Supplementary Material for:

Precision and Robustness of 2D-NMR for structure assessment of filgrastim biosimilars

Houman Ghasriani[1], Derek J. Hodgson[2], Robert G. Brinson[3], Ian McEwen[4], Lucinda F. Buhse[1], Steven Kozlowski[5], John P. Marino[3*], Yves Aubin[2*] and David A. Keire[1*]

[1]*U.S. Food & Drug Administration, Center for Drug Evaluation and Research, Division of Pharmaceutical Analysis, 645 S. Newstead Avenue, St. Louis, MO 63110.*

[2]*Centre for Biologics Evaluation, Biologics and Genetic Therapies Directorate, Health Canada, 251 Sir Frederick Banting Drive, Ottawa, ON, K1A 0K9.*

[3]*Institute for Bioscience and Biotechnology Research, National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850.*

[4]*Medical Products Agency of Sweden, P.O. Box 26 SE-75103, Uppsala, Sweden.*

[5]*U.S. Food & Drug Administration, Center for Drug Evaluation and Research, Office of Biotechnology Products, 10903 New Hampshire Blvd, Silver Spring, MD 20993.*

Corresponding Authors: *David A. Keire, FDA/DPA, 645 S. Newstead Ave., St. Louis, MO, 63110, Phone: 314-539-3850, E-mail: David.Keire@fda.hhs.gov; Yves Aubin BGTD, Health Canada, 251 Sir Frederick Banting Driveway, Ottawa, ON, K1A 0K9, Phone 613-941-6155, E-mail: yves.aubin@hc-sc.gc.ca; John P. Marino, National Institute of Standards and Technology, 9600 Gudelsky Drive, Rockville, MD 20850, Phone: 240-314-6160, E-mail: john.marino@nist.gov

Supplementary Methods

System suitability samples:

Samples of 0.2 mM uniformly ^{15}N-labeled methionyl-granulocyte-colony stimulating factor (^{15}N-Met-G-CSF) that had been produced in *E. coli* and purified into 10 mM sodium acetate-*d3* buffer at pH 4.0 in the Health Canada laboratories, were distributed to all labs in 5 mm NMR tubes (Wilmad Glass #542-PP-7) rated for 1000 MHz. A second isotopically enriched protein sample was also used for instrument calibration with four identical samples containing uniformly labeled ^{13}C,^{15}N-SH3 domain from chicken α-spectrin (9 mg/ml, 10% D_2O, pH 3.5) purchased from Cambridge Isotope Laboratories Inc. (CNLM-6840-S, Lot#: I-17838A, PSO#: 13I-106) and distributed to each laboratory.

Formulated drug products:

Formulated met-G-CSF samples from the following four vendors were prepared for the NMR experiments: Amgen (Neupogen®, Lot 1025275, Expiry: Aug 2013), Biocon (NUFIL Safe™, Batch nr: N010053, Exp: Aug 2013), INTAS (Neukine®, Batch nr: 1010051, Exp: Sep 2014) and Dr. Reddy's Laboratories (Grafeel™, Batch nr: GFAS00312, Expiry: Jan 2014). These same drug product lots had been previously analyzed by LC-MS, NMR, CD and bioassay[1]. NMR samples were prepared as described previously[2]. Briefly, for each NMR sample, 20 vials of formulated met-G-CSF (aqueous solution), each containing 300 μg of the active ingredient, were concentrated (2012-Oct-29) using Amicon Ultra centrifugal filter units with 3000 Dalton molecular weight cutoff (MWCO). No buffer exchange was performed. The centrifuge tubes were spun at 4000 x *g* for 4 x ~20 min at room temperature, and the final sample volume was adjusted to 280 μL using the filtrate. 20 μL D_2O (Acros Organics, Lot: A0310778) was added to

each sample and then transferred into 5 mm Shigemi tubes (Tokyo, Japan) rated for 1000 MHz. **Supplementary Table 1** summarizes the excipients in the formulated met-G-CSF products.

Instruments:

Data were acquired at four different magnetic field strengths (500, 600, 700, and 900 MHz for ^{1}H resonance frequency) using instruments from two different manufacturers (Agilent Technologies or Bruker Biospin Corp.) in four different laboratories. The laboratories were located at the Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER), Division of Pharmaceutical Analysis (DPA) in St Louis, Missouri; the Medical Products Agency of Sweden (MPA Sweden) in Uppsala, Sweden; Health Canada (HC), Centre for Biologics Evaluation, Biologics and Genetic Therapies Directorate, in Ottawa, Canada; and the National Institute of Standards and Technology (NIST), Institute for Bioscience and Biotechnology Research in Rockville, Maryland. **Supplementary Table 2** summarizes the instruments used in this study.

Spectrometer performance and system suitability:

The performances of the NMR spectrometers were assessed for magnetic drift, proton sensitivity, and lineshape and sample temperature (**Supplementary Tables 3** to **7**). The system suitability sample (obtained from Health Canada), a uniformly ^{15}N-labeled met-G-CSF, was made by a different method than that of the filgrastim products, and the 2D-NMR spectrum of this sample showed extra correlation peaks compared to the drug product. These extra peaks were attributed to the presence of a minor component that was found to be an oxidized form of the protein in these samples. As a second non-therapeutic "not similar" model protein, the 62-amino acid SH3 domain from chicken α–spectrin was used and has a tightly-packed beta-barrel structure comprised of five anti-parallel beta strands[3, 4]. The SH3 domain exhibits a simple, well-

resolved set of $^{1}H,^{15}N$-Heteronuclear Single Quantum Coherence (HSQC) amide proton cross peaks (Biological Magnetic Resonance Bank BMRB#3433) compared to the more complex 175 amino acid filgrastim data (BMRB#18291). Of note, for this study, filgrastim amino acids are numbered with the N-terminal methionine as Met0.

NMR Method Comparison:

Considerations for experiment selection:

For folded proteins, backbone amide proton signals are found between 6 ppm and 10 ppm and, for nitrogen, between 100 ppm and 140 ppm. The major proton-nitrogen correlation type in proteins is from the structurally important backbone amides. There are also proton-nitrogen correlations from the side-chains of glutamine, asparagine, arginine and tryptophan present in this spectral region. Because the protein amide signals are sensitive to structure, appear at chemical shifts removed from interfering excipient signals and are of a limited number of types in proteins, the experiment selected for application to filgrastim was the 2D $^{1}H,^{15}N$-HSQC.

Once the nuclei to be measured and the type of filtering to be applied had been selected, the pulse program to acquire the data was carefully chosen from the available options. The standard and sensitivity-enhanced versions of HSQC were tested[5,6]. Moreover, several different solvent suppression techniques were examined: the water-flipback technique[5], the Watergate sequence[7] and the BEST-HSQC (the Bruker implementation of the SOFAST principle in the HSQC experiment)[8]. This BEST method selectively excites spins of interest and only minimally excites water proton spins, enabling protein proton spin magnetization to recover much faster and concomitantly allowing significantly increased pulse repetition rates and improved sensitivity in a given experiment time. Indeed, BEST-HSQC gave greater S/N per time unit of experiment time at higher fields (*i.e.*, 900 MHz) where the longitudinal relaxation rates (T_1s) are longest. At

700 MHz there was little to no gain, while at lower fields (*i.e.,* 500 to 600 MHz), the BEST-HSQC experiments resulted in lower S/N per unit experiment time.

For the filgrastim samples, because spectra with sufficient signal-to-noise ratio (S/N) are required for confident comparison between samples, a minimum S/N target was established to set the experimental times for this study. To achieve the S/N target, a series of performance qualification tests with standards supplied by the instrument vendors were run on each spectrometer to ensure that they were operating within specifications. Then, the system suitability and SH3 domain were used to rapidly test experiments and parameters which gave the best S/N. Based on the S/N measurements made on the standard samples, an average S/N for at least three well resolved signals of greater than 10:1 was selected as the target. (see **Supplementary Table 10** for a description of how S/N was measured).

Another factor that was tested for increasing the S/N ratio was the inter-scan delay. Because T_1s increase with increasing field strength for molecules in a specific molecular correlation time range (*e.g.,* for most proteins), a longer inter-scan delay is needed to minimize signal saturation effects from rapid pulse repetition rates at 900 MHz than at 500 MHz. Across fields we observed that an increase in the inter-scan delay from 1 s to 1.5 s resulted in an increase in S/N of between 20-40% because more complete relaxation is allowed between transients. However, as the experimental time also increased by approximately 50%, the gain in S/N that is due to the increase in inter-scan delay was mitigated by the increased total experiment time. Thus, at 900 MHz a 1.5 s inter-scan delay was used because the higher sensitivity of this instrument allowed less needed signal averaging to obtain sufficient signal to noise in the spectra. By contrast at 700 MHz, 600 MHz and 500 MHz a 1.0 s inter-scan delay was used, and more scans were averaged.

Based on assessment of S/N, spectral information content (the most peaks present) and the fewest spectral artifacts, the water flipback sensitivity enhanced $^{1}H,^{15}N$-HSQC was selected as the best pulse sequence for the 2D NMR method. Finally, the optimal 2D HSQC experimental parameters were applied to collect data on the formulated filgrastim from the US originator company and from the foreign-sourced products. Three to four days of spectrometer time were needed to obtain the targeted 10:1 S/N on selected peaks from the lowest field instrument used in this study (500 MHz). Of note, MPA Sweden used a phase sensitive Watergate-HSQC pulse sequence at 600 MHz on the formulated drug products. The pulse sequences and selected parameters used and the dates on which data were acquired on the formulated filgrastim samples as they were shipped between laboratories are summarized in **Supplementary Table 8**.

Inter-laboratory NMR experimental parameters:

2D sensitivity enhanced $^{1}H,^{15}N$-HSQC experiments with gradient coherence selection were recorded on the NMR spectrometers, which were all equipped with cryogenically-cooled probes[9]. For all experiments (except where noted in parentheses) the spectral widths of the ^{1}H and ^{15}N dimensions were 14 ppm (16 ppm for Health Canada) and 32 ppm, respectively. The center frequency of ^{1}H was set to the water resonance (~4.77 ppm) and that of ^{15}N was set to 117 ppm, such that the resonance for the glycine residue at position 73 (Gly^{73}) amide cross peak would appear at 7.77 ppm and 105.8 ppm in ^{1}H and ^{15}N dimensions, respectively. An inter-scan delay of 1 s was used (1.5 s for NIST 900). Acquisition time was set to 162 ms with the aim of maximizing resolution in the proton dimension (100 ms for Health Canada because during the longer acquisition time, the noise contribution is larger thus resulting in a few percent loss of sensitivity) and 64 complex points were acquired in the indirect dimension. In total, 2048 scans (1024 for NIST 900, and 1536 for NIST 600) were collected and co-added. The polarization

transfer delay (0.5 * J^1) in ^{15}N-HSQC experiment is based on the one-bond scalar coupling, $^1J_{HN}$, which has been accurately determined for proteins by Tjandra's group[10], and found to be -93±1 Hz. Thus, the $^1J_{NH}$ scalar coupling constant for the polarization transfer (*cnst4* for Bruker and *JNH* for Agilent Technologies) was set to 93 Hz (100 Hz was used by Health Canada to compensate for signal loss due to T_2 relaxation, but the difference is not significant). Total experimental time for each experiment was 89 h (62 h for data collected at NIST 900 and 65 h for NIST 600). All experiments were run with the probe air set to 25°C (~4°C and ~2.5°C lower probe air temperature on the Health Canada 600 MHz and the 700 MHz instruments, respectively between the non-calibrated and calibrated variable temperature units).

The same parameters were used for recording the 2D spectra on the system suitability sample (^{15}N-met-G-CSF) except that number of scans, which was 16, and the number of points in the indirect dimension was 128 complex points. For these experiments the measurement times were 40 minutes. For the uniformly labeled ^{13}C,^{15}N-SH3 sample, spectral widths of the ^{1}H and ^{15}N dimensions were 14 ppm and 28 ppm, respectively. 128 complex points were acquired in the indirect dimension, and 8 scans were collected and co-added. The frequency offset for direct dimension was set at water resonance and the frequency offset for the indirect dimension was set at 120.2 ppm, such that valine residue at position 9 amide (Val^9) would have the chemical shifts of 9.18 ppm and 112 ppm in the ^{1}H and ^{15}N dimensions, respectively. A total of 2278 points were collected in the direct dimension and acquisition time was set to 162 ms (99 ms for Health Canada). The inter-scan delay was set to 1.0 s (1.5 s for NIST 900 MHz), and 8 transients were collected and co-added. All other parameters were identical with experiments used for the formulated drug products. For the SH3 experiments the measurement time was 40 minutes.

After sample preparation at the FDA laboratories (October 2012), the samples were shipped cold (~4°C) to NIST where data were collected at 900 MHz and 600 MHz, then to HC where data was collected at 700 MHz and 600 MHz, then to the FDA St Louis laboratory where data was collected at 500 MHz and finally to MPA Sweden where data was collected at 600 MHz. The data collection across the four laboratories took approximately 9 months to complete. The system suitability standard and the model protein samples were shipped separately to each laboratory from the FDA St. Louis laboratory and were retained in each location. The acquisition parameters across laboratories on the ^{15}N-met-G-CSF samples are summarized in **Supplementary Table 9**.

NMR data processing:

All NMR data processing was performed at the FDA St. Louis laboratory with the completed data set from all four laboratories. Solvent suppression by deconvolution was applied in the time-domain and a 90-degree left-shifted sine-squared apodization function was applied in both dimensions. Zero-points were appended to the FID to give a final spectrum size of 4096 and 512 points in ^{1}H and ^{15}N dimensions, respectively. FIDs were Fourier-transformed and phase-corrected, and the resulting frequency-domain data were baseline corrected.

For 2D spectra overlays, data were truncated after 512 total points in the direct dimension. A 45-degree left-shifted sine square function was applied in the direct dimension, and the indirect dimension was multiplied by a 54-degree left-shifted sine square function. The FIDs were zero-filled to 4k and 512 points in the direct and indirect dimension, respectively. FIDs were phase-corrected and Bernstein polynomial base-line correction of the third order was applied in the direct dimension. Solvent suppression through convolution was applied. In the indirect

dimension, 447 total points were added using the MNova Toeplitz linear prediction method with 59 basis points and 4 coefficients.

Data analysis:

Once the data were collected, methods that compared the processed 2D-NMR structural maps collected on reference and test samples were selected and optimized so, ideally, comparability assessments could be performed in a rapid, non-subjective manner across multiple complex data sets. Any significant change in protein structure would be reflected in altered chemical shifts from reporter nuclei near the site of perturbation, thus changing the pattern of the 2D-NMR peaks. Notably, the perturbed residues can then be mapped onto an available tertiary structure (*e.g.*, **Supplementary Fig. 2b**) to assess where the samples differ. For 2D-NMR data the smallest observable shift based on common digital resolution is in the parts per billion (ppb) range.

For SH3 and ^{15}N-met-G-CSF, the measured ^{1}H and ^{15}N chemical shift values from the spectra were used to establish experimental precision across laboratories using the 2D-NMR approach. Because the chemical shifts of the Gly^{73} of ^{15}N-met-G-CSF or Val^{9} of SH3 showed minimal change with pH (2.8 to 5.6) or temperature (15°C to 50°C) these peaks were selected as reference peaks for the 2D spectral comparison. The mean chemical shifts and their standard deviations were calculated across spectra obtained from the standards and different drug products in the four laboratories.

Here, the 2D-NMR spectra acquired in this study were analyzed utilizing two different approaches. The first approach analyzed chemical shifts values across spectra using root square mean deviation (RMSD) calculations or Combined Chemical Shift Differences (CCSD). The

second approach involved the use of principle component analysis (PCA) of a set of spectra, which is a form of multivariate statistical analysis.

For the chemical shift based approaches, the chemical shift assignments for G-CSF[11, 12] and SH3[3] were available so the shift differences could be assigned to individual amino acids. For G-CSF, a total of 127 backbone amide peaks (from a total of 162 non-proline residues) and, for SH3, 59 peaks (out of 59 non-proline residues) could be reliably picked with the automatic peak picking routine in Sparky[13]. With this information, a mean chemical shift across multiple spectra for a given nuclei could be obtained and the deviation of that value across the sample set calculated.

In the initial precision measurements, the direct proton or nitrogen shifts were evaluated as separate values. The inter-laboratory comparison of chemical shifts for the backbone resonances showed that the root-mean-squared deviations (RMSDs) were very small (on the order of single digit ppb for ^{1}H and tens of ppb for ^{15}N) (**Supplementary Table 11**). Thus, the proposed NMR method is highly precise for comparison of the chemical shifts of peaks, which can be reliably picked.

For met-^{15}N-G-CSF, across four spectrometers the average RMSD for the resonances in the ^{1}H or ^{15}N dimensions was 0.003 ± 0.002 ppm (largest value: 0.006 ppm for Gln^{67}) or 0.011 ± 0.006 ppm (largest value: 0.034 ppm for Met^{121}), respectively (**Supplementary Table 11**, 4 fields column). The largest deviations belonged to resonances in the congested zones of the spectra, where peak-picking precision was affected by overlapping resonances. These deviation values can be used as guidance when establishing limits for measurement uncertainty across platforms and laboratories.

Figure 1a also demonstrates that small variations in temperature settings between the

different laboratories could be observed by overlaying the spectra. Including data collected with different temperature offsets (HC 600 and HC 700) increased the chemical shift RMSD values across six spectrometers for a subset of residues (**Supplementary Table 11**, '6 fields columns'). These shifted residues increased the average RMSD for the ^{1}H and ^{15}N dimensions to 0.006 ± 0.004 ppm (largest value: 0.020 ppm for Gln^{70}) and 0.023 ± 0.017 ppm (largest value: 0.108 ppm for Ser^{62}), respectively. The largest deviations belonged to resonances originating from the loop regions of met-G-CSF, indicating that the local structure of the loop regions of met-G-CSF are affected by an approximately 4°C difference in sample temperature. Similar levels of precision in chemical shift determination were obtained from the inter-laboratory comparison of chemical shifts of the SH3 domain protein (**Supplementary Table 11**).

CCSD uses the same chemical shifts from 2D peak picking and compares the drug products based on the similarity of the chemical shifts of the well-resolved peaks[14-16]. The CCSD metric combines the frequency range-weighted proton and nitrogen shifts into one value that can be used for comparison purposes between data sets. For this work the following equation was used:

$$CCSD = \sqrt{ (0.5 * [(\delta_H)^2 + (\alpha * \delta_N)^2])}$$

where α=0.1 and δ_H and δ_N are differences in chemical shifts in ppm for ^{1}H and ^{15}N with respect to a reference cross peak. Here, based on the 4096 x 512 size of the processed data sets over the 14 ppm (^{1}H) by 32 ppm (^{15}N) frequency ranges, at 900 MHz the theoretical limit of precision of the CCSD value is approximately 5 ppb.

Because the chemical shift assignments were known, the CCSD values can be mapped onto the drug amino acid sequence (**Figs. 1b,c**). In addition, the magnitude of these deviations can be mapped onto the 3D rendering of the protein to highlight where in the structural differences arise, which in this case occurred in the solvent exposed regions. Four representative residues in

a loop region that exhibited deviations are shown in **Supplementary Fig. 2b. Figure 1b** shows a plot of CCSD values versus the filgrastim sequence over all spectrometers. The temperature difference for data collected on two spectrometers gave rise to a reproducible pattern of elevated CCSD values (*i.e.* above the 8 ppb limit). Similar results were observed for the SH3 data set (**Supplementary Figure 6**). Importantly, when new data were collected with the spectrometer temperatures set to be the same, all CCSD value deviations above 8 ppb were eliminated (**Fig. 1c**). The green boxes in Figure 1c are from the 500 MHz data (the lowest field strength used) and at higher fields the observed deviations across the sample set are all less than 5 ppb. These data illustrate the sensitivity of the approach for comparison of tertiary structures between drug molecules and highlight the need for system suitability assessment prior to data collection.

By contrast to the CCSD approach, the multivariate statistical analysis approach does not require a peak picking step, includes comparison of peak intensities, and can include the spectral regions that contain non-resolved resonances that are not amenable to peak picking. Whereas CCSD involves a quantitative comparative measure of selected resonances within a spectral region and may potentially miss the appearance of minor peaks, the statistical analysis method intrinsically evaluates an entire spectral region of interest with a full point-by-point evaluation. Our data therefore shows that this analysis method provides a powerful means of lot-to-lot comparability internally for a given laboratory. **Figure 1d** shows that data from each lab clusters together, with the exception of MPA (*see main text for discussion on this exception*).

One issue with 2D-NMR data is that after processing, the transformed spectrum contains two frequency dimensions and one intensity dimension for each point of a matrix which can be over a million points in size (*e.g.,* 2048 x 512 points in size). In addition, most multivariate statistical analysis methods were developed for comparisons of data sets, which contain one frequency

dimension and one intensity dimension (*e.g.*, 1D-NMR, HPLC or optical spectroscopy data sets). Thus the multivariate analysis of 2D-NMR data sets by commercially available software requires choosing a filtering method that reduces the data set dimensionality and size to a reasonable level without removing any of the information content present.

For the statistical analysis portion of this work, each 2D peak pattern was converted into one single row vector, consisting of concatenated 1D traces that were extracted along the ^{1}H axis for each data point in the ^{15}N dimension[17]. The values in the resulting row vector are the chemical shifts for the two nuclei, ^{1}H and ^{15}N, and the spectral intensity for the corresponding map coordinate. Where needed, interpolation between data points was performed such that the spacing of points for comparison across spectra acquired with different spectral widths would be the same.

Subsequently, each data set was individually normalized with respect to the highest intensity in the spectrum. The row vectors were padded together to create a matrix where rows constitute samples, and columns represent intensities at a given chemical shift coordinate [^{1}H, ^{15}N] in the 2D map. Prior to principal component analysis (PCA), the matrix was mean-centered through column-wise mean subtraction of the matrix. In the principal component (PC) space, each 2D spectrum is represented as a single point.

Matlab version 8.1 (The MathWorks, Inc., Natick, Massachusetts) and PLS toolbox (Eigenvector Research, Inc., Wenatchee, Washington) were used to perform the statistical analysis on the data matrix that consists of the 1D-decomposed data set. The region of the 2D correlation map that was used for the PCA analysis was 131 ppm to102 ppm in the ^{15}N dimension, and 10.4 ppm to 5.9 ppm in the ^{1}H dimension. NMRPipe[18] was used for data processing and Sparky_ENREF_13[13] and Mnova NMR (Mestrelab Research, version 8.1,

Santiago de Compostela, Spain) were used for data visualization and chemical shift analysis.

References:

1. Levy, M.J. et al. Analytical techniques and bioactivity assays to compare the structure and function of filgrastim (granulocyte-colony stimulating factor) therapeutics from different manufacturers. *Anal Bioanal Chem* **406**, 6559-6567 (2014).
2. Aubin, Y., Hodgson, D.J., Thach, W.B., Gingras, G. & Sauve, S. Monitoring Effects of Excipients, Formulation Parameters and Mutations on the High Order Structure of Filgrastim by NMR. *Pharm Res* **32**, 3365-3375 (2015).
3. Yu, H. et al. Solution structure of the SH3 domain of Src and identification of its ligand-binding site. *Science* **258**, 1665-1668 (1992).
4. Yu, H., Rosen, M.K. & Schreiber, S.L. 1H and 15N assignments and secondary structure of the Src SH3 domain. *FEBS Lett* **324**, 87-92 (1993).
5. Grzesiek, S. & Bax, A. The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurments. *Journal of the American Chemical Society* **115**, 12593 (1993).
6. Kay, L.E., Keifer, P. & Saarinen, T. Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J. Am. Chem. Soc.* **114**, 10663-10665 (1992).
7. Piotto, M., Saudek, V. & Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2**, 661-665 (1992).

8. Schanda, P., Kupce, E. & Brutscher, B. SOFAST-HMQC experiments for recording two-dimensional heteronuclear correlation spectra of proteins within a few seconds. *J Biomol NMR* **33**, 199-211 (2005).

9. Palmer, A.G.r., Cavanagh, J., Wright, P.E. & Rance, M. Sensitivity improvement in proton detected two dimensional heteronuclear correlation NMR spectroscopy. *Journal of Magnetic Resonance* **93**, 151 (1991).

10. de Alba, E. & Tjandra, N. On the accurate measurement of amide one-bond 15N-1H couplings in proteins: effects of cross-correlated relaxation, selective pulses and dynamic frequency shifts. *J Magn Reson* **183**, 160-165 (2006).

11. Werner, J.M. et al. Secondary structure and backbone dynamics of human granulocyte colony-stimulating factor in solution. *Biochemistry* **33**, 7184-7192 (1994).

12. Zink, T., Ross, A., Ambrosius, D., Rudolph, R. & Holak, T.A. Secondary structure of human granulocyte colony-stimulating factor derived from NMR spectroscopy. *FEBS Lett* **314**, 435-439 (1992).

13. Goddard, T.D. & Kneller, D.G. (Univeristy of California, San Francisco.

14. Clarkson, J. & Campbell, I.D. Studies of protein-ligand interactions by NMR. *Biochem Soc Trans* **31**, 1006-1009 (2003).

15. Hardman, C.H. et al. Structure of the A-domain of HMG1 and its interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* **34**, 16596-16607 (1995).

16. Williamson, M.P. Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy.* **73**, 1-16 (2013).

17. Berglund, A., Brorsson, A.C., Jonsson, B.H. & Sethson, I. The equilibrium unfolding of MerP characterized by multivariate analysis of 2D NMR data. *J Magn Reson* **172**, 24-30 (2005).

18. Delaglio, F. et al. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277-293 (1995).

Supplementary Table 1: Active ingredients and excipients in the formulated met-GCSF products from four different vendors.

	Neupogen® (Amgen)	**Grafeel™ (Dr. Reddy's)**	**Neukine® (Intas)**	**Nufil Safe™ (Biocon)**
Met-GCSF / μg	300	300	300	300
Acetate / mg	0.59	N.A.	0.59	0.295
Sorbitol / mg	50	N.A.	50	25
Polysorbate 80 / mg	0.04	N.A.	0.04	-
Polysorbate 20 / mg	-	-	-	0.02
Sodium / mg	0.035	N.A.	0.035	0.018
H_2O for injection/ ml	1	1	1	0.5
pH	N.A.	N.A.	4.0	N.A.

N.A.: Not announced.

Supplementary Table 2: Equipment type and year of manufacture for the NMR spectrometers used in the inter-laboratory study.

Laboratory	^{1}H Frequency (MHz)	Console	Probe	Year of manufacture
NIST	900	Bruker Avance III	TCI Cold	2011
NIST	600	Bruker Avance III	TXI Cold	2011
HC	700	Bruker Avance	TCI Cold	2001
HC	600	Bruker Avance III	TCI Cold	2007
MPA	600	Bruker Avance I	TCI Cold	2004
FDA	500	Agilent Direct Drive	Cryobay VS	2011

Supplementary Table 3: Line-shape values obtained on the NMR spectrometers used in this study.

Instrument	Width at 50% height (Hz)	Width at 0.55% height (Hz)	Width at 0.11% height (Hz)
FDA 500	0.8	6.9	13.5
MPA 600	0.6	4.6	10.0
HC 600	1.1	11.0	24.0
HC 700	0.9	13.0	24.0
NIST 600	0.7	12.0	21.3
NIST 900	1.2	13.0	26.0

The magnetic homogeneity was controlled by adjusting the shims such that the line width of the ^{1}H resonance of chloroform in deuterated acetone conformed to specifications. The shims were optimized and the width of the chloroform ^{1}H signal was noted. The standard line shape sample, 1% chloroform in 99% deuterated acetone, was used for adjusting the quality of the shims on all magnets.

Supplementary Table 4: Magnet drift rates for the spectrometers used in this study.

Instrument	Temp (°C)	Gain (dB)	Drift Rate (Hz/h)
FDA 500	25	20	0.6
MPA 600	25	3.2	0.4
HC 600	25	5.6	0.3
HC 700	25	16	1.3
NIST 600	25	1	0.2
NIST 900	25	1	0.8

Magnetic drift (Hz/h) was measured by recording a one-scan ^{1}H spectrum of the line shape sample per hour for 12 h with the deuterium lock disengaged. The following settings were used: air flow rate: 15 L/min, sample temperature=25°C, lock power=0, lock gain=0, lock=off.

Supplementary Table 5: Parameters used and line widths obtained on the sucrose samples using the instruments in this study.

Instrument	Gain (dB)	Number of experiments	Resonance (ppm)	Width (Hz) at 100% DSS (W @ 20%)
FDA 500	26	3	~4.77	37 ± 0 *(60 ± 3)*
MPA 600	5	10	~4.8	43 ± 2 *(89 ± 4)*
HC 600	16	5	~4.8	77.4 ± 0.4 *(178.2 ± 0.9)*
HC 700	90.5	10	~4.8	77.4 ± 0.4 *(178.2 ± 0.9)*
NIST 600	32	5	~4.77	70.3 ± 1.5 *(170.4 ± 3)*
NIST 900	32 (4)	5	~4.77	31.2 ± 0.3 *(53.0 ± 2.0)*

The mean and standard deviation of replicate measurements of peak line-widths reflects the magnetic field homogeneity of each instrument. A 2 mM sucrose solution in 90%/10% H_2O/D_2O and 0.25 mM DSS was used for controlling the quality of shims prior to each experiment. Initially the shims were optimized using the standard line-shape sample (1% chloroform in 99% deuterated acetone). Subsequently, the 2 mM sucrose/0.25 mM DSS in 90%/10% H_2O/D_2O sample was used. The width of the water signal was measured at 100% and at 20% of the signal height of DSS. An example of the presat technique settings used on the Agilent spectrometer include: saturation delay=1.5s, saturation power=6 dB (this value is specific for each instrument, *i.e.* on the NIST 600=46.8 dB or NIST 900=44.1 dB), d1=3s (the Bruker acceptance specification default value of 5s was used by NIST), steady state scans=2. Temperature was set at 25°C. The sample was in non-spinning state. Eight scans were co-added.

Supplementary Table 6: Separation and signal to noise ratios obtained using water suppression and the sucrose sample for the instruments used in this study.

Instrument	Gain (dB)	Number of experiments	Resonance (ppm) (Noise Region: 200Hz downfield from peak)	Separation (%)	<S/N>
FDA 500	26	10	~5.40 (~2.65)	80	640 ± 10
MPA 600	5	10	~5.41 (~2.65)	85	320 ± 10
HC 600	32	10	~5.40 (~2.65)	82	720 ± 20
HC 700	16	5	~5.42	68	340 ± 10
NIST 600	32	5	5.41 (2.65)	80	493 ± 10
NIST 900	4	5	5.41 (2.65)	80	860 ± 20

A table of the mean and standard deviation of the S/N values obtained on the 2 mM sucrose sample from replicate measurements. A 2 mM sucrose solution in 90%/10% H_2O/D_2O, and 0.25 mM DSS was used. Prior to the 2 mM sucrose measurements, instrument vendor S/N specifications were met on the 0.1% ethylbenzene in 99.89% deuterated chloroform and 0.01% tetramethylsilane (TMS) sample using the TMS signal. However, as the solvent in our samples is primarily water, we decided to measure the ^1H sensitivity using a water-based sample. Prior to sensitivity experiments, the shimming was optimized using the standard line-shape sample (1% chloroform in 99% deuterated acetone) as shown in **Supplementary Table 3**. Presat technique used and the sample was in non-spinning state. See the legend of **Supplementary Table 5** for a description of the parameters used. For the S/N measurements, the Bruker AU command 'suppcal' or the Agilent Technologies command dsnmax(200) was used.

Supplementary Table 7: An example of the temperature calibration data obtained on the FDA 500 MHz spectrometer used in this study.

Set Temp (°C)	Sample	Actual 1 (°C)	Actual 2 (°C)	Actual 3 (°C)	Average (°C)
20	Methanol-*d4*	19.0	19.2	19.0	19.1
25	Methanol-*d4*	23.8	24.2	23.8	23.9
30	Methanol-*d4*	29.1	29.0	29.1	29.1
20	Ethylene Glycol	19.0	18.8	18.7	18.8
25	Ethylene Glycol	23.7	23.7	23.6	23.7
30	Ethylene Glycol	28.2	28.5	28.3	28.3

1D proton spectra on the FDA 500 MHz spectrometer were acquired on standard samples of methanol or ethylene glycol supplied with the instrument by Agilent. Three spectra were acquired at each temperature for each of the 2 following samples: 100% methanol (A), and 100% ethylene glycol (B). The temperature setting for the probe air was set to 20, 25, or 30°C. The actual value was calculated from the chemical shift of the compound in the standard. Agilent routines tempcal('m') or tempcal('g') were used. The "presat" technique was used and the sample was non-spinning. The following settings were used: air flow rate=11 L/min, FTS pre-cooled air = 15°C, vtc = 25°C, saturation delay=5 s, saturation power=12 dB, d1=3s, gain=0 dB, number of scans=4, steady state scans=2. For the Bruker instruments: Bruker standard samples 99.8% MeOD, and 100% ethylene glycol and the AU command 'calctemp' was used. This command will not calculate temperature below 300K for glycol as solvent.

Supplementary Table 8: The types of 2D $^{1}H,^{15}N$-HSQC experiments and a number of acquisition parameters tested for studying the formulated drug products of met-G-CSF.

Instrument	Manufacturer	Acquisition Date	Experiment	NH coupling constant (Hz)	Temp (°C)
NIST 900	Amgen	2013-Jan-18	Hsqcetfpf3gpsi	93	25
	Biocon	2012-Dec-14	Hsqcetfpf3gp		
	Intas	2013-Feb-04	Hsqcetfpf3gpsi		
	Dr. Reddy's	2013-Feb-15	Hsqcetfpf3gpsi		
NIST 600	Amgen	2013-Jan-22	Hsqcetfpf3gpsi	93	25
	Biocon	2013-Jan-12			
	Intas	2013-Jan-25			
	Dr. Reddy's	2013-Feb-04			
HC 700	Amgen	2013-Feb-21	Hsqcetfpf3gpsi2	100	21
	Biocon	2013-Feb-01			
	Intas	2013-Mar-08			
	Dr. Reddy's	2013-Feb-28			
HC 600	Amgen	2013-Feb-15	Hsqcetfpf3gpsi	100	22.5
	Biocon	2013-Jan-25			
	Intas	2013-Mar-01			
	Dr. Reddy's	2013-Mar-08			
FDA 500	Amgen	2013-May-03	gNhsqc	93	25
	Biocon	2013-May-10			
	Intas	2013-May-17			
	Dr. Reddy's	2013-Apr-23			
MPA 600	Amgen	2013-Jun-14	Hsqcfpf3gpphwg	93	25
	Biocon	2013-Jun-28			
	Intas	2013-Jun-18			
	Dr. Reddy's	2013-Jun-24			

Supplementary Table 8: The types of 2D $^{1}H,^{15}N$-HSQC experiments and acquisition parameters for studying the formulated drug products of ^{15}N-met-G-CSF. Hsqcetfpf3gpsi and gNhsqc employ sensitivity-enhanced techniques (*a.k.a.* echo-antiecho or Rance gradient mode). Hsqcfpf3gpphwg relies on phase cycling of the radio frequency pulse during t_1 for selection of desired coherence transfer pathway and for suppression of undesired artifacts. Hsqcetfpf3gpsi and gNhsqc incorporate a water-flipback technique for solvent suppression, whereas

Hsqcfpf3gpphwg employs the Watergate sequence. The STATES-TPPI technique was used for quadrature detection in the indirect dimension. The pulse sequences can be found in libraries for Topspin v.2.1 (Bruker) or Biopack v.3.0 (Agilent Technologies).

Supplementary Table 9: Acquisition parameters for system suitability sample, ^{15}N-met-G-CSF.

	AT (ms)	**NP (F2 Total)**	**NP (F1 Total)**	**SW (F2 *F1 ppm)**	**F1 digital resol. (Hz/pt)**	**Lowest shift (F2; F1 ppm)**	**$^{1}J_{HN}$ (Hz)**	**NS**	**Gain**
NIST 900	162	2048	256	14x32	11.4	-2.33 ; 101.0	90	16	128
NIST 600	162	1364	256	14x32	7.6	-2.32 ; 101.0	90	16	128
HC 700	140	1366	256	14x32	7.1	-2.27 ; 101.0	94	16	256
HC 600	122	1024	256	14x32	8.2	-3.31 ; 101.1	94	16	90
MPA 600	244	2049	256	14x32	7.6	-2.30 ; 101.0	90	16	161
FDA 500	162	1139	256	14x32	6.2	-2.34 ; 101.0	93	16	20

Supplementary Table 9: Acquisition parameters for system suitability sample, ^{15}N-met-G-CSF. Acquisition time (AT), number of points (NP), direct dimension (F1), indirect dimension (F2), spectral width (SW), ^{1}J coupling constant for polarization transfer delay, number of scans and instrument gain for ^{15}N-HSQC experiment used on system suitability sample.

Supplementary Table 10: S/N from different spectrometers on the cross peaks of three different residues of SH3, ^{15}N-met-GCSF and Amgen filgrastim.

Instrument	Property	SH3	^{15}N-met-G-CSF	Amgen
NIST 900	*Res1*	1190	76.6	23.2
	Res2	1350	74.6	28.7
	Res3	1610	65.4	18.4
	Exp Time (h)	*0.67*	*0.67*	*62*
	<S/N>/√NS	**500 ± 80**	**18 ± 2**	**0.7 ± 0.2**
NIST 600	*Res1*	1540	31.1	19.6
	Res2	1790	26.8	18.6
	Res3	2070	23.3	16.9
	Exp Time (h)	*1.33**	*0.67*	*65*
	<S/N>/√NS	**450 ± 70**	**7 ± 1**	**0.5 ± 0.1**
HC 700	*Res1*	1260	92.1	16.5
	Res2	1180	89.8	15.5
	Res3	1360	119.6	14.2
	*Exp Time (h)***	*0.67*	*0.67*	*89*
	<S/N>/√NS	**450 ± 30**	**25 ± 4**	**0.3 ± 0.1**
HC 600	*Res1*	680	94.5	15.4
	Res2	770	102.4	15.5
	Res3	720	124.1	13.0
	*Exp Time (h)***	*0.67*	*0.67*	*89*
	<S/N>/√NS	**300 ± 20**	**27 ± 4**	**0.3 ± 0.1**
FDA 500	*Res1*	1170	140.2	11.4
	Res2	1230	138.4	10.9
	Res3	1380	183.1***	7.9
	Exp Time (h)	*0.67*	*1.33****	*89*
	<S/N>/√NS	**400 ± 50**	**38 ± 6**	**0.2 ± 0.1**
MPA 600	*Res1*	1110	83.8	13.6
	Res2	1140	92.7	13.6
	Res3	1320	110	10.9
	Exp Time (h)	*0.67*	*0.67*	*89*
	<S/N>/√NS	**400 ± 50**	**24 ± 3**	**0.3 ± 0.1**

Supplementary Table 10: The mean and standard deviations of the S/N from different spectrometers on the cross peaks of three different residues of SH3, ^{15}N-met-G-CSF and Amgen filgrastim. For ^{15}N-met-G-CSF, *Res1*, *Res2* and *Res3* are Leu^{71} (8.58, 126.2 ppm), Gly^{73} (7.77, 105.8 ppm), and Ser^{80} (9.01, 114.6 ppm), and the noise regions for these resonances were ([8.00

- 9.00] [126.0 - 127.0]), ([6.90 - 7.90], [104.0 - 106.0]) and ([8.50 - 9.20], [113.0 - 115.0]) ppm, respectively. For SH3, *Res1*, *Res2* and *Res3* are Val^{9} (9.18, 112.0 ppm), Met^{25} (9.55, 122.1 ppm), and Ala^{55} (7.49, 129.2l ppm), and the noise regions for these resonances were ([9.00 - 10.00], [107.0 - 116.0]), ([9.30 - 10.30], [116.0 - 124.0]) and ([6.00 - 8.00], [126.0 - 132.0]) ppm, respectively. NS stands for number of scans. For SH3 and ^{15}N-met-G-CSF, 8 and 16 scans, respectively, were collected. For formulated drug products, NS=2048, except for NIST 900 and NIST 600 (which were 1024 and 1536, respectively). The notably higher signal-to-noise ratio for the model protein, SH3, is presumably due to the smaller molecular size of SH3, and hence longer T_2 times (and narrower linewidths) for the spin systems in SH3. *Number of scans was twice as many as other labs. **40% shorter acquisition time, which resulted in lower resolution in ^{15}N dimension and hence larger S/N ratio for peaks that encompass several un-resolved resonances (*e.g.*, two populations for Ser^{80} in ^{15}N-met-G-CSF that is resolved in NIST 900, but not in the data from Health Canada). ***Non-resolved peak for Ser^{80} at the low-frequency spectrometer (FDA 500). ****Number of points in the indirect dimension was twice as large as other labs. The program Mnova NMR (Mestrelab Research, Santiago de Compostela, Spain) was used for the signal-to-noise estimations.

Supplementary Table 11

Inter-laboratory variability using system suitability samples		Four fields RMSD (ppm)		Six fields RMSD (ppm)		Total number of backbone resonances	Peak-picked backbone resonances
		Average	Largest deviation	Average	Largest deviation		
^{15}N-met-GCSF	^{1}H	**0.003** ± 0.002	**0.006** (Gln^{67})	**0.006** ± 0.004	0.020 (Gln^{70})	160	127 (79%)
	^{15}N	**0.011** ± 0.006	**0.034** (Met^{121})	**0.023** ± 0.017	0.108 (Ser^{62})		
α-spectrin SH3 domain	^{1}H	**0.003** ± 0.001	**0.008** (Lys^{26})	**0.005** ± 0.003	0.011 (Glu^{17})	59	59 (100%)
	^{15}N	**0.011** ± 0.006	**0.024** (Glu^{17})	**0.028** ± 0.017	0.070 (Ser^{36})		

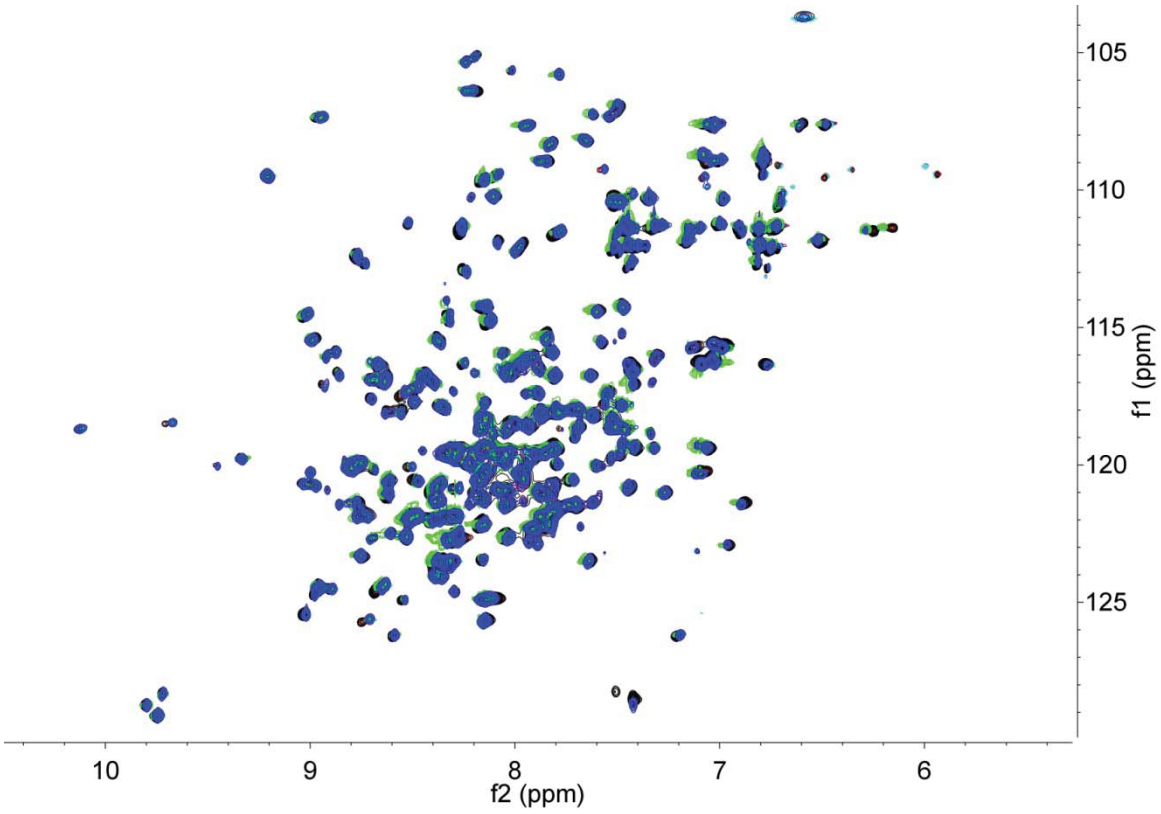
Supplementary Table 11: A table of the root-mean-squared-deviations (RMSDs) and the standard deviations of these values from the mean RMSD values from the inter-laboratory comparison of chemical shifts for the ^{1}H and ^{15}N backbone amide resonances of isotopically enriched ^{15}N-met-GCSF or SH3. Four fields: FDA 500, NIST 900, NIST 600, MPA 600. Six fields: FDA 500, NIST 900, NIST 600, MPA 600**,** HC 600, HC 700. The spectrometers HC 600 and HC 700 had different temperature settings. Total backbone resonances: total number of amino acids excluding the N-terminal residue and the prolines (13 for met-G-CSF, 3 for SH3). For the 4-field comparison, the largest deviations for backbone chemical shifts were represented by residues in the congested zones of the spectra, where precision peak-picking can be obstructed by overlapping and neighboring peaks. However, in the 6-field comparison, the largest deviations for backbone chemical shifts were represented by residues positioned in the loop regions of the proteins.

Supplementary Table 12: The intra-laboratory comparison of chemical shifts for ^{1}H and ^{15}N backbone amide resonances of the formulated drug products from the four different vendors

Intra-lab comparison of 4 drug products	**RMSD** (Amgen, Biocon, Intas, Dr. Reddy's) (ppm)				**CCSD** (Biocon, Intas, Dr. Reddy's) (ppm)		Total number of backbone resonances	Peak-picked backbone resonances
	^{1}H		^{15}N					
	Average	**Largest**	**Average**	**Largest**	**Average**	**Largest**		
NIST 900	**0.003** ± 0.001	**0.008** (Ser^{80})	**0.029** ± 0.014	**0.075** (Ser^{80})	**0.003** ± 0.002	**0.014** (Ser^{80}, Dr.Reddy's)	161	124 (77%)
NIST 600	**0.004** ± 0.002	**0.010** (Ile^{95})	**0.030** ± 0.016	**0.091** (Gln^{145})	**0.004** ± 0.002	**0.015** (Cys^{42}, Dr.Reddy's)	161	117 (73%)
HC 700	**0.004** ± 0.002	**0.010** (Gln^{20})	**0.022** ± 0.010	**0.062** (Asp^{104})	**0.004** ± 0.002	**0.018** (Gln^{20}, Intas)	161	119 (74%)
HC 600	**0.004** ± 0.002	**0.009** (Gly^{94})	**0.021** ± 0.020	**0.162** (Gln^{158})	**0.004** ± 0.002	**0.021** (Gln^{158}, Dr.Reddy's)	161	120 (75%)
FDA 500	**0.004** ± 0.002	**0.009** (Asp^{104})	**0.026** ± 0.016	**0.086** (Asp^{104})	**0.004** ± 0.003	**0.014** (Ser^{80}, Dr.Reddy's)	161	114 (71%)
MPA 600	**0.006** ± 0.002	**0.011** (Glu^{98})	**0.023** ± 0.015	**0.087** (Leu^{161})	**0.005** ± 0.003	**0.014** (Met^{126}, Dr.Reddy's)	161	119 (74%)

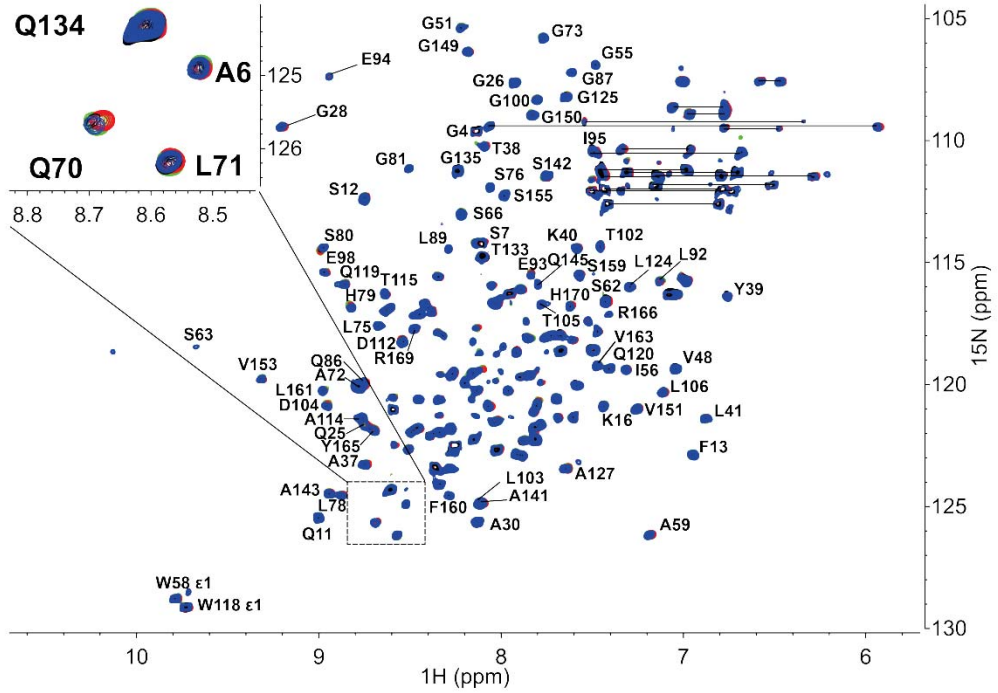
Supplementary Table 12: A table of the mean root-mean-squared-deviations (RMSDs) and the standard deviations of these values from the inter-laboratory comparison of chemical shifts for ^{1}H and ^{15}N backbone amide resonances of the formulated drug products from the four different vendors (Amgen, Biocon, Intas and Dr. Reddy's). All 2D ^{1}H,^{15}N-HSQC spectra of ^{15}N-met-G-

CSF were referenced through Gly[73]. *Root Mean Square Deviation*, $RMSD_i=\sqrt{(1/4)}\sqrt{(\delta_i^2{}_{Amgen} + \delta_i^2{}_{Biocon} + \delta_i^2{}_{Intas +} \delta_i^2{}_{DrReddy's})}$, where *i* is either ^1H or ^{15}N, has units of ppm. *Combined Chemical Shift Difference* is calculated for each of the biosimilar product purchased from overseas market, $CCSD_i=\sqrt{[0.5\ (\Delta_i\delta_H^2 + (\alpha\ \Delta_i\delta_N)^2)]}$, where $\Delta_i\delta$ is the chemical shift difference between the originator product (Amgen filgrastim) and the biosimilar product.

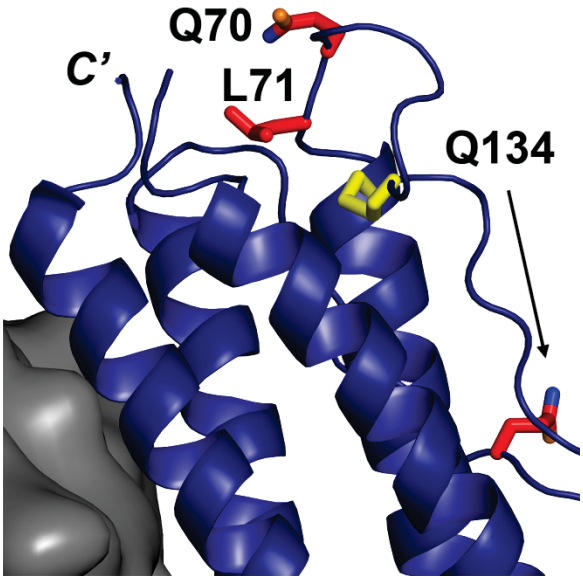


Supplementary Figure 1: A plot of the overlay of 0.2 mM ^{15}N-met-G-CSF 2D-^{1}H,^{15}N-HSQC NMR data. ^{1}H,^{15}N-HSQC spectra are from four fields and six instruments, NIST 900 (purple), NIST 600 (cyan), Health Canada 700 (black), Health Canada 600 (red), FDA 500 (green) and MPA Sweden 600 MHz (blue).

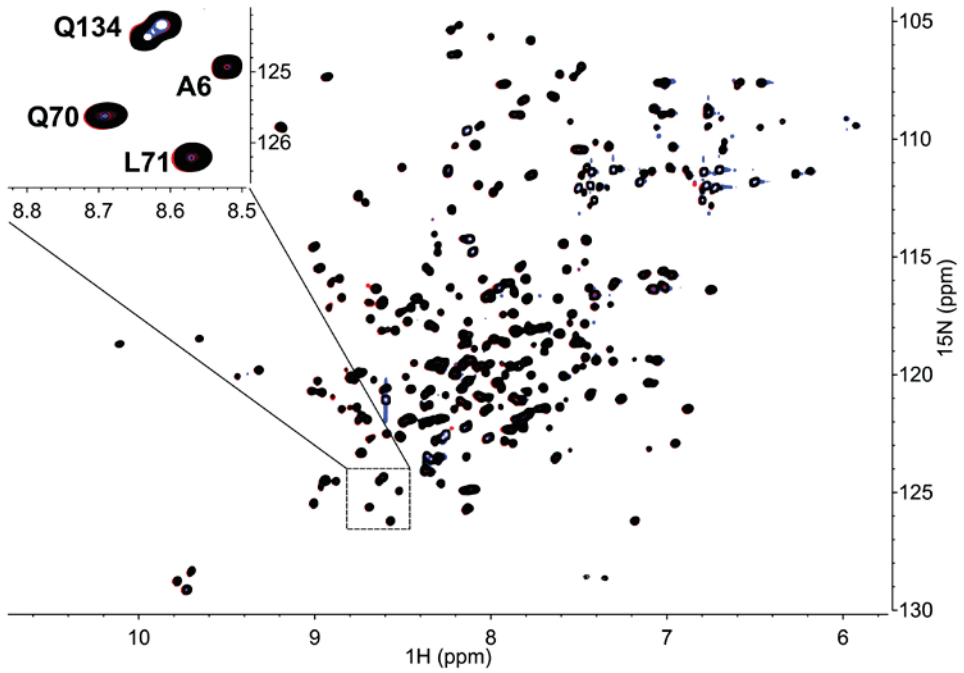
Supplementary Figure 2a



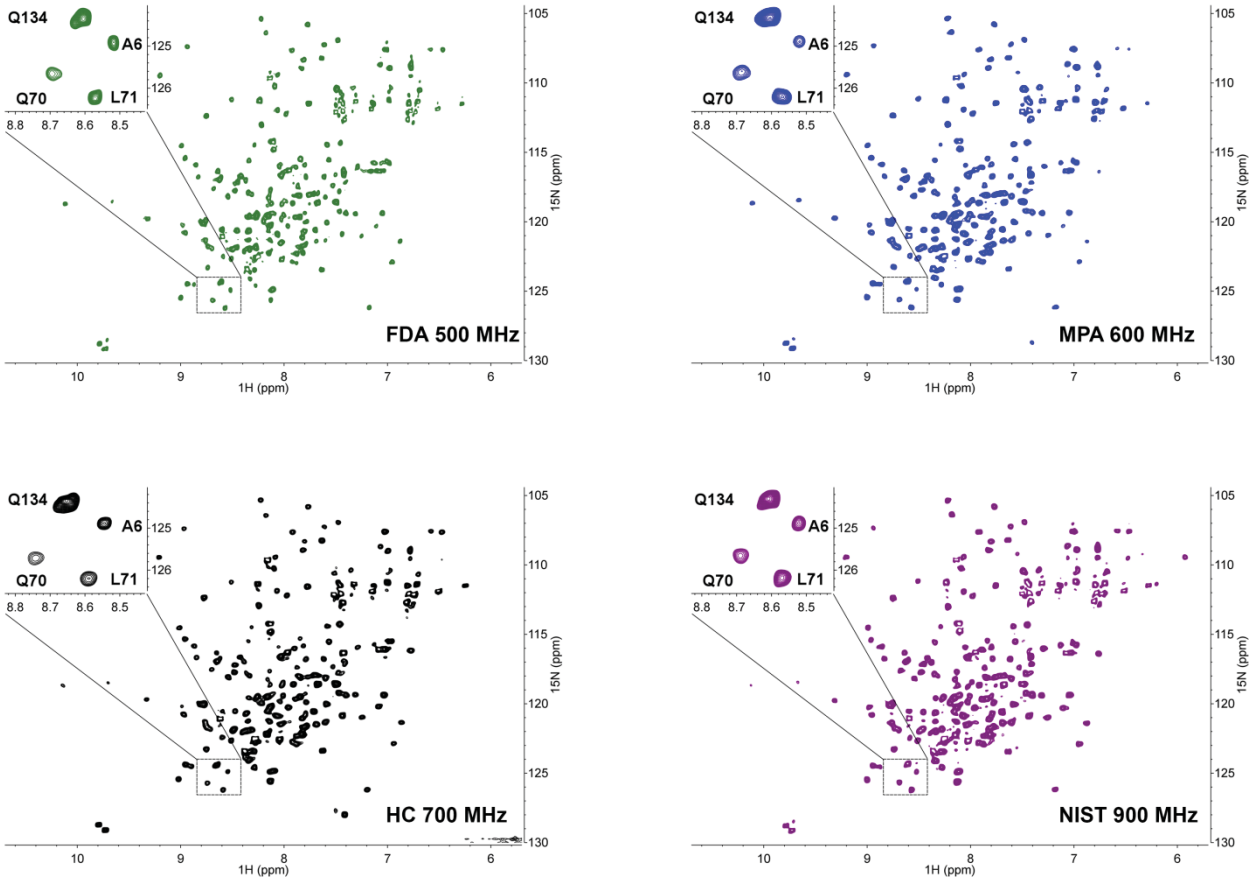
Supplementary Figure 2b



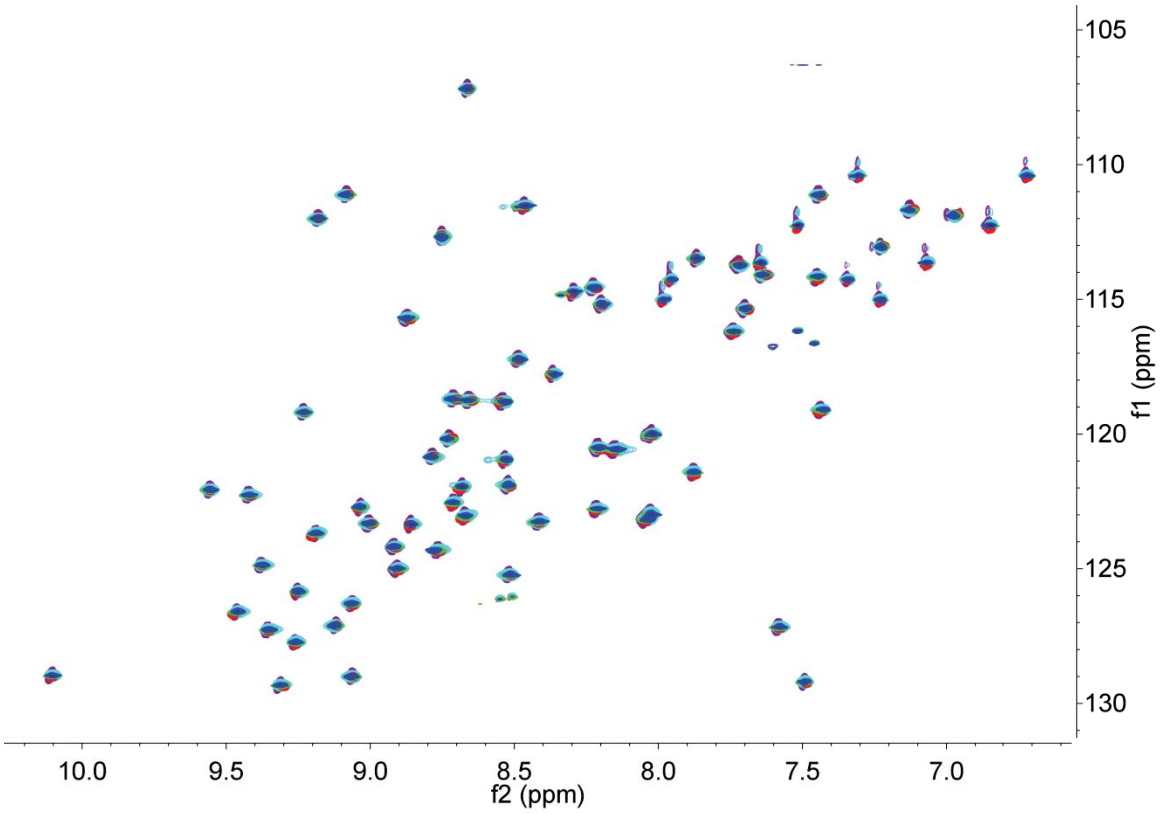
Supplementary Figure 2c



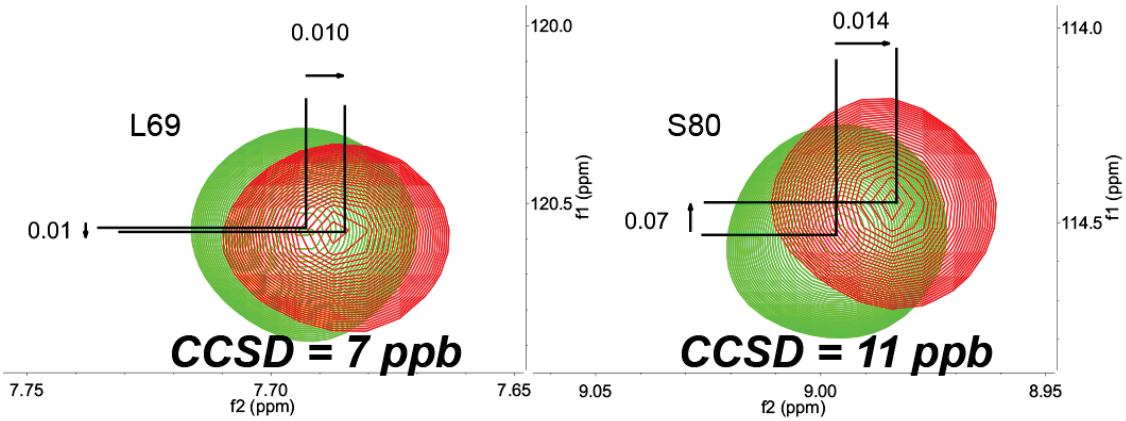
Supplementary Figure 2: Panel (**a**) shows a plot of the overlay of the 900 MHz $^{1}H,^{15}N$-HSQC spectra for the 6 ppm to 10 ppm (^{1}H) and 100 ppm to 130 ppm (^{15}N) region of spectra acquired on four formulated drug products at approximately 1 mM concentration: Neupogen® (Amgen, green), Nufil Safe™ (Biocon, red), Neukine® (Intas, black), and Grafeel™ (Dr. Reddy's, blue). Signals from selected residues are annotated. Panel **(b)** shows the four representative residues that displayed deviations from average chemical shifts mapped onto the structure of filgrastim (pdb accession code: 2D9Q). In general, residues in loops regions deviated due to the temperature mis-calibration. The disulfide bond is shown in yellow sticks (Cys^{64}-Cys^{74}), and gray surface signifies the binding site on the receptor. Panel **(c)** gives the temperature corrected map of the $^{1}H,^{15}N$-HSQC spectra from Health Canada overlaid with the NIST 900 data for the ^{15}N-labeled filgrastim system suitability sample.



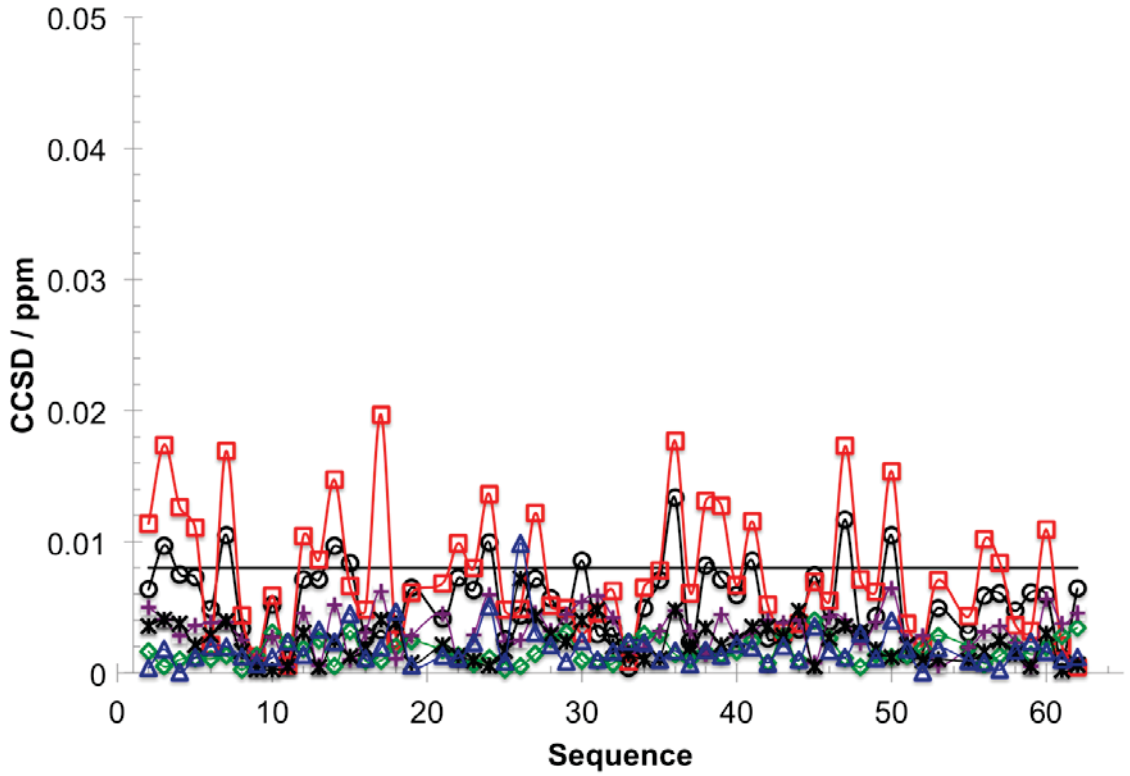
Supplementary Figure 3: 2D $^{1}H,^{15}N$-HSQC NMR spectra of the originator formulated drug product (Amgen filgrastim) from four fields; NIST 900 (purple), Health Canada 700 (black), FDA 500 (green) and MPA Sweden 600 (blue). The region containing peaks Q134, A6, Q70 and L70 is boxed and shown in an expanded view.



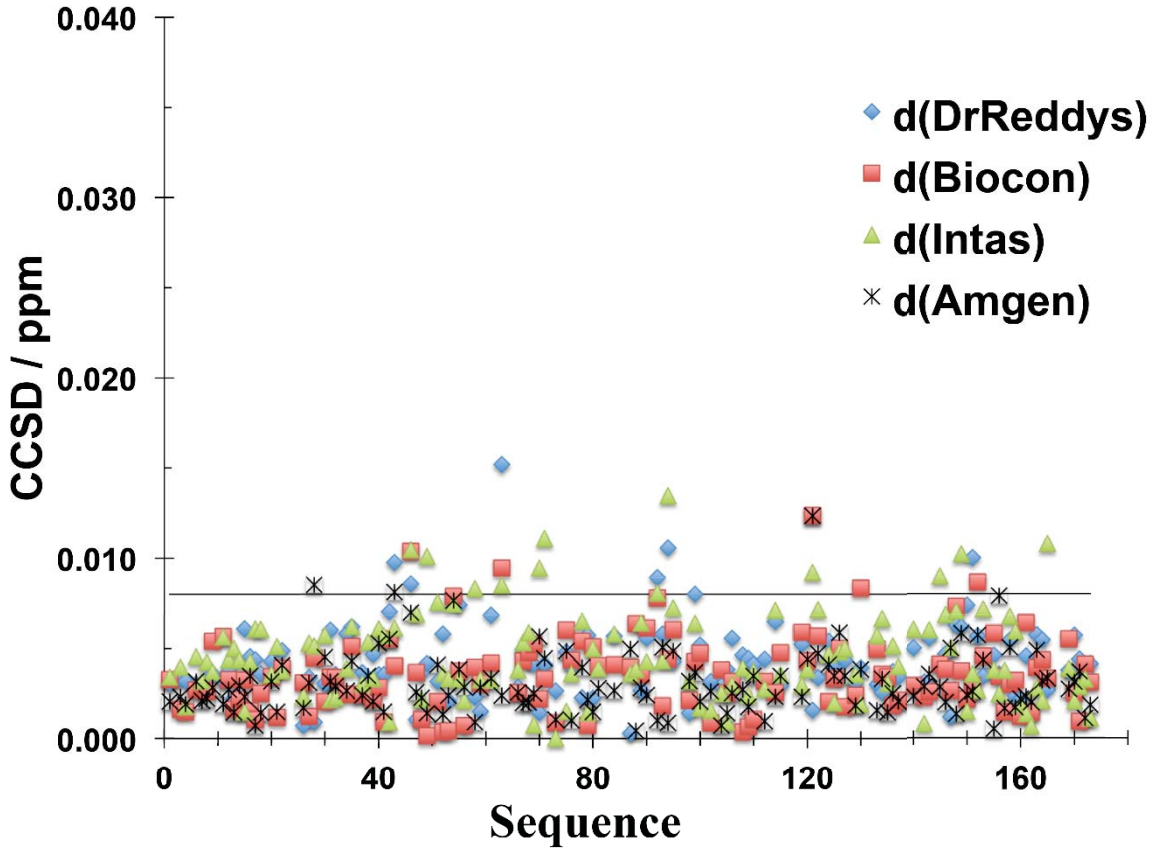
Supplementary Figure 4: A plot of the overlay of $^{1}H,^{15}N$-HSQC spectra of SH3 domain from five instruments, NIST 900 (purple), NIST 600 (cyan), Health Canada 700 (red), MPA Sweden 600 (blue), and FDA 500 MHz (green).



Supplementary Figure 5: An illustration of $^{1}H,^{15}N$-HSQC cross peak chemical shift differences and the corresponding CCSD values obtained from two cross peaks (Leu^{69} or Ser^{80}) from the overlay of Neupogen® (Amgen, green) and Nufil Safe™ (Biocon, red) from data collected on a 600 MHz spectrometer. Left: The overlay of peaks for Leu^{69} where the deviation of the CCSD value between the two cross peaks is approximately 7 ppb. Right: The overlay of peaks for Ser^{80} where the CCSD deviation between the two cross peaks is approximately 11 ppb. CCSD values greater than 8 ppb are deemed statistically significant.



Supplementary Figure 6: Combined chemical shift difference (CCSD) plot for the model protein, SH3 domain, as a function of sequence. NIST 900 (purple plus), NIST 600 (black star), HC 700 (black circle), HC 600 (red square), FDA 500 (green diamond), MPA 600 (blue triangle). Reference chemical shift was based on average shift from NIST, FDA and MPA. The horizontal bar indicated the established experimental precision limit of 8 ppb. Here, the SH3 protein is less sensitive (*i.e.* the CCSD deviations are smaller) to the 2°C or 4°C temperature difference of the HC 700 and HC 600 data than that observed for the data collected on the ^{15}N-met-G-CSF sample with the same two spectrometers.



Supplementary Figure 7: CCSD analysis from NIST 900 data recollected after 1 year. The horizontal line reflects the experimental precision at 8 ppb. Very few signals are above the experimental precision despite the age of the samples.