# Characterization of PEGylated Asparaginase: New Opportunities from NMR Analysis of Large PEGylated Therapeutics

# Characterization of PEGylated asparaginase: new opportunities from NMR analysis of large pegylated therapeutics

Linda Cerofolini[a], Stefano Giuntini[a,b], Azzurra Carlon[a], Enrico Ravera[a,b,*], Vito Calderone[a,b], Marco Fragai[a,b,*], Giacomo Parigi[a,b], Claudio Luchinat[a,b,*]

[a] Magnetic Resonance Center (CERM), University of Florence and Consorzio Interuniversitario Risonanze Magnetiche di Metallo Proteine (CIRMMP), Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

[b] Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

[*]To whom correspondence should be addressed. Email: luchinat@cerm.unifi.it, fragai@cerm.unifi.it, ravera@cerm.unifi.it

**Abstract**

Resonance assignment and structural characterization of pharmacologically-relevant proteins promise to improve their understanding and safety by rational design. However, the PEG coating that is used to evade the immune system also causes these molecules to "evade" the standard structural biology methodologies. We here demonstrate that it is possible to obtain the resonance assignment and a reliable structural model of large PEGylated proteins through an integrated approach encompassing NMR and X-ray crystallography.

## Introduction

The advances in biotechnology and chemical biology have caused a revolution in drug discovery, increasing the number of biological drugs (biologics hereafter) that are approved every year, encompassing a wide range of diseases.[1,2] Several biologics (e.g.: enzymes, antibodies, cytokines and growth factors) have been structurally characterized by X-ray crystallography,[3–5] NMR spectroscopy[6] and, more recently, by cryo-electron microscopy (Cryo-EM).[7] Far more problematic is the structural characterization and the analysis of the biological fate of "stealth biologics", where the protein component is coated with polyethylene glycol (PEG) chains to evade immune recognition and renal clearance, improving pharmacokinetics and safety. The PEG coating prevents the growth of crystals for X-ray diffraction and brings the hydrodynamic volume of the protein beyond the size limits for solution NMR.[8–10] Indeed, only small monopegylated proteins escape these limitations for either X-ray diffraction[9] or NMR.[9,11] These problems notwithstanding, the characterization of large PEGylated biologics is becoming a requirement to bring new therapeutics to the market, and a challenge for toxicology and structural pharmacology.[12–15] It has been shown that: i) protein, ii) PEG typology, and iii) linkage affect the biological fate of PEGylated proteins in vivo.[16–18] In particular, the protein component of PEGylated biologics undergoes proteolytic degradation mainly after cellular uptake allowing the presentation of the fragments to the immune system with the subsequent generation of specific antibodies.[16] The availability of the resonance assignment of the protein component and a reliable experimentally-derived structural model are becoming important steps to evaluate the degree of conjugation of the target residues[19] and the shielding of the protein surface. In turn, this information, integrated with biological activity data, may help the rational design of more proteolysis-resistant mutants/conjugates with the same structural and functional properties but better safety profiles.

We have recently shown that PEGylated proteins, large polysaccharide-protein conjugates and protein grafted onto PEGylated gold nanoparticles, for which solution NMR characterization is difficult, yield high quality solid-state NMR spectra (SSNMR hereafter).[19–21] Our analysis, based on the simple comparison of $^{13}C$-$^{13}C$ two-dimensional SSNMR spectra, confirmed that the three-dimensional structure of the protein is maintained after PEGylation on a set of pharmacologically-relevant proteins, including *E.coli* L-asparaginase II (ANSII hereafter).[20] The latter is one of the oldest biologics approved for clinical use, being administered intravenously against acute leukemia.[22] The native enzyme consists of four identical subunits, forming a dimer of dimers of 138 kDa with $D_2$ symmetry.[5] The active site of each subunit sits at the interface between the two monomers forming each of the two intimate dimers (A/C and B/D). Each subunit consists of two

domains, totaling 326 amino acids (34.5 kDa), and contains 22 lysine residues. When PEGylated, this system does not crystallize and also falls out of the boundaries of solution NMR. Therefore, it is an ideal candidate for attempting extensive resonance assignment and collection of structural restraints by SSNMR.

With this complex experimental task ahead, we have decided to pursue a highly integrated approach, as integration is the only way to derive sensible structural information from sparse and potentially unrelated experimental data.[23,24]

For the structural characterization of large PEGylated proteins/assemblies we will rely upon an integrated use of solid-state and solution NMR and X-ray crystallography. The proposed strategy is the following: i) SSNMR spectra are recorded on crystals of the same crystal form that is used to generate the X-ray structure of the native protein (ANSIIcrystal), as well as on the PEGylated protein in the form of a sediment (PEG-ANSIIsediment), and the two datasets are compared; ii) a combination of solid-state and solution NMR experiments is used for extensive resonance assignment; iii) structural restraints are obtained by NMR, and used to confirm the fold of both the native protein in solution and of the sedimented PEGylated protein, and iv) a structural model is obtained by restraints-driven docking calculation, validated against RDCs measured on the native protein.

## Results and discussion

*Biophysical characterization and enzymatic activity.*

After PEGylation the protein retains a tetrameric structure as shown by analytical gel filtration (Figure S1) and size-exclusion chromatography equipped with multiangle light scattering and quasi-elastic light scattering detectors (SEC-MALS-QELS). The molecular weight estimated by SEC-MALS-QELS analysis for the PEGylated protein is consistent with the conjugation of about 29 PEG chains per tetramer (about 7 per monomer, Figure S2). The tetrameric assembly of the PEGylated ANSII is also preserved after the freeze-drying process (Figure S2). Finally, the enzymatic activity of the PEGylated protein is retained under the conditions used (see materials and methods and supporting information section, Figure S3).

*Solution NMR spectroscopy of native ANSII.*

The reorientation correlation time of the native ANSII tetramer (~ 140 kDa) is expected to be of the order of 50 ns[25] (hydroNMR) and found to be 60 +/- 2 ns by FFC NMRD[26] (see supporting information, Figure S4). Therefore, native ANSII would be expected to be beyond the practical limits

of conventional solution NMR. However, NMRD shows a component in the spectral density corresponding to a correlation time of 9 +/- 1 ns, which may reflect an intrinsic flexibility consistently with the presence of many loop regions in the structure of the native enzyme. The presence of relatively fast internal motions explains the intensity and resolution of the 2D $^{1}H$-$^{15}N$ TROSY-HSQC solution NMR spectrum of perdeuterated [U- $^{2}H$-$^{13}C$-$^{15}N$] native ANSII (ANSIIsolution) (Figure 1A). The high quality of the 2D $^{1}H$-$^{15}N$ TROSY-HSQC solution NMR spectrum notwithstanding, only 250 spin systems out of the expected 313 for the non-Proline residues are observed (~ 80%).

*Solid-state NMR spectra of ANSII under different conditions.*

We had previously obtained high quality SSNMR spectra for PEG-ANSIIsediment as well as for ANSIIcrystal.[20] We have thus extended the collection of available 2D and 3D correlation spectra for both proteins and verified an almost complete spectral identity. The quality of the 2D $^{1}H$-$^{15}N$ CP-HSQC SSNMR spectrum acquired on [U- $^{2}H$-$^{13}C$-$^{15}N$] ANSIIcrystal (Figure 2B) is similar to that of the 2D $^{1}H$-$^{15}N$ TROSY-HSQC spectrum recorded for ANSIIsolution. Also in this case the number of identified spin systems is about 80% of the total number of expected signals.

*Assessment of tertiary structure and tetrameric assembly in crystals of native ANSII used for SSNMR.*

To prove the preservation of the tetrameric assembly, assessed in solution by FFC NMRD, the crystals, used to prepare the sample for solid-state NMR measurements (ANSIIcrystal), were investigated by X-ray crystallography. The X-ray structure of L-asparaginase II in space group $P2_12_12_1$ was solved in the absence of any ligand of the protein such as asparagine, glutamine or their respective acids (see supporting information). The overall fold and quaternary structure are superimposable to those of the other structures deposited in the protein data bank; the only significant difference with respect to several of the equivalent deposited structures is the absence, in all the four molecules present in the asymmetric unit, of residues A36 to L57, which have not been modeled since a very faint density or no density at all is present for this region.

The main chain RMSD of Cα atoms between the structure solved in our laboratory (PDB code: 6EOK) and three (1NNS,[27] 3ECA[5] and 1JAZ[28]) of the several structures, available in the protein data bank, is generally low (Figure S5). The above mentioned region, not accounted by any electron density, corresponds to a long loop (the so called “lid”) which closes the catalytic site and clearly experiences a large flexibility in solution.[27] It is interesting to point out that our protein and 1JAZ have been crystallized in the absence of the substrate/product (asparagine/aspartate), while the structures 1NNS and 3ECA have been obtained from ANSII co-crystallized with aspartate. The

different experimental conditions used to crystallize the protein explain the different degree of definition of the electron density in the "lid" region, as the interaction with the ligand likely contributes to the stabilization of the loop. In this respect, the 1NNS structure shows a very well-defined density for all the residues of the loop in both molecules in the asymmetric unit, while for the structure 3ECA the electron density of this region is significantly less well-defined. Finally, in our structure (6EOK) and in 1JAZ the electron density for this loop region is often missing, in agreement with the presence of a crystallographic disorder. Also the different space group observed in these structures and the related different packing interactions seem to play a role in these observed structural differences. These interactions stabilize the "lid" in 1NNS, and to a minor extent in 3ECA (where some sort of stabilization could be granted by the presence of aspartate). Conversely, in our crystal structure (6EOK) neither such packing interactions occur with symmetry-mate molecules, nor aspartate is present with its stabilizing effect, thus the loop shows a very high structural heterogeneity; in this respect, our structure is probably closer to the situation present in solution. A very similar situation applies to 1JAZ where neither favorable crystal packing contacts nor the presence of aspartate contribute to stabilize the "lid".

The structure solved in our laboratory also shows the clear presence of a metal bound to D122 in each of the four molecules of the asymmetric unit (as also observed for 1JAZ). This metal has been accounted for as zinc, since $ZnCl_2$ is present in the crystallization mother liquor. The binding seems quite specific with a full occupancy, despite the absence of a biological meaning for this.

As far as successive modelling steps are concerned, the 3ECA structure was used because it is more complete than the structure obtained in this study and displays less crystal packing interactions with respect to 1NNS. In particular, chain A in the 3ECA structure was used as input coordinates for docking calculations.

*A combination of solution and solid-state NMR spectroscopy yields extensive resonance assignment.*

Assignment of the spectra of native and PEGylated ANSII was achieved through the combined analysis of solution and SSNMR spectra acquired on both preparations of the native and PEGylated protein (see Table S3).

We started from the analysis of the triple resonance solution NMR spectra acquired on perdeuterated samples of the native protein (ANSIIsolution). However, the lack of 20% of the spin systems and the overlap of some signals made the analysis difficult and prevented the complete signal assignment (BMRB code: 27588).

Having established the substantial identity between the NMR spectra of ANSIIsolution and as ANSIIcrystal (see above), and the preservation of the tetrameric assembly, the NMR assignment obtained from the inspection of the solution NMR spectra was thus complemented with the analysis of the three-dimensional proton-detected SSNMR spectra (Figure S6) acquired on ANSIIcrystal (BMRB code: 27589). The use of SSNMR spectra allowed for the identification of some spin systems missing in the solution NMR spectra. These spin systems correspond mainly to residues that are located on rigid secondary structure elements, and are therefore lost in solution NMR spectra because of unfavorable relaxation properties. Conversely, few other residues, located in flexible regions, could be identified only in solution NMR spectra.

The analysis of the difference in chemical shift between solution and SSNMR for the non-PEGylated protein indicates that the residues that experience the largest perturbations in the SSNMR spectra are located on the protein surface or on loops (Figure S7), as expected because of the effect of crystal packing forces[29] on the X-ray structure. Some of the residues exhibiting the largest perturbations (C127, K284, and the C-terminus T333-Q343) are indeed located at the interface between different tetramers (Figure S8).

The $H^N$ signals of the backbone atoms of the PEGylated protein are not detectable by solution NMR performed on the perdeuterated sample, because of the increase in hydrodynamic volume of the protein after functionalization; however, 13 new signals, corresponding to the amide groups generated after PEGylation, are visible in the 2D $^{1}H$-$^{15}N$ TROSY-HSQC solution NMR spectrum of PEG-ANSII (Figure S9).[20] This confirms that solution NMR and SSNMR are really complementary for the investigation of PEGylated systems of this size. The $^{1}H$-$^{15}N$ CP-HSQC SSNMR of PEG-ANSIIsediment (Figure 1C) is of high quality. The assignment of the spectra of the PEGylated protein was obtained by comparison of the 2D spectra with those of ANSIIcrystal, and confirmed by the analysis of the 3D (H)CANH acquired on PEG-ANSIIsediment. Carbon-detected SSNMR spectra (2D NCA, 2D DARR, 3D NCACX and 3D CANCO), acquired on a sample of PEGylated ANSII [U-$^{13}C$, $^{15}N$] were also used in this analysis to extend the assignment (BRMB code: 27590). Collectively, 74% of the resonances of ANSII (in both native and PEGylated forms) have been assigned by combining the information obtained from solution and solid state spectra.

The analysis of the difference in chemical shift between ANSIIcrystal and PEG-ANSIIsediment indicates that the residues that experience the largest perturbations are located on the protein surface close to the exposed lysine residues that are most likely functionalized with PEG moieties (Figure S10, panel A and B). In PEG-ANSIIsediment, residues experiencing crystal packing contacts in ANSIIcrystal do not revert to the values observed for native ANSII in solution. This behavior could be associated to high compaction in the sediment.

The analysis of the conjugation pattern and the evaluation of the conjugation degree of each lysine residue has been achieved by integrating solution and solid-state NMR spectroscopy.[19] The spectra were collected on a PEGylated sample of ANSII where only Lys residues were isotopically enriched in $^{13}C$ and $^{15}N$ [U- $^{2}H$, Lys-$^{1}H$, $^{13}C$, $^{15}N$]. The $H_\zeta$-$N_\zeta$ new cross-peaks corresponding to eight lysine residues functionalized with PEG chains could be assigned in the 2D $^{1}H$-$^{15}N$ TROSY-HSQC of PEG-ANSII in solution, from the analysis of 3D NOESY spectra. The intensities of the signals have been used as a parameter to evaluate the functionalization degree, and combined with the changes in the relative intensity of the cross-peaks correlating the $C_\alpha$ and the $C_\varepsilon$ of the lysine sidechains in the 2D $^{13}C$-$^{13}C$ DARR SSNMR spectrum of the native and PEGylated protein. The analysis, now reported in the supporting information and summarized in Figure S10, reveals that K71, K101, K251, K336 and an unassigned lysine (see supporting information) experience the largest conjugation degree. This approach allows only an estimate of PEGylation degree as dynamic properties influencing signal intensities are not taken into account.

*RDC-based analysis of the quaternary organization of native ANSII in solution.*

To prove the preservation of the reciprocal orientation of the monomers within the tetrameric assembly present in the X-ray structure, a set of residual dipolar couplings have been collected in solution on a sample of deuterated protein before PEGylation (ANSIIsolution), using filamentous phages Pf1 as external orienting medium. The alignment tensor was determined from the best fit of selected RDCs from secondary structure elements (18 RDCs) to the 3ECA[5] structure using the FANTEN web-interface[30] (Table S4). Because of the $D_2$ symmetry of the protein, the tensor axes must be directed along the $C_2$ symmetry axes, and this has been experimentally verified (Figure S11) [footnote: The tensor was calculated both for the individual monomer and for the tetrameric structure. The fit of RDCs appears of good quality (Qfactor of 0.14 in both cases, Figure S12) and the tensor parameters and orientation appear very similar for the monomeric and the tetrameric structures], thus confirming the preservation in solution of the inter-domain arrangement of the monomer as well as the symmetric quaternary organization present in the X-ray structure.

*Calculation of the structural model of the PEGylated ANSII.*

The number of resonances present in the spectra, and the fact that no signal splitting is observed, are both consistent with the preservation of the $D_2$ symmetry of the ANSII tetramer. The availability of the resonance assignment of SSNMR spectra yielded a set of long-range interdomain/intermonomer restraints from the cross-peaks in the $^{13}C$-$^{13}C$ correlation spectra of [U-$^{13}C$, $^{15}N$] PEG-ANSIIsediment (Figure 2A). The number of restraints is however significantly limited (29) by the signal crowding that affects the 2D $^{13}C$-$^{13}C$ correlation spectra. Eight of the assigned long-

range restraints are between the two domains of each monomer, while the other twenty-one can only be attributed to distances between different monomers. Indeed, this number of collected restraints is not enough to calculate the structure of the PEGylated ANSII *de novo*. However, because SSNMR has already confirmed the preservation of the tertiary fold, SSNMR inter-subunit restraints can be confidently used in an experimentally driven docking calculation to assess the preservation of the native tetrameric assembly of the enzyme after PEGylation.

Among the 21 inter-monomer distance restraints, 10 define the interface between the two monomers (chain A with C, and chain B with D) forming the intimate dimer; whereas the remaining 11 define the contacts between the dimers (chain A with D, and chain B with C) (Figure 2B). These inter-monomer experimental restraints confirm that the tetrameric organization of the protein is maintained after PEGylation. Since most of the residues, for which the assignment is missing, pertain to the interface between chains A and B (and between chains C and D), no direct contact between these two monomers could be identified. The HADDOCK calculation, performed using these experimental restraints and the imposition of the symmetry between the monomers, clearly yielded only one cluster fulfilling the distance restraints (see Table S5). The four best models (in terms of HADDOCK score) in this cluster are reported in Figure 3A. The global RMSD between the structure with the lowest energy and the crystal structure 3ECA (used as input coordinates in the docking) is 0.97 Å, and 1.18 Å with respect to the crystal structure solved in the current paper (6EOK).

The structural model generated by HADDOCK is fully consistent with the tetrameric arrangement observed in the crystal structure of the native enzyme, and proves that the structural features of ANSII are fully preserved upon PEGylation.

Since the preservation of the quaternary organization of the PEGylated ANSII is confirmed by NMR, it is not surprising that the RDCs calculated from the structural model with the lowest HADDOCK-score (Table S6 and Figure S13) are in good agreement with the RDCs measured in solution on the non-PEGylated enzyme.

**Conclusions**

The NMR assignment and the structural characterization of a very large PEGylated protein assembly have been here achieved by integrating solution and solid state NMR data, and complementing the NMR information with X-ray crystallography.

The structural characterization of large non-crystallizable proteins remains a challenge in structural biology. In the last years solid state NMR has emerged as a powerful technique to obtain structural information on non-crystalline protein samples such as fibrils, bioinspired materials, and membrane proteins.[20,31–36] However, only 112 structures on a total of more than 100000 protein

structures available in the Protein Data Bank have been solved by SSNMR as of October 2018. Problems in sample preparation and the relatively low sensitivity of the experiments are the main limiting factors for the assignment of the residues and the collection of the structural restraints in the solid state. By far, the best conditions to obtain a workable SSNMR dataset is to rely upon microcrystals or upon sediments. In the present work we have tested our strategy on microcrystalline non-PEGylated and sedimented PEGylated ANSII, for which we were able to collect proton-detected and carbon-detected solid state NMR spectra without the use of expensive and time-consuming selective labeling strategies. The presence in large proteins of regions experiencing a fast-motion regime may be beneficial for an extended assignment of the protein resonances, and for a more detailed characterization of the system. In the present case, the different sample conditions notwithstanding, the structural features of the native protein are preserved in the crystalline preparation. This preservation of structural features also allows for tracking the minor perturbations that result from the conjugation with the PEG chains and explains the preservation of the biological activity of the PEGylated enzyme. The possibility of verifying structural restraints across different conditions and functionalization states ensures that a more accurate model of the protein in agreement with all the experimental data can be obtained. We expect that the resonance assignment and the structural model will represent a better starting point for forthcoming research aimed to improve the safety of the investigated biologics, and that this methodology will be easily ported to many PEG-functionalized protein formulations.

## Experimental Section

### *Sample preparation*

The protein was expressed, purified and conjugated with PEG chains as described in the supporting information.

### *NMR measurements*

Solution NMR experiments for backbone resonance assignment with a TROSY scheme [37] [3D tr-HNCA, tr-HNCACB, tr-HNCO and 3D tr-HN(CA)CO] were performed on perdeuterated [U-$^{2}$H-$^{13}$C-$^{15}$N] samples of native ANSII (at concentrations ranging between 0.5 and 1 mM) in water buffer solutions [20 mM sodium phosphate, pH 7.5, 0.02% $NaN_3$, 0.1 mg/mL protease inhibitors (Pefabloc)]. For 3D tr-HNCO and 3D tr-HN(CA)CO NUS at 33% was used. All the spectra were recorded at 310 K on Bruker AVANCE MHD NMR spectrometers operating at 900 and 950 MHz and equipped with triple resonance cryo-probes. A 3D $^{1}$H-$^{15}$N NOESY-TROSY spectrum (mixing time 100 ms) was also acquired. Two-dimensional carbon-detected solution NMR spectra [$^{13}$C-$^{15}$N

CON, $^{13}C$-$^{15}N$ CACO, $^{13}C$-$^{15}N$ CBCACO and $^{13}C$-$^{13}C$ NOESY (mixing time of 1 sec)][38] were acquired on a Bruker AVANCE 700 spectrometer equipped with a triple-resonance Cryo-Probe optimized for $^{13}C$-direct detection.[39]

One-bond $^{1}H$-$^{15}N$ couplings were measured as the difference in the peak position between the TROSY spectrum and the conventional decoupled $^{1}H$-$^{15}N$ HSQC spectrum,[40,41] and the RDCs were calculated as the difference in the splitting of the one-bond $^{1}H$-$^{15}N$ couplings in partially aligned and isotropic samples of native ANSII [U- $^{2}H$-$^{13}C$-$^{15}N$] at 304 K. Filamentous phages Pf1,[42] (ASLA biotech), at the concentration of 22.5 mg/mL, were used as alignment medium. Deuteration removes the artifacts due to cross-correlation.[43]

3D tr-HNCA (NUS 30%), 3D tr-HNCACB (NUS 40%) and 2D $^{13}C$-$^{13}C$ FLOPSY spectra were acquired on a perdeuterated sample of PEGylated ANSII [U- $^{2}H$-$^{13}C$-$^{15}N$] in the same experimental conditions on 950 and 700 MHz spectrometers, respectively.

3D $^{1}H^{15}N$ NOESY (mixing time 100 ms) and 3D $^{1}H^{1}H^{1}H$ NOESY-NOESY (100 ms for both mixing times) NMR spectra have been also acquired on solutions of [U $^{2}H$, Lys $^{1}H$, $^{13}C$, $^{15}N$] ANSII as native protein (1 mM) and functionalized with the PEG chains (0.5 mM) on 900 and 950 MHz spectrometers.

NMRD profiles were acquired on a Stelar FFC Spinmaster 2000 Relaxometer as described previously[26].[26]

SSNMR experiments were performed both on the crystalline preparation of native ANSII and on the PEGylated preparation of the protein, obtained with rehydration of freeze-dried material as previously reported.[19,20] All the spectra were recorded at ~ 280 K on a Bruker Avance III 850 MHz wide-bore spectrometer (20 T, 213.6 MHz $^{13}C$ Larmor frequency), equipped with 3.2 and 1.3 mm DVT MAS probe heads in triple-resonance mode. The inter-scan delay was set to 2.2 s in all the experiments.

Three pairs of $^{1}H$-detected SSNMR spectra [3D (H)CONH/3D (H)CO(CA)NH, 3D (H)CANH/3D (H)CA(CO)NH and 3D (H)(CA)CB(CA)NH/(H)(CA)CB(CACO)NH] were acquired at fast MAS frequency of 60 kHz, using the standard parameters reported in literature.[45,46] For some of these experiments [3D (H)CO(CA)NH, 3D (H)CA(CO)NH, 3D (H)(CA)CB(CA)NH and (H)(CA)CB(CACO)NH] the NUS approach (NUS at 33%) was implemented to reduce the total acquisition time. The nonselective 90° pulses were set to 2.6 μs at 96 kHz rf-field amplitude ($^{1}H$), 7.2 μs at 35 kHz rf-field amplitude ($^{15}N$), and 3.5 μs at 71 kHz rf-field amplitude ($^{13}C$). Experimental details are given in table S7.

Standard $^{13}C$-detected SSNMR spectra were acquired at MAS frequency of 14 kHz on the sample of PEG-ANSIIsediment [U- $^{13}C$, $^{15}N$], using the pulse sequences reported in the literature.[47]

Pulses were 2.6 μs for $^{1}H$, 4.4 μs for $^{13}C$ and 7.2 μs for $^{15}N$. Experimental details are given in table S8.

All the spectra were processed with the Bruker TopSpin 3.2 software package and the NMRPipe system,[48] and analyzed with the program CARA.[49]

*Structural modelling of ANSII tetramer.*

Models of the tetrameric arrangement of ANSII were obtained performing multi-body docking calculations with the program HADDOCK 2.2.[50] The structure of chain A in the 3ECA[5] structure (which is more complete than the structure obtained in this study and displays less crystal packing interactions with respect to 1NNS) was used as input coordinates, for each monomeric unit in the calculations. The four monomers were docked to one another using all the assigned long-range experimental restraints between the A/C (10 restraints) and A/D (11 restraints) interfaces. The lower distance cut-off was set to 3.0 Å, and the upper to 6.0 Å for all the distance restraints. Non-crystallographic symmetry restraints and $C_2$ symmetry were imposed for all the dimeric interfaces (A/B, A/C, A/D, B/C, B/D, C/D), using the same protocol for structural calculations of symmetric protein dimers. The histidine protonation states were automatically determined by the Molprobity module embedded in the HADDOCK server. During the rigid docking calculations, 1000 complexes were generated, then 200 structures were selected for the semi-flexible simulated annealing in torsion angle space, and finally refined in Cartesian space with explicit solvent.

**Acknowledgements**

This work has been supported by Fondazione Cassa di Risparmio di Firenze, MIUR PRIN 2012SK7ASN and COST Action CA15209 “European Network on NMR Relaxometry”. The authors acknowledge the support and the use of resources of Instruct-ERIC, a landmark ESFRI project, and specifically the CERM/CIRMMP Italy center, as well as the EC Contracts iNext No. 653706. The authors acknowledge the support of the University of Florence CERM-TT and the *Recombinant Proteins JOYNLAB*. The FP7 WeNMR (project No. 261572) and H2020 West-Life (project No. 675858) European e-Infrastructure projects are acknowledged for the use of their web portals. These portals make use of the EGI infrastructure and DIRAC4EGI service with the dedicated support of CESNET-MetaCloud, INFN-PADOVA, NCG-INGRID-PT, RAL-LCG2, TW-NCHC, IFCA-LCG2, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, South Africa, Malaysia, Taiwan and the US Open Science Grid.

**Keywords:** Protein structures, Protein modifications, Structural Biology, PEGylation, biopharmaceuticals

[1] R. A. Rader, *Nat. Biotechnol.* **2008**, *26*, 743–751.
[2] G. Dranitsaris, E. Amir, K. Dorward, *Drugs* **2011**, *71*, 1527–1536.
[3] T. F. Lerch, P. Sharpe, S. J. Mayclin, T. E. Edwards, E. Lee, H. D. Conlon, S. Polleck, J. C. Rouse, Y. Luo, Q. Zou, *mAbs* **2017**, 0.
[4] S. Li, K. R. Schmitz, P. D. Jeffrey, J. J. W. Wiltzius, P. Kussie, K. M. Ferguson, *Cancer Cell* **2005**, *7*, 301–311.
[5] A. L. Swain, M. Jaskólski, D. Housset, J. K. Rao, A. Wlodawer, *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90*, 1474–1478.
[6] H. Ghasriani, D. J. Hodgson, R. G. Brinson, I. McEwen, L. F. Buhse, S. Kozlowski, J. P. Marino, Y. Aubin, D. A. Keire, *Nat. Biotechnol.* **2016**, *34*, 139–141.
[7] Q. Wang, H. Yang, X. Liu, L. Dai, T. Ma, J. Qi, G. Wong, R. Peng, S. Liu, J. Li, et al., *Sci. Transl. Med.* **2016**, *8*, 369ra179.
[8] D. J. Hodgson, Y. Aubin, *J. Pharm. Biomed. Anal.* **2017**, *138*, 351–356.
[9] G. Cattani, L. Vogeley, P. B. Crowley, *Nat. Chem.* **2015**, *7*, 823–828.
[10] B. Plesner, P. Westh, S. Hvidt, A. D. Nielsen, *J. Pharm. Biomed. Anal.* **2011**, *55*, 597–602.
[11] Y.-S. Wang, S. Youngster, M. Grace, J. Bausch, R. Bordens, D. F. Wyss, *Adv. Drug Deliv. Rev.* **2002**, *54*, 547–570.
[12] A. Beck, H. Diemer, D. Ayoub, F. Debaene, E. Wagner-Rousset, C. Carapito, A. Van Dorsselaer, S. Sanglier-Cianférani, *TrAC Trends Anal. Chem.* **2013**, *48*, 81–95.
[13] Y. Lu, S. E. Harding, A. Turner, B. Smith, D. S. Athwal, J. G. Grossmann, K. G. Davis, A. J. Rowe, *J. Pharm. Sci.* **2008**, *97*, 2062–2079.
[14] S. A. Berkowitz, J. R. Engen, J. R. Mazzeo, G. B. Jones, *Nat. Rev. Drug Discov.* **2012**, *11*, 527–540.
[15] A. Beck, E. Wagner-Rousset, D. Ayoub, A. Van Dorsselaer, S. Sanglier-Cianférani, *Anal. Chem.* **2013**, *85*, 715–736.
[16] V. L. Elliott, G. T. Edge, M. M. Phelan, L.-Y. Lian, R. Webster, R. F. Finn, B. K. Park, N. R. Kitteringham, *Mol. Pharm.* **2012**, *9*, 1291–1301.
[17] T. Parton, L. King, J. van Asperen, S. Heywood, A. Nesbitt, *J. Crohns Colitis Suppl.* **2008**, *2*, 26–26.
[18] M. W. Modi, J. S. Fulton, D. K. Buckmann, T. L. Wright, D. J. Moore, *Hepatology* **2000**, *32*, 371A.
[19] S. Giuntini, E. Balducci, L. Cerofolini, E. Ravera, M. Fragai, F. Berti, C. Luchinat, *Angew. Chem. Int. Ed.* **2017**, *56*, 14997–15001.
[20] E. Ravera, S. Ciambellotti, L. Cerofolini, T. Martelli, T. Kozyreva, C. Bernacchioni, S. Giuntini, M. Fragai, P. Turano, C. Luchinat, *Angew. Chem. Int. Ed Engl.* **2016**, *55*, 2446–2449.
[21] S. Giuntini, L. Cerofolini, E. Ravera, M. Fragai, C. Luchinat, *Sci. Rep.* **2017**, *7*, 17934.
[22] J. M. Hill, J. Roberts, E. Loeb, A. Khan, A. MacLellan, R. W. Hill, *JAMA* **1967**, *202*, 882–888.
[23] H. van den Bedem, J. S. Fraser, *Nat. Methods* **2015**, *12*, 307–318.

[24] J. L. MacCallum, A. Perez, K. A. Dill, *Proc. Natl. Acad. Sci.* **2015**, *112*, 6985–6990.
[25] J. García de la Torre, M. L. Huertas, B. Carrasco, *J. Magn. Reson.* **2000**, *147*, 138–146.
[26] E. Ravera, G. Parigi, A. Mainz, T. L. Religa, B. Reif, C. Luchinat, *J. Phys. Chem. B* **2013**, *117*, 3548–3553.
[27] M. Sanches, J. A. R. G. Barbosa, R. T. de Oliveira, J. Abrahão Neto, I. Polikarpov, *Acta Crystallogr. D Biol. Crystallogr.* **2003**, *59*, 416–422.
[28] M. Kozak, D. Borek, R. Janowski, M. Jaskólski, *Acta Crystallogr. Sect. D* **2002**, *58*, 130–132.
[29] A. Carlon, E. Ravera, J. Hennig, G. Parigi, M. Sattler, C. Luchinat, *J. Am. Chem. Soc.* **2016**, *138*, 1601–1610.
[30] M. Rinaldelli, A. Carlon, E. Ravera, G. Parigi, C. Luchinat, *J. Biomol. NMR* **2014**, 1–14.
[31] S. H. Park, B. B. Das, F. Casagrande, Y. Tian, H. J. Nothnagel, M. Chu, H. Kiefer, K. Maier, A. A. De Angelis, F. M. Marassi, et al., *Nature* **2012**, *491*, 779–783.
[32] E. Ravera, T. Schubeis, T. Martelli, M. Fragai, G. Parigi, C. Luchinat, *J. Magn. Reson. San Diego Calif 1997* **2015**, *253*, 60–70.
[33] M. Fragai, C. Luchinat, T. Martelli, E. Ravera, I. Sagi, I. Solomonov, Y. Udi, *Chem. Commun. Camb. Engl.* **2014**, *50*, 421–423.
[34] T. Martelli, E. Ravera, A. Louka, L. Cerofolini, M. Hafner, M. Fragai, C. F. W. Becker, C. Luchinat, *Chem. Weinh. Bergstr. Ger.* **2015**, DOI 10.1002/chem.201503613.
[35] E. Ravera, L. Cerofolini, T. Martelli, A. Louka, M. Fragai, C. Luchinat, *Sci. Rep.* **2016**, *6*, 27851.
[36] A. Louka, I. Matlahov, S. Giuntini, L. Cerofolini, A. Cavallo, S. Pillozzi, E. Ravera, M. Fragai, A. Arcangeli, A. Ramamoorthy, et al., *Phys. Chem. Chem. Phys. PCCP* **2018**, *20*, 12719–12726.
[37] Y. Xu, S. Matthews, *Top. Curr. Chem.* **2013**, *335*, 97–119.
[38] W. Bermel, I. Bertini, I. C. Felli, Y.-M. Lee, C. Luchinat, R. Pierattelli, *J. Am. Chem. Soc.* **2006**, *128*, 3918–3919.
[39] W. Bermel, I. Bertini, L. Duma, I. C. Felli, L. Emsley, R. Pierattelli, P. R. Vasos, *Angew. Chem. Int. Ed Engl.* **2005**, *44*, 3089–3092.
[40] N. C. Fitzkee, A. Bax, *J. Biomol. NMR* **2010**, *48*, 65–70.
[41] G. Kontaxis, G. M. Clore, A. Bax, *J. Magn. Reson.* **2000**, *143*, 184–196.
[42] M. R. Hansen, P. Hanson, A. Pardi, *Methods Enzymol.* **2000**, *317*, 220–240.
[43] L. Yao, J. Ying, A. Bax, *J. Biomol. NMR* **2009**, *43*, 161–170.
[44] M. Kadkhodaie, O. Rivas, M. Tan, A. Mohebbi, A. J. Shaka, *J. Magn. Reson. 1969* **1991**, *91*, 437–443.
[45] L. B. Andreas, T. Le Marchand, K. Jaudzems, G. Pintacuda, *J. Magn. Reson.* **2015**, *253*, 36–49.
[46] E. Barbet-Massin, A. J. Pell, J. S. Retel, L. B. Andreas, K. Jaudzems, W. T. Franks, A. J. Nieuwkoop, M. Hiller, V. Higman, P. Guerry, et al., *J. Am. Chem. Soc.* **2014**, *136*, 12489–12497.
[47] A. Schuetz, C. Wasmer, B. Habenstein, R. Verel, J. Greenwald, R. Riek, A. Böckmann, B. H. Meier, *ChemBioChem* **2010**, *11*, 1543–1551.
[48] F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, A. Bax, *J. Biomol. NMR* **n.d.**, *6*, 277–293.
[49] R. Keller, *The Computer Aided Resonance Assignment Tutorial (CARA)*, CANTINA Verlag, **2004**.
[50] S. J. de Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. J. J. Bonvin, *Proteins* **2007**, *69*, 726–733.

**Figure Legends**

**Figure 1.** (A) 2D $^{1}H$-$^{15}N$ TROSY-HSQC solution NMR spectrum of [U- $^{2}H$-$^{13}C$-$^{15}N$] native ANSII, (B) 2D $^{1}H$-$^{15}N$ CP-HSQC SSNMR spectrum of crystalline ANSII [U- $^{2}H$-$^{13}C$-$^{15}N$] and (C) 2D $^{1}H$-$^{15}N$ CP-HSQC SSNMR spectrum of PEG-ANSII [U- $^{2}H$-$^{13}C$-$^{15}N$]. The 2D $^{1}H$-$^{15}N$ TROSY-HSQC solution NMR spectrum was recorded at 310 K on a Bruker AVANCE MHD NMR 950 spectrometer and the 2D $^{1}H$-$^{15}N$ CP-HSQC SSNMR spectra were recorded in 1.3 mm rotor, at 60 kHz and ~ 280 K on a Bruker Avance III 850 MHz wide-bore spectrometer.

**Figure 2**. A) Enlargement of two regions of the 2D $^{13}C$-$^{13}C$ DARR SSNMR spectrum (mixing time 200 ms) displaying long-range inter-residue restraints (blue interdimer and green interdomain/intramonomer). B) Representation of the 21 intermonomer distance restraints used in HADDOCK calculations. Ten of these restraints define the interface between the two monomers (chains A and C in wheat and pink, respectively) forming the intimate dimer (contacts between the stretches K184-T187 and R294-T300, and residue K184 with N268, D82 with Y272, M83/W88 with A243); whereas eleven define the interface between other two monomers (chains A and D, in wheat and light-blue, respectively) (contacts between the stretches Q212-A216 and A216-Q212, Y198-G199 and R213, Y198 and Y203, N206/K208 and D303). The restraints have been replicated according to the protein symmetry. C) Structural model with the lowest energy obtained with HADDOCK calculations, implementing SSNMR experimental restraints. The PEG chains are schematized as white spheres.

**Figure 3**. A) Four best structural models of the clusters of PEG-ANSII with the lowest HADDOCK-scores, obtained implementing in HADDOCK 2.2 the experimental SSNMR distance restraints. B) Main chain RMSD per residue among the best structural models of the family calculated with HADDOCK. The RMSD and the standard deviation are calculated on the four chains of the structure.

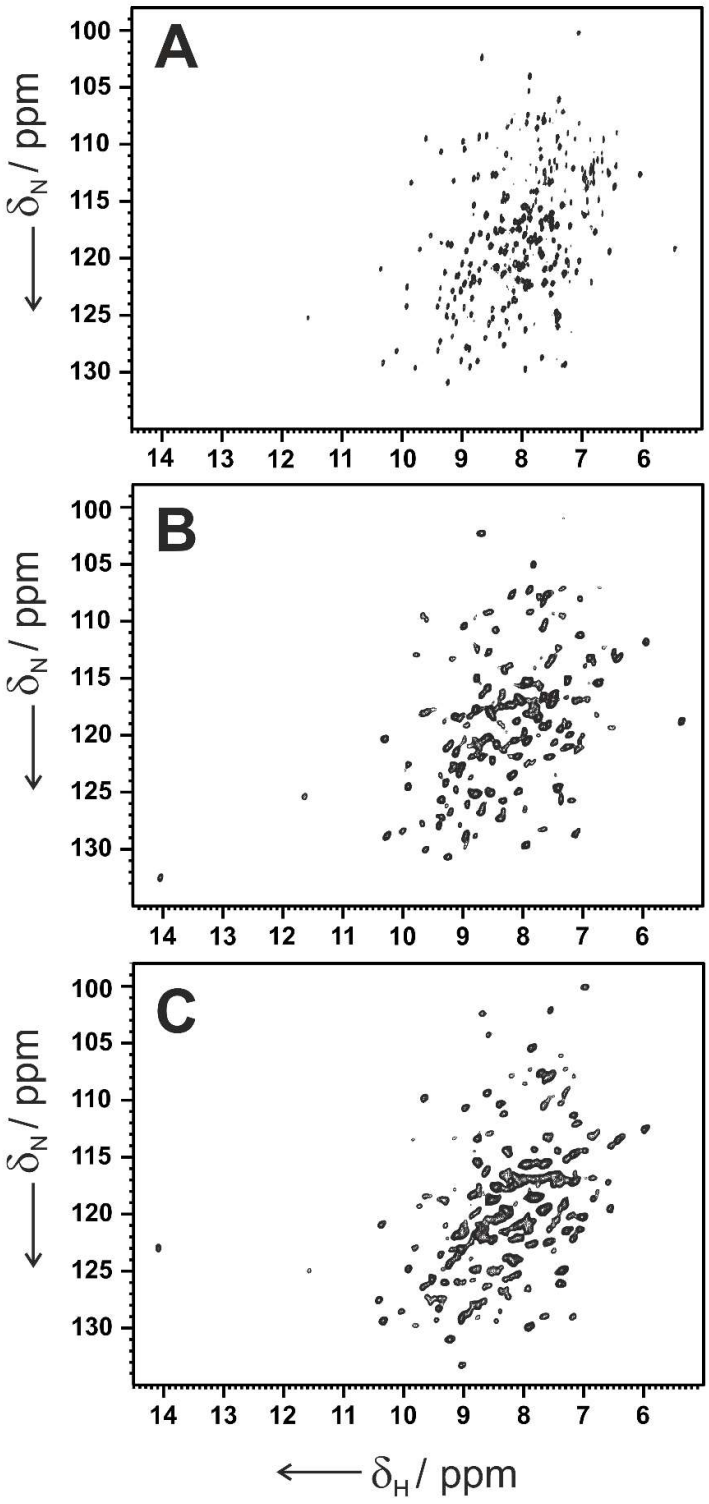


**Figure 1.**

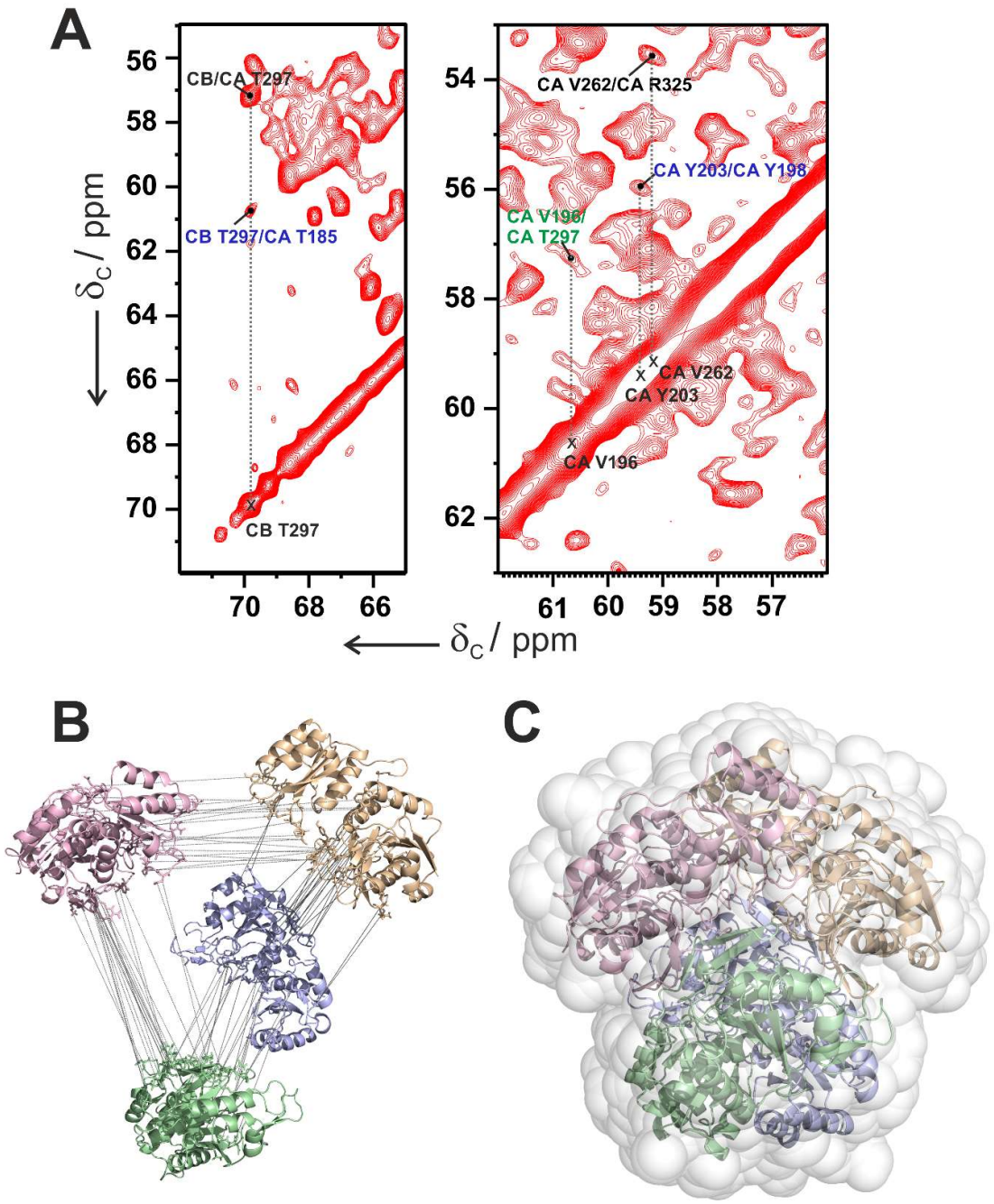
A
CB/CA T297
CB T297/CA T185
CB T297
CA V262/CA R325
CA Y203/CA Y198
CA V196/
CA T297
CA V262
CA Y203
CA V196
δc / ppm
δc / ppm
B
C

**Figure 2.**

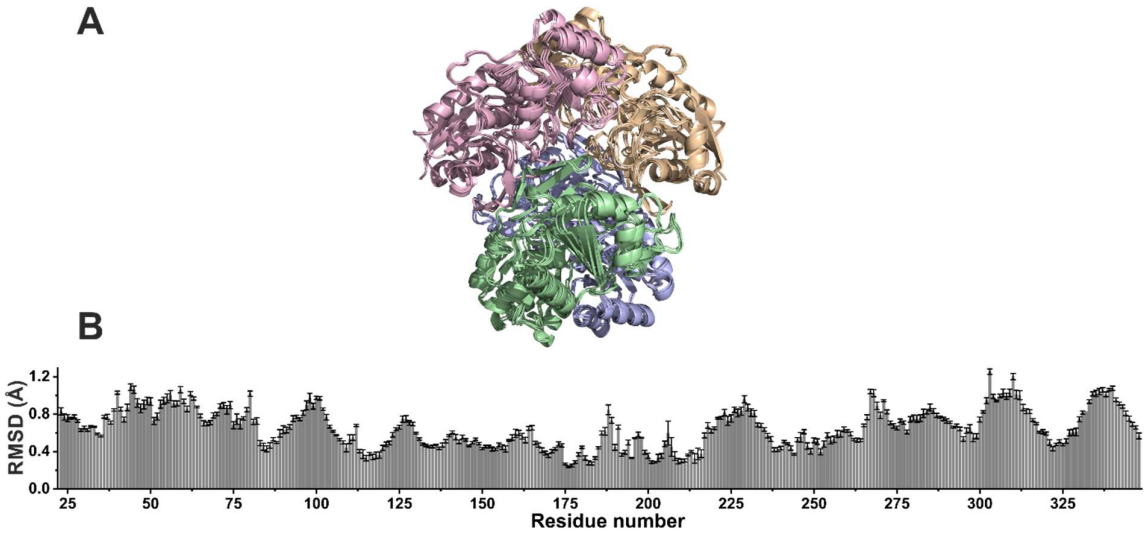
A
B
RMSD (Å)
1.2
0.8
0.4
0.0
25
50
75
100
125
150
175
200
225
250
275
300
325
Residue number

**Figure 3**.

**TOC.** Coating a protein with PEG may have structural consequences but, no single biophysical methodology can spot the possible modifications. This limitation has been overcome by the integration of solution and solid state NMR data, relaxometry and X-ray crystallography for the characterization of a PEGylated enzyme used in clinical practice.