceutical formulations, and was found to inhibit G-CSF aggregation under physiological conditions in the liquid state [39, 40]. Trehalose was also included as an alternative sugar to sucrose [67].

Detergents, such as polysorbate 80, were not included as they are difficult to model computationally, and were not previously found to interact with G-CSF or change any higher order structure, as measured by ^{15}N -NMR [41].

Each excipient was added at three different concentrations ranging from 12.5-200 mM, while all included the 50 mM condition. These were all within the ranges found in other biopharmaceutical formulations. Two different G-CSF concentrations of 1 mg/mL and 0.5 mg/mL in 10 mM sodium acetate, pH 4.25 were formulated initially to assess whether the protein concentration affected the relative impact of the excipients. Samples were pre-equilibrated in the buffers for 1 hour, and then for T_m and T_{agg} measurement, a more typical temperature ramp rate of 1 °C/min was used, with measurements taken every 1 °C. As expected, the increased ramp rate led to a decrease in both T_m and T_{agg} compared to the initial scans above carried out at 2 °C/min, because aggregation kinetics become apparent earlier when using a slower ramp rate. Increasing the concentration of G-CSF from 0.5 to 1.0 mg/mL was also found to decrease T_{agg} as expected, and by an average of $1.2\text{ °C} \pm 0.4\text{ °C}$, but the rank order of excipients as measured by T_{agg} , remained unchanged, and so we focussed further analysis on the 0.5 mg/mL conditions.

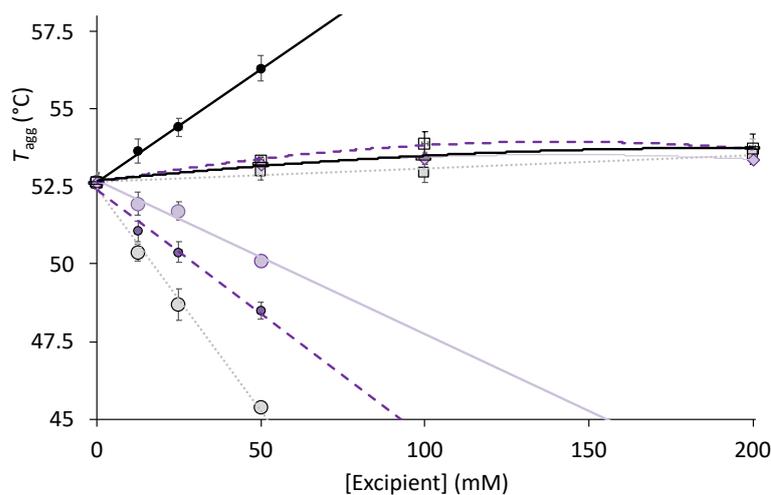
The SLS and FLI profiles for 0.5 mg/mL G-CSF with varying types and concentrations of excipient compared to the control are shown in Figure S3 (Supplementary Information), while the excipient-concentration dependence of the extracted T_{agg} and T_m values are shown in Figure 1. T_{agg} and T_m values, at 50 mM of each excipient, are also shown in Table S3 (Supplementary information) for direct

comparison. T_{agg} was linearly dependent upon the excipient concentrations up to 50 mM at least. The amino acids at 50 mM led to an increase in T_{agg} of 3.7 °C for phenylalanine, but a decrease in T_{agg} of 7.3 ± 0.3 °C, 4.2 ± 0.4 °C and 2.5 ± 0.3 °C for arginine, histidine and glutamic acid, respectively. The sugars, mannitol, sorbitol, sucrose, and trehalose at 50 mM, had no significant effects, or modest increases, in T_{agg} of 0.35 ± 0.4 °C, 0.55 ± 0.3 °C, 0.7 ± 0.3 °C and 0.6 ± 0.35 °C, respectively. The impact of these sugars appeared to plateau at between 100 mM and 200 mM, with T_{agg} increased respectively by 0.9 ± 0.4 °C, 0.7 ± 0.3 °C, 1.1 ± 0.55 °C and 1.1 ± 0.4 °C, at 200 mM mannitol, sorbitol, sucrose, and trehalose. Phenylalanine was unusual in that it not only increased T_{agg} , but also decreased the scattering intensity until it had almost disappeared at the highest concentration, although with a relatively large standard deviation. The behaviour of arginine was different to that previously observed in many studies, particularly for monoclonal antibodies at pH 5.5-8, where it is typically stabilising against aggregation, leading to an increase in T_{agg} . The current study for G-CSF is at pH 4.25, and so the positively charged arginine excipient may interact less with glutamic acids on the protein surface, than at higher pH, as their pKa distribution will also be centred at around 4.25. Also, the protein, which has a pI of 6.1, is already highly positively charged at pH 4.25, so adding more positive charge through the binding of arginine, would be expected to have less impact on colloidal stability.

The impact of excipient concentration on T_m could be grouped by excipient type. For all amino acids studied except arginine, T_m increased initially, reached an optimum at 12.5 or 25 mM, and then decreased again at 25 or 50 mM. Arginine had only a

minimal increase in T_m initially, before decreasing. By contrast, the sugars increased T_m to a plateau at typically 100 mM and above. Except for one condition, $T_m > T_{agg}$, implying that the native or near-native protein interacts and aggregates before any global unfolding becomes significant, consistent with previous observations [10]. The fraction unfolded at T_{agg} ($f_{T_{agg}}$), was determined to be between $10^{-5}\%$ and 0.5% over all these conditions at pH 4.25, consistent with previous observations at 0.2 mg/mL that aggregation at $\leq T_{agg}$, was dominated by a bimolecular diffusion-limited aggregation from a native-like state N^* [10, 39].

A



B

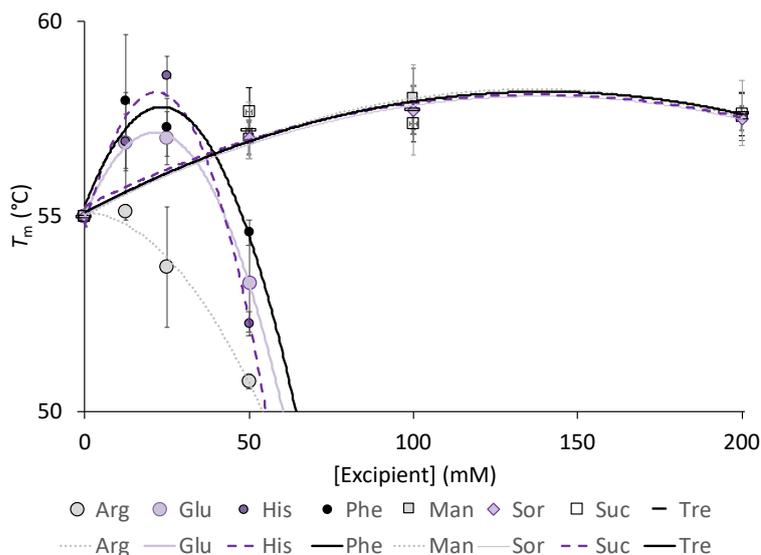


Figure 1 Impact of excipients on aggregation and thermal transition-midpoint temperatures. G-CSF at 0.5 mg/ml in 10 mM sodium acetate, pH 4.25, was subjected to a thermal ramp from 20 to 90 °C at a rate of 1 °C/min. Effect of increasing excipient concentrations on A) T_{agg} as measured by SLS. B) T_m as measured by FLI. Each formulation was made in triplicate and their data averaged. Trendlines are indicative only, obtained as linear or second-order polynomial fits.

The distinctly different behaviours of the amino acids compared to the sugars suggests that there is at least one fundamental difference in mechanism by which the two excipient types influence the stability of G-CSF at pH 4.25. Phenylalanine presents one further difference from the other amino acids in that the T_{agg} increased with its addition. T_m initially increased but then decreased at above 12.5 mM, and so at 50 mM phenylalanine, T_{agg} became greater T_m . At the same time, the maximum light scattered decreased almost to zero at 50 mM phenylalanine. Therefore, the

higher concentration of phenylalanine began to promote unfolding of G-CSF, but also suppressed aggregation, either through a general solubilising effect, or at least through selective solubilisation of the unfolded state.

Given that the sugars were all stabilising, as measured by both T_m and T_{agg} , these provided the main focus for further analysis by docking and HDX. Arginine and phenylalanine were also retained to represent the two different behaviours from amino acids, with arginine destabilising to both T_m and T_{agg} , and phenylalanine stabilising to both up to 12.5 mM, but then destabilising to T_m at 50 mM.

***In silico* excipient docking**

The use of computational predictions for stabilising excipients could potentially alleviate the formulation-screening burden during early-phase development, by reducing the number of candidates in a short space of time. Docking software is primarily used to determine enzyme-substrate and protein-ligand interactions [60]. However, it has recently been applied to biopharmaceutical formulation design [68, 46].

Of the eight available crystal structures for G-CSF in the Protein Data Bank (PDB), only two (PDB: 2D29Q and 1CD9) had the same primary sequence as our human G-CSF, and no unresolved loop sections. Both had the same 2.8 Å resolution, but the crystallographic B-factors of 2D9Q correlated slightly better to our HDX-MS measurements described below ($R^2 = 0.4$), than did those of 1CD9 ($R^2 = 0.3$) (Figure

S4, supplementary information). Therefore, 2D9Q was selected as the most appropriate for *in silico* docking.

The four sugars and two selected amino-acid excipients selected above were computationally docked onto the G-CSF surface using iGEMDOCK. This generated 10 “best docking” poses for each excipient, that were clustered by their location on the protein surface. The binding energies for each excipient pose, and within each cluster, were averaged as shown in Table S4 (Supplementary information) where the lower values indicate stronger interactions between protein and excipient. Sucrose, followed closely by trehalose, had the lowest (strongest) total binding energy and sorbitol was found to have the highest (weakest) binding energy.

iGEMDOCK provided an in-depth analysis of the G-CSF residues interacting with each excipient as detailed in Figure S5 (Supplementary Information). These were also clustered, and as shown in Figure 2, two hotspots were identified for excipient docking onto G-CSF. All excipients docked onto the first hotspot, located at the C-terminus and the start of loop AB2. Arginine, phenylalanine and mannitol only docked into this first hotspot. However, sorbitol, sucrose and trehalose also docked into a second hotspot, located at the surface of the short helix, the end of loop ABII and the start of helix α A. Similar hotspots for excipient binding were identified previously by docking excipients into the antibody fragment Fab A33, and into the *Drosophila* Su(dx) protein (WW34) [46], which suggests that there are particular features of protein surfaces that have a greater propensity to form interactions with excipients. These are likely to be small cavities or pockets that enable a greater number of interactions with the

excipients. Both of the hotspots identified in G-CSF were located in surface pockets, with the first hotspot more deeply enclosed than the second. As the software ranked and sorted all docked conformations into the top 10 most energetically favourable conformations, it does not rule out the presence of weaker interactions at other sites on the G-CSF surface. In addition, many excipients such as the sugars, are also known to stabilise proteins at least in part through preferential exclusion from the protein surface, which does not involve direct interactions between protein and excipients. However, preferential exclusion of excipient molecules from the majority of the protein surface does not rule out a small number of preferential stabilizing or destabilizing interactions with the protein surface, which could have a major influence on overall conformational stability or aggregation propensity.

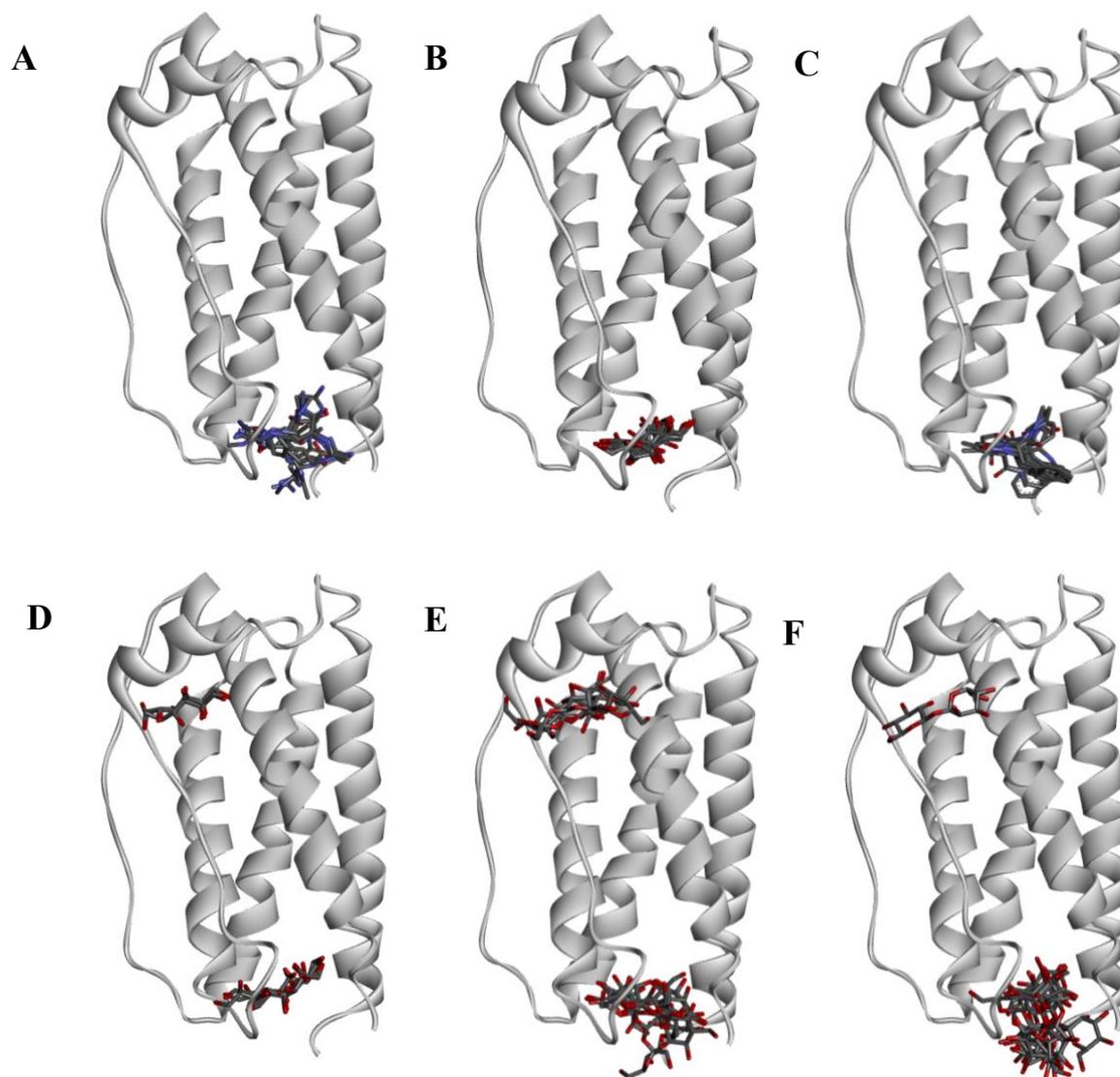


Figure 2 Docking of excipients onto G-CSF. The 10 best poses are displayed docked onto G-CSF from PDBID: 2D9Q, for each excipient: A) Arginine, B) Mannitol, C) Phenylalanine, D) Sorbitol, E) Sucrose, F) Trehalose. Images were created in Discovery Studio 2016 (Dassault Systemes Biovia, San Diego, CA).

Protein-excipient T_m values were found previously to correlate with iGEMDOCK total binding energy outputs, where excipients that increased the T_m also had lower predicted total binding energies [46]. Here we found a similar correlation between the

predicted binding energies for the sugar excipients with G-CSF, and their experimental T_m ($R^2=0.988$) values, obtained for each sugar excipient at 50 mM (Figure 3), indicating that direct interactions of sugar excipients with the protein surface may indeed play a significant role in stabilising proteins. By contrast, the range of T_{agg} values obtained with the sugars was too small relative to the measurement errors, to be able to accurately evaluate the correlation with *in silico* docking.

The two amino acids arginine and phenylalanine did not fit within the correlations obtained for the sugars, reflecting their different impacts on T_m compared to the sugars, and probable differences in mechanisms. Arginine was destabilising as often observed for this amino acid. Previous crystallography [69], molecular dynamics simulations [35], and HDX-MS [42] studies have highlighted the direct interactions of arginine with protein surfaces, that can mediate its destabilising effect. By contrast, phenylalanine was stabilising initially at 12.5 mM, but eventually destabilising to T_m at 50 mM. Interestingly, inclusion of phenylalanine at 12.5 mM where it was maximally stabilising to T_m , brought this excipient into line with the correlations obtained for the sugar excipients, giving a new R^2 of 0.88. While promising, this and the previous correlations are only indicative given that they assume a “zero” docking energy for the no-excipient case. Also, the docking models the binding of an excipient onto a static protein structure, while each excipient may also influence conformational stability (T_m) differentially through multiple mechanisms, including preferential exclusion and interaction. At present, docking does not model or predict any potential influence from global protein dynamics, solvent-mediated effects (eg. preferential exclusion), or

specific aggregation-prone regions (APRs) within the protein structure. Nevertheless, it does provide a useful tool for identifying the potential preferential excipient interactions, that could be exploited as one mechanism for stabilising proteins.

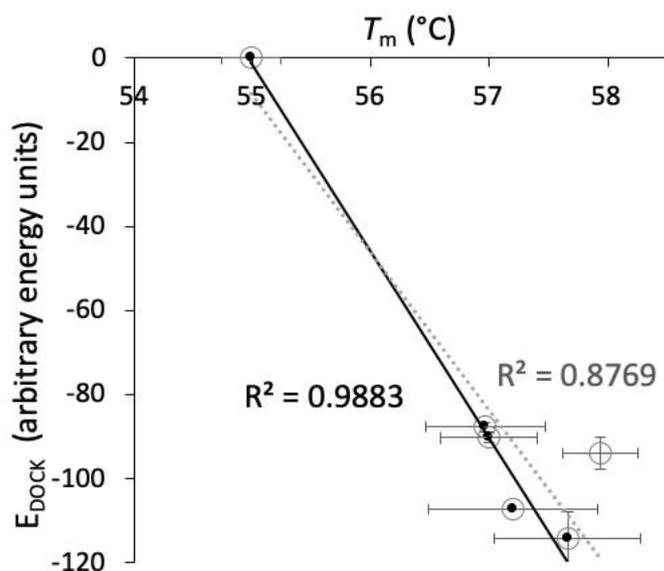


Figure 3 Correlation between docking total energy (E_{Dock}) and the experimental thermal transition mid-point T_m . Experimental data were obtained at 0.5 mg/ml G-CSF, and with (●) 50 mM of sugar excipient formulations only (— linear fit), and (○) including the 12.5 mM phenylalanine formulation (⋯ linear fit).

Hydrogen-deuterium exchange mass spectrometry (HDX-MS)

To validate the predictions of binding locations, HDX-MS analysis was carried out in the presence and absence of excipients at pH 4.25. Initially, the analysis was compared for G-CSF without excipients at pH 7.4 and pH 4.25 to optimise and validate

the approach, as well as to enable a comparison of the pH 4.25 conditions, with previous HDX for G-CSF carried out at pH 7 (Zhang 2015).

Effect of pH on G-CSF HDX-MS

The structure and dynamics of G-CSF was compared in 50 mM acetate, pH 4.25, and 10 mM PBS pH 7.4 by HDX-MS, by comparison of HDX-MS measurements of 59 peptides common to both pH 4.25 and pH 7.4 experimental datasets. As expected, due to the higher intrinsic rate of HDX at pH 7.4 compared to pH 4.25, the total relative deuterium uptake was higher for the pH 7.4 dataset (Figure 4) when considering several incubation time points. The greatest relative uptake was found in Loop CD of G-CSF, consistent with previous observation that Loop CD is the most flexible region [41].

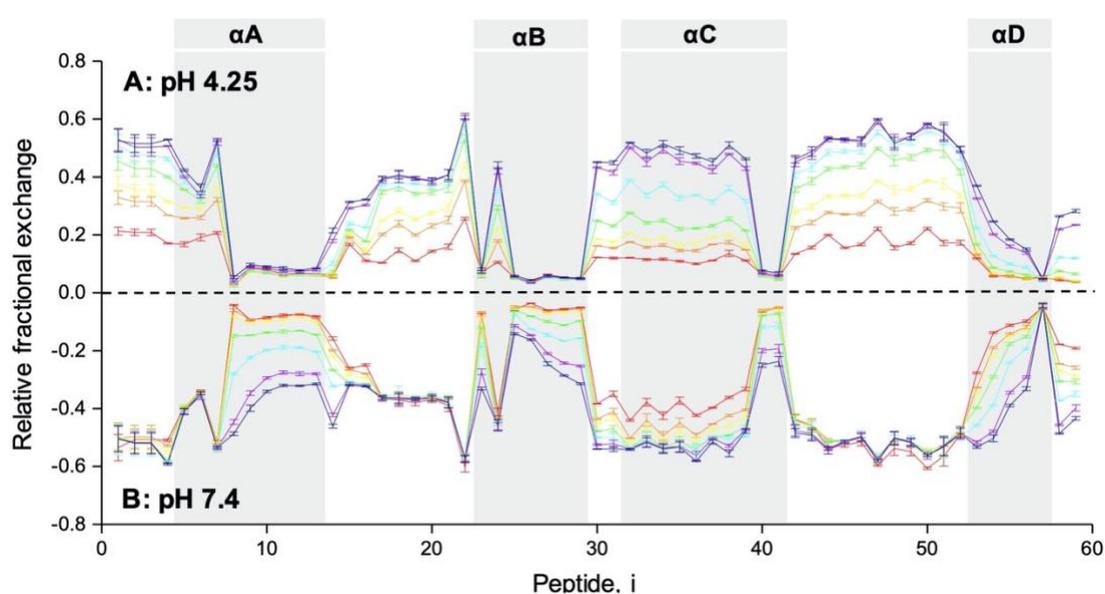


Figure 4 Mirror plot representing the relative fractional exchange for G-CSF at pH 4.25 (A) and pH 7.40 (B) for each G-CSF peptide. Measurements are an average of triplicate measurements of G-CSF in either 50 mM sodium acetate pH 4.25, or 10 mM PBS pH

7.4, following incubation in deuterium buffer for 30 seconds, 2.6, 5, 15, 60, 240, and 480 minutes and indicated by red, orange, yellow, green, cyan, violet and navy-blue lines, respectively. The x-axis labels the identified peptides 1-59, ordered according to their midpoint residue. See Table S5, supplementary information, for peptide sequences. The average relative fractional exchange was calculated as the observed deuterium level (in Da) divided by the total number of exchangeable backbone amide hydrogens. To form a mirror plot for visual comparison purposes, the y-axis values in (B) were obtained by multiplying the Relative uptake (Da) by -1. The helical regions (α A-D) of G-CSF are shaded in the background, with the connecting loop regions unshaded.

At pH 4.25 the uptake rate of peptides was slower, and therefore measurable in the loop regions of the protein, whereas the structured helical regions (α A-D) showed minimal measurable exchange across all time points, indicating significant structural protection. Conversely, for pH 7.4, the uptake rate was increased for the structural helical regions, whereas the loop regions had reached their maximum exchange level already by the first labelling time point ($t= 30$ s), as seen by an overlay of uptake at all time points in these regions (Figure 4 B). The N-terminal regions of α C and α D, and to some extent α B, were flexible at both pHs, exchanging over time to a high level of relative uptake, comparable with the loop regions, suggesting fraying of helices at the N-termini. Peptides within helices α C and α D showed higher exchange levels compared to those within the α A and α B helices, consistent with previous HDX-MS

measurements at pH 7.0 [40], and their suggestion that this could be due to stronger structural constraints on helices α B and α C.

Comparison of HDX kinetics at the two pH conditions, for the different regions of secondary structure, highlights the power of exploiting two pH solutions to increase the time window for HDX-MS, such as deuterium labelling at low pH to study unstructured fast exchanging amide hydrogens [57]. The change in protein dynamics, or K_{op} , in different solution conditions can be determined by plotting deuterium uptake curves with time window correction. If the data overlays and is continual protein dynamics can be assumed to be unchanged. Therefore, to account for the differences in intrinsic exchange rates at the two pHs, the exchange times for pH 4.25 samples were converted into corresponding exchange times for pH 7.4 at 22 °C using Equation 5 in Methods (Table S2, supplementary information). Following correction, the longest exchange time of 8 hrs at pH 4.25 was calculated to be \sim 28.8 s, similar to the shortest exchange time of 30 s at pH 7.4. Differential analysis of G-CSF peptide uptake values from the equivalent time points at pH 4.25 and pH 7.4 showed a large number of peptides with significant negative differentials ($\Delta D(t) = m_{pD\ 7.4}(t=0.5) - m_{pD\ 4.25}(t=480)$), as shown in Figure 5, suggesting less structure, or faster backbone-motion dynamics, at the lower pH.

In theory, differential uptake values of zero can come from regions where the G-CSF protein dynamics did not change, or those for which exchange was complete before the shortest time point at both pH. Regions for which there were significant negative differentials in exchange included peptides 15-17 in the first half of Loop AB (including

the short helix), peptides 33-36 covering loop BC, peptides 37-43 in the first half of α C, peptides 49 and 50 at the boundary of α C in loop CD, peptides 63, 67 and 68 spanning α D, and the final two peptides of the C-terminus. The negative differentials indicated that the pH 4.25 sample had a higher relative uptake of deuterium, and therefore that the dynamics of G-CSF in these regions had increased. Modest positive differentials for some peptides of the loop CD region, indicated that this region may be slightly more stable at pH 4.25. The increased dynamics at pH 4.25 was surprising given previous comparisons of G-CSF by CD, which indicated a slight decrease in helical content at pH 7.0 (56%) compared to at pH 4.0 (61%) [49], and where the T_m is typically 4-6 °C higher at pH 4 than at pH 7 [10, 49]. Previous NMR studies revealed only minor conformational shifts between high and low pH, through re-orientation of the α C helix [41], consistent with the main differences in uptake rates being located in this region and in loops BC and CD, but also suggesting that while this formed new stabilising contacts observed in the NMR study, it also leads to increased flexibility in the α C helix itself.

Some caution must be taken in the above comparison, as the extrapolation of kinetics between the two pH values potentially introduced a systematic error which affects the interpretation of their relative exchange kinetics. An additional key difference could be that measurements of T_m are convoluted with aggregation kinetics, and the faster kinetics at pH 7 appear to be dependent on additional factors to differences in conformational stability [10]. Furthermore, G-CSF is known to aggregate slowly but extensively at above pH 6.2 [41, 50], but not in solutions at less than pH 5 [47]. The

presence of any aggregates at pH 7.4 would be protective, and so potentially also decrease the relative uptake of deuterium in the HDX-MS analysis.

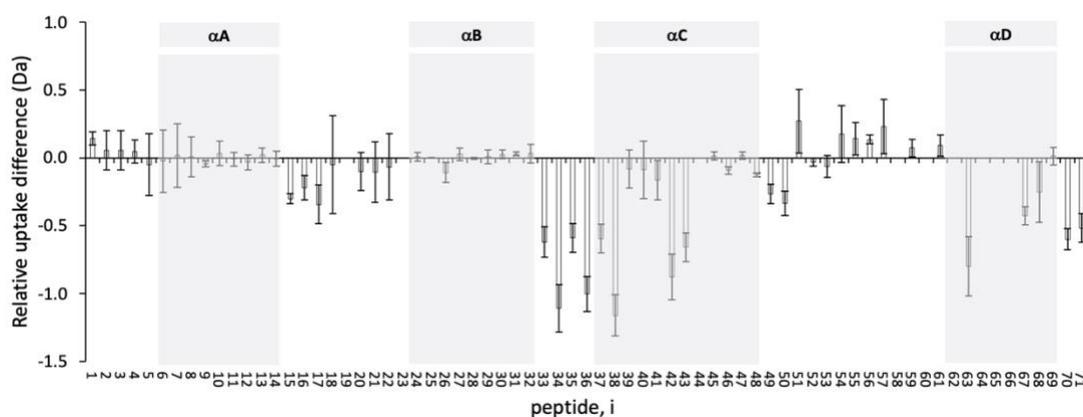


Figure 5 Differential HDX-MS plots of G-CSF peptide uptake in differing pH solutions, corrected to a standard exchange time, as measured by HDX-MS. The relative uptake values were taken from G-CSF peptide-level HDX-MS with pH 4.25 deuterium solution with 8 hrs exchange time and pH 7.4 deuterium solution with 30 s exchange time. The y-axis denotes $\Delta D(t) = m_{\text{pH } 7.4}(t=0.5) - m_{\text{pH } 4.25}(t=480)$, where m denotes the mass of the peptide as a function of deuterium exposure time, t . Each figure is an independent HDX-MS experiment comparing the two pH solutions as shown by the run order displayed in the dark grey box in the top left-hand corner of each figure. The x-axis labels the identified peptides 1-71 of G-CSF, ordered according to their midpoint residue. The different helical regions of G-CSF are coloured in the background. The non-coloured regions represent the connecting loop regions. G-CSF peptide sequences, residue numbers and locations are shown in the supplementary information (Figure S5).

Effect of deuterium-exchanged excipients on G-CSF HDX-MS

Comparisons of T_m values for formulations above, were carried out at pH 4, where the deuterium uptake kinetics measured by HDX-MS were unlikely to have been affected by aggregation, compared to at pH 7.4. Also, to precisely match the formulations above, the lower buffer concentration of 10 mM sodium acetate, pH 4.25, was used in further HDX-MS measurements. As noted earlier, the T_m in 10 mM sodium acetate, pH 4.25, was only marginally lower than in 50 mM acetate (Figure S1 in supplementary information). With this alteration, a peptide map was generated from a GCSF sample in 10 mM sodium acetate, pH 4.25, containing 67 peptides. The relative deuterium uptake at seven different labelling time points for each of these peptides (Figure S6 in supplementary information) showed a similar profile to the one at 50 mM acetate, pH 4.25 (Figure 4 A), with a general increase over time, for most regions of G-CSF, apart from the solvent inaccessible hydrophobic-core forming regions of α -helices, which underwent little or minimal exchange.

To shed light on protein-excipient interactions experimentally, we compared the relative peptide deuterium uptake rates for G-CSF incubated in 10 mM sodium acetate, pH 4.25, with and without the deuterium-exchanged excipients. Arginine, mannitol, sucrose and phenylalanine were selected for HDX-MS experiments to represent different classes and behaviours of excipients: amino acids (arginine and phenylalanine), saccharides (sucrose) and sugar alcohols (mannitol), and the full range of T_m observed. Sorbitol was not selected as it formed a gel in the bottom of the vial during the freeze-drying step required for deuteration. Sucrose was selected over

trehalose because, while they had similar T_{agg} values, sucrose was studied previously with G-CSF at pH 7.0, by HDX-MS [40], allowing for a direct comparison with published data. Little formulation data has been published previously for phenylalanine, and so this work would provide novel insights into its suppression of G-CSF aggregation. The final excipient concentrations of 10 % (w/v) (549 mM) mannitol, 50 mM arginine, 10 % (w/v) (292 mM) sucrose, and 50 mM phenylalanine, were chosen to maximise the potential to observe structural effects on exchange rates.

Relative uptake was compared by sampling and analysis of fast exchange (2.6 min), mid exchange (1 hr) and slow exchange (8 hrs) conditions. A typical G-CSF peptide map generated with a deuterium-exchanged excipient, is shown in Figure S7 (supplementary information), and achieved sequence 96% coverage and a redundancy of 3.81. The internal reference peptide (IRP) with sequence PPPI was included at 3 μ M in G-CSF samples, prior to the labelling step, as described previously [40], to monitor and correct for any changes to the intrinsic rate of uptake caused by differences in formulations. The IRP was found to have no effect on T_m or T_{agg} values determined for 0.3 mg/mL G-CSF in 10 mM sodium acetate pH 4.25, indicating that it did not interact with the G-CSF or affect its stability (Figure S8 supplementary information). The relative uptake for the IRP was measured alongside those for the G-CSF peptides in the control without excipients, and in the G-CSF containing deuterium-exchanged excipients, as shown in Table S6 (supplementary information). The differential uptake for each peptide was obtained using Equation 4, by measuring the difference in deuterium uptake between the excipient and control G-CSF samples, and correcting for

any difference in intrinsic uptake, using the differential uptake between the IRP in excipient and control G-CSF samples. Any differences in intrinsic exchange for the IRP between excipient solutions and the control was either negligible or very small, suggesting that the excipients had been fully deuterium-exchanged.

The impact of exchange time, ranging from 2.6 min to 8 hrs, on the IRP-adjusted differential uptake plots for each excipient, is shown in Figure S9 (supplementary information), where a positive differential indicates higher deuterium uptake in the presence of excipient. Under fast exchange (2.6 min), the majority of the G-CSF peptides had positive differential uptake values with mannitol, phenylalanine and sucrose. This indicated that the protein had higher exchange, and hence an initial loss of structure in the presence of these excipients, than in their absence at pH 4.25. As exchange continued with the same three excipients, the majority of the G-CSF peptides shifted to negative differential uptake values, indicating that after 8 hrs, the protein had lower exchange, and increased structure, in the presence of these excipients. At mid exchange (1 hr), differential uptake was mid-way between the fast and slow exchange data sets, with values ~ 0 Da (no difference) for mannitol and sucrose, and with a weak negative differential for phenylalanine. Therefore, these excipients gave a clear transition from high to low deuterium uptake over time, relative to the control. By contrast, arginine gave a negative differential uptake value across the majority of the G-CSF peptides for all time points. The behaviour of mannitol, phenylalanine and sucrose, indicated that during the initial equilibration, these excipients promoted exchange in G-CSF peptides, perhaps as the excipients had not yet interacted directly

with the G-CSF surface. Over time, these excipients became more stabilising to G-CSF, as indicated by the decreased exchange relative to G-CSF in the absence of excipient. The long equilibration time suggested a slow binding kinetic (k_{on}) for the formation of protein-excipient interactions, perhaps due to the requirement for a relatively rare protein conformation in the bound state. By contrast, arginine led to protection from exchange more rapidly, initially at a few specific sites, but then more globally after 8 hrs. Overall, the time-dependence of the HDX was consistent with that observed also for the thermal unfolding profiles (Figure S2), which showed that equilibration with excipients was achieved only after at least 1 hour.

Alignment of HDX protection with predicted binding sites

The exchange data at 8 hrs, showed the most established interactions with excipient, and were compared to the iGEMDOCK residue-interaction predictions. The peptides containing the residues predicted as hotspots by iGEMDOCK are highlighted in the differential uptake plots after 8 hours of exchange in Figure 6. In all cases, the peptides for which iGEMDOCK predicted residue-level interactions, were highly correlated to sites with the largest negative differential uptake values. Therefore, sites predicted to bind to excipients were also the sites for which the protein had a lower exchange in the presence of the excipients, notably in loop AB, helix α D, and loop D (C-terminus). Most differential uptake values within hotspots decreased, indicating protection due to binding of the excipients. However, for sucrose there were also 3 peptides with significant positive differentials located within the hotspot involving the C-terminal residues 142-175, indicating that specific ligand interactions can also be locally

destabilising, perhaps due to localised shifts in structure that expose backbone amides in compensation while others are protected. The same effect appears to have been induced by phenylalanine, even it was not predicted to interact directly with these neighbouring residues. The effects of each excipient on the peptides in this region had very different kinetic trends, suggesting that the local structure was sensitive to small differences in binding to this region. Most of the G-CSF peptides outside of the hotspots showed small decreases in differential uptake. These smaller effects in HDX are most likely due to an additional global protection mechanism such as from the preferential exclusion of excipient.

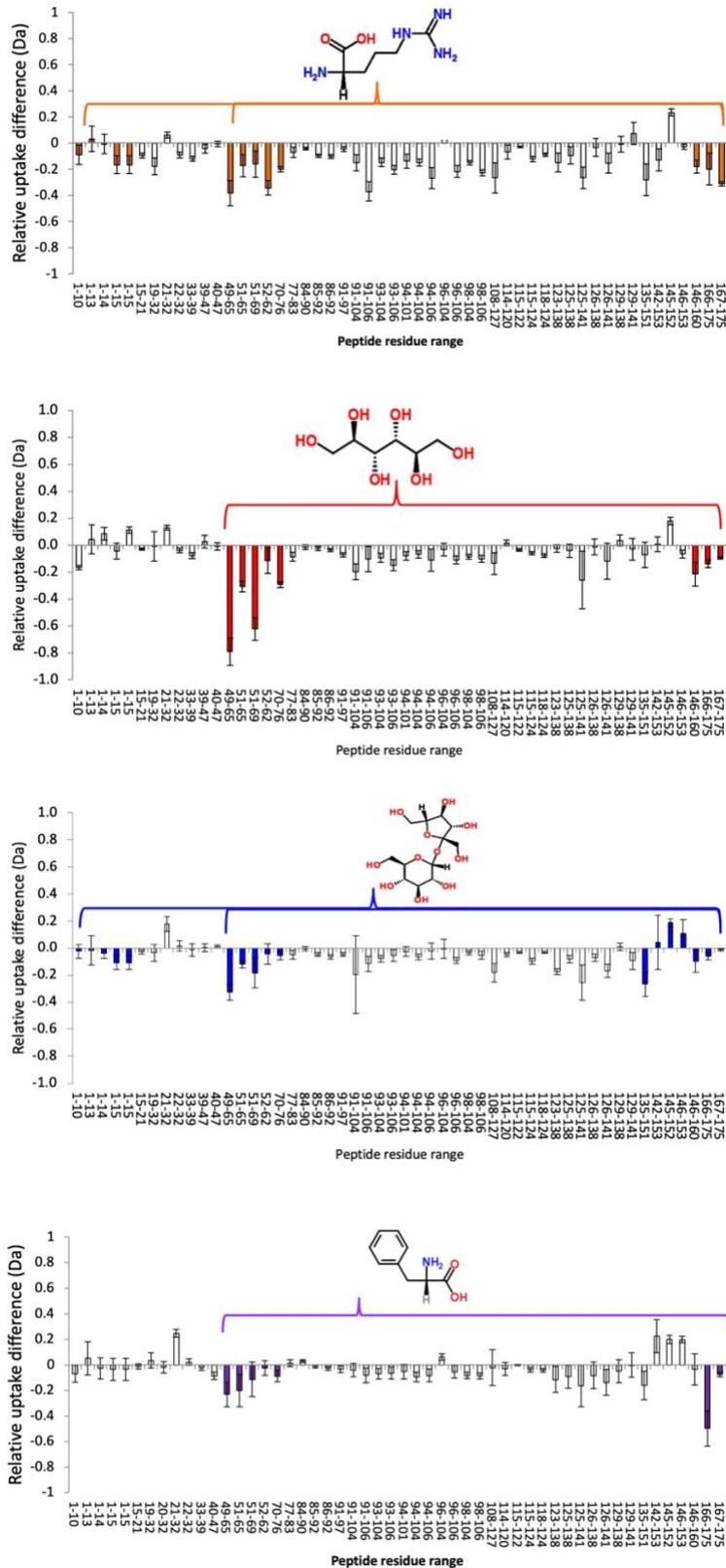


Figure 6 Alignment of HDX-MS differential uptake with in silico docking data for G-CSF and excipients. Differential uptake values ($\Delta D(t) = m_{\text{excipient}} - m_{\text{no excipient}}$) for G-CSF

peptides calculated, from HDX-MS measurements of G-CSF incubated with deuterium-exchanged excipients minus measurements in a purely deuterated buffer, following an 8 hr incubation. Peptides containing iGEMDOCK docked residues are highlighted in A) Orange for arginine; B) Red for mannitol; C) Blue for sucrose; D) Purple for phenylalanine. See Table S7, supplementary information, for peptide sequences.

In addition to the main hotspot spanning residues 47-71 and 152-174 (loop AB, helix α D, and loop D), sucrose and arginine both slightly protected the peptides at residues 1-15 (loop A), consistent with their second hotspot predicted in that location. By contrast, mannitol and phenylalanine were not predicted to bind in that location, and accordingly did not protect these regions from exchange.

At pH 4.25 in the absence of excipient, we observed no exchange in some helical region peptides including α A (residues 18-36), half of α B (res 80-89), and the last few residues of α C, and so any protection by sucrose would be unobservable in these peptides. However, this did not account for the absence of protection by excipients in most peptides outside of the hotspots. For example, residues 77-153 with mannitol, sucrose and phenylalanine, were largely unaffected by the presence of excipient, and yet these spanned residues 91-141 (in Loop BC, α C, Loop CD) for which exchange kinetics were mostly observable at pH 4.25 (Figure 4). Our result thus contrasts with a previous HDX-MS study with G-CSF at pH 7.0, that showed only global protection by

1 M sucrose [40]. This was concluded from a fast exchange study at 37 °C, and a slow exchange study at 4 °C that showed that all regions were protected by sucrose in at least one condition. However, a lower concentration of sucrose (292 mM sucrose) and also the lower pH 4.25, were used in our study. The study at the higher pH 7.0 would have had a convoluting influence from aggregate formation at 37 °C, and a stronger preferential exclusion effect at the higher sucrose concentration. Furthermore, it is very possible that sucrose simply does not bind preferentially to the hotspots at pH 7, but does at pH 4.25.

Docking for arginine identified one single docked pose also highlighted in Figure 2. However, from the deuterium-exchanged excipient differential measurements, it appears that arginine provides much greater global protection from deuterium uptake than the other excipients. Thus, arginine shows a weaker match between the peptides with the largest differences in uptake, and the residues predicted by excipient docking. Given that arginine is destabilising to T_m , and is known to be mildly chaotropic, it appears that arginine was able to bind to G-CSF in such a way as to be protecting under HDX conditions at 22 °C, and yet destabilising thermodynamically to the global structure as the mid-point transition temperature is approached. In other words, arginine binds weakly to many sites over the protein surface and protects the amides from the solvent, while also then weakening the structure as the temperature is increased.

Equilibrium dialysis and vapour pressure osmometry (VPO) previously found that <0.5 M arginine was neither strongly bound nor excluded from protein surfaces,

whereas >0.5 M arginine became increasingly excluded, in studies with proteins at physiological pH [63]. This indicated that the protein surface was initially saturated with arginine, which then excluded further arginine from interacting with the surface. Later molecular dynamics simulations on lysozyme and ovalbumin at physiological pH, revealed the formation of Arginine cation clusters at specific loci on the protein surface [35]. The low pH, in our G-CSF formulation is likely to have altered the binding of arginine with protein surfaces, which at physiological pH was primarily through simultaneous interactions between i) the arginine carboxylate and protonated amine or guanidinium groups of the protein, and ii) arginine guanidinium and carboxylate groups of the protein. At pH 4.25 in our study, the carboxylate groups of the protein and also of the arginine excipient, will be largely neutralised, potentially disrupting the cluster formation of arginine. Under these conditions, the low arginine concentration used (0.05 M) was likely to have enabled many weak interactions with the protein surface, but not yet leading to preferential exclusion effects or protein denaturation. While, these conditions decreased the T_m of G-CSF by 4.2 °C to 51 °C, little impact on protein flexibility and unfolding would actually be expected under the HDX-MS conditions of 22 °C. Therefore, 0.05 M arginine can decrease the T_{agg} for heat-induced aggregation at 50 °C, without revealing any destabilising effects by HDX-MS at 22 °C.

Comparison of the effects of excipients

Overall, HDX-MS experiments confirmed that specific interaction of excipients with surface hotspots was most likely occurring in addition to a background of preferential hydration (exclusion of excipient). The stability and HDX-MS experiments both

revealed a slow equilibration after generating the protein formulation. HDX-MS measurements, revealed that apart from arginine, the excipients increased exchange globally during equilibration, but then became stabilising through specific interactions when fully equilibrated after 8 hours. By contrast, arginine was protective at all time-points, initially in specific regions of the protein corresponding to the hotspots predicted by docking. However, over time arginine became protective in some additional regions. These results could be explained by the two-step thermodynamic process for equilibration of protein with excipients [70]. In step 1, immediately after mixing protein and excipients, the excipient is located in the bulk solvent and at the protein surface in equal concentrations. Equilibration occurs in step 2 where protein-excipient interactions replace interactions between protein and water if they are thermodynamically more favourable. Otherwise, the protein is preferentially hydrated and the excipient preferentially excluded. For mannitol, phenylalanine and sucrose, HDX suggests that there is initially a thermodynamically unfavourable interaction between the protein surface and the excipients which may cause local destabilisation. During equilibration, this was reduced over time as the excipient became preferentially excluded from the protein surface, except at one or two specific hotspots for which interactions were favourable. Most other G-CSF peptides retained a low level of protection in the presence of mannitol, sucrose and phenylalanine, most likely due to compaction of the native state and reduction of protein flexibility [71]. Therefore, it is possible that both preferential exclusion and specific hotspot interactions are occurring simultaneously between excipients and G-CSF in solution.

Our results, which stem from two surface pockets on G-CSF, partly explain why the excipients studied are not always protective to HDX in other proteins. For example, mannitol has been found previously to increase the T_m of an IgG4 mAb, and yet slightly increased the average global HDX, indicating mAb destabilisation [43]. Previous studies have suggested that the main mechanism of stabilisation by mannitol is via preferential exclusion [25]. Mannitol and other polyhydric compounds, may also exhibit varying levels of preferential exclusion depending on the mannitol concentration and solution pH.

Phenylalanine also exhibited very different SLS behaviour to the control and other excipient formulations, as the scattering intensity remained very low even at the elevated temperatures of >70 °C, where the protein usually denatures and then increases the scattering due to aggregation. It appears that phenylalanine can suppress protein-protein interactions considerably, potentially through interactions with other hydrophobic amino-acid residues in proteins. For G-CSF, increasing the phenylalanine concentration gradually increased T_{agg} but also decreased the total intensity of scattering. Therefore, it is reasonable to assume that specific interactions between the excipient and G-CSF blocked protein-protein interactions at temperatures that may expose hydrophobic regions.

Preferential interaction measurements have been performed on a number of strong denaturants (chaotropes) including urea and guanidine hydrochloride, which showed weak preferential exclusion, and a high tendency to bind to proteins [72]. As a milder chaotrope, arginine is often added to minimise aggregation, and yet it lowered the T_{agg}

and T_m of G-CSF at all concentrations studied (50 - 200 mM). This destabilising effect is consistent with studies on other proteins including Fab A33 [46] and an IgG4 [43]. Recent work using density functional theory (DFT) calculations found that arginine increased the propensity for heat-induced aggregation in proteins with the highest proportion of acidic residues, with which arginine could preferentially interact [73]. Compared to the high (16.3 % and 14.6%), and low (6.4%) acidic content of proteins used in that study, G-CSF contains 12.1% acidic residues, more consistent with the high range, and thus the observed destabilisation. In contrast to the stabilising excipients which revealed specific regions of protection after 8 hours of HDX, arginine led to a clear decrease in uptake for the majority of G-CSF peptides, with no distinct regional bias, and at all labelling time points. This indicates an immediate protective effect of arginine through rapid preferential binding to G-CSF. By contrast, in previous studies 0.3-0.5 M arginine at pH 6-7.4, substantially increased the uptake for specific regions of IgG1 and IgG4 MAbs [42, 43], due to increased backbone flexibility and destabilisation.

CONCLUSION

Using G-CSF as a model protein we have shown that both HDX-MS and *in silico* molecular docking are useful tools for identifying potential sites of proteins that can interact with excipients. This provides a powerful approach to facilitate and accelerate the selection of excipients to include in formulation screens for liquid biologics. While stable G-CSF formulations were all at pH 4.25, our findings were consistent with recent

work at physiological pH in which specific interactions of excipients with proteins, have been found to play an important role in parallel with preferential exclusion effects. Interestingly, the HDX-MS revealed that while the system was still equilibrating, the excipients used often promoted amide exchange in some regions of structure, but these became stabilising through specific preferential interactions when fully equilibrated. With all excipients studied by HDX-MS, a background of global protection was also present, likely due to preferential exclusion effects. However, arginine showed a more broadly protective effect from the shortest exchange times onwards, and suggested that arginine interactions can stabilise the structure of proteins at low storage temperatures through many weak interactions, even though arginine was destabilising according to a decreased thermal transition temperature.

ASSOCIATED CONTENT

Supporting Information

Detailed mass spectrometry methods and data analysis; Effect of buffer concentration on G-CSF T_m ; Effect of equilibration time on thermal denaturation; Effect of excipient concentrations on G-CSF T_m ; Comparison of B-Factors and HDX-MS exchange rates; Average residue docking energies from IGEMDOCK; Relative fractional exchange for G-CSF peptides in 10 mM sodium acetate, pH 4.25; G-CSF peptide map from HDX-MS; Effect of IRP on G-CSF T_m ; Differential uptake plots of G-CSF with excipients; Freeze-drying cycle parameters; HDX reaction condition and exchange time corrections; Thermal denaturation parameters for GCSF with excipients; Pose-energies

for excipients docked with GCSF using iGEMDOCK; G-CSF peptide sequences in HDX-MS measurements from Fig 4; Relative deuterium uptake for IRP with excipients; G-CSF peptide sequences in HDX-MS measurements from Fig 6.

ACKNOWLEDGEMENT

We thank Dr Adrian Bristow (NIBSC) for the provision of the G-CSF expression system and for helpful advice on G-CSF purification. We also thank Peter Smith and Stephen Coales from Leap Technologies (now Trajan Scientific and Medical) for their assistance in writing a customised script for HDx Director software which facilitated using multiple deuteration sources during HDX sample handling. Finally, we thank the Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training in Emergent Macromolecular Therapies (EP/L015218/1), and the EPSRC Future Targeted Healthcare Manufacturing Hub (EP/P006485/1, EP/I033270/1) for financial support.

REFERENCE

1. Moussa, E. M.; Panchal, J. P.; Moorthy, B. S.; Blum, J. S.; Joubert, M. ; Narhi, L. O.; Topp, E, M. Immunogenicity of Therapeutic Protein Aggregates. *J Pharm Sci.* **2016**, *105*, 417-430.
2. Espargaró, A.; Castillo V.; de Groot, N. S.; Ventura, S. The in vivo and in vitro aggregation properties of globular proteins correlate with their conformational stability: The SH3 case. *J. Mol. Biol.* **2008**, *378*, 1116-1131.
3. Shukla, A. A.; Gupta, P.; Han X. Protein aggregation kinetics during Protein A chromatography: Case study for an Fc fusion protein. *J. Chrom. A.* **2007**, *1171*, 22-28.

4. Callahan, D. J.; Stanley, B.; Li, Y. Control of Protein Particle Formation During Ultrafiltration/Diafiltration Through Interfacial Protection. *J. Pharm. Sci.* **2014**, *103*, 862-869.
5. Voynov, V.; Chennamsetty, N.; Kayser, V.; Helk, B.; Trout, B. L. Predictive tools for stabilization of therapeutic proteins. *mAbs.* **2009**, *1*, 580-582
6. Grant, Y.; Matejtschuk, P.; Bird, C.; Wadhwa, M.; Dalby, P. A. Freeze drying formulation using microscale and design of experiment approaches: a case study using granulocyte colony-stimulating factor. *Biotechnol. Lett.* **2012**, *34*, 641-648.
7. Robinson, M. J.; Matejtschuk, P.; Longstaff, C.; Dalby, P. A. Selective stabilisation and destabilisation of protein domains in tissue-type plasminogen activator using formulation excipients. *Mol. Pharma.* **2019**, *16*, 744-755.
8. Maddux, N. R.; Iyer, V.; Cheng, W.; Youssef, A. M.; Joshi, S. B.; Volkin, D. B.; Ralston, J. P.; Winter, G.; Middaugh, C. R. High throughput prediction of the long-term stability of pharmaceutical macromolecules from short-term multi-instrument spectroscopic data. *J. Pharm. Sci.* **2014**, *103*, 828-839
9. Chakroun, N.; Hilton, D.; Ahmad, S.; Platt, G.; Dalby, P. A. Mapping the aggregation kinetics of a therapeutic antibody fragment. *Mol. Pharma.* **2016**, *13*, 307-319.
10. Robinson M. J.; Matejtschuk, P.; Bristow, A. F.; Dalby, P. A. T_m -values and unfolded fraction can predict aggregation rates for GCSF variant formulations, but not under predominantly native conditions. *Mol. Pharma.* **2018**, *15*, 256-267.
11. Dobson J.; Kumar, A.; Willis, L. F.; Tuma, R.; Higazi, D. R.; Turner, R.; Lowe, D. C.; Ashcroft, A. E.; Radford, S. E.; Kapur, N.; Brockwell, D. J. Inducing protein aggregation by extensional flow. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4673-4678.
12. Authelin, J. R.; Rodrigues, M. A.; Tchessalov, S.; Singh, S. K.; McCoy, T.; Wang, S.; Shalaev, E. Freezing of biologicals revisited: Scale, stability, excipients, and degradation stresses. *J. Pharma. Sci.* **2020**, *109*, 44-61.
13. Wöll, A.K.; Hubbuch, J. Investigation of the reversibility of freeze/thaw stress-induced protein instability using heat cycling as a function of different cryoprotectants. *Bioproc. Biosys. Eng.* **2020**, *43*, 1309-1327.
14. Philo, J. S.; Arakawa, T. Mechanisms of protein aggregation. *Curr. Pharm. Biotechnol.* **2009**, *10*, 348-351.
15. Wang, W.; Nema, S.; Teagarden, D. Protein aggregation-pathways and influencing factors. *Int. J. Pharm.* **2010**, *390*, 89-99.
16. Kim, N.; Remmele Jr, R. L.; Liu, D.; Razinkov, V. I.; Fernandez, E. J.; Roberts, C. J. Aggregation of anti-streptavidin immunoglobulin gamma-1 involve Fab unfolding and competing growth pathways mediated by pH and salt concentration. *Biophys.*

Chem. **2013**, *172*, 26-36.

17. Davis-Searles, P.R.; Saunders, A.J.; Erie, D.A.; Winzor, D.J.; Pielak, G.J. Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 271-306.

18. Patel, C.; Noble, S. M.; Weatherly, G. T.; Tripathy, A.; Winzor, D. J.; Pielak, G. J. Effects of molecular crowding by saccharides on α -chymotrypsin dimerization. *Protein Sci.* **2002**, *11*, 997-1003.

19. Lee, J. C.; Timasheff, S. N. Partial specific volumes and interactions with solvent components of proteins in guanidine hydrochloride. *Biochemistry* **1974**, *13*, 257-265.

20. Lee, J. C.; Timasheff, S. N. The stabilization of proteins by sucrose. *J. Biol. Chem.* **1981**, *256*, 7193-7201.

21. Gekko, K.; Morikawa, T. Preferential hydration of bovine serum albumin in polyhydric alcohol-water mixtures. *J. Biochem.*, **1981A**, *90*, 39-50.

22. Gekko, K.; Morikawa, T. Thermodynamics of polyol-induced thermal stabilization of chymotrypsinogen. *J. Biochem.*, **1981B**, *90*, 51-60.

23. Barnett, G. V.; Razinkov, V. I.; Kerwin, B. A.; Blake, S.; Qi, W.; Curtis, R. A.; Roberts, C. J. Osmolyte Effects on Monoclonal Antibody Stability and Concentration-Dependent Protein Interactions with Water and Common Osmolytes. *J. Phys. Chem. B* **2016**, *120*, 3318-3330.

24. Sudrik, C.; Cloutier, T.; Pham, P.; Samra, H. S.; Trout, B. L. Preferential interactions of trehalose, L-arginine.HCl and sodium chloride with therapeutically relevant IgG1 monoclonal antibodies. *MAbs* **2017**, *9*, 1155-1168.

25. Sudrik, C.; Cloutier, T.; Mody, N.; Sathish, H. A.; Trout, B. L. Understanding the Role of preferential exclusion of sugars and polyols from native state IgG1 monoclonal antibodies and its effect on aggregation and reversible self-association. *Pharm. Res.* **2019**, *36*, 109.

26. Jorgensen, L.; Hostrup, S.; Moeller, E. H.; Grohganz, H. Recent trends in stabilising peptides and proteins in pharmaceutical formulation—considerations in the choice of excipients. *Expert Opin. Drug Deliv.* **2009**, *6*, 1219-1230.

27. Xie, G.; Timasheff, S. N. Mechanism of the stabilization of ribonuclease A by sorbitol: preferential hydration is greater for the denatured than for the native protein. *Protein Sci.*, **1997**, *6*, 211-221.

28. Myers, J. K.; Pace, C. N.; Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **1995**, *4*, 2138-2148.

29. Shearwin, K. E.; Winzor, D. J. Effect of sucrose on the dimerization of α -

chymotrypsin: Allowance for thermodynamic nonideality arising from the presence of a small inert solute. *Biophys. Chem.* **1988**, *31*, 287-294.

30. Kornblatt, J. A.; Kornblatt, M. J.; Bon Hoa, G. H.; Mauk, A. G. 1993. Responses of two protein-protein complexes to solvent stress: Does water play a role at the interface? *Biophys. J.* **1993**, *65*, 1059-1065.

31. Xavier, K. A.; Shick, K. A.; Smith-Gill, S. J.; Wilson, R. C. Involvement of water molecules in the association of monoclonal antibody Hy-HEL-5 with bovine quail lysozyme. *Biophys. J.* **1997**, *73*, 2116-2125.

32. Morar, A.S., Wang, X., and Pielak, G.J. Effects of crowding on mono-, di-, and tetrasaccharides on cytochrome c-cytochrome c peroxidase binding: Comparing experiment to theory. *Biochemistry* **2001**, *40*, 281-285.

33. Kayser, V.; Chennamsetty, N.; Voynov, V.; Helk, B.; Forrer, K.; Trout, B. Evaluation of a Non-Arrhenius Model for Therapeutic Monoclonal Antibody Aggregation. *J. Pharm. Sci.* **2011**, *100*, 2526-2542.

34. Roberts C.J.; Darrington, R. T.; Whitley, M. B. Irreversible aggregation of recombinant bovine granulocyte-colony stimulating factor (bG-CSF) and implications for predicting protein shelf life. *J. Pharm. Sci.* **2003**, *92*, 1095-1111.

35. Vagenende, V.; Han, A. X.; Mueller, M.; Trout, B. L. Protein-associated cation clusters in aqueous arginine solutions and their effects on protein stability and size. *ACS Chem. Biol.* **2013**, *8*, 416-422.

36. Shukla, D.; Trout, B. L. Preferential interaction coefficients of proteins in aqueous arginine solutions and their molecular origins. *J. Phys. Chem. B* **2011**, *115*, 1243-1253.

37. Kendrick, B. S.; Chang, B. S.; Arakawa, T.; Peterson, B.; Randolph, T. W.; Manning, M. C.; Carpenter, J. F. Preferential exclusion of sucrose from recombinant interleukin-1 receptor antagonist: role in restricted conformational mobility and compaction of native state. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11917-11922.

38. Hill, C. P.; Osslund, T. D.; Eisenberg, D. The structure of granulocyte-colony-stimulating factor and its relationship to other growth factors. *Proc. Natl. Acad. Sci. USA.* **1993**, *90*, 5167-5171.

39. Krishnan, S.; Chi, E. Y.; Webb, J. N.; Chang, B. S.; Shan, D.; Goldenberg, M.; Manning, M.C.; Randolph, T.W.; Carpenter, J. F. Aggregation of granulocyte colony stimulating factor under physiological conditions: characterization and thermodynamic inhibition. *Biochemistry* **2002**, *41*, 6422-6431.

40. Zhang, J.; Banks, D. D.; He, F.; Treuheit, M. J.; Becker, G. W. Effects of sucrose and benzyl alcohol on G-CSF conformational dynamics revealed by hydrogen deuterium exchange mass spectrometry. *J. Pharm. Sci.* **2015**, *104*, 1592-1600.

41. Aubin, Y.; Hodgson, D. J.; Thach, W. B.; Gingras, G.; Sauvé, S. Monitoring effects

of excipients, formulation parameters and mutations on the high order structure of filgrastim by NMR. *Pharm. Res.* **2015**, *32*, 3365-3375.

42. Manikwar, P.; Majumdar, R.; Hickey, J. M.; Thakkar, S. V.; Samra, H. S.; Sathish, H. A.; Bishop, S. M.; Middaugh, C. R.; Weis, D. D.; Volkin, D. B. Correlating excipient effects on conformational and storage stability of an IgG1 monoclonal antibody with local dynamics as measured by hydrogen/deuterium-exchange mass spectrometry. *J Pharm. Sci.* **2013**, *102*, 2136-2151.

43. Illes-Toth, E.; Rempel, D. L.; Gross, M. L. Pulsed Hydrogen-Deuterium Exchange Illuminates the Aggregation Kinetics of α -Synuclein, the Causative Agent for Parkinson's Disease. *ACS Chem. Neurosci.* **2018**, *9*, 1469-1476.

44. Gentiluomo, L.; Roessner, D.; Augustijn, D.; Svilenov, H.; Kulakova, A.; Mahapatra, S.; Winter, G.; Streicher, W.; Rinnan, A.; Peters, G. H. J.; Harris, P.; Friess, W. Application of interpretable artificial neural networks to early monoclonal antibodies development. *Eur. J. Pharma. Biopharma.* **2019**, *141*, 81-89.

45. Cloutier, T.; Sudrik, C.; Mody, N.; Sathish, H. A.; Trout, B. L. Molecular computations of preferential interaction coefficients of IgG1 monoclonal antibodies with sorbitol, sucrose, and trehalose and the impact of these excipients on aggregation and viscosity. *Mol. Pharma.* **2019**, *16*, 3657-3664.

46. Barata, T. S., Zhang, C., Dalby, P. A., Brocchini, S., & Zloh, M. Identification of protein-excipient interaction hotspots using computational approaches. *Int. J. Mol. Sci.* **2016**, *17*, 853.

47. Herman, A. C.; Boone, T. C.; Lu, H. S. Characterization, formulation, and stability of Neupogen®(Filgrastim), a recombinant human granulocyte-colony stimulating factor. In *Formulation, characterization, and stability of protein drugs: case histories* Pearlman, Rodney, Wang, Y. John (Eds) Springer, Boston, MA. **2002**; pp. 303-328.

48. Chi, E. Y.; Krishnan, S.; Kendrick, B. S.; Chang, B. S.; Carpenter, J. F.; Randolph, T. W. Roles of conformational stability and colloidal stability in the aggregation of recombinant human granulocyte colony stimulating factor. *Protein Sci.* **2003**, *12*, 903-913.

49. Kolvenbach, C. G.; Narhi, L. O.; Philo, J. S.; Li, T.; Zhang, M.; Arakawa, T. Granulocyte-colony stimulating factor maintains a thermally stable, compact, partially folded structure at pH 2. *J. Pept. Res.* **1997**, *50*, 310-318.

50. Raso, S. W.; Abel, J.; Barnes, J. M.; Maloney, K. M.; Pipes, G.; Treuheit, M. J.; King, J.; Brems, D. N. Aggregation of granulocyte-colony stimulating factor in vitro involves a conformationally altered monomeric state. *Protein Sci.* **2005**, *14*, 2246-2257.

51. Bristow, A. F.; Bird, C.; Bolgiano, B.; Thorpe, R. Regulatory requirements for therapeutic proteins: the relationship between the conformation and biological activity of filgrastim. *Pharmeur. Bio. Sci. Notes* **2012**, 103-17.

52. Santoro, M. M.; Bolen, D. W. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **1988**, *27*, 8063-8068.
53. Consalvi, V.; Chiaraluce, R.; Giangiacomo, L.; Scandurra, R.; Christova, P.; Karshikoff, A.; Knapp, S.; Ladenstein, R. Thermal unfolding and conformational stability of the recombinant domain II of glutamate dehydrogenase from the hyperthermophile *Thermotoga maritima*. *Protein Eng.* **2000**, *13*, 501-507.
54. Glasoe, P. K.; Long, F. A. Use of glass electrodes to measure acidities in deuterium oxide. *J. Phys. Chem.* **1960**, *64*, 188-189.
55. Ahn, J.; Cao, M. J.; Yu, Y. Q.; Engen, J. R. Accessing the reproducibility and specificity of pepsin and other aspartic proteases. *Biochim. Biophys. Acta - Proteins Proteomics*, **2013**, *1834*, 1222-1229.
56. Groves, K.; Cryar, A.; Cowen, S.; Ashcroft, A. E.; Quaglia, M. Mass Spectrometry Characterization of Higher Order Structural Changes Associated with the Fc-glycan Structure of the NISTmAb Reference Material, RM 8761. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 553-564.
57. Coales, S. J.; Lee, J. E.; Ma, A.; Morrow, J. A.; Hamuro, Y. Expansion of time window for mass spectrometric measurement of amide hydrogen/deuterium exchange reactions. *Rapid Comm. Mass Spectr.* **2010**, *24*, 3585-3592.
58. Bai, Y.; Milne, J. S.; Mayne, L.; Englander, S. W. Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Bioinf.* **1993**, *17*, 75-86.
59. Tamada, T.; Honjo, E.; Maeda, Y.; Okamoto, T.; Ishibashi, M.; Tokunaga, M.; Kuroki, R. Homodimeric cross-over structure of the human granulocyte colony-stimulating factor (G-CSF) receptor signaling complex. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 3135-3140.
60. Yang, J. M.; Chen, C. C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins: Struct. Funct. Bioinf.* **2004**, *55*, 288-304.
61. Ablinger, E.; Hellweger, M.; Leitgeb, S.; Zimmer, A. Evaluating the effects of buffer conditions and extremolytes on thermostability of granulocyte colony-stimulating factor using high-throughput screening combined with design of experiments. *Int. J. Pharm.* **2012**, *436*, 744-752.
62. Michaelis, U.; Rudolph, R.; Winter, G.; Woog, H. Stable lyophilized pharmaceutical preparations of G-CSF **1999**, U.S. Patent No. 5,919,443.
63. Schneider, C. P.; Trout, B. L. (2009). Investigation of cosolute– protein preferential interaction coefficients: New insight into the mechanism by which arginine inhibits aggregation. *J. Phys. Chem. B* **2009**, *113*, 2050-2058.
64. Rowe, R. C.; Sheskey, P. J.; Owen, S. C. (Eds.). Handbook of pharmaceutical

excipients (Vol. 6). London: Pharmaceutical press., **2006**.

65. Alebouyeh, M.; Tahzibi, A.; Yaghoobzadeh, S.; Zahedy, E. T.; Kiumarsi, S.; Soltanabad, M. H.; Shahbazi, S.; Amini, H. Rapid formulation assessment of filgrastim therapeutics by a thermal stress test. *Biologicals* **2016**, *44*, 150-156.

66. Sato, Y. Preparations stabilized over long time **2000**, WO2000051629A1

67. Carpenter, J. F.; Pikal, M. J.; Chang, B. S.; Randolph, T. W. Rational design of stable lyophilized protein formulations: some practical advice. *Pharm. Res.* **1997**, *14*, 969-975.

68. Li, C.; Wang, J. X.; Le, Y.; Chen, J. F. Studies of Bicalutamide–Excipients Interaction by Combination of Molecular Docking and Molecular Dynamics Simulation. *Mol. Pharm.* **2013**, *10*, 2362-2369.

69. Ito, L.; Shiraki, K.; Matsuura, T.; Okumura, M.; Hasegawa, K.; Baba, S.; Yamaguchi, H.; Kumasaka, T. High-resolution Xray analysis reveals binding of arginine to aromatic residues of lysozyme surface: implication of suppression of protein aggregation by arginine. *Protein Eng. Des. Sel.* **2011**, *24*, 269–274.

70. Timasheff, S. N. Protein-solvent preferential interactions, protein hydration, and the modulation of biochemical reactions by solvent components. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9721-9726.

71. Carpenter, J. F.; Chang, B. S.; Garzon-Rodriguez, W.; Randolph, T. W. Rational design of stable lyophilized protein formulations: theory and practice. *Pharm. Biotechnol.* **2002**, *13*, 109-133.

72. Arakawa, T.; Prestrelski, S. J.; Kenney, W. C.; Carpenter, J. F. Factors affecting short-term and long-term stabilities of proteins. *Adv. Drug Deliv. Rev.* **2001**, *46*, 307-326.

73. Shah, D.; Li, J.; Shaikh, A. R.; Rajagopalan, R. Arginine–aromatic interactions and their effects on arginine-induced solubilization of aromatic solutes and suppression of protein aggregation. *Biotechnol. Prog.* **2012**, *28*, 223-231.