

FROM PRIMARY SEQUENCE TO HIGHER-ORDER STRUCTURE: AN ORTHOGONAL ANALYTICAL APPROACH TO SEMAGLUTIDE CHARACTERIZATION WITH EMPHASIS ON NMR SPECTROSCOPY

Dr. G. Jithender Reddy*, Kandi Subhashini¹

*Scientist E, Department of Analytical and Structural Chemistry, CSIR- Indian Institute of Chemical Technology,
Hyderabad, India.

¹Department of Pharmacy, University College of Technology (A), Osmania University, Hyderabad, 500007,
Telangana, India.

Article Received: 25 August 2026 || Article Revised: 15 September 2026 || Article Accepted: 05 October 2026

***Corresponding Author: Dr. G. Jithender Reddy**

Scientist E, Department of Analytical and Structural Chemistry, CSIR- Indian Institute of Chemical Technology, Hyderabad, India.

DOI: <https://doi.org/10.5281/zenodo.23416259>

How to cite this Article: Dr. G. Jithender Reddy, Kandi Subhashini (2026) FROM PRIMARY SEQUENCE TO HIGHER-ORDER STRUCTURE: AN ORTHOGONAL ANALYTICAL APPROACH TO SEMAGLUTIDE CHARACTERIZATION WITH EMPHASIS ON NMR SPECTROSCOPY. World Journal of Pharmaceutical Science and Research, 5(10), 592-619.



Copyright © 2026 Dr. G. Jithender Reddy | World Journal of Pharmaceutical Science and Research.

This work is licensed under creative Commons Attribution-NonCommercial 4.0 International license (CC BY-NC 4.0).

ABSTRACT

Semaglutide is a long-acting, lipidated glucagon-like peptide-1 (GLP-1) receptor agonist whose structural characterization requires information extending beyond primary sequence and molecular mass. Its non-proteinogenic α -aminoisobutyric acid (Aib) substitution and lipid modification contribute to its stability, albumin binding, conformational behavior, and potential for concentration-, pH-, and time-dependent molecular association. This review presents an orthogonal analytical framework for the characterization of semaglutide, integrating chromatographic, mass spectrometric, spectroscopic, biophysical, and nuclear magnetic resonance (NMR) approaches. HPLC/UHPLC provides information on purity, related substances, and degradation, whereas high-resolution mass spectrometry, tandem mass spectrometry, and peptide mapping provide molecular-mass, sequence, and modification-site information. Circular dichroism and Fourier-transform infrared spectroscopy provide complementary information on global secondary-structure characteristics and conformational transitions, while small-angle X-ray scattering and cryogenic transmission electron microscopy provide information on supramolecular organization and aggregate morphology. Particular emphasis is placed on multidimensional NMR spectroscopy, including COSY, TOCSY, HSQC, HSQC-TOCSY, HMBC, and NOESY, which enables residue-specific characterization of chemical environments, through-bond and through-space interactions, conformational behavior, and molecular association in solution. NMR-derived chemical shifts can further support backbone torsion-angle prediction using TALOS-N, while NOE-derived distance restraints and dihedral restraints can be integrated into 3D-structure calculations using CYANA. Recent studies demonstrating semaglutide oligomerization, micelle formation, and pH-dependent fibrillation highlight the importance of controlling experimental conditions during structural characterization. Collectively, the available evidence supports a complementary analytical strategy in which each technique addresses a different structural question, while NMR provides an important residue-resolved connection between molecular characterization and solution-state 3D-structural analysis. Future studies integrating residue-specific NMR measurements with CD, SAXS, cryo-TEM, and molecular-dynamics simulations may provide a more comprehensive description of the conformational and higher-order structural landscape of semaglutide.

KEYWORDS: Semaglutide; GLP-1 receptor agonist; peptide characterization; higher-order structure; aggregation; NMR spectroscopy; NOESY; chemical shift assignment; TALOS-N; CYANA; SAXS; cryo-TEM; mass spectrometry.

INTRODUCTION

Peptide pharmaceuticals can be viewed as a class of highly valuable pharmaceutical compounds due to their high biological specificity and ability to bind selective receptors. However, peptide characterization is much more complicated than small molecule characterization as peptides exhibit a number of different structural levels, such as primary peptide structure, secondary structure, tertiary conformation, peptide molecular interactions, and higher order structure. Hence, a good peptide characterization strategy would require the utilization of several analytical methods in parallel, with each method contributing to its own information about the structure of peptides. For example, regulatory documents like ICH Q6B stress the importance of employing physicochemical, structural, and biological peptide characterization techniques to define the identity, purity, and structure of peptide and protein medicines.^[1,2]

Semaglutide is a chemical modification of human glucagon-like peptide-1 (GLP-1) and a GLP-1 receptor agonist. It is made up of a 31-amino-acid chain and possesses certain chemical modifications like substitution of Alanine at position 8 by α -aminoisobutyric acid (Aib), substitution of lysine at position 34 with arginine, and modification of lysine at position 26 with a fatty acid side chain.^[3] The purpose behind these modifications was to make it resistant to enzymatic degradation, enhance binding to albumin, and extend the biological half-life of the compound.^[3] Therefore, a precise characterization of semaglutide not only entails establishing its amino-acid sequence but also involves evaluating its chemical modifications and conformational properties.

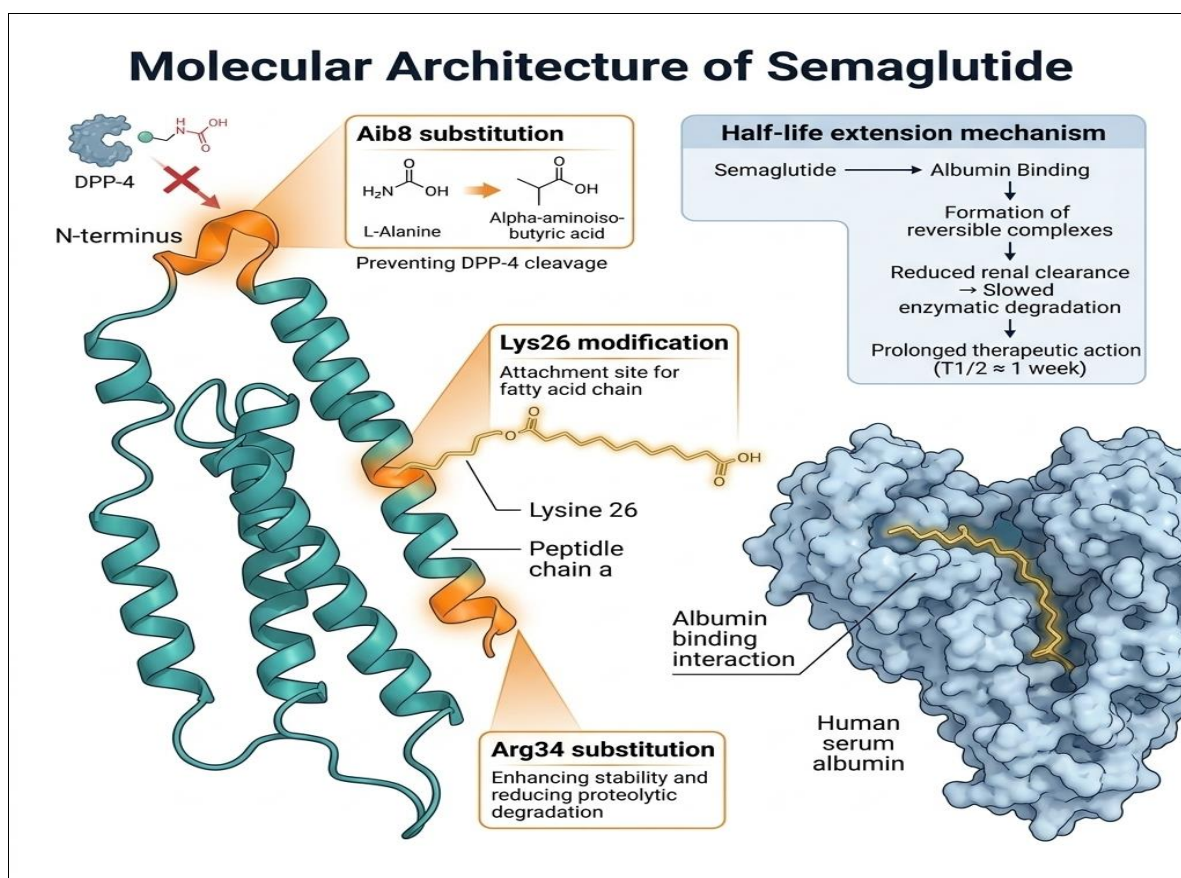


Figure 1: Molecular structure of semaglutide showing the GLP-1-derived peptide backbone, Aib substitution, Arg substitution and lipidated Lys26 side chain. Adapted from Lau et al. (2015).

Semaglutide possesses certain specific issues during the analysis due to the peculiar structural complexity of the drug, which includes both hydrophilic peptide backbone and hydrophobic lipid side chain. It may impact the physicochemical characteristics of the molecule, such as its conformation, self-association, oligomerization, aggregation, and stability. The studies using spectroscopy and molecular dynamics have shown the association of semaglutide molecules in dependence on their concentration; it involves the formation of monomers and dimers, as well as the formation of larger aggregates and structural alterations after a long-time incubation.^[4] Recent studies using circular dichroism (CD) spectroscopy, small-angle X-ray scattering (SAXS), cryogenic transmission electron microscopy (cryo-TEM), and molecular-dynamics simulation have shown the ability of semaglutide to create different oligomeric, micellar, and fibrillar structures depending on environment.^[5,6] This indicates that the characterization of semaglutide should not be limited to its primary structure alone.

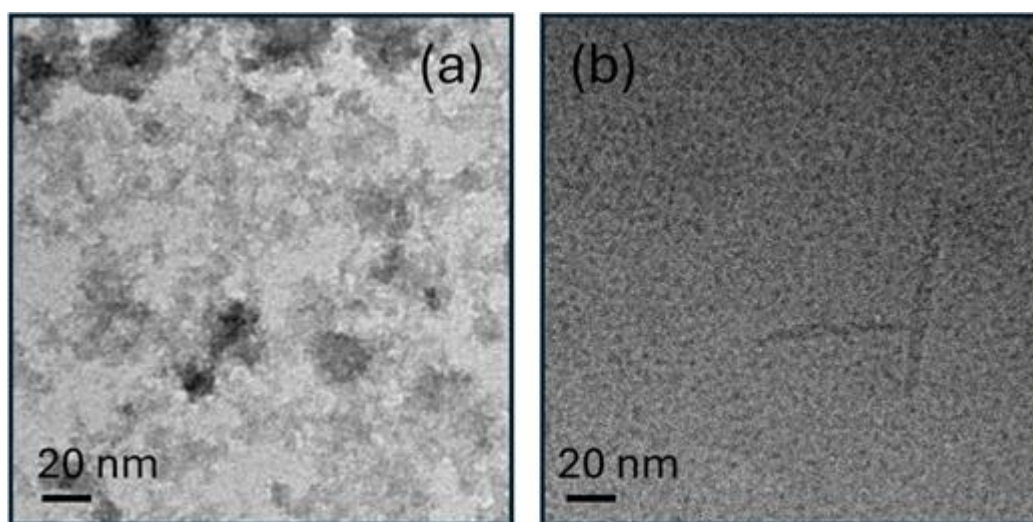


Figure 2: Cryo-TEM images for 1 wt % solutions of semaglutide: (a) freshly prepared solution and (b) solution aged for 40 days. Adapted from Hamley et al. (2025), *Biomacromolecules*, 26, 3786–3794.

For cryo-TEM analysis, semaglutide was dissolved in ultrapure water to prepare a 1 wt% aqueous solution. Freshly prepared samples were analyzed directly, while aged samples were stored at 4 °C under light-protected conditions for 40 days. For grid preparation, 3 μ L of the sample was applied to a plasma-cleaned Quantifoil 3.5/1 holey carbon copper grid, blotted twice for 5 s, and vitrified in a 1:1 liquid ethane/propane mixture at approximately -180 °C. The vitrified grids were maintained at liquid-nitrogen temperature and cryo-transferred to a JEOL JEM-3200FSC cryo-electron microscope for imaging at 200 kV and approximately -187 °C.^[64]

Thus, several analytical approaches have been used to confirm the identity and structural properties of semaglutide. Chromatographic approaches such as reversed-phase high-performance liquid chromatography (RP-HPLC) and ultra-high-performance liquid chromatography (UHPLC/UPLC) have been predominantly useful for purity analysis, separation of related compounds, and analysis of degradation products. While separation using chromatographic approaches is quite effective, retention behavior alone usually does not give enough structural information regarding the isolated compound.^[1,2]

Mass spectrometry (MS) offers a more advanced level of structural information through molecular weight determination and peptide fragment characterization. High resolution mass spectrometry (HRMS), tandem mass

spectrometry (MS/MS), and peptide mapping techniques may be employed to validate the molecular weight, amino acid sequence, post-translational modifications, and structural variants of the peptides used as drug therapies. Semaglutide structural characterization studies have found UPLC-HRMS based peptide mapping to be an efficient strategy in providing sequence coverage and verifying the fatty acid modification.^[7] Therefore, MS based techniques are very efficient in primary structure determination and chemical modification detection, but they are limited with respect to describing the complete three-dimensional structure and solution-state dynamics of the intact peptide molecule.

The optical and vibrational spectroscopies techniques give complementary data regarding the peptide structure. The technique of circular dichroism spectroscopy is especially suitable for the determination of peptide secondary structure and conformational changes, while Fourier transform infrared spectroscopy (FTIR) provides data about the peptide backbone conformation and structural transitions. The fluorescence and UV-visible spectroscopy techniques are suitable for studying the molecular interactions and the processes of peptide aggregation and association.^[4-6] But usually, these techniques provide only ensemble averaged structural data and do not easily yield residue-specific assignments.

On the other hand, NMR spectroscopy offers molecular and residue-specific information in solution state and is especially useful for characterizing the structure of peptide-based therapies. Information regarding the chemical environment, assignment of residues, secondary structure propensity, molecular interactions, conformational change, and structural dynamics can be obtained by NMR. Overall NMR information on peptides can be obtained from one-dimensional ¹H NMR spectra, while detailed residue assignment and through-bond/through-space interactions can be performed with multidimensional NMR experiments like COSY, TOCSY, NOESY, ¹H-¹³C HSQC, ¹H-¹⁵N HSQC, HMBC, and HSQC-TOCSY. Moreover, NMR-based chemical shift data, nuclear Overhauser effect (NOE) information, and other structural constraints may be applied to computer algorithms for resolving three-dimensional structures of peptides.

The key benefit of using NMR in this case is not only because it confirms that semaglutide is a peptide, but because it also offers the ability to give specific information about structure in solution. Modulations in specific resonances can offer valuable information about changes in the environment of these residues, preferred conformation, molecular interactions, and aggregation state. This makes NMR a valuable technique for complementing LC-MS, peptide mapping, CD, FTIR, and other analytical methods in studying the higher order structure of semaglutide.

The analysis of semaglutide could thus be a multilevel approach where each technique has a specific role to play in answering questions regarding the structure. Chromatography is more useful in answering questions related to purity and separability; while MS and MS/MS techniques would answer questions on molecular mass, sequencing and modifications. Peptide mapping is useful in answering questions on the primary structure. CD and FTIR would be helpful in answering questions related to secondary structure and conformational changes while NMR would be useful in answering residue specific information about structure, conformation, interactions and dynamics.^[1,2,4-8] Instead of viewing such methods as competing methodologies, they ought to be seen as a complementary set of orthogonal analysis tools.

In this respect, this review will be centered around the analysis and structural elucidation of semaglutide with the use of several instrumental methods of analysis. In this review, the structure of semaglutide will be described first, followed by the explanation of the principles and application of different instrumental techniques such as chromatography, mass spectrometry, spectroscopy, and NMR. Special attention will be paid to the kind of structural information which is provided by each technique as well as to the analytical advantages and limitations of each technique, the capability of each technique to provide information about higher-order structures of semaglutide. Finally, the capabilities of NMR are evaluated against other methods to determine what contribution NMR can make in determining the structure of semaglutide.

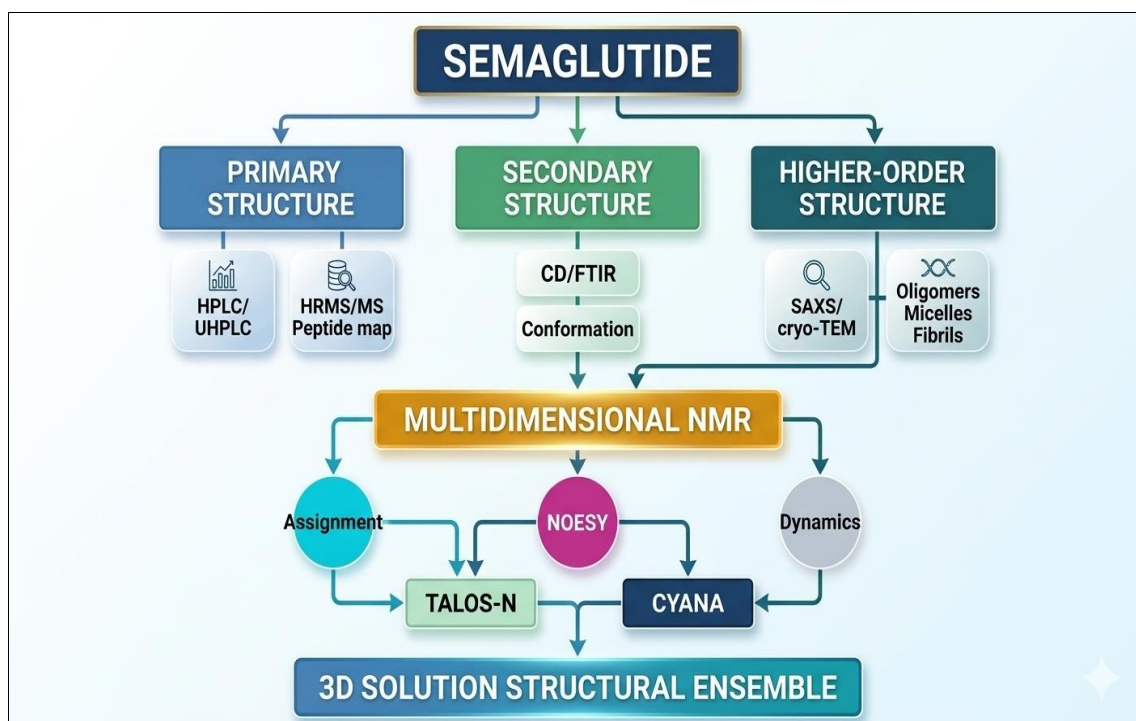


Figure 3: Overall orthogonal analytical characterization workflow.

HIGHER-ORDER STRUCTURE AND AGGREGATION OF SEMAGLUTIDE

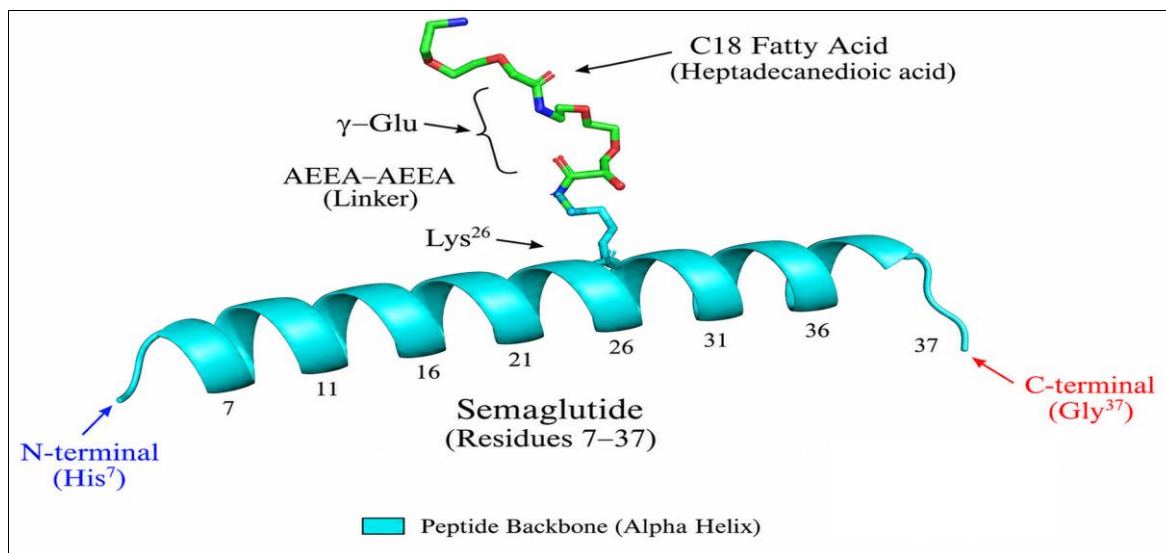


Figure 4: 3D Structure of Semaglutide.

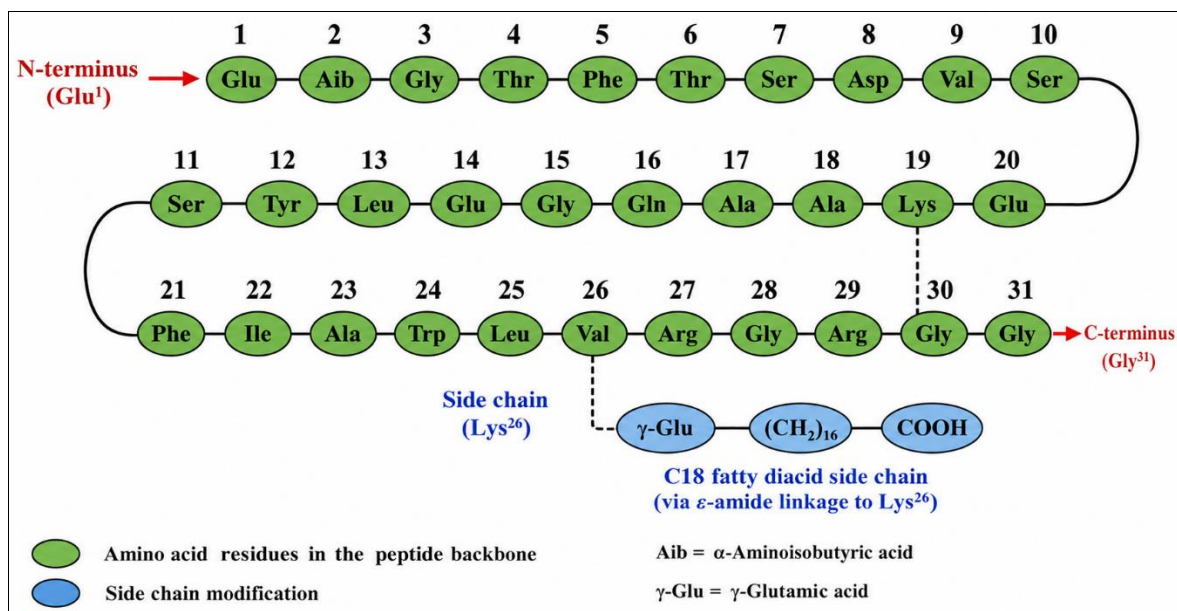


Figure 5: Primary Amino Acid Sequence and Backbone Modification Map of the Semaglutide Peptide.

The higher order structure (HOS) of any therapeutic peptide is one of the key factors in characterizing a molecule because alteration of molecular association and conformation may affect physical stability, biological activity, performance of formulation, and possibly immunogenicity of such a peptide. However, the HOS characterization of semaglutide is especially significant as the molecule contains a hydrophobic C18 fatty-acid substitution at the peptide backbone. The lipidation of semaglutide creates amphiphilic characteristics of the molecule and results in the formation of intermolecular associations via hydrophobic interactions. Therefore, semaglutide cannot be a molecule that is only capable of existence as one rigid molecular species. On the contrary, the structural state of semaglutide may depend on the concentration, pH, temperature, solvent, and the period passed since the preparation of semaglutide. Recent research using combined spectroscopic, scattering, microscopic, and computational methods showed that semaglutide may form several different levels of molecular assembly, from dimers and trimers up to micellar and fibrillar structures.^[9,10]

An early study carried out by Venanzi et al. investigated the self-assembly properties of semaglutide by employing techniques such as UV-Vis absorption spectroscopy, steady-state and time-resolved fluorescence spectroscopy, and electronic circular dichroism (ECD) in addition to molecular dynamics simulations. In the study, it was discovered that semaglutide is capable of self-associating based on its concentration with monomers and dimers being seen under the studied aqueous conditions. It was found that long incubation times resulted in larger aggregation as well as increased structural features related to the β -sheet structure. This study was significant since it demonstrated that the structure of semaglutide is a variable one and may change over time following dissolution.^[9]

More recent work has described a more in-depth structure of semaglutide's supramolecular organization. Specifically, Hamley and colleagues examined semaglutide dissolved in water via cryogenic transmission electron microscopy (cryo-TEM), small-angle X-ray scattering (SAXS), and molecular dynamics simulations. In solution at low concentration, semaglutide existed in its dimeric and trimeric forms, but at higher concentration, the peptide gradually aggregated into spherical micelles. Fibrils in the form of needles were also observed. Computer modeling confirmed a

model where the lipid tails were in the hydrophobic core, and the peptides themselves were more exposed to the solvent.^[10]

This process is especially intriguing since the formation of a supramolecular assembly does not mean that the secondary structure of the polypeptide has been totally disrupted. According to the circular dichroism study from 2025, aged semaglutide samples have presented a typical α -helical CD spectrum characterized by a positive maximum at 190 nm and two negative minima at 208 nm and 222 nm. This indicates that polypeptide parts of the micelle assembly are characterized by a considerable level of α -helices while hydrophobic lipid tails participate in the micelle structure formation.^[10]

This is significant from an analytical point of view in that it is evident that one method alone will not be sufficient to characterize the aggregates of semaglutide. The CD can confirm that the secondary structure is α -helical while cryo-TEM confirms the existence of the supramolecular assemblies through morphological observations, and SAXS gives information on their size and structure. Molecular dynamics simulation, in turn, can give an atomistic explanation of how the peptides and lipids aggregate. Therefore, the combination of techniques leads to data acquisition at different structural levels.^[10]

INFLUENCE OF PH AND TIME ON THE HIGHER-ORDER STRUCTURE OF SEMAGLUTIDE

Even more complicated becomes the picture regarding the structure of semaglutide when we discuss its dependency on pH. According to research performed in 2026 by Castelletto et al., semaglutide can exhibit a variety of supramolecular structures depending on whether pH is acidic or nearly neutral. While at pH of 2.4 the protein assumes narrow fibril-like assemblies with dominant β -sheet conformation, at higher pH levels it exhibits α -helix conformations as an oligomer and a micelle. Cryo-TEM, SAXS, fiber diffraction, FTIR, CD, fluorescence spectroscopy, and molecular-dynamics modeling were used for characterization of these states.^[11]

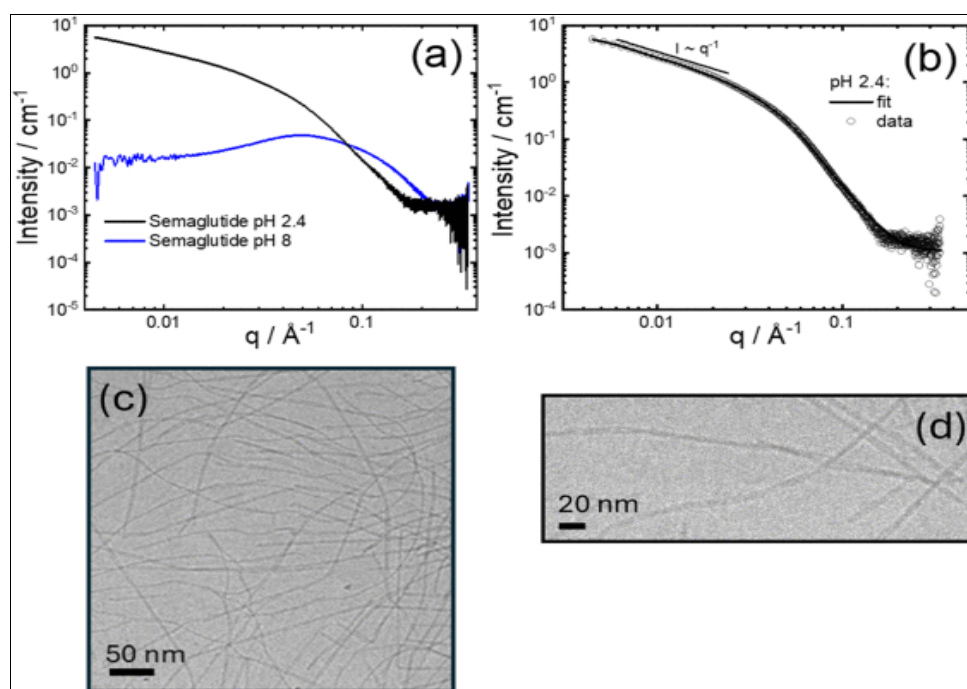


Figure 6: pH-dependent self-assembly of semaglutide revealed by SAXS and cryo-TEM.

(a) SAXS comparison of aged 1 wt % semaglutide at pH 8 and pH 2.4; (b) SAXS data and model fitting for semaglutide at pH 2.4; (c) cryo-TEM image showing fibrils at pH 2.4; and (d) enlarged image showing the twisted internal morphology of the fibrils. The study demonstrates the distinct self-assembled states formed at acidic and higher pH. Adapted from Castelletto et al. (2026), *Biomacromolecules*.^[65]

In particular, the fibrillar and micellar aggregation behavior investigation of semaglutide is highly relevant since it showed a considerable α -helix to β -sheet conformational change upon aging in an acidic environment. Fresh samples of semaglutide in approximately pH 2.4 had CD bands typical for the α -helix conformation, such as a double minimum in 208 and 222 nm range. Aging of semaglutide in about 40 days resulted in considerable changes in CD spectra and the formation of a wide minimum at 216 nm. FTIR analysis confirmed this conformational change as well.^[11]

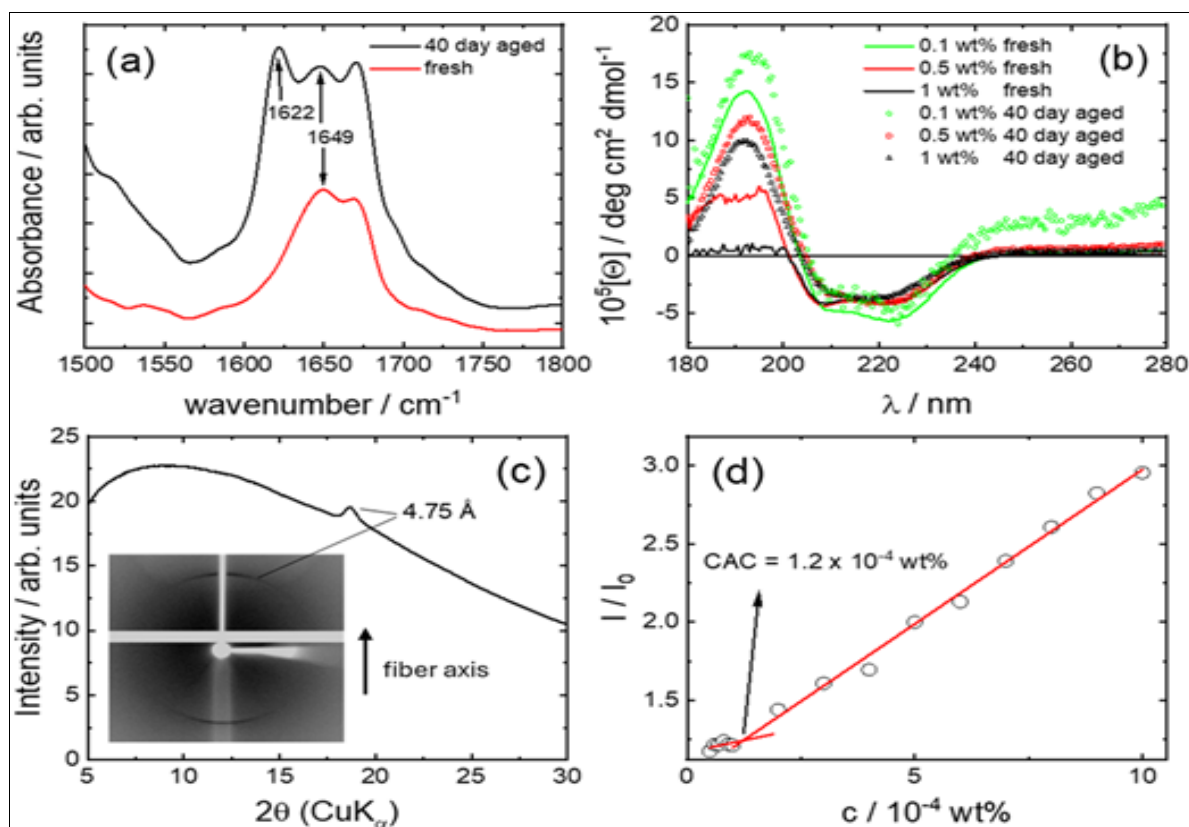


Figure 7: (a) FTIR spectra from 1 wt % solutions for fresh and 40-day-aged samples of semaglutide (TFA) (pH 2.4). (b) CD spectra for semaglutide (TFA) at the concentrations indicated that the pH was measured to be 1 wt % (pH 2.4), 0.5 wt % (pH 2.6), and 0.1 wt % (pH 3.1). (b), (c) Fiber XRD intensity profile obtained from the 2D image inset, which shows meridional orientation of peaks due to β -strand spacing. (d) Determination of critical aggregation concentration (CAC) using ThT fluorescence (intensity at $\lambda = 486$ nm) for a semaglutide (TFA) pH 2.4 sample. Adapted from Castelletto et al. (2026), *Biomacromolecules*.^[65]

Fibrillation was also seen during the structural conversion process. Twisted fibrillar structure was identified using cryo-TEM analysis, whereas fiber X-ray diffraction indicated that there is a possibility of ordered β -strand structure. Based on the experiment findings as well as the molecular dynamics simulation, the authors suggested a model based on fibrils with curved antiparallel β -strands surrounding the core where lipid chains corresponding to the modified lysine

were found. The mechanism helps understand the significance of amphiphilicity of semaglutide in contributing to different supramolecular structures under varying physicochemical conditions.^[11]

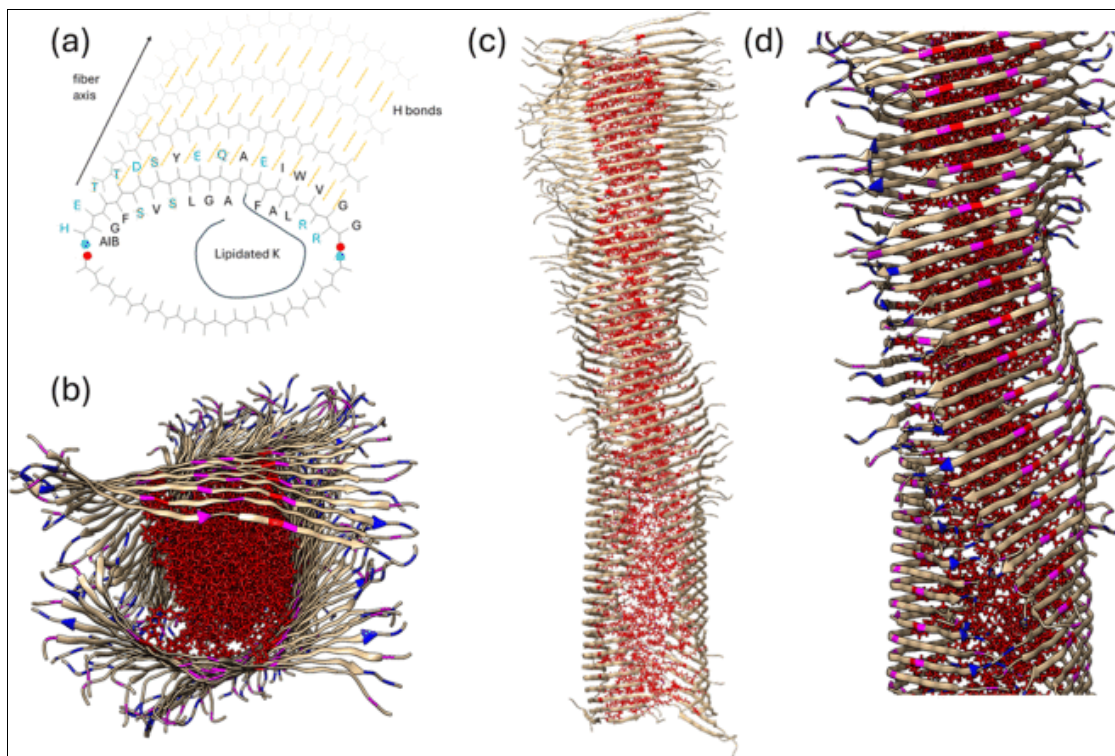


Figure 8: Molecular model of the semaglutide fibril obtained from molecular-dynamics simulations. (a) Schematic representation of the β -strand and residue distribution; (b) view along the fibril axis showing the lipidated lysine (Lyc) core; (c) twisted fibril structure showing the β -strands and Lyc-rich interior; and (d) enlarged view of the twisted fibril structure. Adapted from Castelletto et al.

It is clear from the above that any pH-dependent experiments need to be carried out within well-controlled conditions, since the structure of semaglutide measured at pH 8 will not always give the same structural population as a spectrum at pH 2.4. The freshly prepared sample may not have the same conformational profile as a sample kept for some weeks. Hence, the pH, concentration, temperature, solvent, buffer composition, and age of the sample should all be mentioned in the interpretation of structure measurements.^[10,11]

From an analysis perspective, this observation also represents a good argument in favor of the orthogonal approach. CD and FTIR spectroscopy can detect changes in the structure population, SAXS can measure the size and shape of assemblies, TEM can give information on their morphology, and fluorescence spectroscopy can provide information on the kinetics of aggregation. The NMR technique could be helpful here since it has the ability to detect which residues undergo a change in their chemical environment when undergoing this structural transition. This is especially important when trying to understand the molecular mechanism of aggregation rather than merely proving that it occurs.

USE OF NMR SPECTROSCOPY FOR ELUCIDATING THE STRUCTURE OF SEMAGLUTIDE

NMR spectroscopy has a unique role in peptide structural analysis since it gives information at the atomic and residue level as the peptide is still in its solution phase. As opposed to chromatographic methods, which offer separation data, and spectroscopic methods such as CD, which give ensemble conformational data, NMR can separate individual

signals and identify their specific nuclei in the peptide structure. NMR spectroscopy has therefore been widely used to analyze the identity, purity, conformation, and higher structural organization of therapeutically relevant peptides and proteins.^[12,13]

One of the main strengths of NMR in peptide characterization is the possibility of obtaining different types of structural information in one sample. While one-dimensional ^1H NMR produces a spectral fingerprint for the entire molecule, multidimensional experiments like COSY and TOCSY give information on proton spin system connections. Information on the residue specificity can be achieved using heteronuclear experiments such as ^1H - ^{15}N HSQC and ^1H - ^{13}C HSQC. Long range information can be provided by HMBC, while NOESY provides through space connections, which are crucial for the three-dimensional structure identification.^[12,13]

ONE-DIMENSIONAL ^1H NMR SPECTROSCOPY

In one-dimensional ^1H NMR spectroscopy, all the signals are obtained from the amide protons, α -protons, side chain protons, aromatic residues, aliphatic groups, and lipid chain. Semaglutide has many distinct proton environments in a small molecular structure that causes many overlapping resonances in the ^1H NMR spectrum. For this reason, while one-dimensional ^1H NMR spectroscopy is valuable for comparing samples and understanding the general structure of molecules, it is not sufficient for assigning each individual residue.^[14,15]

Nonetheless, any alterations in the one-dimensional ^1H NMR spectrum can be considered because of changes in the state of molecules. Any change in linewidth, signal intensity, chemical shift position, or the appearance of new resonances indicates any changes in the molecular state including conformational change, association with other molecules, degradation, or aggregation.^[16]

COSY AND TOCSY

The correlation spectroscopy (COSY) determines correlations between the protons that are connected via the scalar coupling. Therefore, it determines the local through-bond correlations and can help to determine the coupled protons in the amino acid residue. Nevertheless, as there are many overlapping resonances in the peptide, COSY alone is not sufficient to determine all assignments.^[17,18]

Total correlation spectroscopy (TOCSY) is a technique that gives more information about the spin system since the magnetization can be transferred through the whole chain of the scalar-coupled nuclei. Thus, TOCSY can connect the amide/ α -hydrogen region with several protons of the side chain of the amino-acid residue. This spin-system pattern can be matched with the expected pattern of the individual amino acid residue. Thus, TOCSY helps to establish the identity of individual amino-acid spin systems prior to the sequential assignment.^[18,19]

While COSY mainly reveals direct proton connectivity, TOCSY gives information on a wider range of the coupled spin system. Together, they make up an essential basis for peptide resonance assignments.^[17,19]

^1H - ^{15}N HSQC AS A RESIDUE SPECIFIC FINGERPRINT

One of the multidimensional NMR experiments that prove to be useful in peptide/protein analysis is the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) experiment. In the ^1H - ^{15}N HSQC experiment, the chemical shifts of directly coupled ^1H and ^{15}N are correlated, and usually, there is only one peak per each observable backbone amide

resonance. Hence, the spectrum generated by this experiment can be used as a fingerprint of a specific residue in the peptide.^[20,21]

For semaglutide, the ^1H - ^{15}N HSQC spectrum gives very valuable information, as any changes in the cross-peaks could correspond to changes in the chemical environment of certain residues. Thus, displacement of a cross-peak indicates changes in the electronic environment or conformation. Intensity or linewidth changes may indicate changes in the dynamics or exchange of molecules. Broadening or disappearance of the peaks could indicate that residues start participating in the process of aggregation/conformational exchange.^[22,23]

This capability of residue-specific analysis offers an added advantage over CD. If the CD proves that the semaglutide molecule has experienced any change in its overall secondary structure, it does not necessarily mean that the CD spectrum will be able to determine the residues involved in such changes. On the other hand, HSQC spectra have the ability to indicate which particular amide environments have undergone any change.^[21,22]

Hence, NMR may be able to identify the molecular parts influenced by changes in pH, concentration, temperature, formulation, and interactions between molecules.^[22,23]

^1H - ^{13}C HSQC AND CHARACTERIZATION OF THE LIPIDATED REGION

The ^1H - ^{13}C HSQC method correlates directly bonded hydrogen and carbon nuclei. It is especially useful for making assignments to aliphatic and aromatic side-chain nuclei and expanding assignments beyond the region of backbone amide nuclei.^[24,25]

For semaglutide, ^1H - ^{13}C HSQC is particularly significant since the compound includes a relatively large modification in the form of C18 fatty acid. The presence of such a modification adds additional aliphatic carbon-proton environments that can be differentiated from the environments of the peptide backbone. Thus, ^{13}C -based NMR experiments could add additional information about the modified region and allow us to check if the lipid part of the molecule experiences any change in the chemical environment due to molecular association.^[25,26]

Such an approach could be especially useful for studies of higher order structure. In case, the lipid chains participate in hydrophobic interactions in the process of micelle formation, one may observe changes in the chemical shift, linewidth or intensity of the ^{13}C signals.^[26,27]

HSQC-TOCSY AND HMBC

The HSQC-TOCSY technique is a combination of HSQC, which offers heteronuclear correlation and residue-specificity information, and TOCSY, which offers more spin-system connectivities. In this experiment, the magnetization is transferred from a directly observed heteronuclear nucleus to the coupled proton spin system so that there will be correlation between a particular resonance of the backbone amide group and other resonances of the same amino acid residue. The HSQC-TOCSY technique will be helpful in assigning peptide resonance assignment since it helps in extending the information from one cross-peak of an amide group to α -proton and side-chain resonances.^[28,29]

In the case of semaglutide, HSQC-TOCSY can thus serve as an extra connection between the ^1H - ^{15}N HSQC backbone fingerprint and the proton spin systems generated by TOCSY. After matching an amide peak with its corresponding side-chain peaks, the spin system pattern obtained can then be matched against the expected amino acid residue type.

This gives extra assurance that the residue is correctly assigned prior to establishing sequential connections through NOESY experiments.^[28,29]

Heteronuclear multiple bond correlation (HMBC) constitutes an additional form of structural information. Unlike HSQC, which is mostly focused on detecting direct proton-carbon (^1H - ^{13}C) correlations, HMBC is especially designed to detect long-range heteronuclear correlations, typically over two or three bond lengths but, under appropriate experimental conditions, even longer coupling pathways. In this regard, HMBC is especially valuable for providing connectivity information about protons and carbon atoms not bonded to protons and for confirming structural relationships that cannot be observed in a conventional one-bond HSQC spectrum.^[30,31]

For structural analysis of a chemically modified peptide, like semaglutide, the use of HMBC technique can provide further confirmation regarding connectivity involving carbonyls, modified amino acids, linkers, and the lipid portion of the compound. Long-range correlations may assist in finding the correlations between nuclear spins that are far apart in chemical connectivity through several chemical bonds, thus providing additional data to the information obtained using ^1H - ^{13}C HSQC and other directly bonded correlation spectroscopy techniques. In spite of that, the lack of HMBC correlation peak alone cannot be considered a proof that there is no connectivity between nuclei, since HMBC correlation sensitivity relies heavily on the size of the relevant long-range coupling and proper experimental optimization.^[30-32]

For semaglutide, the complementary technique becomes especially pertinent because the molecule contains peptide moiety together with a long lipophilic chain modification. The role of the lipid chain in the formation of molecular associations and aggregation of semaglutide behavior in aqueous solution has been proven by structural and spectral investigations; thus, heteronuclear correlation experiments may become a source of chemical information regarding structural features related to the modified part of the molecule, although interpretation should be based on the experimentally observed correlations rather than assuming that every observed change represents aggregation.^[33]

The combined use of HSQC, HSQC-TOCSY, and HMBC techniques thus allow the chemist to move from a cluster of multidimensional resonances to a chemically defined molecular framework. HSQC offers the initial heteronuclear spectrum, HSQC-TOCSY expands individual backbone resonances to their associated spin systems, and HMBC adds additional information regarding the long-range correlations. When combined with COSY, TOCSY, NOESY, ^1H - ^{13}C HSQC, and sequential assignment techniques, such NMR methods offer a complete foundation for the structural analysis and resonance assignment of semaglutide.^[28-33]

NOESY AND 3D STRUCTURE INFORMATION

The nuclear Overhauser effect spectroscopy (NOESY) experiment signifies a shift from connectivity-based analysis to 3D structural studies. In COSY and TOCSY experiments, the primary information one gets is about nuclei which have a connection through scalar coupling, but in NOESY, information is collected about nuclei which have through-space magnetization. This implies that NOESY cross-peaks contain information regarding the spatial organization of a peptide, as they indicate proximity of protons to protons, which may not be necessarily connected through covalent bonds.^[34,35]

The strength of an NOE correlation depends on the cross-relaxation characteristics of the nuclei involved, which, in turn, depend on the spatial separation of the nuclei and the effects of molecular dynamics. Accordingly, under controlled experimental conditions, the NOEs may be used to derive interproton distance restraints. In standard protein NMR structure analysis, these restraints are usually expressed in terms of distance range values rather than distance values, due to the effects of spin diffusion, conformational averaging, molecular motions, and resonance overlap on NOE intensities.^[35,36]

The distance restraints obtained from experimental studies are basic to the process of obtaining an NMR structure because they limit the range of possible conformations for the molecule. An adequate number of NOE-based restraints may be used in structure calculation methods such as distance geometry, simulated annealing, CYANA, and other restrained molecular dynamics programs. The larger the number and better the quality of the experimentally determined interproton distance restraints will improve the precision and accuracy of the resulting structural ensemble.^[36,37]

For Semaglutide, NOESY information can be obtained beyond the validation of the amino acid sequence of the peptide. Sequential NOEs, especially the NOEs between neighboring residues, can help to assign the sequence-specific resonance assignments. In contrast, the medium-range and longer-range NOEs can give an indication about the spatial contact of residues which are separated in the primary structure. Such contacts are important for checking local folding and overall three-dimensional organization of the peptide.^[37,38]

The interpretation of NOESY experiments is particularly crucial for developing the structure of semaglutide since NOE constraints are not merely used for defining individual residues; they define geometry of pairs of nuclei. In combination with the sequence-specific resonance assignment and other information obtained from NMR experiments, these constraints can be used for building up an ensemble of structures consistent with experimental data. The resulting ensemble reflects the conformational space defined by NMR experiments rather than a single conformation of a molecule.^[35,38]

This contrast becomes all the more relevant within the context of this current review because it shows how NMR moves from chemistry to structure. While the assignment studies determine which resonances are associated with which atoms or amino acid residues, NOESY provides spatial data by which these assignments can then be linked in three dimensions. Therefore, the combination of COSY, TOCSY, HSQC, HSQC-TOCSY, HMBC, and NOESY results in a sequential progression from spin system analysis to experimental determination of 3D structure.^[34-38]

ASSIGNMENT OF CHEMICAL SHIFTS AND STRUCTURE DETERMINATION

The assignment of chemical shifts can be considered as one of the most crucial steps of structure determination by NMR spectroscopy. It involves assigning a particular resonance signal to a particular atom of a particular residue. Instead of labeling a resonance signal as a “peak,” a particular resonance can be referred to as H α , H β , NH, or another atom of a particular residue.^[39,40]

The assignment of semaglutide is dependent on integration of information from several experiments. TOCSY and COSY will identify the spin systems, HSQC will assign heteronuclear coupling, HSQC-TOCSY will assign side chains regions and finally NOESY will provide sequential and spatial couplings. This chemical shift table will serve as the basis of structure calculation.^[39-41]

This is a crucial difference between NMR and mass spectrometry. In MS peptide mapping, peptides are identified based on their mass and fragmentation pattern, while in NMR, individual atoms in the molecule are assigned based on their environments within the intact molecule in solution. These approaches are therefore complementary rather than interchangeable.^[42,43]

USE OF TALOS-N FOR PREDICTION OF BACKBONE TORSION-ANGLES USING NMR CHEMICAL SHIFT ASSIGNMENTS

Once a sufficiently complete set of NMR chemical shift assignments, one may use these experimentally observed chemical shifts to get an idea about the local backbone conformation of the peptide. TALOS-N (Torsion Angle Likelihood Obtained from Shift and Sequence Data-Neural Network) is a neural network-based approach designed by Shen and Bax to predict protein backbone ϕ (phi) and ψ (psi) torsion angles using NMR chemical shifts. TALOS-N can also predict side chain χ^1 torsion angles and secondary structure elements. The main principle behind TALOS-N is that the chemical shifts depend on the local electronic and structural environment of the residues. TALOS-N uses the characteristic pattern of ^1H , ^{13}C and ^{15}N chemical shifts to identify the parts of conformational space compatible with the experimentally observed chemical shift pattern.^[44]

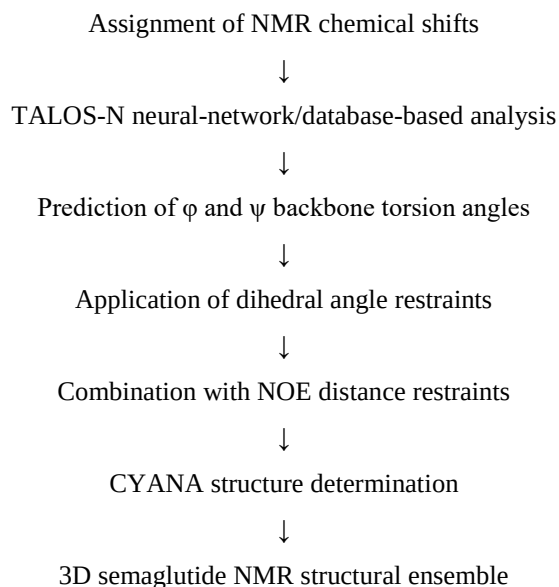
The key advantage of TALOS-N relative to the previous versions of TALOS and TALOS+ is the employment of artificial neural network analysis along with structural database knowledge. In TALOS-N, the Ramachandran space of conformations is defined in terms of 324 regions of conformations, while the chemical shift profile of a residue is assessed based on the data obtained from known protein structures. In other words, TALOS-N does not merely classify the residue into α -helices and β -sheets but instead assesses the likelihood of compatibility between particular sets of backbone torsion-angles combinations and observed chemical shifts. This provides a more detailed description of local backbone conformational preferences.^[44,45]

The primary components that go into TALOS-N are the amino acid sequence and the NMR chemical shifts. The chemical shifts can be of the backbone $^1\text{H-N}$, ^{15}N , ^{13}C , $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, and $1\text{H}\alpha$. Chemical shifts carry structural information since changes in backbone geometry affect the magnetic field around the nuclei. TALOS-N then compares the chemical shift data obtained experimentally with the chemical shift data obtained from the neural network model and the database to give the ϕ and ψ dihedral angles of the backbone. Therefore, the method provides an important connection between chemical-shift assignment and conformational analysis.^[44]

A significant aspect of TALOS-N is its provision of confidence values regarding predicted torsion angles such that the predictions can be assessed in terms of their reliability. Strong predictions suggest high structural conformational consistency while ambiguous predictions suggest the need for additional caution when making subsequent structure predictions. The initial validation of TALOS-N showed reliable backbone angle predictions of over 90% of residues, with an error rate of less than 3.5% under the specified validation criteria.^[44] TALOS-N predictions should therefore not be seen as proof of one particular global structure but rather as local conformational evidence only.

These dihedral angle predictions can then be used as restraints in NMR structure determinations using algorithms like CYANA in combination with distance restraints obtained from NOE. TALOS-N thus acts as a bridge between chemical shift assignments and NMR structure determinations.^[45,46]

For semaglutide, the overall process can thus be presented as:



CYANA STRUCTURE DETERMINATION

CYANA (Combined Assignment and NMR Analysis) is often employed for determining three-dimensional structures from structural constraints derived from NMR. CYANA makes use of torsion-angle dynamics to produce conformations that fulfill the distance and dihedral angle constraints that have been acquired from experimental results. In the NMR structure determination process, CYANA incorporates data from NOESY and chemical-shift-derived backbone torsion-angle predictions.^[47,48]

Input Data for CYANA

In case of semaglutide, the primary input data files include:

1. .seq - semaglutide amino acid sequence.
2. .prot - chemical shifts (¹H, ¹³C and ¹⁵N) assigned to semaglutide.
3. .upl - NOE distance restraints upper limits.
4. .lol - NOE distance restraints lower limits.
5. .aco - dihedral angle restraints from TALOS-N.
6. .peaks - NOESY peak lists, for automatic NOE assignment.
7. .cya - CYANA macro containing the structure-calculation commands.

CYANA can use assigned chemical shifts together with NOESY cross-peak information for automated NOE assignment and structure calculation.^[47,48]

CYANA STRUCTURE CALCULATION

The calculation is controlled using a CYANA macro. The restraint files are read into CYANA and torsion-angle dynamics is then used to generate multiple structural conformers. A typical calculation conceptually follows:

1. read upl semaglutide.upl
2. read aco semaglutide.aco
3. calc_all
4. overview semaglutide

The exact macro depends on the CYANA version and calculation protocol. The calculation generates multiple conformers, which are subsequently evaluated according to their agreement with the experimental restraints.^[47,48]

Output Obtained from CYANA

The major output is a PDB structural ensemble containing several calculated conformers. CYANA can also generate an overview file containing structural statistics. Important information obtained includes:

1. number of calculated conformers;
2. target-function values;
3. distance-restraint violations;
4. dihedral-angle restraint violations;
5. backbone and heavy-atom RMSD;
6. residue range used for structural superposition;
7. structural convergence;
8. Ramachandran/dihedral-angle information.

The CYANA overview procedure ranks structures according to their target function and can calculate RMSD values relative to the mean coordinates.^[47,48]

STRUCTURAL ENSEMBLE AND RMSD ANALYSIS

The generation of a structural ensemble rather than a single structure is a characteristic feature of solution-state NMR structure determination. Unlike X-ray crystallography, where a crystal lattice can constrain molecules into a particular structural arrangement, solution-state NMR measurements represent an ensemble of molecules in solution. Consequently, the experimental restraints obtained from NOESY, chemical shifts, scalar couplings, and other NMR experiments generally define a family of conformations that are compatible with the experimental observations. Structure-calculation programs such as CYANA therefore generate multiple conformers that satisfy the available experimental restraints, and the resulting ensemble provides information about both the well-defined structural regions and regions that remain conformationally variable.^[49,52]

The calculated conformers are commonly superimposed to evaluate their structural convergence. One of the most widely used quantitative parameters for this purpose is the root-mean-square deviation (RMSD), which describes the average positional difference between corresponding atoms after optimal superposition. For two structures containing N corresponding atoms, RMSD can be expressed as:

$$RMSD = \sqrt{\frac{1}{N} \sum_{i=1}^N (r_i - r'_i)^2}$$

Symbol	Meaning
$RMSD$	Root-mean-square deviation between two structures
N	Number of corresponding atoms being compared
i	Index of the atom being considered
r_i	3D coordinates of atom i in structure 1

r'_i	3D coordinates of the corresponding atom i in structure 2
$r_i - r'_i$	Difference in position between the two corresponding atoms
$(r_i - r'_i)^2$	Squared positional difference
$\sum$	Adds the squared differences for all atoms
$1/N$	Calculates the average squared difference

Table 1: Mathematical Variables and Structural Interpretation of the Root-Mean-Square Deviation (RMSD) Formula.

For peptide structure determination, however, RMSD should generally be calculated separately for the well-defined structural regions and for the entire molecule. Flexible terminal residues or poorly restrained loops can contribute disproportionately to the overall RMSD even when the central, experimentally well-defined region is highly converged. Consequently, reporting the RMSD of the complete molecule without considering residue-specific structural variability can give an incomplete picture of the quality of an NMR ensemble. NMR structure-validation approaches therefore commonly distinguish between the structurally well-defined core and regions exhibiting greater conformational variability.^[52,54]

In the context of semaglutide, this distinction is particularly important because different regions of the peptide may not be constrained to the same degree by the available NMR data. Regions containing numerous unambiguous NOE restraints and reliable TALOS-N dihedral predictions are expected to show greater convergence among the calculated conformers. Conversely, regions with fewer experimental restraints, ambiguous NOE assignments, missing resonances, or increased conformational flexibility may display greater positional variation. Such variation should not automatically be interpreted as an error in the calculation; it may indicate that the experimental data permit several conformations for that region.^[49,50,52]

A crucial point in interpreting RMSD is that a low RMSD does not by itself establish that an NMR structure is experimentally correct. Two structures can have a low RMSD because they were generated from similar starting conditions or because strong computational restraints forced them toward a common geometry, while the underlying experimental evidence may still be insufficient. Conversely, a relatively high RMSD can occur in a genuinely flexible region even when the structure calculation is completely consistent with the experimental data. Therefore, RMSD should be interpreted together with the experimental restraint statistics and structural-validation parameters.^[52-54]

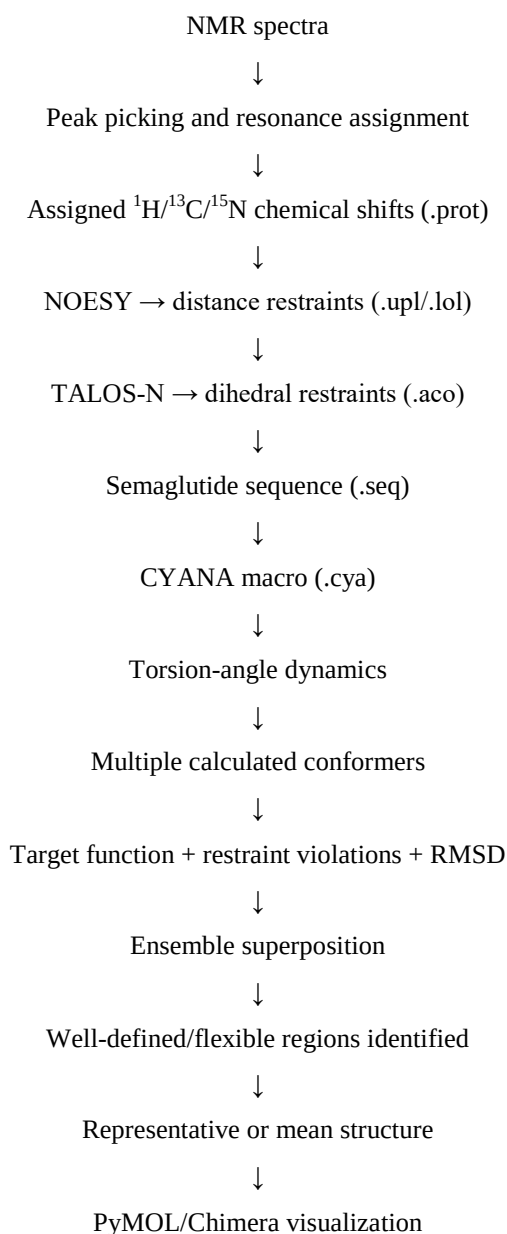
A comprehensive assessment of an NMR-derived structure should therefore consider the number and distribution of NOE-derived distance restraints, the number of dihedral-angle restraints, distance-restraint violations, dihedral-angle violations, covalent geometry, Ramachandran statistics, structural convergence, and other stereochemical parameters. Programs and databases developed for NMR structure validation use these complementary criteria to determine whether the calculated ensemble is consistent with both the experimental restraints and accepted stereochemical principles.^[52,54]

The relationship between RMSD and experimental restraints can be particularly useful when interpreting a CYANA-generated semaglutide ensemble. If ten or more calculated conformers are superimposed and the backbone atoms of the central region show very similar positions while the terminal or flexible regions diverge, this suggests that the central region is more strongly defined by the experimental restraints. In contrast, if the entire peptide displays substantial structural dispersion, this may indicate that the available restraints are insufficient to define a unique global conformation or that the peptide exists as a conformationally heterogeneous system under the experimental conditions.

Thus, RMSD can be used not only as a numerical measure of similarity but also as a way of identifying which portions of the peptide are structurally well defined.^[52,53]

For your semaglutide study, it is also important to distinguish pairwise RMSD from the RMSD of an ensemble relative to an average or reference structure. Pairwise RMSD compares individual conformers with one another, whereas ensemble RMSD can describe the deviation of individual structures from a representative structure. These values answer slightly different questions. Pairwise RMSD is useful for assessing how similar the calculated conformers are to one another, while residue-specific RMSD or coordinate variability can identify precisely where structural convergence is achieved or lost.^[52,53]

Therefore, a high-quality semaglutide NMR structure should be described using a combination of RMSD + restraint statistics + stereochemical validation + residue-specific convergence, rather than using RMSD as a standalone indicator of structural quality.^[52-54]



COMPARATIVE ASSESSMENT OF ANALYTICAL METHODS

Structural data from various analytical methods should be viewed as complementary rather than competing data sources. The separation, purity determination, and analysis of impurities is where chromatography methods excel. However, they yield little information about the structure of the molecules directly. HRMS and MS/MS yield extremely useful data related to the molecular weight, sequence-dependent fragments, chemical modifications, and degradation products. Peptide mapping expands this potential to sequence mapping and enabling confirmation of modification sites. A new study conducted in 2025 showed that UPLC-HRMS peptide mapping with complementary Glu-C and chymotrypsin digestions yielded 100% sequence coverage of semaglutide and could be used for the identification of synthetic peptide compounds.^[63]

CD and FTIR give us another kind of structural information. They are especially useful when it comes to tracking secondary structure changes and recognizing transitions between α -helices and β -sheet-rich states. The recently published semaglutide aggregation studies illustrate the utility of such methods for tracking transitions from α -helices to β -sheet-rich fibrils in acidic and prolonged-aged environments.^[56,57]

It is also possible to use the NMR technique to obtain complementary information because it resolves individual atom environments. The advantage of NMR is that it can offer information regarding residue specificity, through space interactions, conformational information, and dynamic information. It has been shown that the NMR technique is highly sensitive for detecting small structural differences in therapeutic proteins and for evaluating higher-order structure.^[58]

Table 2: Different Analytical Techniques and Principal Structural Information for Peptide Characterization.

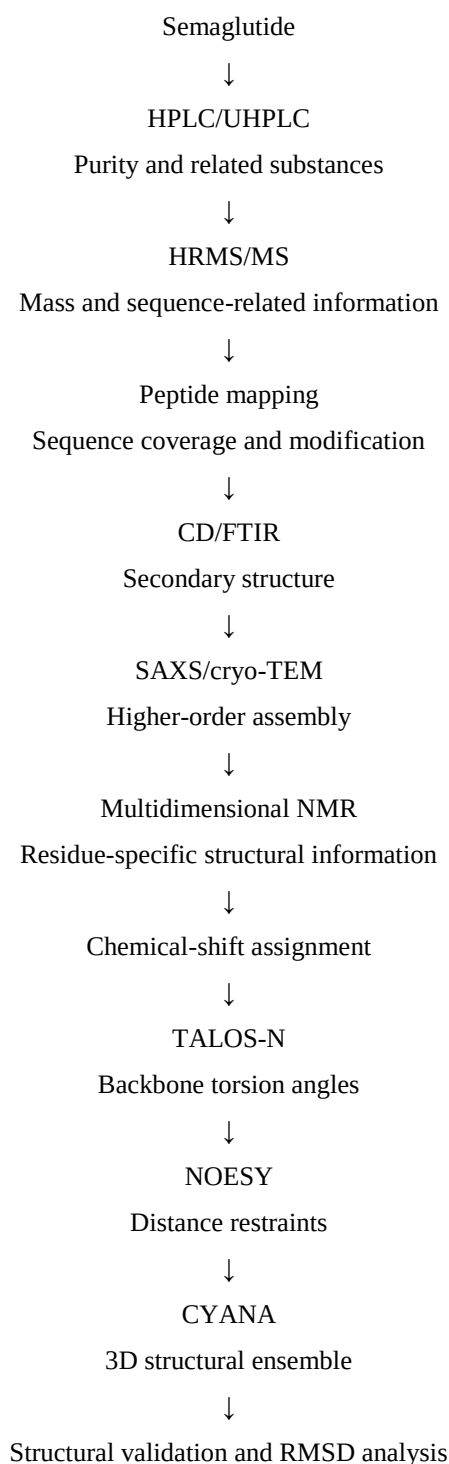
Technique	Principal structural information
High-Performance Liquid Chromatography, Ultra-High-Performance Liquid Chromatography (HPLC/UHPLC)	Purity, separation, related substances and degradation
High-Resolution Mass Spectrometry (HRMS)	Accurate molecular mass and molecular composition
Tandem Mass Spectrometry (MS/MS)	Sequence-related fragmentation and modification information
enzymatic digestion followed by chromatographic/MS analysis	Sequence coverage and modification-site confirmation
Circular Dichroism Spectroscopy (CD)	Overall secondary structure
Fourier-Transform Infrared Spectroscopy (FTIR)	Secondary-structure-associated vibrational information
Small-Angle X-ray Scattering (SAXS)	Size, shape and supramolecular organization
Cryogenic Transmission Electron Microscopy (Cryo-TEM)	Morphology of aggregates and assemblies
Nuclear Magnetic Resonance Spectroscopy (NMR)	Residue-specific chemical environment, conformation, interactions and dynamics
Torsion Angle Likelihood Obtained from Shift and Sequence Data-Neural Network (TALOS-N)	Backbone torsion-angle prediction
Combined Assignment and NOE Analysis (CYANA)	Three-dimensional structure calculation from NMR restraints

STRATEGY TO CHARACTERIZE SEMAGLUTIDE USING INTEGRATED ORTHOGONAL APPROACH

An approach that combines all techniques based on the specific structural questions that can be solved by each is well documented in literature. HPLC/UHPLC may be used to define chromatographic purity and distinguish related compounds. Next, HRMS may determine the exact molecular weight, whereas MS/MS and peptide mapping techniques may provide information about sequences and chemical modification. Finally, CD and FTIR may analyze the general features of secondary structure of the peptide. SAXS and cryo-TEM can also provide additional information about supramolecular assemblies and aggregate morphology.^[56,57,63]

NMR occupies a particularly important position within this workflow since it can serve as a link between molecular identification and structure elucidation. After establishing the molecular identity, the use of multidimensional NMR will allow for the assignment of the resonances and description of the chemical environment of each residue. The chemical shift data will help in predicting the torsion angles, whereas NOESY restraints can help in the determination of the three-dimensional structure of the molecules. TALOS-N and CYANA are the computational connections between NMR data and the final structural ensemble.^[59-62]

The final analytical methodology can therefore be described as follows:



This question-driven approach is preferable to presenting the analytical methods as independent techniques because it demonstrates how the information from one method complements the limitations of another.

CRITICAL COMPARISON OF NMR WITH CD SPECTROSCOPY AND MS SPECTROMETRY

The comparison of NMR, cd and MS is particularly relevant to semaglutide.

MS is a very sensitive technique and is extremely valuable in terms of molecular mass determination, sequence analysis, modifications and impurities detection. The gas phase character of MS experiments, however, implies that the state of the molecule after ionization and desolvation is not necessarily the same as the solution state one. Thus, MS is very useful when studying the primary structure but offers less detailed information about the solution state three-dimensional structure.

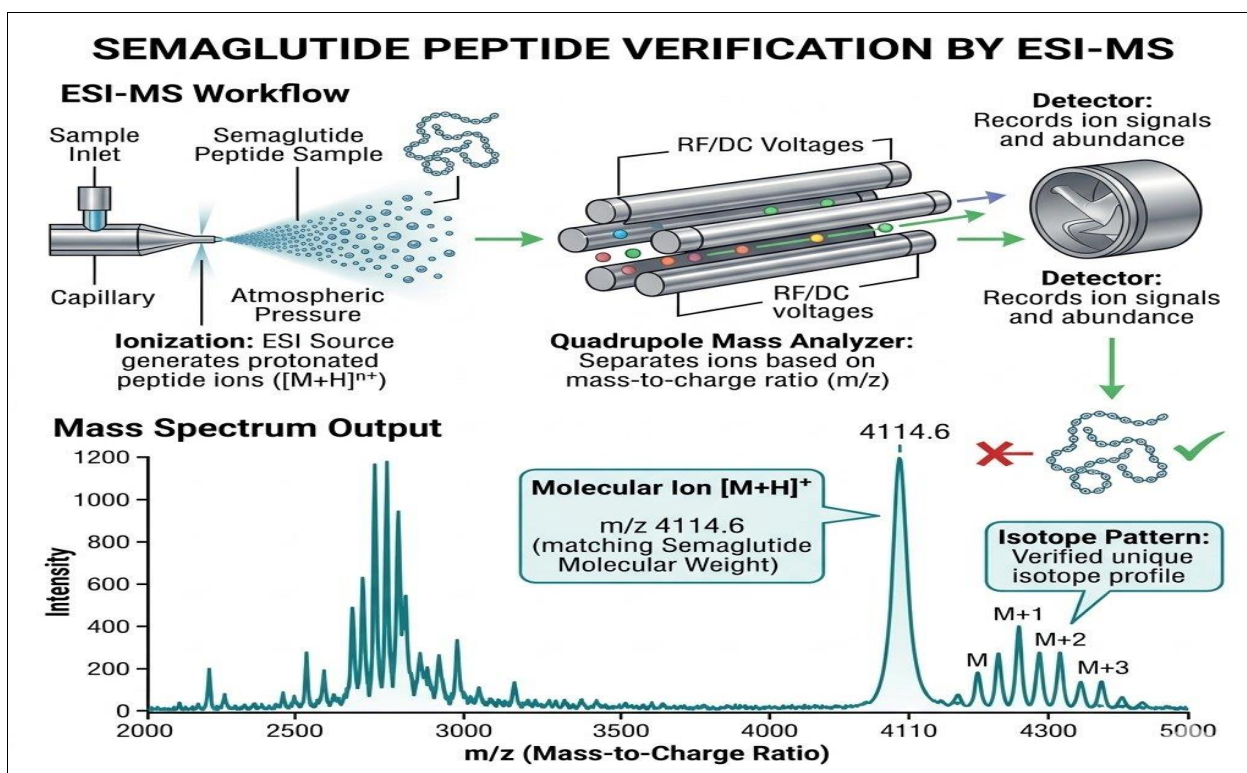


Figure 9: High-Resolution Characterization and Structure Verification of Semaglutide Peptide Using Electrospray Ionization Mass Spectrometry (ESI-MS).

However, unlike circular dichroism, the CD study involves analysis of the optical property of the entire molecular population within solution, making it very effective in analyzing overall secondary structure properties. But it usually fails to assign structures to individual residues. For instance, although an α -helix to β -sheet transition can easily be identified using CD studies, it is unable to show which residues trigger or stabilize that particular transition.

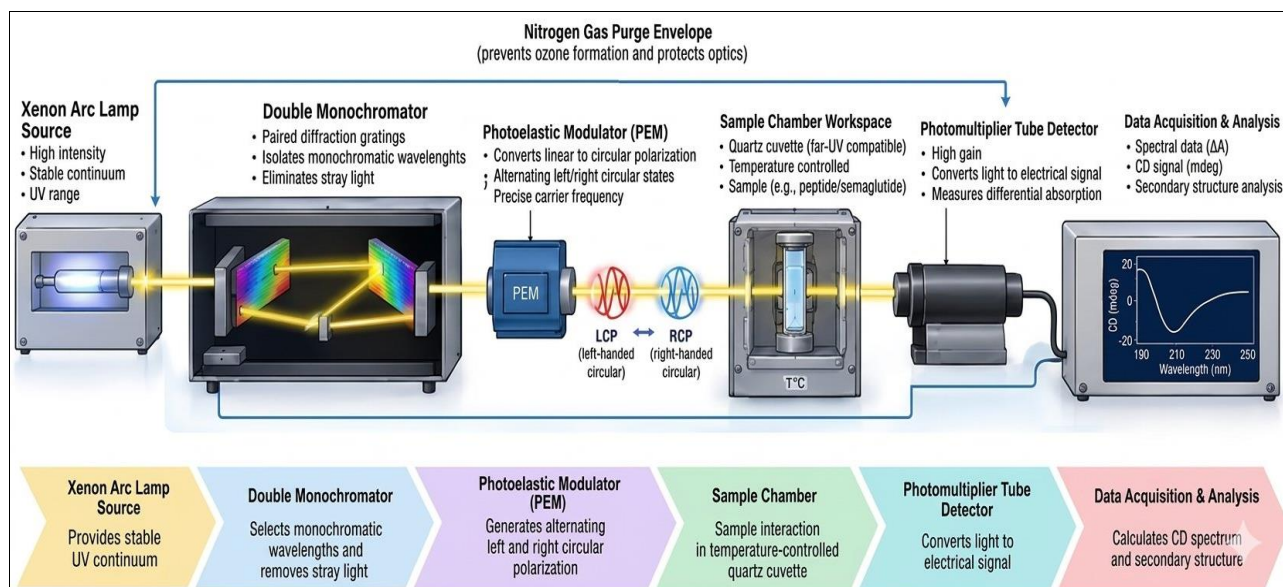


Figure 10: Schematic Optical Layout and Instrument Train of a Far-UV Circular Dichroism (CD) Spectrometer.

NMR occupies an intermediate and highly informative position because it provides information directly from individual nuclei in solution. Chemical shifts, coupling patterns, heteronuclear correlations and NOE interactions can collectively provide residue-specific and spatial information. Therefore, the techniques should be considered complementary.



Figure 11: Instrumentation Layout of a High-Resolution Bruker Avance NEO 600 MHz Superconducting NMR Spectrometer.

This analytical installation features a high-resolution nuclear magnetic resonance (NMR) platform configured for the structural characterization of therapeutic peptides like semaglutide. On the left, the **Bruker Ascend™ 600**

Superconducting Magnet houses a niobium-based coil cooled to 4.2 K, generating a highly homogeneous **14.1 Tesla** magnetic field to polarize nuclear spins. Adjacent to its base is the pneumatic **SampleCase Autosampler Tower**, which manages automated sample queues and tube injection into the top magnet bore. Positioned centrally is a portable **Liquid Nitrogen Cryogen Storage Dewar** (77 K), used to routinely replenish the magnet's thermal radiation shield and minimize liquid helium boil-off. On the right, the **Avance NEO Spectrometer Control Console** acts as the electronic core, housing the high-precision digital transmitters and receivers required to execute complex multi-dimensional radiofrequency (RF) pulse sequences and digitize weak spin-echo emissions. Together, this integrated system delivers the high-resolution spectral datasets essential for downstream **multidimensional NMR assignment**, NOESY mapping, and 3D solution structural ensemble modeling.

CURRENT CHALLENGES AND FUTURE PERSPECTIVES

Despite its structural advantages, NMR characterization of semaglutide presents several challenges. The relatively large number of resonances generated by the 31-residue peptide produces substantial spectral overlap, particularly in one-dimensional proton spectra. In addition, lipidation introduces additional resonances and hydrophobic interactions that can influence molecular association. The concentration required for multidimensional NMR experiments may also approach conditions under which self-association becomes significant. Consequently, sample preparation and experimental conditions must be carefully controlled.

Solvent selection is another important consideration. Exchangeable amide protons can be affected by solvent composition, pH and temperature. In aqueous solutions, exchange with deuterium can reduce or eliminate signals from exchangeable protons. Consequently, solvent and buffer optimization is not merely an experimental convenience but an essential component of structural characterization. This is particularly relevant when comparing NMR spectra obtained under different experimental conditions.

Future studies should increasingly integrate NMR with other biophysical methods and computational approaches. Combining residue-specific NMR information with CD-derived secondary structure, SAXS-derived dimensions, cryo-TEM morphology and molecular-dynamics simulations could provide a more complete description of the semaglutide conformational landscape. Such an integrated strategy could move characterization beyond the simple detection of aggregation toward identification of the molecular regions involved in oligomerization, micellization and fibril formation.

A particularly valuable future direction would be the development of residue-resolved higher-order structure analysis of semaglutide under different formulation and stress conditions. Such studies could examine changes in HSQC chemical shifts, peak intensities and linewidths as a function of pH, concentration, temperature and storage time and correlate those changes with CD, SAXS and cryo-TEM observations.

CONCLUSION

Semaglutide represents a structurally complex therapeutic peptide whose characterization requires evaluation at multiple molecular levels. Its GLP-1-derived peptide backbone is combined with a non-proteinogenic Aib residue and a long C18 fatty-acid-containing modification, giving the molecule distinctive physicochemical and supramolecular properties. Recent studies have demonstrated that semaglutide can undergo concentration-, pH- and time-dependent self-association, resulting in dimeric, trimeric, micellar and fibrillar structures.^[55-57]

Chromatographic methods provide essential information about purity and related substances, while HRMS/MS and peptide mapping provide accurate molecular-mass, sequence-related and modification information. CD and FTIR provide complementary information about global secondary-structure changes, while SAXS and cryo-TEM provide information about the dimensions and morphology of supramolecular assemblies.^[56,57,63]

NMR provides a distinctive contribution because it enables investigation of the molecular structure at residue-specific resolution in solution. Multidimensional experiments such as COSY, TOCSY, ^1H - ^{15}N HSQC, ^1H - ^{13}C HSQC, HSQC-TOCSY, HMBC and NOESY can collectively establish resonance assignments, local structural information and spatial relationships. Chemical-shift analysis using TALOS-N can provide backbone torsion-angle restraints, while CYANA can integrate NMR-derived restraints into 3D structural calculations.^[58-62]

Thus, the structural characterization of semaglutide should not be viewed as a competition between analytical techniques. Rather, the most informative strategy is an orthogonal analytical framework in which each technique addresses a specific structural question. HPLC establishes chromatographic purity, MS defines molecular mass and sequence-related chemical information, CD and FTIR characterize global conformational behaviour, SAXS and cryo-TEM characterize supramolecular organization, and NMR provides residue-specific solution-state structural information. Within this framework, NMR provides an essential link between molecular identity and three-dimensional molecular structure.

REFERENCE

1. ICH Q6B. Specifications: Test Procedures and Acceptance Criteria for Biotechnological/Biological Products. International Council for Harmonisation, 1999.
2. Beck A, et al. Analytical tools for characterizing biopharmaceuticals and the implications for biosimilars. *mAbs*, 2013.
3. Lau J, Bloch P, Schäffer L, et al. Discovery of the once-weekly glucagon-like peptide-1 (GLP-1) analogue semaglutide. *J Med Chem*, 2015; 58: 7370–7380. DOI: 10.1021/acs.jmedchem.5b00726.
4. Venanzi M, Savioli M, Cimino R, et al. A spectroscopic and molecular dynamics study on the aggregation process of a long acting lipidated therapeutic peptide: the case of semaglutide. *Soft Matter*, 2020; 16: 10122–10131. DOI: 10.1039/D0SM01011A.
5. Hamley IW, et al. Semaglutide Aggregates into Oligomeric Micelles and Short Fibrils in Aqueous Solution. *Biomacromolecules*, 2025. DOI: 10.1021/acs.biomac.5c00342.
6. Hamley IW, et al. Fibrillar and Micellar Aggregation of Semaglutide and Formation of a Chiral-Imprinted Glass. *Biomacromolecules*, 2026. DOI: 10.1021/acs.biomac.5c02669.
7. *Peptide mapping analysis of synthetic semaglutide and liraglutide for generic development of drugs originating from recombinant DNA technology. J Pharm Biomed Anal*, 2025.
8. General NMR/biopharmaceutical structural-characterization literature.
9. Venanzi, M.; Savioli, M.; Cimino, R.; Gatto, E.; Palleschi, A.; Ripani, G.; Cicero, D.; Placidi, E.; Orvieto, F.; Bianchi, E. A Spectroscopic and Molecular Dynamics Study on the Aggregation Process of a Long-Acting Lipidated Therapeutic Peptide: The Case of Semaglutide. *Soft Matter*, 2020; 16: 10122–10131. DOI: 10.1039/D0SM01011A.

10. Hamley, I. W.; de Mello, L. R.; Castelletto, V.; et al. Semaglutide Aggregates into Oligomeric Micelles and Short Fibrils in Aqueous Solution. *Biomacromolecules*, 2025; 26: 3786–3794. DOI: 10.1021/acs.biomac.5c00342.
11. Castelletto, V.; de Mello, L. R.; Hamley, I. W.; et al. Fibrillar and Micellar Aggregation of Semaglutide and Formation of a Chiral-Imprinted Glass. *Biomacromolecules*, 2026; 27: 2818–2827. DOI: 10.1021/acs.biomac.5c02669.
12. Haxholm, G. W.; Petersen, B. O.; Malmstrøm, J. Higher-Order Structure Characterization of Pharmaceutical Proteins by 2D Nuclear Magnetic Resonance Methyl Fingerprinting. *J. Pharm. Sci.*, 2019; 108: 3029–3035. DOI: 10.1016/j.xphs.2019.04.032.
13. Kellenbach, E.; Rundlöf, T. Determination of the Identity, Content and Purity of Therapeutic Peptides by NMR Spectroscopy. In *Peptide Therapeutics: Strategy and Tactics for Chemistry, Manufacturing and Controls*; Srivastava, V., Ed.; Royal Society of Chemistry, 2019; pp 381–420. DOI: 10.1039/9781788016445-00381.
14. Kwan, A. H.; Mobli, M.; Gooley, P. R. Macromolecular NMR Spectroscopy for the Non-Spectroscopist. *FEBS J.*, **2011**; 278: 3492–3507. <https://doi.org/10.1111/j.1742-4658.2011.08004.x>.
15. Frueh, D. P. Practical Aspects of NMR Signal Assignment in Larger and Challenging Proteins. *Prog. Nucl. Magn. Reson. Spectrosc.*, **2014**; 78: 47–75. <https://doi.org/10.1016/j.pnmrs.2013.12.001>.
16. Venanzi, M.; Savioli, M.; Cimino, R.; Gatto, E.; Palleschi, A.; Ripani, D.; Cicero, D.; Placidi, E.; Orvieto, F.; Bianchi, E. A Spectroscopic and Molecular Dynamics Study on the Aggregation Process of a Long-Acting Lipidated Therapeutic Peptide: The Case of Semaglutide. *Soft Matter*, **2020**; 16: 5571–5581. <https://doi.org/10.1039/D0SM01011A>.
17. Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.
18. Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Academic Press: San Diego, 2007.
19. Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. A General Method for Assigning NMR Spectra of Denatured Proteins Using 3D HC(CO)NH-TOCSY Triple Resonance Experiments. *J. Biomol. NMR*, **1993**; 3: 225–231. <https://doi.org/10.1007/BF00178264>.
20. Ozawa, K.; Wu, P. S. C.; Dixon, N. E.; Otting, G. N-Labelled Proteins by Cell-Free Protein Synthesis: Strategies for High-Throughput NMR Studies of Proteins and Protein-Ligand Complexes. *FEBS J.*, **2006**; 273: 4154–4169. <https://doi.org/10.1111/j.1742-4658.2006.05433.x>.
21. Barile, E.; Pellecchia, M. NMR-Based Approaches for the Identification and Optimization of Inhibitors of Protein–Protein Interactions. *Chem. Rev.*, **2014**; 114: 4749–4763. <https://doi.org/10.1021/cr500043b>.
22. Zuiderweg, E. R. P. Mapping Protein–Protein Interactions in Solution by NMR Spectroscopy. *Biochemistry*, **2002**; 41: 1–7. <https://doi.org/10.1021/bi011870b>.
23. Krichevsky, O.; Sirota, L.; Levy, Y.; et al. NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chem. Rev.*, **2022**.
24. Kay, L. E.; Ikura, M.; Tschudin, R.; Bax, A. Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins. *J. Magn. Reson.*, **1990**; 89: 496–514. [https://doi.org/10.1016/0022-2364\(90\)90333-2](https://doi.org/10.1016/0022-2364(90)90333-2).
25. Frueh, D. P. Practical Aspects of NMR Signal Assignment in Larger and Challenging Proteins. *Prog. Nucl. Magn. Reson. Spectrosc.*, **2014**; 78: 47–75. <https://doi.org/10.1016/j.pnmrs.2013.12.001>.

26. Venanzi, M.; Savioli, M.; Cimino, R.; Gatto, E.; Palleschi, A.; Ripani, D.; Cicero, D.; Placidi, E.; Orvieto, F.; Bianchi, E. A Spectroscopic and Molecular Dynamics Study on the Aggregation Process of a Long-Acting Lipidated Therapeutic Peptide: The Case of Semaglutide. *Soft Matter*, **2020**; *16*: 5571–5581. <https://doi.org/10.1039/D0SM01011A>.
27. Hamley, I. W.; et al. Semaglutide Aggregates into Oligomeric Micelles and Short Fibrils in Aqueous Solution, **2025**.
28. Logan, T. M.; Olejniczak, E. T.; Xu, R. X.; Fesik, S. W. A General Method for Assigning NMR Spectra of Denatured Proteins Using 3D HC(CO)NH-TOCSY Triple Resonance Experiments. *J. Biomol. NMR*, **1993**; *3*: 225–231. <https://doi.org/10.1007/BF00178264>.
29. Buchler, N. E.; Zuiderweg, E. R. P.; Wang, H.; Goldstein, R. A. Protein Heteronuclear NMR Assignments Using Mean-Field Simulated Annealing. *J. Magn. Reson*, **1997**; *125*: 34–42. <https://doi.org/10.1006/jmre.1997.1106>.
30. Williamson, R. T.; Buevich, A. V.; Martin, G. E.; Parella, T. LR-HSQMBC: A Sensitive NMR Technique to Probe Very Long-Range Heteronuclear Coupling Pathways. *J. Org. Chem*, **2014**; *79*: 3887–3894. <https://doi.org/10.1021/jo500333u>.
31. Furrer, J. A Robust, Sensitive, and Versatile HMBC Experiment for Rapid Structure Elucidation by NMR: IMPACT-HMBC. *Chem. Commun*, **2010**; *46*: 3396–3398. <https://doi.org/10.1039/C000964D>.
32. Saurí, J.; Liu, Y.; Parella, T.; Williamson, R. T.; Martin, G. E. Selecting the Most Appropriate NMR Experiment to Access Weak and/or Very Long-Range Heteronuclear Correlations. *J. Nat. Prod*, **2016**; *79*: 1400–1406. <https://doi.org/10.1021/acs.jnatprod.6b00139>.
33. Venanzi, M.; Savioli, M.; Cimino, R.; Gatto, E.; Palleschi, A.; Ripani, D.; Cicero, D.; Placidi, E.; Orvieto, F.; Bianchi, E. A Spectroscopic and Molecular Dynamics Study on the Aggregation Process of a Long-Acting Lipidated Therapeutic Peptide: The Case of Semaglutide. *Soft Matter*, **2020**; *16*: 10122–10131. <https://doi.org/10.1039/D0SM01011A>.
34. Wüthrich, K.; Billeter, M.; Braun, W. Pseudo-Structures for the 20 Common Amino Acids for Use in Studies of Protein Conformations by Measurements of Intramolecular Proton-Proton Distance Constraints with Nuclear Magnetic Resonance. *J. Mol. Biol*, **1983**; *169*: 949–961. [https://doi.org/10.1016/S0022-2836\(83\)80144-2](https://doi.org/10.1016/S0022-2836(83)80144-2).
35. Zaccai, G.; Wüthrich, K. NMR: The Nuclear Overhauser Effect and Its Applications to Structural Biology. *Q. Rev. Biophys*, **1998**; *31*: 145–237.
36. Vögeli, B. The Nuclear Overhauser Effect from a Quantitative Perspective. *Prog. Nucl. Magn. Reson. Spectrosc*, **2014**; *78*: 1–46. <https://doi.org/10.1016/j.pnmrs.2013.11.001>.
37. Clore, G. M.; Gronenborn, A. M. Determining the Structures of Small Proteins by NMR. *Protein Eng*, **1987**; *1*: 275–288.
38. Schmidt, E.; Güntert, P. Reliability of Exclusively NOESY-Based Automated Resonance Assignment and Structure Determination of Proteins. *J. Biomol. NMR*, **2013**; *57*: 193–204. <https://doi.org/10.1007/s10858-013-9779-x>.
39. Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*, 2nd ed.; Academic Press: Burlington, MA, 2007.
40. Frueh, D. P. Practical Aspects of NMR Signal Assignment in Larger and Challenging Proteins. *Prog. Nucl. Magn. Reson. Spectrosc*, **2014**; *78*: 47–75. <https://doi.org/10.1016/j.pnmrs.2013.12.001>.

41. Williamson, M. P. Using Chemical Shift Perturbation to Characterise Ligand Binding. *Prog. Nucl. Magn. Reson. Spectrosc*, **2013**; 73: 1–16. <https://doi.org/10.1016/j.pnmrs.2013.02.001>.
42. Aebersold, R.; Mann, M. Mass Spectrometry-Based Proteomics. *Nature*, **2003**; 422: 198–207. <https://doi.org/10.1038/nature01511>.
43. Kaltashov, I. A.; Eyles, S. J. *Mass Spectrometry in Biophysics: Conformation and Dynamics of Biomolecules*; Wiley: Hoboken, NJ, 2005.
44. Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR*, **2013**; 56: 227–241. <https://doi.org/10.1007/s10858-013-9741-y>.
45. Shen, Y.; Bax, A. Protein Structural Information Derived from NMR Chemical Shifts with the Neural Network Program TALOS-N. *Methods Mol. Biol.*, **2015**; 1260: 17–32. https://doi.org/10.1007/978-1-4939-2239-0_2.
46. Shen, Y.; Delaglio, F.; Cornilescu, G.; Bax, A. TALOS+: A Hybrid Method for Predicting Protein Backbone Torsion Angles from NMR Chemical Shifts. *J. Biomol. NMR*, **2009**; 44: 213–223. <https://doi.org/10.1007/s10858-009-9333-z>.
47. Güntert, P. Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol*, **2004**; 278: 353–378. DOI: 10.1385/1-59259-809-9:353.
48. Güntert, P.; Buchner, L. Combined Automated NOE Assignment and Structure Calculation with CYANA. *J. Biomol. NMR*, **2015**; 62: 453–471. DOI: 10.1007/s10858-015-9924-9.
49. Güntert, P. Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol*, **2004**; 278: 353–378. DOI: 10.1385/1-59259-809-9:353.
50. Güntert, P.; Buchner, L. Combined Automated NOE Assignment and Structure Calculation with CYANA. *J. Biomol. NMR*, **2015**; 62: 453–471. DOI: 10.1007/s10858-015-9924-9.
51. Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* **2013**, 56, 227–241. DOI: 10.1007/s10858-013-9741-y.
52. Nabuurs, S. B.; Spronk, C. A. E. M.; Krieger, E.; Maassen, H.; Vriend, G.; Vuister, G. W. Quantitative Evaluation of Experimental NMR Restraints. *J. Am. Chem. Soc*, **2003**; 125: 12026–12034. DOI: 10.1021/ja035440f.
53. Czerwinski, M.; Kowalski, K.; et al. Structural Analysis and Validation of NMR-Derived Molecular Ensembles. *J. Biomol. NMR*. [Use the specific validation reference corresponding to the software used for your final analysis.]
54. Laskowski, R. A.; Rullmann, J. A. C.; MacArthur, M. W.; Kaptein, R.; Thornton, J. M. AQUA and PROCHECK-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR*, **1996**; 8: 477–486. DOI: 10.1007/BF00228148.
55. Venanzi, M.; Savioli, M.; Cimino, R.; Gatto, E.; Palleschi, A.; Ripani, G.; Cicero, D.; Placidi, E.; Orvieto, F.; Bianchi, E. A Spectroscopic and Molecular Dynamics Study on the Aggregation Process of a Long-Acting Lipidated Therapeutic Peptide: The Case of Semaglutide. *Soft Matter*, **2020**; 16: 10122–10131. DOI: 10.1039/D0SM01011A.
56. Hamley, I. W.; de Mello, L. R.; Castelletto, V.; Zinn, T.; Seitsonen, J.; et al. Semaglutide Aggregates into Oligomeric Micelles and Short Fibrils in Aqueous Solution. *Biomacromolecules* **2025**, 26, 3786–3794. DOI: 10.1021/acs.biomac.5c00342.
57. Castelletto, V.; de Mello, L. R.; Seitsonen, J.; Hamley, I. W.; et al. Fibrillar and Micellar Aggregation of Semaglutide and Formation of a Chiral-Imprinted Glass. *Biomacromolecules*, **2026**; 27: 2818–2827. DOI: 10.1021/acs.biomac.5c02669.

58. Haxholm, G. W.; Petersen, B. O.; Malmstrøm, J. Higher-Order Structure Characterization of Pharmaceutical Proteins by 2D Nuclear Magnetic Resonance Methyl Fingerprinting. *J. Pharm. Sci*, **2019**; *108*: 3029–3035. DOI: 10.1016/j.xphs.2019.04.032.
59. Kellenbach, E.; Rundlöf, T. Determination of the Identity, Content and Purity of Therapeutic Peptides by NMR Spectroscopy. In *Peptide Therapeutics: Strategy and Tactics for Chemistry, Manufacturing and Controls*; Srivastava, V., Ed.; Royal Society of Chemistry: Cambridge, 2019; pp 381–420. DOI: 10.1039/9781788016445-00381.
60. Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR*, **2013**; *56*: 227–241. DOI: 10.1007/s10858-013-9741-y.
61. Güntert, P. Automated NMR Structure Calculation with CYANA. *Methods Mol. Biol*, **2004**; *278*: 353–378. DOI: 10.1385/1-59259-809-9:353.
62. Güntert, P.; Buchner, L. Combined Automated NOE Assignment and Structure Calculation with CYANA. *J. Biomol. NMR*, **2015**; *62*: 453–471. DOI: 10.1007/s10858-015-9924-9.
63. Kim, S. H.; Kim, S. S.; Kim, H. J.; Park, E. J.; Na, D. H. Peptide Mapping Analysis of Synthetic Semaglutide and Liraglutide for Generic Development of Drugs Originating from Recombinant DNA Technology. *J. Pharm. Biomed. Anal*, **2025**; *256*: 116682. DOI: 10.1016/j.jpba.2025.116682.
64. Hamley, I. W.; de Mello, L. R.; Castelletto, V.; Zinn, T.; Cowieson, N.; Seitsonen, J.; Bizien, T. Semaglutide Aggregates into Oligomeric Micelles and Short Fibrils in Aqueous Solution. *Biomacromolecules*, **2025**; *26*(6): 3786–3794.
65. Valeria Castelletto, Lucas R. de Mello, Jani Seitsonen, Ian W. Hamley; Fibrillar and Micellar Aggregation of Semaglutide and Formation of a Chiral-Imprinted Glass. *Biomacromolecules*, 13 April 2026; *27* (4): 2818–2827. <https://doi.org/10.1021/acs.biomac.5c02669>