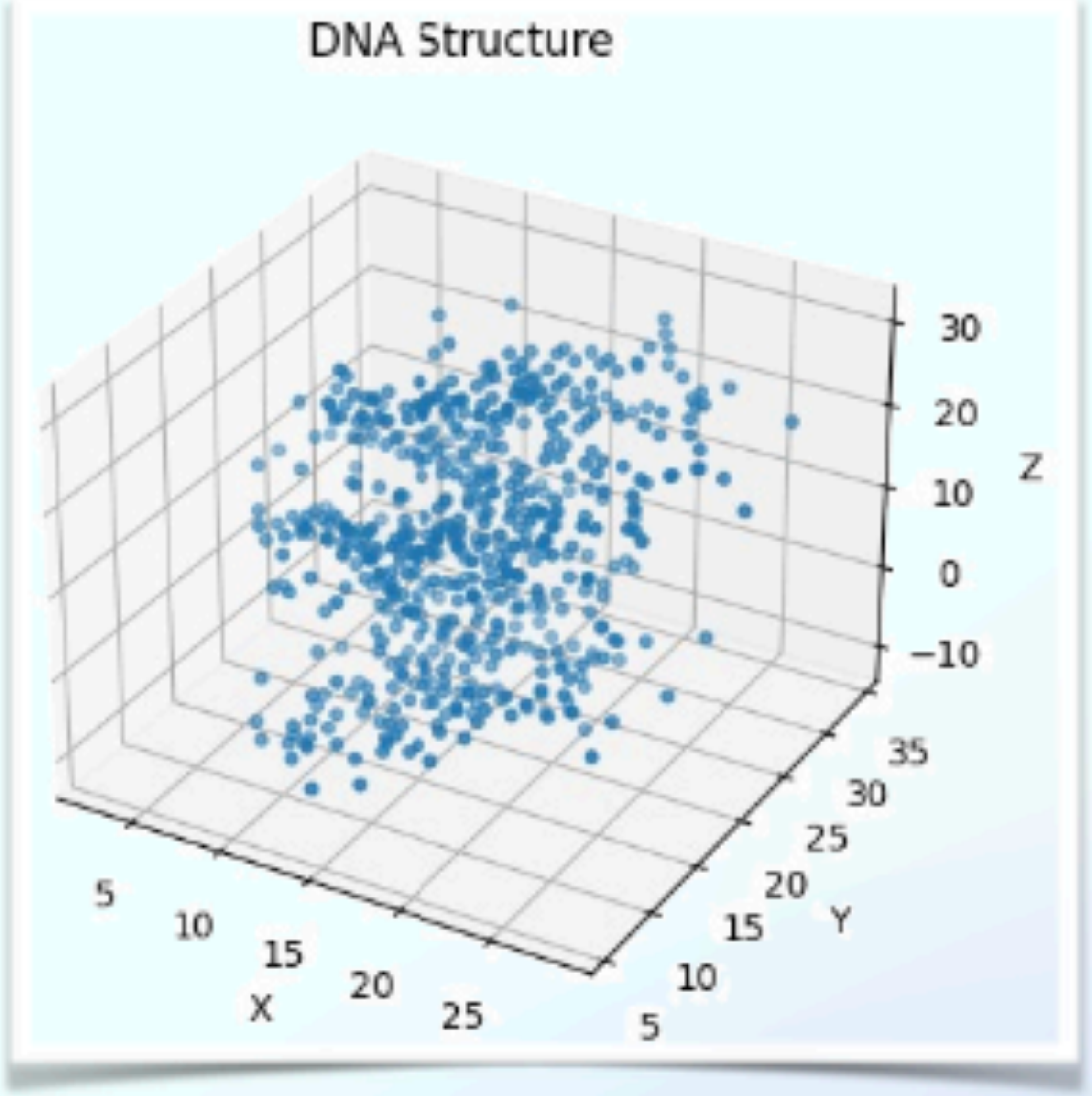
LINEAR ALGEBRA IN BIO INFORMATICS

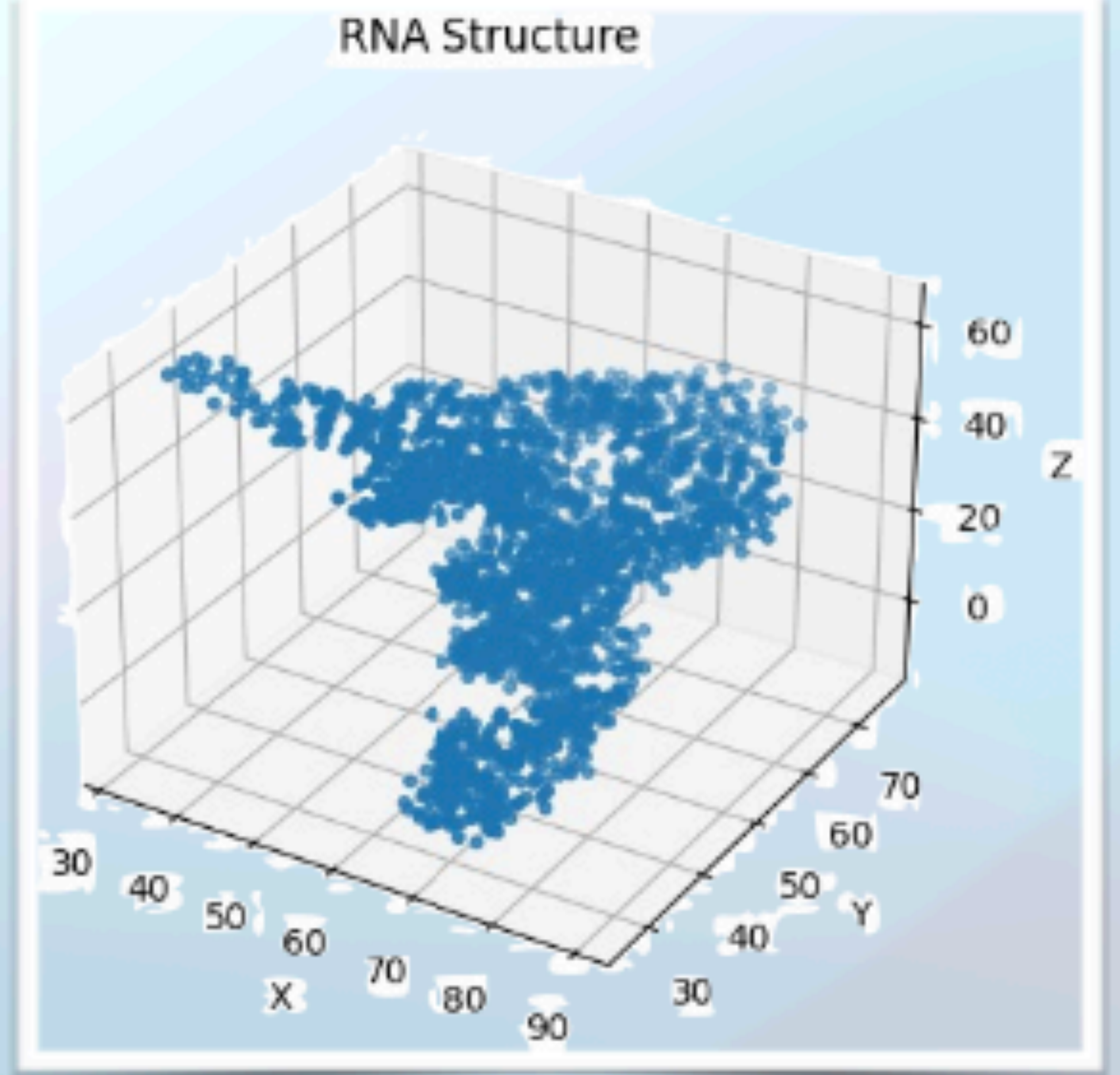
Linear algebra plays a crucial role in various applications within bioinformatics.

- **Sequence Alignment**-Comparing biological sequences such as DNA, RNA, or protein sequences
- **Protein Structure Prediction**: Predicting the three-dimensional structure of proteins.
- **Studying biological networks:** such as protein-protein interaction networks or metabolic pathways.
- **Structural Bioinformatics:** Analysing and predicting the structure of biomolecules like proteins and RNA.
- **Machine Learning in Bioinformatics:** Developing machine learning models for tasks like classification, clustering, and prediction in biological data.

Genome Structures

- Genome structures are complex arrangements of genetic material within the cell.
- The primary components of genomes are DNA molecules.
- Their structures are represented like the linear sequence of nucleotides, the 3D arrangement of chromatin, etc
- Linear algebra can be used to model certain aspects of genome structures, particularly when dealing with large-scale genomic data analysis.





Here are a few ways linear algebra can be applied to model aspects of genome structures:

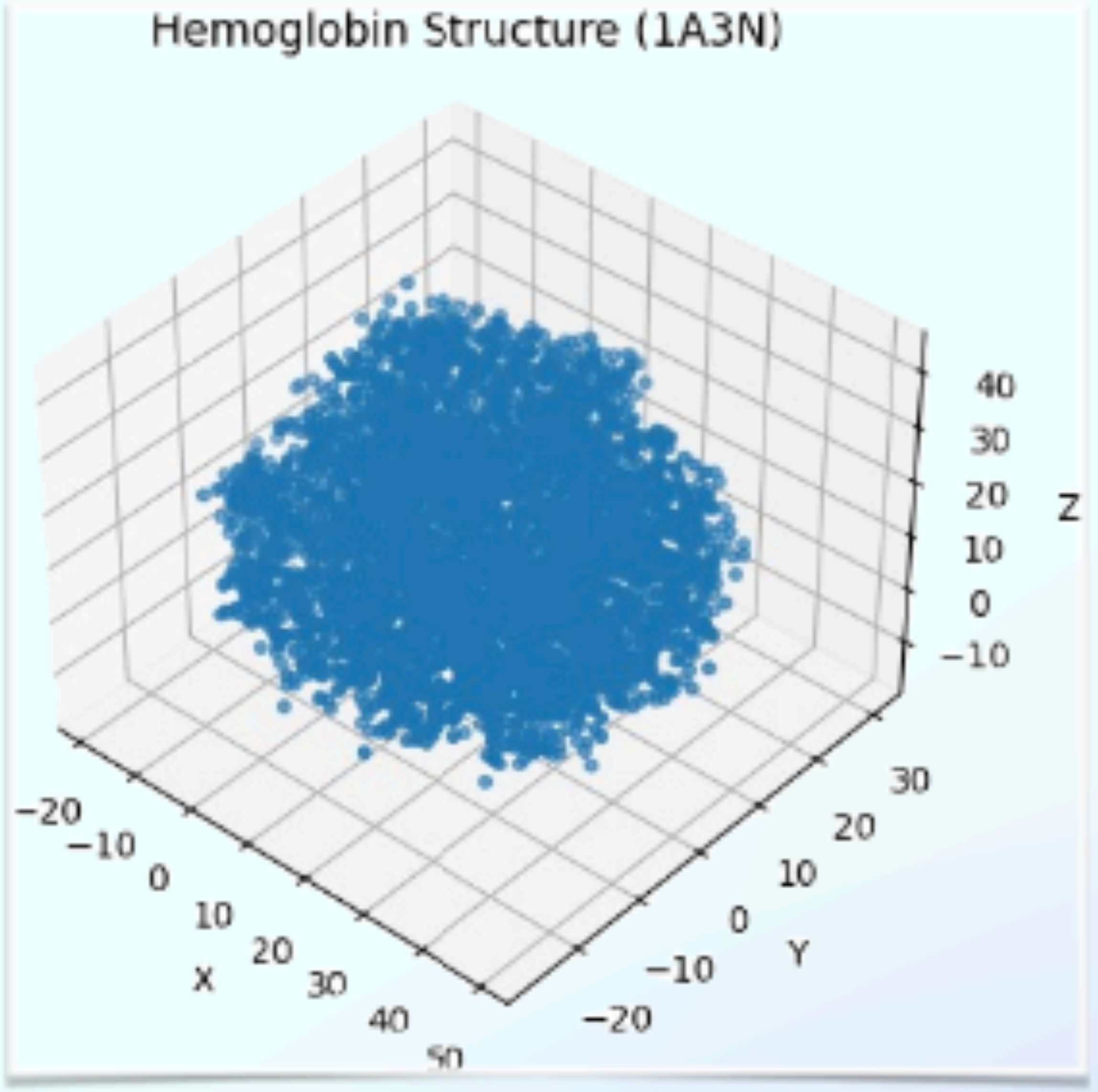
- **Vector Representation**: Assign numerical values to each nucleotide (e.g., A=1, C=2, G=3, T=4), and then represent a DNA sequence as a vector of these values.
- **One-Hot Encoding:** Each position in the vector corresponds to a specific nucleotide, and only one position has a value of 1 while the others are 0.
- **Principal Component Analysis (PCA):** Apply linear algebra techniques like PCA to reduce the dimensionality of high-dimensional genomic data.
- linear algebra can be used to analyse and visualise **chromatin interaction matrices**, which represent the frequency of interactions between different genomic regions.
- Model interactions between genes or genomic elements as a graph and use linear algebra to analyse properties of the graph, such as **eigenvectors and eigenvalues.**

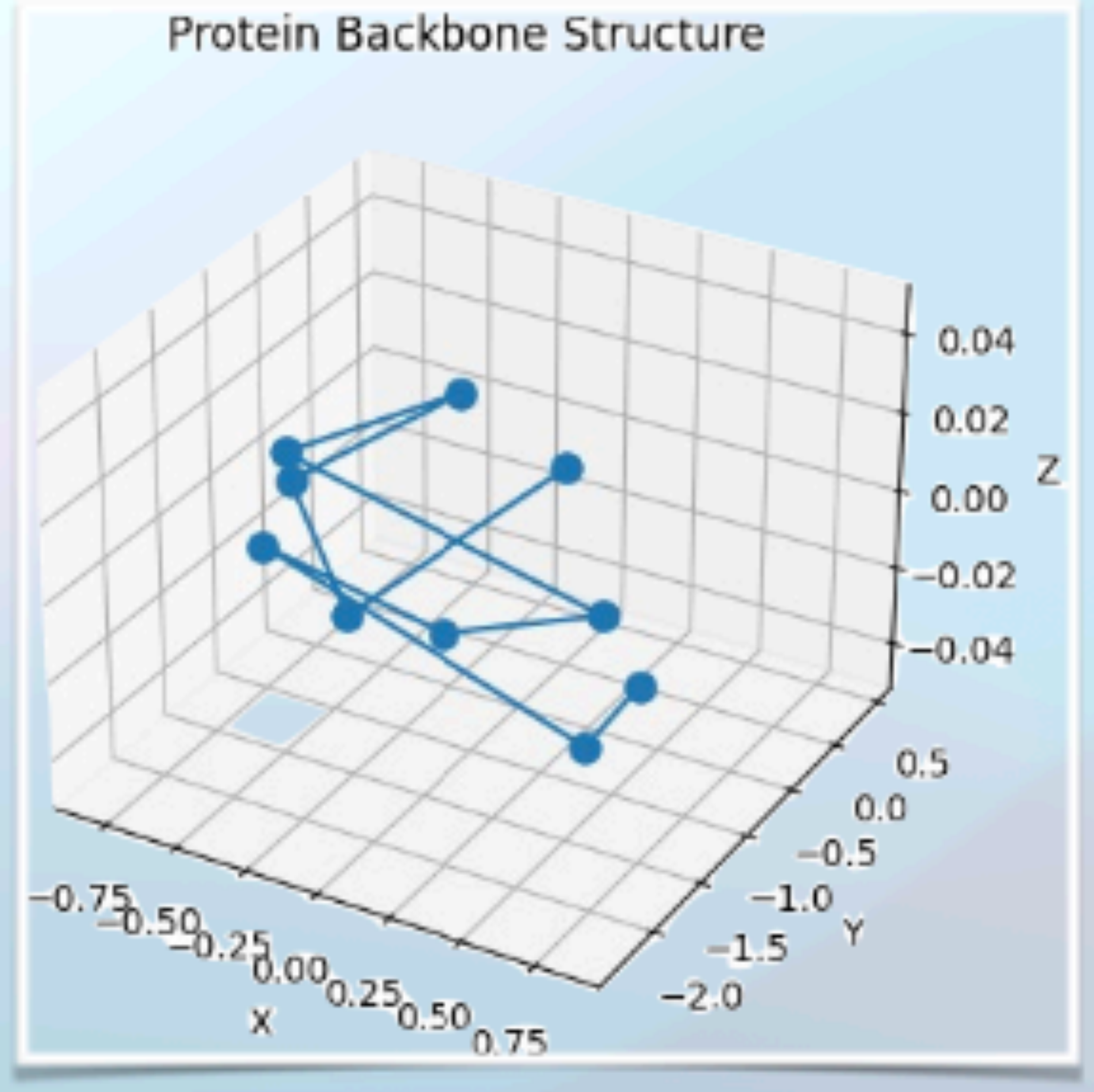
```
ACGT -> [1, 2, 3, 4]
```

```
A: [1, 0, 0, 0]
C: [0, 1, 0, 0]
G: [0, 0, 1, 0]
T: [0, 0, 0, 1]
```

Protein Structures

- The backbone structure is often represented as a series of 3D coordinates for each node. Linear algebra operations can be used to manipulate these coordinates.
- The application of protein structure analysis and manipulation using tools like MDAnalysis and linear algebra has various real-life applications in bioinformatics, structural biology, drug discovery, and related fields.
- These applications highlight the importance of structural analysis in various biological and medical contexts, ultimately contributing to advancements in healthcare, biotechnology, and our understanding of fundamental biological processes.





Original Image
Translated Image
Z
X
Y
Protein Backbone Structure
X
Y
Protein Structure
X
Y
Z
Original Protein Structure
X
Y
Z
Transformed Protein Structure
X
Y
Z

- Drug Discovery and Design: Analysing protein structures helps identify binding sites where drugs can interact. Analysing protein structures helps identify binding sites where drugs can interact.
- Enzyme Mechanism Elucidation: Analysing how proteins interact helps in understanding cellular processes and developing therapeutics.
- Vaccine Development: Analysing the structure of antigens helps in designing vaccines that mimic the virus or pathogen.
- Biotechnology and Protein Engineering.
- Molecular Dynamics Simulations.
- Disease Understanding and Treatment.

Let's break down the linear algebra concepts involved.

- **Coordinates as Vectors:** The coordinates matrix can be considered as a collection of column vectors, where each column represents the position vector of an atom.

$$\mathbf{C} = \begin{bmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_n & y_n & z_n \end{bmatrix}$$

Here, n is the number of atoms.

- **Masses as vetor :** $\mathbf{M} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$

- **Center of Mass Calculation**: $\mathbf{CM} = \frac{\sum(\mathbf{C} \cdot \mathbf{M})}{M_{\text{total}}}$

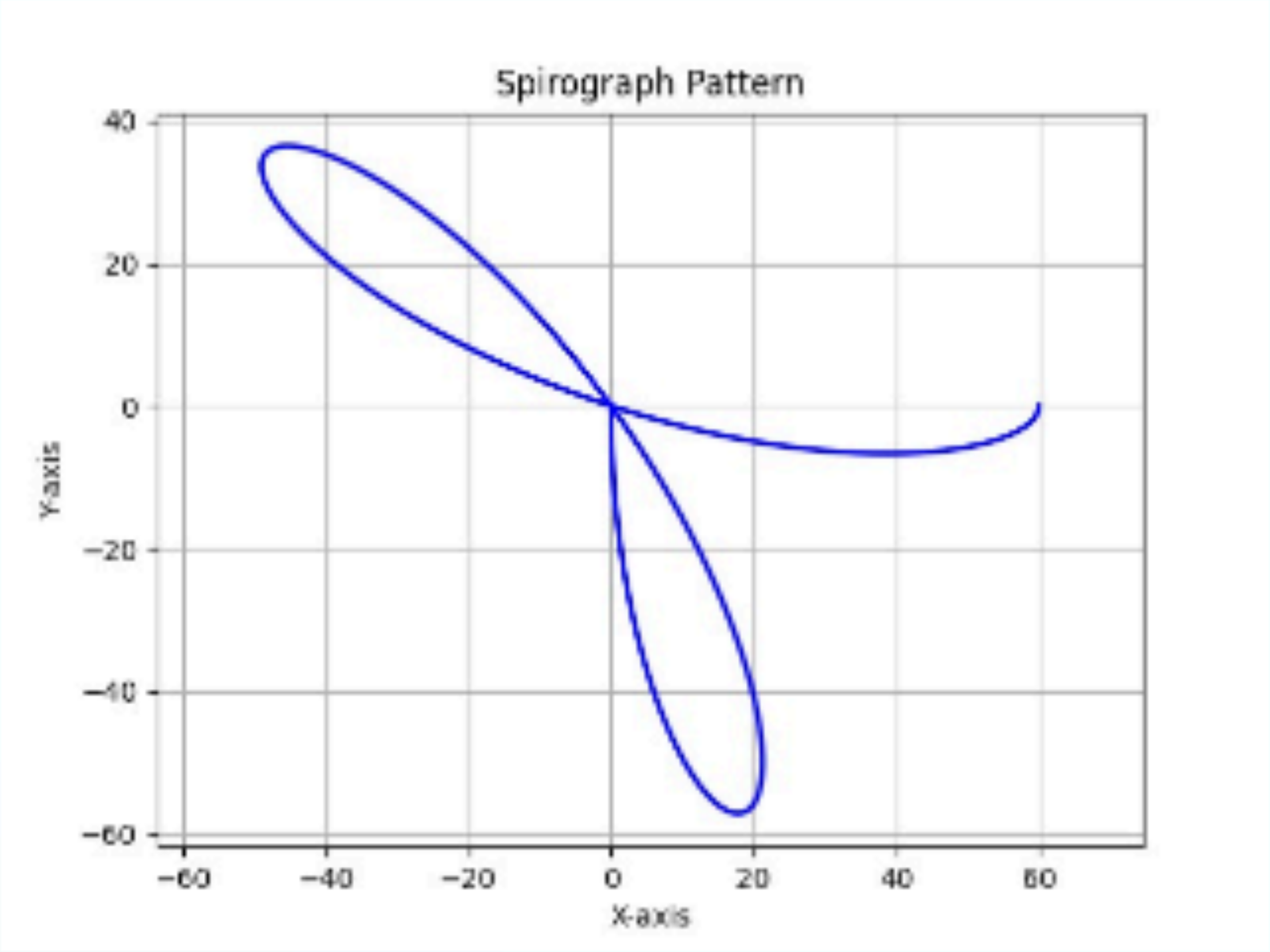
$$\mathbf{C} \cdot \mathbf{M} = \begin{bmatrix} x_1 m_1 + y_1 m_2 + z_1 m_3 \\ x_2 m_1 + y_2 m_2 + z_2 m_3 \\ \vdots \\ x_n m_1 + y_n m_2 + z_n m_3 \end{bmatrix}$$

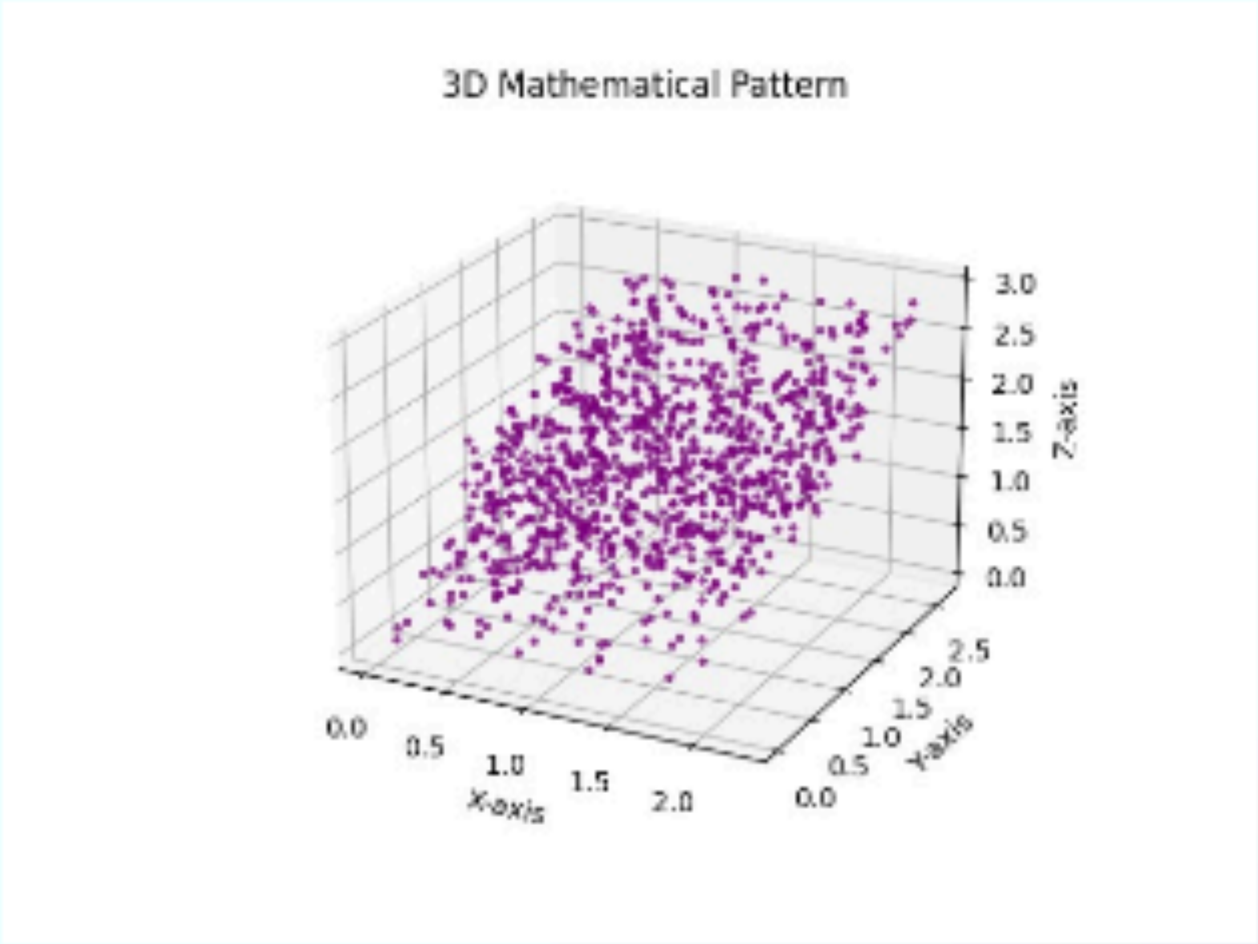
- **Translation of Coordinates:** The translation vector is added element-wise to each corresponding element in the coordinates matrix. This is a simple vector addition operation.

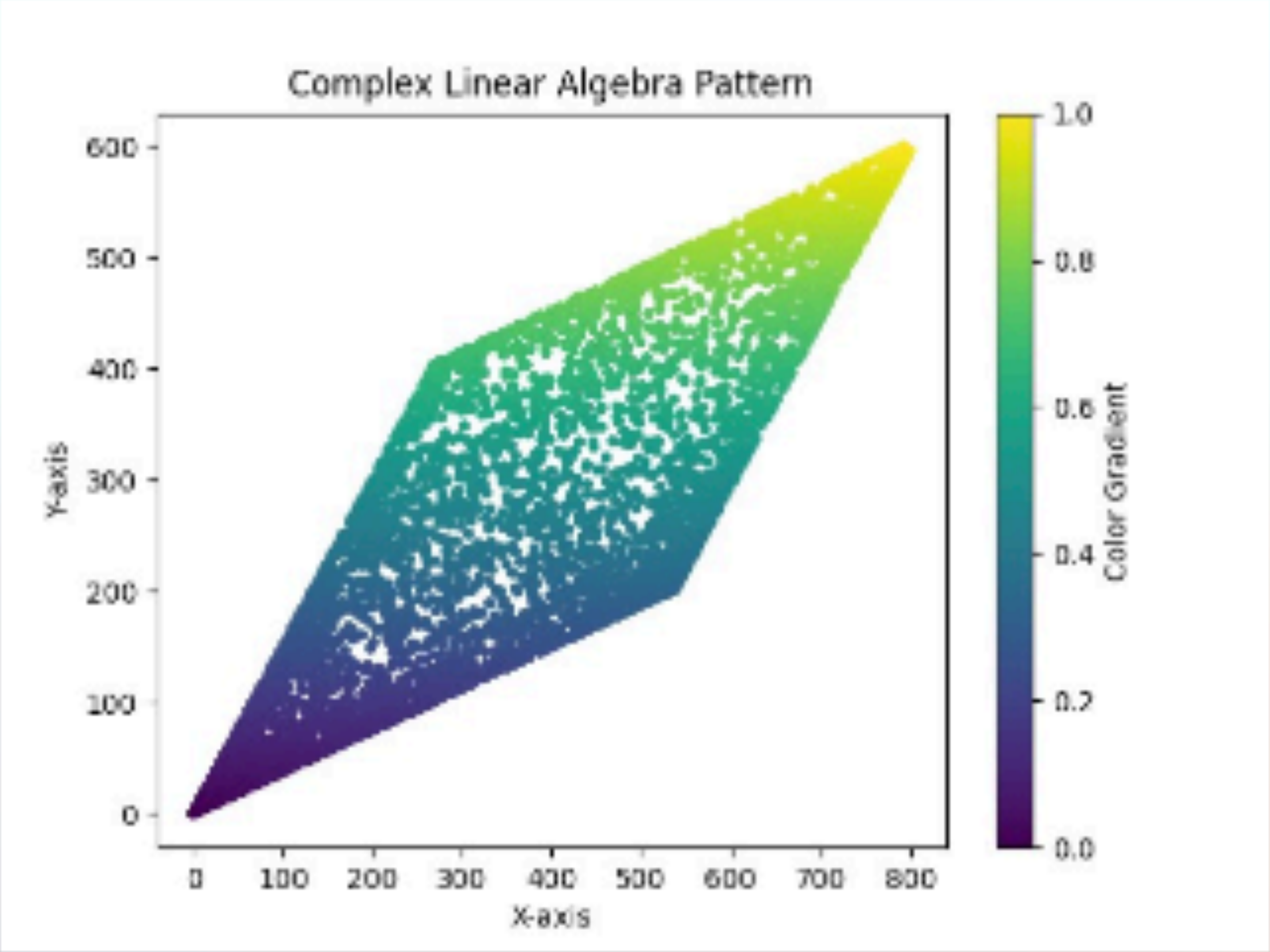
$$\sum(\mathbf{C} \cdot \mathbf{M}) = \begin{bmatrix} \sum(x_i m_i) \\ \sum(y_i m_i) \\ \sum(z_i m_i) \end{bmatrix}$$

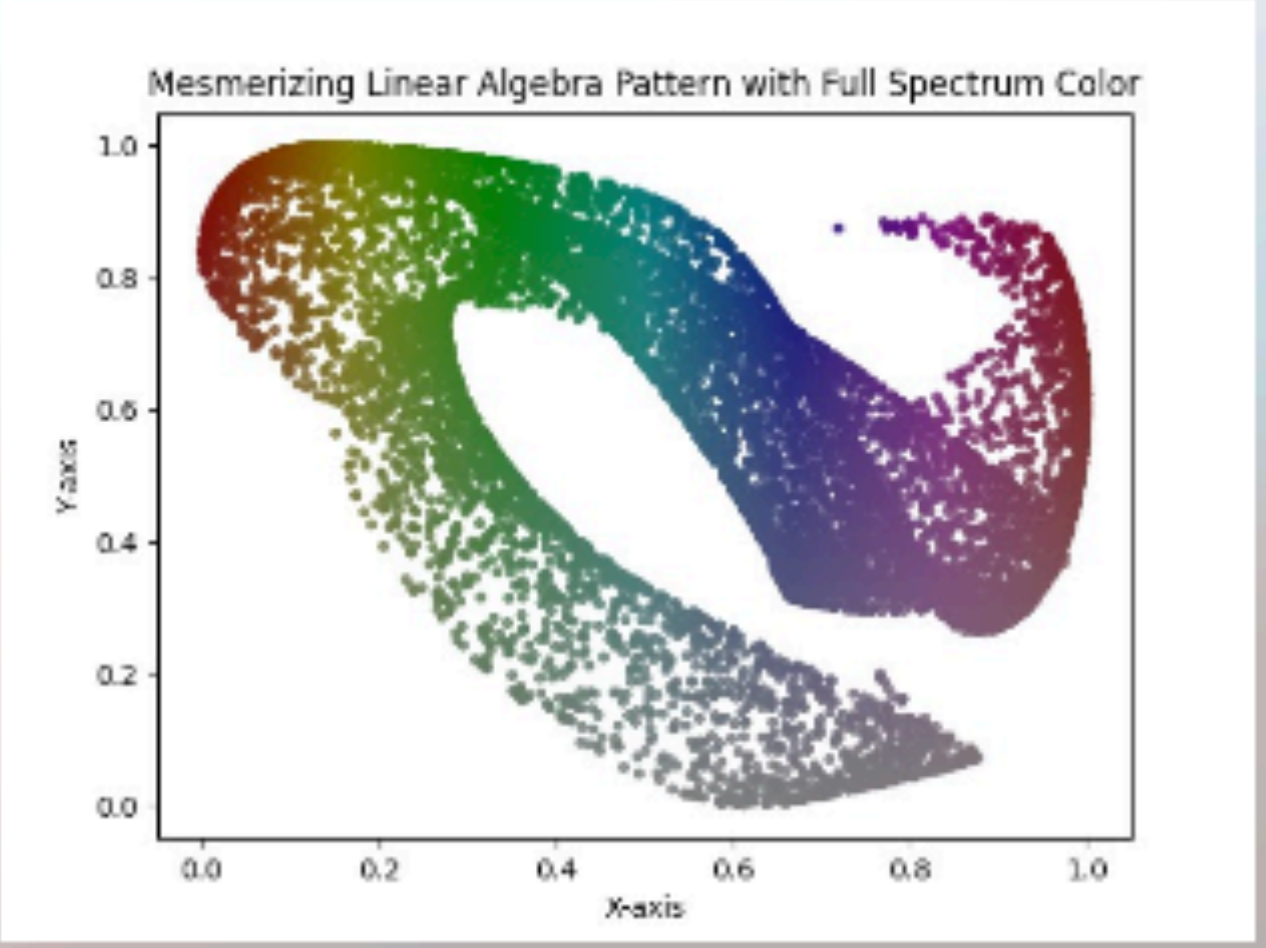
- e linear algebra concepts include treating coordinates as vectors, calculating the centre of mass using vector operations (dot product and scalar division), and translating coordinates using vector addition.

MATHEMATICAL PATTERNS GENERATED USING LINEAR ALGEBRA









CONCLUSION

- These operations are fundamental linear algebra concepts applied to the manipulation of 3D coordinates in the context of a synthetic protein-like structure.
- These applications highlight the importance of structural analysis in various biological and medical contexts.
- Ultimately contributing to advancements in healthcare, biotechnology, and our understanding of fundamental biological processes
- Creating a realistic representation of a protein structure involves using more advanced methods, such as molecular dynamics simulations, which go beyond basic linear algebra

THANK YOU