Analysis of a novel insulin lisargine at the molecular level using X-ray crystallography and biophysical techniques

Zhu Zhu
University of Science and Technology of China

Han Wang
Chinese Academy of Sciences

Li Wang
Chinese Academy of Sciences

Qingrong Yang
Chinese Academy of Sciences

Wei Zhou
Hefei Tianmai Biotechnology Development Co., Ltd

Peng Wang (✉ pengwang@ipp.ac.cn)
Chinese Academy of Sciences

Zhiming Zheng
Chinese Academy of Sciences

Article

Keywords:

Posted Date: July 31st, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3191793/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

The structural characterization of insulin analogs is crucial for their clinical applications. In previous study, we developed a novel long-acting insulin analog called insulin lisargine, which exhibits superior sustained-release properties. However, its structure remains unclear and requires thorough investigation. We determined the amino acid sequence of insulin lisargine using mass spectrometry and analyzed its secondary structure using circular dichroism spectroscopy. Furthermore, we employed X-ray crystallography experiments at the Shanghai Synchrotron Radiation Facility to elucidate the crystal structure of the insulin lisargine analog, achieving a resolution of approximately 2.0Å. Comparison with the human insulin structure in the Protein Data Bank revealed structural similarities between the two. These findings suggest promising prospects for the insulin lisargine analog and provide essential foundational information for subsequent drug design and optimization. In summary, we conducted a comprehensive structural study on the insulin lisargine analog, shedding light on its potential applications in the field. The obtained insights pave the way for future drug development and optimization efforts.

1. Introduction

Insulin lisargine is a novel insulin analogue used for the treatment of diabetes¹. It exhibits structural modifications compared to human insulin, where asparagine at position B3 in the A chain is replaced by lysine, and lysine and arginine are introduced at positions B31 and B32 in the B chain, respectively. These structural changes confer insulin lisargine with unique pharmacological characteristics and advantages.

Insulin lisargine is a rapid-acting insulin analogue that exerts its effects more quickly after subcutaneous injection, achieving higher blood drug concentrations on a per-unit basis compared to traditional human insulin. It demonstrates superior performance in terms of onset time and peak concentration compared to human insulin. These properties make insulin lisargine a promising long-acting insulin substitute, offering more effective treatment options for individuals with diabetes.

Despite extensive research and application of the pharmacokinetics and pharmacodynamics of the newly developed insulin lisargine, its crystal structure has not been fully elucidated. This study aims to provide a detailed analysis and investigation of the primary and higher-order structures of insulin lisargine using mass spectrometry, circular dichroism spectroscopy, and X-ray diffraction techniques. By obtaining crystals of insulin lisargine and conducting crystallographic analysis, we can acquire precise three-dimensional structural information and understand its molecular organization and folding at the atomic level. These research findings are of significant importance for understanding the pharmacological characteristics, molecular interactions, and binding mechanisms of insulin lisargine with insulin receptors. Furthermore, through structural comparisons with other insulin analogues, we can further optimize drug design to enhance stability, efficacy, and bioavailability.
In summary, as a novel insulin analogue, insulin lisargine holds promising clinical prospects. In-depth investigations into its structure will contribute to a better understanding of its mechanisms of action and provide essential foundations for the development of more efficient and stable insulin-like medications.

2. Materials and methods

2.1 Materials

Recombinant insulin Lisargine injection (Lot No. M201605003) (Lisargine) was obtained from Hefei Tianmai Biotechnology Development Co., Ltd. CSA standards are purchased from Sigma Corporation, while protein crystallization reagent kits and additive screening kits are purchased from Hampton Corporation.

2.2 Liquid Chromatography Mass Spectrometry (LC-MS)

Liquid chromatography conditions: ACQUITY UPLC peptide BEH C18 column (100mm×2.1 mm, 1.7 µm), with 0.1% formic acid/water solution as mobile phase A and 0.1% formic acid/acetonitrile solution as mobile phase B. Gradient elution was performed at a flow rate of 0.3 mL/min, column temperature of 50°C, and an injection volume of 1µL. Mass spectrometry conditions: Positive ion detection mode; capillary voltage 1-2.5kV; cone voltage 35V; cone gas flow 50L•h-1; collision voltage 30-45V; MCP detector voltage 2300 V; source temperature 80°C; scan range: m/z 50-2000.

2.2.1 Determination of complete molecular weight by mass spectrometry

Weigh 10 mg of insulin lisargine and add 10 mL of 10mM hydrochloric acid solution to prepare a 1mg/mL sample solution. Take 2 µL of the solution and directly inject it into a high-resolution mass spectrometer. Collect and process the data to calculate the molecular weight.

2.2.2 Mass spectrometry determination of amino acid sequence

Weigh 10 mg of insulin lisargine and add 5 mL of Tris-EDTA buffer (pH 8.0) to prepare a 2 mg/mL sample solution. Add 100 µL of the sample solution to an equal volume of 12 mol/L urea solution, followed by the addition of 2 µL of 0.1mol/L DTT solution, mix well, and incubate at 37°C for 30 minutes. Then add 4 µL of 1mol/L iodoacetamide solution for alkylation treatment in the dark for 30 minutes. Take an appropriate amount of the above sample and dilute it 30-fold with ultrapure water to prepare the test solution. Use 100 µL of Tris-EDTA buffer (pH 8.0) instead of the sample solution as a blank control solution for the same treatment.

2.3 Circular Dichroism (CD) spectroscopy

Near-ultraviolet scanning and far-ultraviolet scanning were performed as follows:
For near-ultraviolet scanning, soak the cuvette in 2mol/L HNO3 overnight, rinse it thoroughly with deionized water, and let it dry. First, collect the background spectrum, followed by the spectrum of the blank buffer solution. Then, add 200µL of the 0.2mg/mL test sample to the cuvette and perform scanning from 190–260 nm according to the parameters mentioned above, and collect the data.

For far-ultraviolet scanning, soak the cuvette in 2mol/L HNO3 overnight, rinse it thoroughly with deionized water, and let it dry. First, collect the background spectrum, followed by the spectrum of the blank buffer solution. Then, add 400µL of the 1.0mg/mL test sample to the cuvette and perform scanning from 250–340 nm according to the parameters mentioned above, and collect the data.

After scanning, use the Pro-Data Viewer software to subtract the baseline and apply smoothing to all spectra. Calculate the ratio of CD values between the peak and valley of the standard sample. The acceptable range for the valid ratio is 2.08 ± 0.06.

2.4 X-ray crystallography

2.4.1 Crystallisation

Weigh 10mg of insulin lisargine and dissolve it in 1mL of 10mM hydrochloric acid solution. Place the solution in a refrigerator at 4°C overnight to obtain a sample solution with a concentration of 10mg/mL. Then, centrifuge the sample solution at 13200rpm for 10 minutes and collect the supernatant as the crystallization sample liquid. Crystallization was performed using the vapor diffusion hanging drop method, and two protein crystallization kits, Crystal Screen™ and Crystal Screen 2™ from Hampton, were used under a total of 98 conditions. The crystallization was carried out at a temperature of 12°C for 2–3 days for the preliminary screening of crystallization conditions.

Once the initial conditions were determined, the Additive Screen HT™ additive screening kit was used to optimize the crystal shape. Following the instructions, the drop was carefully added to the sitting drop plate, which was then sealed and placed in a 12°C incubator. The crystal shape was periodically observed. After obtaining the optimal crystals through screening, the crystals were gently picked using a loop with a diameter of φ = 0.2mm under a microscope. They were quickly immersed in a cryoprotectant solution, which consisted of the above-mentioned crystallization mother liquor containing 22.5% ethylene glycol, and then transferred to liquid nitrogen for storage and further use.

2.4.2 Data collection

The crystals from the liquid nitrogen were placed on the Shanghai Synchrotron Radiation Facility beamline 19U1 equipped with a 100K low-temperature device for diffraction data collection. The diffraction conditions were set as follows: detector-crystal distance of 400mm, exposure time of 0.0977 seconds, Δφ = 1.0°, temperature of 100K, total number of frames collected was 360, and the wavelength was 0.97774Å.
The collected single-crystal diffraction data were processed using the specialized protein crystallography software HKL2000. After indexing each individual frame, the data were scaled and merged according to the H3 (R3) crystallographic space group with the following unit cell parameters: \( a = 78.966 \, \text{Å} \), \( b = 78.966 \, \text{Å} \), \( c = 40.420 \, \text{Å} \), \( \alpha = 90^\circ \), \( \beta = 90^\circ \), \( \gamma = 120^\circ \). The integrated intensities for each reflection were obtained and then processed to obtain the original diffraction intensity data.

### 2.4.3 Structure determination of glulisine

The processed diffraction data were used in an attempt to solve the crystal structure by molecular replacement. The Phenix software’s molecular replacement program was employed, using the PDB ID 1EV3, which corresponds to the human insulin model. The program performed automated rotation and translation operations to find a unique solution, resulting in successful structure determination. The structure was further refined using the Phenix software package\(^2\).

### 2.4.4 Model building and refinement of glulisine

The initial model of insulin obtained through molecular replacement was further refined using iterative cycles of model adjustment with the crystallographic electron density map modeling software, Coot, and structure refinement with Refmac5. The refinement process resulted in a final insulin crystal structure model that closely matched the electron density map at a resolution of 3.20 Å, with reasonable model parameters. The refined model exhibited an R-factor of 0.143, R-free factor of 0.244, a root mean square deviation (r.m.s.d) of 0.014 Å for bond lengths, and an r.m.s.d of 1.451° for bond angles.

### 3. Results

#### 3.1 Determination of the complete molecular weight of insulin lisargine

Accurate determination of the complete molecular weight is a prerequisite for accurate protein sequencing, making the determination of the complete molecular weight of insulin lisargine the first step in structural analysis. In this study, high-resolution mass spectrometry was employed to measure the molecular weight of insulin lisargine. The results, as shown in Fig. 1, indicate that the observed average molecular weight \([M + H]^+\) is 6036.2 m/z, which closely matches the theoretical molecular weight of 6034.97.

#### 3.2 Determination of the amino acid sequence of insulin lisargine

Insulin lisargine contains three disulfide bonds, which can impact the peptide fragmentation in tandem mass spectrometry, leading to insufficient ion fragments for sequence analysis. To enhance fragmentation efficiency, in this study, the disulfide bonds were reduced using dithiothreitol (DTT) prior to sequence determination. The inter-chain disulfide bonds were cleaved, resulting in the reduction of insulin
into A and B chains. The reduced cysteine residues were then alkylated before separate sequence analysis.

The A chain of insulin lisargine consists of 21 amino acids, including 20 amide bonds. The tandem mass spectrum and fragmentation pattern are shown in Fig. 2. Both b ions and y ions are abundant, and all 20 amide bonds are cleaved, achieving 100% sequence coverage. The determined sequence is GIVEQCCTSICSLYQLENYCG, which matches the theoretical sequence.

The B chain of insulin lisargine consists of 32 amino acids, including 31 amide bonds. The tandem mass spectrum and fragmentation pattern are shown in Fig. 3. Fragment ions are abundant, and all 31 amide bonds are cleaved, achieving 100% sequence coverage. The determined sequence is FVNQHLCGSHLVEALTLVCGERGFFYTPKTKR, which matches the theoretical sequence.

3.3 Determination of insulin secondary structure by circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy in the far-ultraviolet region reflects the arrangement of protein peptide bonds. Through analysis using CDNN software, the relative content of different protein secondary structures, including α-helix, fold (antiparallel + parallel), β-turn, and random coil, can be predicted and calculated. The CD spectra in the near-ultraviolet region reflect the arrangement of chromophores in protein side chains, such as tryptophan and phenylalanine residues. Disulfide bonds also exhibit circular dichroism, but their CD signals appear in the near-ultraviolet region (250–340 nm). These pieces of information can be used as spectral probes to investigate their asymmetric microenvironment disturbances without interfering with the CD signals of peptide bonds in the far-ultraviolet region. Figure 4 shows the circular dichroism spectra of insulin lisargine in the far-ultraviolet region (A) and near-ultraviolet region (B).

The predicted secondary structure by CDNN software was used for fitting calculations to obtain the relative content of α-helix, fold, β-turn, and random coil at different wavelengths, as shown in Table 1. The α-helix content is the highest, at around 40% in all wavelength regions. The random coil is the next prominent component, accounting for approximately 27%, while the content of fold and β-turn is similar, both around 15%.
Table 1
Relative content of secondary structure of insulin lysine in different wavebands

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>α-Helix</td>
<td>n.d</td>
<td>n.d</td>
<td>45.3%</td>
<td>45.6%</td>
<td>42.0%</td>
<td>36.0%</td>
<td>40.1%</td>
</tr>
<tr>
<td>Antiparallel</td>
<td>n.d</td>
<td>n.d</td>
<td>1.2%</td>
<td>7.3%</td>
<td>7.9%</td>
<td>8.5%</td>
<td>9.5%</td>
</tr>
<tr>
<td>Parallel</td>
<td>n.d</td>
<td>n.d</td>
<td>8.5%</td>
<td>7.0%</td>
<td>5.6%</td>
<td>5.3%</td>
<td>5.6%</td>
</tr>
<tr>
<td>β-turn</td>
<td>n.d</td>
<td>n.d</td>
<td>15.1%</td>
<td>15.1%</td>
<td>15.4%</td>
<td>21.0%</td>
<td>16.9%</td>
</tr>
<tr>
<td>Rndm.Coil</td>
<td>n.d</td>
<td>n.d</td>
<td>24.6%</td>
<td>22.6%</td>
<td>27.0%</td>
<td>29.2%</td>
<td>29.8%</td>
</tr>
<tr>
<td>Total Sum</td>
<td>-</td>
<td>-</td>
<td>94.7%</td>
<td>97.7%</td>
<td>97.8%</td>
<td>100.0%</td>
<td>101.9%</td>
</tr>
</tbody>
</table>

3.4 X-ray Crystallography

3.4.1 Crystals

After 24 hours, crystals of typical size 0.2 × 0.2 × 0.2 mm were obtained. Preliminary screening of crystallization conditions revealed that the crystallization solution containing isopropanol was able to facilitate insulin crystallization. To improve the quality of crystal diffraction, additive screening kits were used in the isopropanol-based crystallization system to further optimize the crystal morphology. Eventually, larger crystals were obtained by adding thiocyanate, as shown in Fig. 5.

3.4.2 Crystal Diffraction and Electron Density

From the single X-ray diffraction image (Fig. 6), it can be observed that the crystal diffraction resolution is approximately 2.0 Å. This resolution provides information about the arrangement of atoms within the crystal lattice. By analyzing the diffraction pattern, the electron density distribution within the crystal can be determined, allowing for the determination of the three-dimensional structure of the molecule.

The formation of disulfide bonds contributes to protein folding and stabilization and is essential for maintaining protein structure and function. The position and length of disulfide bonds can be observed through electron density cloud maps. The sentence describes a dimer model with a total of 6 pairs of disulfide bonds. The pairs are labeled as A6-A11, A7-B7, A20-B19, C6-C11, C7-D7, and C20-D19. The electron density map, shown in Fig. 7, illustrates the distribution of electron clouds. Among the disulfide bonds, A6-A11 and C6-C11 are intra-chain bonds with bond lengths of 2.01Å and 2.02Å, respectively. The inter-chain bonds, A7-B7, A20-B19, C7-D7, and C20-D19, have bond lengths of 2.07Å, 2.04Å, 2.07Å, and 2.03Å, respectively.

3.4.3 Structure Analysis of Lysine and Arginine Insulin

The crystal structure of lysine and arginine insulin is depicted in Fig. 8. Each crystallographic asymmetric unit contains a dimer of insulin molecules, comprising two A chains and two B chains. In the spatial
coordinates shown in Fig. 8, the A chains are denoted as A and C, while the B chains are denoted as B and D. Additionally, the dimeric structure includes two Zn ions and Cl ions, with each Zn and Cl ion occupying approximately one-third of the sites, along with 18 water molecules. It should be noted that due to the flexibility and lower resolution of crystal diffraction, the C-terminal residues of the B chains could only be modeled up to residue 29, which is a lysine residue.

The obtained crystal structure model exhibits favorable geometric and stereochemical parameters, meeting the established academic criteria. The majority of amino acid residues show excellent agreement with the electron density map, and there are no significant unexplained residual densities in the protein molecule and its surrounding regions.

To further analyze the lysine and arginine insulin structure, a comparison was made with the publicly available insulin model (PDB entry: 1EV3) under the same crystallographic space group. The comparison, as shown in Fig. 9, revealed a remarkable similarity between the two structures, indicating a conserved spatial arrangement between lysine and arginine insulin and human insulin.

4. Discussion

Lisargine is a novel insulin analogue designed for extended duration of action. In this study, a recombinant form of lisargine was created through genetic engineering. The molecular structure of lisargine involves substituting the 21st position Asn with Gly in the A chain, as well as introducing Lys and Arg at the 31st and 32nd positions, respectively, in the B chain. This substitution strategy effectively prevents the deamidation of Asn at the A chain's C-terminus, which can lead to the formation of covalently linked dimers under acidic conditions. Additionally, the addition of amino acids in the B chain modifies its isoelectric point, facilitating the formation of "microcrystals" upon subcutaneous injection and thus prolonging its therapeutic effect.

The development of insulin analogues holds significant importance, encompassing several key aspects. Firstly, it improves diabetes treatment by offering more effective blood glucose control and convenient therapeutic options, ultimately enhancing patients’ quality of life. Secondly, it increases medication selectivity by providing a range of insulin analogues to cater to the varying needs of different diabetes patients. These include rapid-acting, intermediate-acting, and long-acting insulin, among others. Developing diverse insulin analogues allows for personalized treatment that meets the specific requirements of individual patients. Thirdly, it advances pharmaceutical technology. The development of insulin analogues has been a driving force behind advancements in pharmaceutical technology. The research and development process involves the application of various advanced techniques and methods, such as genetic engineering, protein engineering, and crystallography\(^3\). These technologies not only serve as platforms for insulin analogue development but also provide valuable experience and insights for research and development of other medications.
In our study, we found that the dimeric model contains a total of six pairs of disulfide bonds: A6-A11, A7-B7, A20-B19, C6-C11, C7-D7, and C20-D19. Human insulin, specifically insulin lisargine, possesses three disulfide bonds. These bonds play a crucial role in connecting different parts of the insulin molecule, stabilizing its structure, and maintaining its biological activity. More specifically, in insulin lisargine, the disulfide bonds link Cys7 and Cys19 of the A chain, Cys20 of the A chain with Cys6 of the B chain, and Cys46 of the A chain with Cys30 of the B chain. The presence of these disulfide bonds is of utmost importance for the structure and function of the insulin molecule. Disulfide bonds in the insulin structure are significant as they not only stabilize the three-dimensional structure but also facilitate insulin folding, which affects its stability and half-life. The presence of six pairs of disulfide bonds in insulin lisargine contributes to its longer half-life and plays a crucial role in preserving its biological activity.

Insulin lisargine is a novel long-acting insulin analog that can be evaluated for its pharmacological efficacy and quality using relevant assessment criteria. These protein-based drugs are typically produced through genetic engineering techniques, involving the synthesis of active substances within organisms and subsequent complex processes such as refolding, enzymatic cleavage, and purification. They primarily consist of proteins, peptides, or peptide derivatives. Compared to traditional chemical drugs, protein-based drugs exhibit highly intricate two-dimensional and three-dimensional conformations. The complexity of their production processes and structures presents new challenges for quality research and control, necessitating multidimensional analytical techniques to effectively characterize their physicochemical properties.

Proteins possess a higher-order structure comprising secondary, tertiary, and quaternary structures, with secondary structures including α-helices, β-sheets, β-turns, and random coils. The higher-order structure of a protein significantly influences its therapeutic efficacy, functionality, bioavailability, and safety.

Therefore, the analysis and control of the higher-order structure play a vital role in the development process of protein-based biosimilars.

X-ray diffraction is a scientific discipline that utilizes X-rays to study the structural characteristics of substances. By analyzing the scattering of X-rays by electrons, information about the electron density distribution within crystals can be obtained, enabling the determination of the molecular structure within the crystal. In the early 20th century, several Chinese scientists pioneered the application of X-ray diffraction in protein research using synthesized bovine insulin as their experimental material. This groundbreaking work led to the successful acquisition of China's first high-resolution protein crystal structure and laid the foundation for the widespread utilization of X-ray diffraction in the study of protein higher-order structures. In recent decades, the integration of X-ray crystallography and life sciences has revolutionized our understanding of protein structures and their functional mechanisms. By analyzing the coordinates of individual atoms within protein spatial structures, valuable insights into the three-dimensional organization of proteins have been gained, making X-ray diffraction an indispensable tool for investigating protein structures. The overall process of X-ray analysis involves three key steps. First, obtaining protein crystals, which is a prerequisite for conducting X-ray research. Second, measuring
various geometric parameters of the crystal diffraction to determine the crystal system, cell parameters, and the number of independent parameters within the unit cell. Finally, measuring the intensity of diffraction spots and employing Fourier transformation to reconstruct the molecular structure based on these intensities. Since each part of the molecule contributes to diffraction, it is essential to measure the intensity of all diffraction spots to accurately reconstruct the molecular structure.

This study provides the first novel structural information on insulin lisargine analogs. This information reveals the dissociation of this insulin analogue in solution, which may be important for optimizing the drug formulation and reducing the side effects of this drug. In addition, knowledge of the structure is necessary and important when considering the docking/binding of insulin lisargine to its receptor.

5. Conclusion

This study provides the first novel structural information on insulin lisargine analogs. This information reveals the dissociation of this insulin analogue in solution, which may be important for optimizing the drug formulation and reducing the side effects of this drug. In addition, knowledge of the structure is necessary and important when considering the docking/binding of insulin lisargine to its receptor.

Declarations

Contributions

P.W. and ZM.Z. (Zhiming Zheng) conceived the idea and supervised the research. Z.Z. (Zhu Zhu) and H.W. designed and performed the experiments, analyzed and interpreted data, performed statistical analysis and wrote the manuscript. L.W. and Z.W. provided technical assistance in the experiments. All authors have read and agreed to the published version of the manuscript.

Ethics declarations

Competing interests

The authors declare no competing interests.

Rights and permissions

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission
Funding: This work was sponsored by financial support from China National Key Research and Development Program 2019YFA0904304); the Major Projects of Science and Technology of Anhui Province (202103a06020003; 202004b11020014).

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Data availability

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to a partner wanting to continue the research.

References


Figures
Figure 1

The mass spectrum of insulin lysine
Figure 2

The MS2 spectrum and the cleavage pattern of chain A
Figure 3

The MS2 spectrum and the cleavage pattern of chain B
Figure 4

CD spectrum of insulin lisargine Far-UV Region (a) and Near-UV Region (b)

Figure 5

Protein crystal of insulin lisargine. a: workflow of X-ray Crystallography; b: Crystals.
Figure 6

X-ray diffraction pattern of insulin lisargine
Figure 7

Electron density cloud map of six disulfide bonds

Figure 8

Crystal structure streamer diagram of insulin lisargine
Figure 9

Comparison of crystal structure between insulin Lysin and human insulin. (Red: insulin lisargine; Green: human insulin)