مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٩) العدد (١٢) ٢٠٢٥
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (9) No. (12) 2025

كشف تطابق التتابعات البروتينية في منطقة الشفق باستخدام تحويل فورييه

م. علي رستم•

(تاريخ الإيداع ٢٠٢٥/٩/٣ . قُبِل للنشر في ٢٠٢٥/١٠/١٥)

□ ملخّص □

يؤدي التطابق الجيني، لاسيما بين تسلسلات الأحماض الأمينية دوراً حاسماً في مختلف مجالات البحوث البيولوجية والطبية، حيث يساعد تحديد أوجه التشابه والاختلاف بين هذه التسلسلات في فهم العلاقات التطورية والتنبؤ بوظيفة البروتينات، بالإضافة إلى أن التطابق العالي في تسلسل الحموض الأمينية المكونة للبروتينات المراد دراسة التطابق بينها يشير إلى أوجه تشابه وظيفي أو بنيوي وهو ما يمكن أن يوفر رؤى حول أدوارها البيولوجية.

نقدم في هذا البحث منهجية مقترحة تعتمد على تطبيق تحويل فورييه على تسلسل الأحماض الأمينية المكونة للبروتينات للكشف عن الأنماط وأوجه التشابه في مجال التردد وذلك من خلال تحويل ترددات ورود الأحماض الأمينية إلى المجال الترددي، حيث مكّن هذا النهج من تحديد البنى الدورية والأنماط المتكررة التي قد لا تكون واضحة في بيانات التسلسل.

تعد هذه الطريقة مفيدة بشكل خاص لتحديد أوجه التشابه البنيوية بدقة أكبر حتى عندما تكون نسبة تطابق التسلسلات منخفضة.

كلمات المفتاحية: المعلوماتية الحيوية، التطابقات الجينية، تحليل التسلسل البروتيني، محاذاة التسلسل الزوجي، تحويل فورييه، ترددات الأحماض الأمينية، منطقة الشفق.

•ماجستير في قسم هندسة النظم الحاسوبية والالكترونية ، كلية هندسة تكنولوجيا المعلومات والاتصالات، جامعة طرطوس، طرطوس، سوريا
alirostom5798@gmail.com

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٩) العدد (١٢) ٢٠٢٥
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (9) No. (12) 2025

Detection of Protein Sequence Identity in the Twilight Zone Using Fourier Transform

Eng. Ali Rostom*

(Received 3/9/2025 . Accepted 15/10/2025)

□ ABSTRACT □

Genetic matches, especially between amino acid sequences, plays a crucial role in various fields of biological and medical research, where identifying similarities and differences between these sequences helps in understanding evolutionary relationships and predicting the function of unknown proteins. In addition, high similarity in the amino acid sequences of the proteins to be studied indicates functional or structural similarities, which can provide insights into their biological roles.

In this study, we present a proposed methodology based on applying Fourier transform to the amino acid sequences of proteins to detect patterns and similarities in the frequency domain by converting the frequencies of amino acid occurrences to the frequency domain. This approach allows the identification of periodic structures and recurring patterns that may not be apparent in the sequence domain sequences.

This method is particularly useful for identifying structural similarities more accurately even when the sequence match ratio is low.

Keywords: bioinformatics, Genetic matches, protein sequence analysis, pairwise alignment, Fourier transform, amino acid frequencies, twilight zone.

*Postgraduate in Department of Computer and Electronic Systems Engineering, Faculty of Information Technology and Communications Engineering, Tartous University, Tartous, Syria
alirostom5798@gmail.com.

١- مقدمة

في مجالات الدراسات الأكاديمية أصبحت المعلوماتية الحيوية ضرورة لا بد منها، فهي يمكن أن تشارك في تحليل تركيب الأحماض الأمينية والبروتينات.

إحدى التقنيات الأساسية في مجال المعلوماتية الحيوية هي محاذاة التسلسل الزوجي Pairwise Alignment، والتي تتضمن مقارنة تسلسلين لتحديد مناطق التشابه التي قد تشير إلى علاقات وظيفية أو هيكلية أو تطورية مشتركة بينهما ويوجد العديد من الخوارزميات لدراسة المحاذاة لتسلسلين بروتينيين تعتمد على البرمجة الديناميكية مثل خوارزمية الاصطفاف العامة Global alignment والاصطفاف المحلي Local alignment.

تستخدم المحاذاة الزوجية في أكثر أشكالها صرامةً طريقة تسمى البرمجة الديناميكية وهي طريقة دقيقة للغاية ولكن حسابها مكلف بحيث يجب أن تعرف الخوارزمية ما الذي تبحث عنه وكيف يمكنها تقييم ما تجده [1,2]، وتحقيقًا لهذه الغاية تم إنشاء مصفوفات المقارنة Comparison Matrices وهي تحدد درجة لكل إمكانية تطابق ممكنة.

تكون التتابعات الجينية عموماً والبروتينية خاصةً متجانسة عندما يكون لها سلف تطوري مشترك، ويمكن تحديد هذه الصفة من خلال التشابه في البنية أو التركيب، وهو ما يعكس غالبًا علاقاتها الوظيفية.

يعد فهم التتابعات المتجانسة أمرًا بالغ الأهمية لأنها توفر نظرة ثاقبة للعمليات التطورية التي تشكل الجينوم وتساعد في التنبؤ بوظيفة الجينات المجهولة استناداً إلى التتابعات المتجانسة المعروفة التي تعد حجر الزاوية في علم الجينوم المقارن Comparative genomics [3].

في محاذاة التسلسلات تمثل منطقة الشفق "Twilight Zone" تحدي كبير بسبب الغموض الموجود فيها حول تحديد التماثل بين التسلسلات التي تُظهر نتيجة تطابق منخفضة والتي عادة ما تكون بين ٢٥-٣٠٪ حيث تكافح خوارزميات الاصطفاف لتمييز التسلسلات المتجانسة في هذه المنطقة [4].

قدم الباحثان saul B.Needleman و Christian D.wunsch عام ١٩٧١ خوارزمية لمقارنة التتابعات البيولوجية تستخدم البرمجة الديناميكية لمحاذاة التسلسل العام وتحديد المحاذاة المثلى من بين تسلسلات البروتين من خلال النظر في جميع المحاذاة الممكنة واختيار المحاذاة التي لها اعلى درجة تشابه [5].

تعد هذه الخوارزمية مكلفة حسابياً وتطلب موارد حسابية كبيرة للبروتينات الطويلة كما أنها تفرض وجود عقوبة فجوة خطية وهو لا يعكس دائماً السيناريوهات البيولوجية بدقّة حيث يمكن أن تختلف عقوبة الفجوات gap penalty [1,2].

بالإضافة إلى ذلك، فإن اعتماد هذه الخوارزمية على مصفوفات تسجيل الدرجات يعني أن دقتها تعتمد على جودة هذه المصفوفات، والتي قد لا تمثل العلاقات التطورية بشكل مثالي[6].

طور الباحثان T.F smith و M.S waterman خوارزمية لإجراء الاصطفاف الثنائي المحلي مما يجعلها مفيدة بشكل خاص في تحديد المجالات الوظيفية والمناطق الأخرى ذات الأهمية البيولوجية داخل التسلسلات [7].

وتعد أيضاً مكلفة من الناحية الحسابية، وذلك بسبب تعقيدها التربيعي من حيث الزمان والمكان وكذلك الأمر تعتمد بشكل أساسي على اختيار مصفوفات تسجيل الدرجات وغرامات الفجوات التي قد لا تلتقط السياق البيولوجي بشكل مثالي [1,2].

تطرّقت الباحثة الإنكليزية Lisa Mullan في عام ٢٠٠٥ إلى التحديات الحسابية التي تواجهها خوارزميات البرمجة الديناميكية المحلية والعامة، بما في ذلك القدر الكبير من الذاكرة المطلوبة لتسلسلات البروتين.

أكّد الباحثون Jiannan Chao, Furong Tang, and Lei Xu أنّ أهمية ودقّة المحاذاة تعتمد اعتماداً كبيراً على اختيار مصفوفة تسجيل الدرجات حيث يمكن أن يؤدي اختلافها إلى نتائج متفاوتة، ولا توجد مصفوفة واحدة مثالية بشكل عام لجميع أنواع التسلسلات أو المسافات التطورية مما يخلق تحدّي في تحقيق محاذاة متّسقة وذات معنى بيولوجي.

قدم الباحثون M.O. Dayhoff, R.M. Schwartz, and B.C. Orcutt مصفوفات PAM بإصداراتها المختلفة وهي أداة أساسية لتسجيل تسلسل البروتين بناءً على بدائل الأحماض الأمينية المرصودة مما أحدث ثورة في المعلوماتية الحيوية [8,6].

كشف روست في عام ١٩٩٩ تحديات محاذاة تسلسل البروتينات في "منطقة الشفق"، حيث يكون التشابه في التسلسل منخفضاً، مما يجعل اكتشاف التماثل صعباً [9].

تعد هذه المنطقة حاسمة لفهم حدود خوارزميات المحاذاة، مما يسلط الضوء على الحاجة إلى أساليب متقدمة لتنفيذ المحاذاة وتحسين الدقة.

أظهرت النماذج اللغوية البروتينية الحديثة تقدماً ملحوظًا في مجالات المحاذاة والتنبؤ البنيوي، إلا أن فعاليتها تنخفض بشكل واضح عند التعامل مع البروتينات الواقعة ضمن منطقة الشفق، حيث تقل الهوية التسلسلية إلى مستويات يصعب معها التحقق من التشابه البنيوي أو الوظيفي، فقد طُوّر نموذج PEbA خصيصاً لتحسين المحاذاة في هذه المنطقة عبر تمثيل شعاعي للتسلسلات، لكنه لا يزال يواجه تحديات في استنتاج العلاقات البنيوية الدقيقة [10].

وبالمثل، يُظهر نموذج ESMFold أداءً عاليًا في التنبؤ البنيوي على نطاق واسع، إلا أن دقته تتراجع بشكل ملحوظ في حالات انخفاض التشابه التسلسلي [11].

٢- هدف وأهمية البحث

يقدم هذا البحث منهجية مقترحة للكشف عن مدى تطابق البروتينات بالاعتماد على تحويل فورييه، وهذه المنهجية قادرة على كشف الأنماط الدورية والتراكيب المتكررة التي تكون غير واضحة في طرق محاذاة التسلسل التقليدية.

تسعى المنهجية إلى تجاوز القيود التي تواجهها خوارزميات المحاذاة التقليدية، مثل خوارزميات الاصطفاف المحلي والعام، في التعرف على العلاقات البنيوية بين هذه البروتينات.

تكمن أهمية هذا البحث في تقديم مقاربة جديدة لمعالجة إحدى أبرز التحديات في علم الأحياء الحاسوبي، وهي صعوبة الكشف عن التطابق البنيوي بين البروتينات في منطقة الشفق والتي تكون فيها نسبة التطابق التسلسلي منخفضة جداً.

تعتمد معظم الخوارزميات التقليدية المستخدمة في محاذاة البروتينات على مصفوفات تسجيل الدرجات مثل PAM وBLOSUM، والتي تفقد فعاليتها في هذه المنطقة.

المنهجية المقترحة، باستخدام تحويل فورييه ومعامل الارتباط التبادلي، توفر أداة بديلة قادرة على التقاط الأنماط الدورية والمكونات الترددية التي تعكس التشابه البنيوي، مما يساهم في تحسين دقة التنبؤات البنيوية والوظيفية للبروتينات.

٣- طرائق البحث وموارده

تم الاعتماد على الطرائق الآتية لإجراء البحث:

٣-١ منطقة الشفق Twilight Zone

تُعد منطقة الشفق من أكثر المفاهيم تعقيداً في مجال المعلوماتية الحيوية، وهي تشير إلى المرحلة التي تصبح فيها نسبة التطابق بين البروتينات منخفضة جداً، بحيث يصعب على خوارزميات المحاذاة التقليدية اكتشاف العلاقات البنيوية أو الوظيفية بينها.

غالباً ما تكون نسبة التطابق في هذه المنطقة منخفضة (~٢٠-٣٥٪)، مما يجعل من الصعب اكتشاف التطابقات الحقيقية باستخدام طرق المحاذاة التقليدية مثل BLAST أوNeedleman-Wunsch [3].

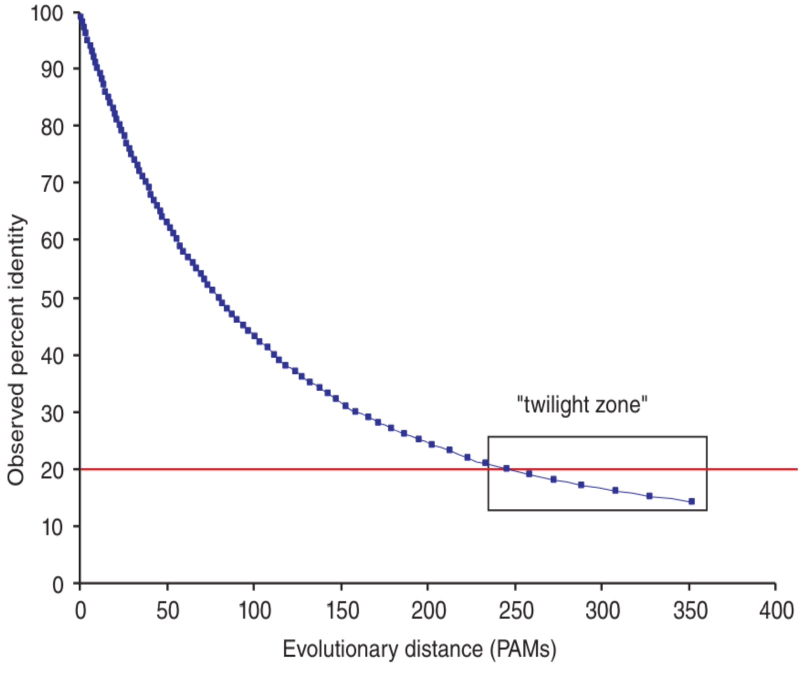
على الرغم من ضعف التطابق التسلسلي، قد تحتفظ البروتينات بالوظائف الأساسية أو البنية الثلاثية الأبعاد التي تسمح لها بأداء نفس الدور الحيوي.

المنحنى البياني المبين في الشكل (١) يُظهر علاقة عكسية بين المسافة التطورية ونسبة التطابق؛ فكلما زادت المسافة التطورية (عدد الطفرات)، انخفضت نسبة التطابق بين التسلسلات.

المحور الأفقي في المخطط يمثل المسافة التطورية بين البروتينات، ويُقاس بواحدة تعرف باسم "نقاط الطفرات المقبولة" (PAM)، وهي مقياس لعدد الطفرات التي حدثت خلال عملية التطور.

المحور العمودي يمثل نسبة التطابق المرصودة بين تسلسلين بروتينيين، أي النسبة المئوية للأحماض الأمينية المتطابقة عند إجراء محاذاة مباشرة.

الخط الأحمر الأفقي يحدد عتبة 20%، وهي النقطة التي تبدأ عندها منطقة الشفق.



الشكل (١): منطقة الشفق. مأخوذ من .Pevsner (2015, p. 95)

٣-٢ البروتينات المتباعدة Divergent Proteins

البروتينات المتباعدة هي البروتينات التي تنشأ من سلف مشترك ولكنها تطوّرت وتباعدت لتؤدي وظائف مختلفة في كائنات مختلفة أو داخل نفس الكائن الحي، نتيجة لتغيرات في التعبير الجيني، رغم احتفاظها أحياناً ببعض الخصائص البنيوية أو الوظيفية المشتركة[12] .

تم في هذا البحث دراسة زوجين من البروتينات المتباعدة تطورياً ينتميان الى منطقة الشفق وهما:

١. Myoglobin isoform 1 [Homo sapiens] و Hemoglobin subunit beta [Homo sapiens]
٢. Lysozyme C precursor(Human Lysozyme) و Lysozyme T4 (from Bacteriophage)

٣-٣ توصيف البروتينات المدروسة

تعرض هذه الفقرة تسلسلات الأحماض الأمينية الخاصة بمجموعة من البروتينات التي تم اختيارها للتحليل ضمن هذا البحث، وذلك من خلال بيانات مستخرجة من قواعد بيانات بيولوجية موثوقة، تشمل كلًا من المركز الوطني لمعلومات التقانة الحيوية(NCBI) وقاعدة GenBank وPDB (protein data bank)، حيث تم عرض هذه التسلسلات بصيغة FASTA القياسية، وهي الصيغة المعتمدة في تحليل التسلسلات الجينية باستخدام أدوات المعلوماتية الحيوية.

تتضمن الأشكال المرفقة معلومات تعريفية لكل بروتين، مثل الاسم العلمي، ومعرف الدخول المرجعي (Accession Number)، بالإضافة إلى مصدر التسلسل .

تُستخدم هذه البيانات كمدخلات أولية لتحويل التسلسلات إلى إشارات تشابهية، تمهيداً لتحليلها في المجال الترددي باستخدام تحويل فورييه.

تم اختيار البروتينات بحيث تكون ذات تطابق تسلسلي منخفض، مما يسمح باختبار فعالية المنهجية في منطقة الشفق.

٣-٣-١ توصيف بروتينات الزوج الأول

يُعد بروتين الميوغلوبين إيزوفورم ١ والوحدة الفرعية بيتا من الهيموغلوبين (Hb β) بروتينين متماثلين Homologous، وكلاهما عضو في عائلة الغلوبين ويشتركان في أصل تطوري مشترك، ويتضح هذا من تشابه بنيتهما ووظيفتهما المتعلقة بربط الأوكسجين ونقله.

على الرغم من طبيعتهما المتماثلة ووظيفتهما المتشابهة المتعلقة بربط الأكسجين، إلا أن الميوغلوبين والهيموجلوبين لهما أدوار مختلفة حيث تتركز وظيفة الميوغلوبين بشكل أكبر في تلبية احتياجات الأنسجة العضلية، في حين يخدم الهيموغلوبين احتياجات الجسم التنفسية بأكمله.

يوضح الشكل (٢) ملف FASTA الخاص ببروتين myoglobin isoform 1[Homo sapiens].

myoglobin isoform 1 [Homo sapiens]

NCBI Reference Sequence: NP_005359.1

GenPept Identical Proteins Graphics

```
>NP_005359.1 myoglobin isoform 1 [Homo sapiens]
MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVL
TALGGILKKKGHHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFR
KDMASNYKELGFQG
```

الشكل(2): ملف FASTA لبروتين ميوغلوبين إيزوفورم ١.

يوضح الشكل (٣) ملف FASTA الخاص ببروتين بيتا غلوبين البشري.

beta-globin [Homo sapiens]

GenBank: AAA16334.1

GenPept Identical Proteins Graphics

```
>AAA16334.1 beta-globin [Homo sapiens]
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKVKAHGKKVLG
AFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVAN
ALAHKYH
```

الشكل(3): ملف FASTA لبروتين بيتا غلوبين البشري.

٣-٣-٢ توصيف بروتينات الزوج الثاني

يمثل بروتين Lysozyme C البشري أحد الإنزيمات الأساسية في المناعة الفطرية، حيث يُفرز في سوائل الجسم مثل الدموع واللعاب ويقوم بتحليل الببتيدوغليكان(Peptidoglycan) في جدران البكتيريا، أما Lysozyme T4 chain Aفهو إنزيم من العاثية T4 (فيروس يصيب البكتيريا)، ويُستخدم على نطاق واسع كنموذج بنيوي في علم الأحياء الجزيئي، ويؤدي وظيفة مشابهة في مهاجمة البكتيريا.

رغم اختلاف المصدر التطوري بين البروتينين (بشري وفيروسي)، فإن التشابه الوظيفي بينهما يعكس حفاظاً تطورياً على آلية التحلل البكتيري ومع ذلك فإن نسبة التشابه التسلسلي بينهما منخفضة، مما يضعهما ضمن ما منطقة الشفق.

يوضح الشكل (٤) ملف FASTA الخاص ببروتين Lysozyme C precursor.

lysozyme C precursor [Homo sapiens]

NCBI Reference Sequence: NP_000230.1

GenPept Identical Proteins Graphics

```
>NP_000230.1 lysozyme C precursor [Homo sapiens]
MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLANWMCLAKWESGYNTRATNYNAGDRST
DYGIFQINSRYWCNDGKTPGAVNACHLSCSALLQDNIADAVACAKRVVRDPQGIRAWVAWRNRCQNRDVR
QYVQGCGV
```

الشكل(٤): ملف FASTA لبروتين Lysozyme C precursor.

يوضح الشكل (٥) ملف FASTA الخاص ببروتين Lysozyme T.

Chain A, T4 LYSOZYME

PDB: 6LZM_A

GenPept Identical Proteins Graphics

```
>pdb|6LZM|A Chain A, T4 LYSOZYME
MNIFEMLRIDEGLRLKIYKDTEGYYTIGIGHLLTKSPSLNAAKSELDKAIGRNCNGVITKDEAEKLFNQD
VDAAVRGILRNAKLKPVYDSLDAVRRCALINMVFQMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYN
QTPNRAKRVITTFRTGTWDAYKNL
```

الشكل(5): ملف FASTA لبروتين Lysozyme T.

٣-٤ تحليل المحاذاة التسلسلية بين البروتينات

تهدف هذه الخطوة إلى قياس درجة التطابق التسلسلي بين البروتينين، وتحديد ما إذا كانا يقعان ضمن منطقة الشفق، والتي يصعب فيها الكشف عن العلاقة البنيوية باستخدام أدوات المحاذاة التقليدية.

تم إجراء محاذاة تسلسلية عامة Global Alignment بين بروتين myoglobin isoform1 و beta-globin، باستخدام خوارزمية Needleman-Wunsch عبر منصة EMBOSS Needle مع اعتماد مصفوفة تسجيل الدرجات PAM250.

٣-٤-١ تحليل المحاذاة التسلسلية للزوج الأول من البروتينات

أظهرت نتائج المحاذاة نسبة تطابق بلغت %٢٣.٢، مع ٣٦ موضعاً متطابقاً من أصل ١٥٥، ونسبة فجوات بلغت ٥.٨٪. هذه القيم تُشير بوضوح إلى أن العلاقة التسلسلية بين البروتينين ضعيفة، مما يُعزز الحاجة إلى منهجيات تحليلية بديلة، مثل التحليل الترددي، للكشف عن التشابه البنيوي غير الظاهر تسلسلياً، كما هو مقترح في هذا البحث.

يوضح الشكل (٦) نتيجة المحاذاة الثنائية العامة للبروتينين بيتا غلوبين وميوغلوبين إيزوفورم ١.

```
# Aligned_sequences: 2
# 1: 1-154
# 2: 1-147
# Matrix: EPAM250
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 155
# Identity:      36/155 (23.2%)
# Similarity:    74/155 (47.7%)
# Gaps:           9/155 ( 5.8%)
# Score: 172.5
#
#
#=======================================

1-154              1 -MGLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKH     49
                      :.|:..|...|..:||||  ::...|.|.|.||:...|.|...|:.|.:
1-147              1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

1-154             50 LKSEDEMKASEDLKKHGATVLTALGGILKKKGHHEAEIKPLAQSHATKHK     99
                     |.:.|.:.::..:|.||..||.|:::.|...::..:.:..|::.|..|..
1-147             49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH     98

1-154            100 IPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKE    149
                     :....:.::::.::.||......:|.:..|:|.:|.:....:.:|.:|.
1-147             99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH-    147

1-154            150 LGFQG    154

1-147            148 -----    147
```

الشكل(٦) : نتيجة المحاذاة الثنائية العامة للبروتينين بيتا غلوبين وميوغلوبين إيزوفورم ١.

٣-٤-٢ تحليل المحاذاة التسلسلية للزوج الثاني من البروتينات

أظهرت نتائج المحاذاة نسبة تطابق بلغت ١٠.٥٪، مع 22 موضعًا متطابقًا من أصل ٢١٠، ونسبة فجوات بلغت ٥١.٤٪. هذه القيم تُشير أيضا إلى أن العلاقة التسلسلية بين البروتينين ضعيفة.

يوضح الشكل(7) نتيجة المحاذاة الثنائية العامة للبروتينين Lysozyme C precursor وLysozyme T .

```
# Matrix: EPAM250
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 210
# Identity:      22/210 (10.5%)
# Similarity:    64/210 (30.5%)
# Gaps:         108/210 (51.4%)
# Score: 52.0
#
#
#=======================================

1-148              1 MKALIVLGLVLLSVTVQGKVFERCELARTLKRLGMDGYRGISLANWMCLA     50
                        :.::.::.:...::.|::.           :.:||..|::::.:
1-164              1 ---MNIFEMLRIDEGLRLKIYK-----------DTEGYYTIGIGHLL---     33

1-148             51 KWESGYNTRATNYNAGDRSTDYGIFQINSRYWCNDGKTPGAVNACHLSCS    100
                            |::.:.||:....|.:|     :|. ||:..|.:..:
1-164             34 -------TKSPSLNAAKSELDKAI----GRN-CNGVITKDEAE-------     64

1-148            101 ALLQDNIADAVACAKRVVRDP---------QGIRAWVAWRNRCQNRDVRQ    141
                     .|::::::..||   :.::|::          :::|       ||...::..
1-164             65 KLFNQDVDAAV---RGILRNAKLKPVYDSLDAVR-------RCALINMVF    104

1-148            142 YVQGCGV-------------------------------------------    148
                     .:...||
1-164            105 QMGETGVAGFTNSLRMLQQKRWDEAAVNLAKSRWYNQTPNRAKRVITTFR    154

1-148            149 ----------    148

1-164            155 TGTWDAYKNL    164
```

الشكل(7): نتيجة المحاذاة الثنائية العامة للبروتينين **Lysozyme C precursor** و**Lysozyme T** .

٣-٥ تحويل البروتينات إلى إشارات تشابهية

بعد أن تم استيراد البروتينات المراد دراستها من قاعدة البيانات وباستخدام لغة البرمجة python تم الحفاظ على تسلسل الأحماض الأمينية المشكلة للبروتين وجعلها تمثل المحور الأفقي، أما المحور العمودي فهو يمثل الترددات المعيارية لورود الأحماض الأمينية في البروتين أي عدد مرات ورود كل حمض أميني داخل البروتين.

تم بتحويل المحور الأفقي إلى محور زمني بعد تحديد عدد الأحماض الأمينية التي يمكن للخلية تخليقها في الثانية الواحدة [13].

٣-٦ تحويل فورييه المتقطع

علاقة فورييه تتيح تحويل الإشارات من المجال الزمني إلى المجال الترددي، مما يساعد على تحليل مكوناتها الترددية، وتُستخدم لفهم توزيع الطاقة والترددات داخل الإشارة [14].

الصيغة الرياضية لتحويل فورييه المتقطع:

$$X_k = \sum_{n=0}^{N-1} X_n e^{-\frac{i2\pi kn}{N}} \quad (١)$$

$$f_k = \frac{k}{N*T} \quad (٢)$$

$$T = \frac{1}{f_s} \quad (٣)$$

- X_k: مطال الإشارة في المجال الترددي.
- X_n : مطال الإشارة في المجال الزمني.
- f_k: تردد العينة k.
- k: رقم العينة في المجال الترددي.
- n: رقم العينة في المجال الزمني.
- N: عدد العينات.
- T: دور التقطيع.
- f_s: تردد لتقطيع.

تقوم المعادلة (١) بتحويل الإشارة من المجال الزمني إلى المجال الترددي أما المعادلة رقم (٢) فتقوم بربط رقم المركّبة الترددية k والتردد الحقيقي الذي تمثله.

المعادلة (٣) تربط بين الزمن والتردد، وهي أساسية لفهم كيف تؤثر سرعة أخذ العينات على دقة التحليل الترددي.

٣-٧ معامل الارتباط التبادلي

معامل الارتباط التبادلي هوي مقياس احصائي يستخدم لتحديد قوة واتجاه العلاقة الخطية بين متغيرين أو إشارتين x و y.

تبين العلاقة (٤) معامل الارتباط التبادلي حيث يعبر البسط عن التباين covariance بين المتغيرين x و y، أما المقام فهو حاصل ضرب الانحراف المعياري لكل من x و y.

تتراوح قيمة معامل الارتباط التبادلي (بيرسون) يأخذ قيماً تتراوح بين -١ و +١**[15]** .

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \quad (4)$$

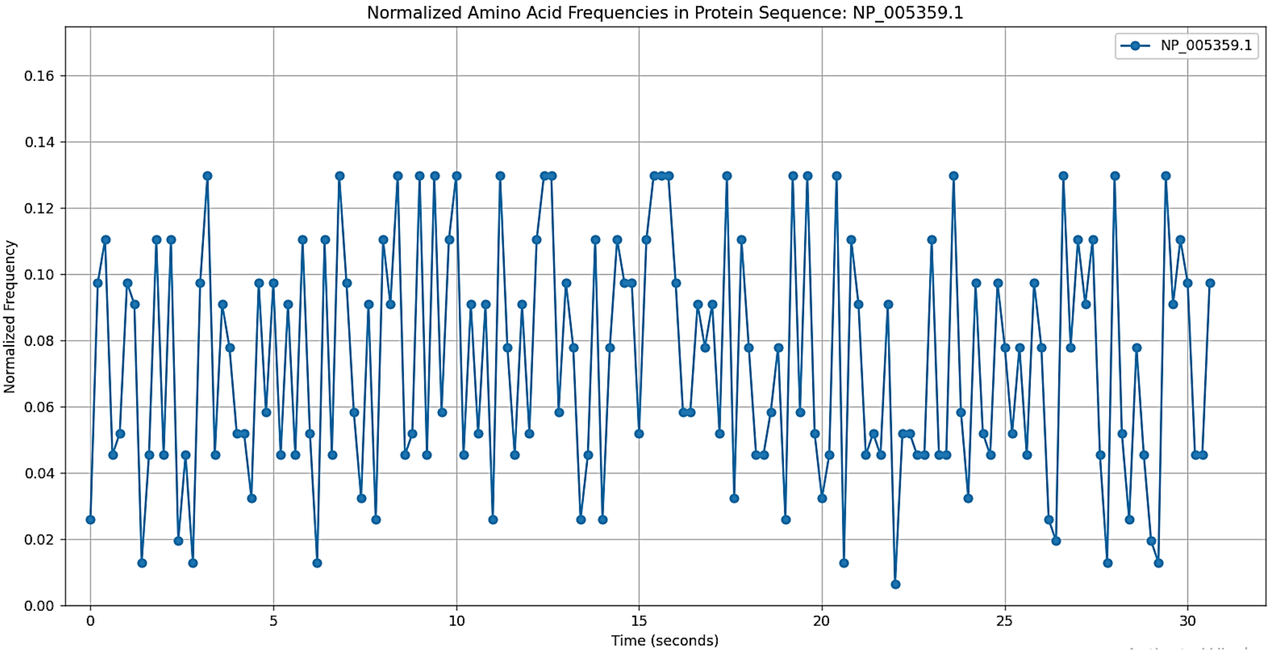
إذا كانت القيمة +١ فهذا يعني وجود علاقة خطية موجبة مثالية أي كلما زادت قيمة أحد المتغيرين، زادت الأخرى بنفس النمط أما إذا كانت القيمة -١، فالعلاقة خطية سالبة مثالية أي كلما زادت قيمة أحد المتغيرين، قلت الأخرى بنفس النمط. إذا كانت القيمة قريبة من الصفر، فهذا يشير إلى عدم وجود علاقة خطية واضحة بين المتغيرين.

٤- النتائج والمناقشة

٤-١ تحويل البروتينات إلى إشارات تشابهية

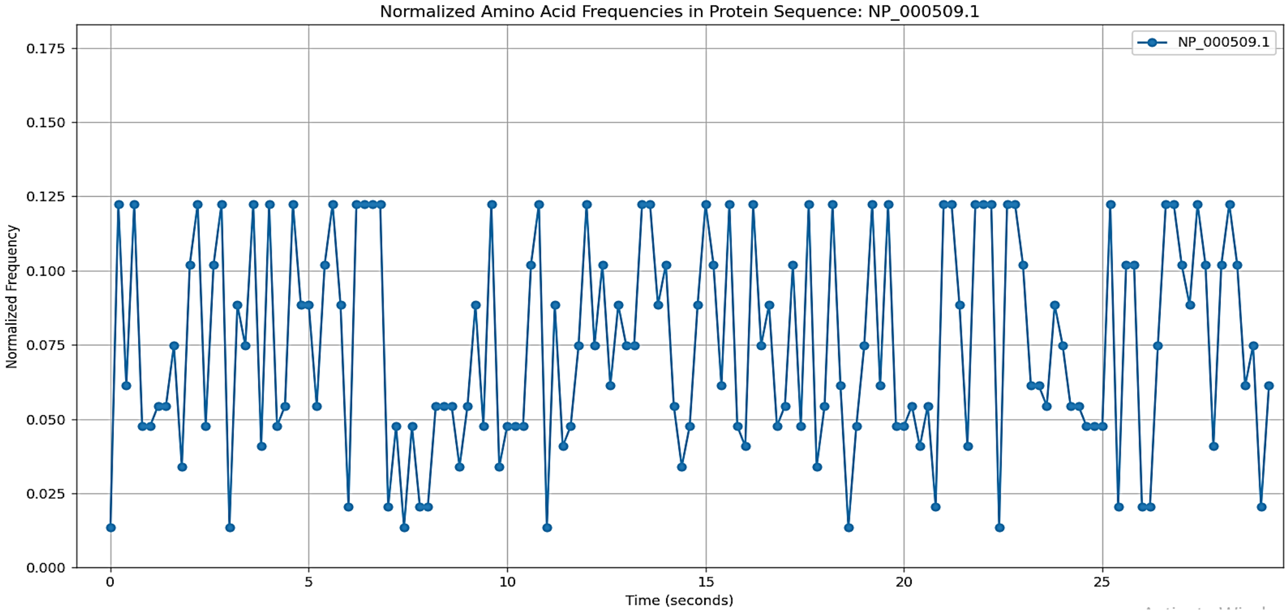
٤-١-١ تحويل بروتينات الزوج الأول إلى إشارات تشابهية

يوضح الشكل (٨) تحويل البروتين الأول ميوغلوبين إيزوفورم ١ إلى إشارة تشابهية.



الشكل(8):الإشارة التشابهية الخاصة ببروتين ميوغلوبين إيزوفورم ١

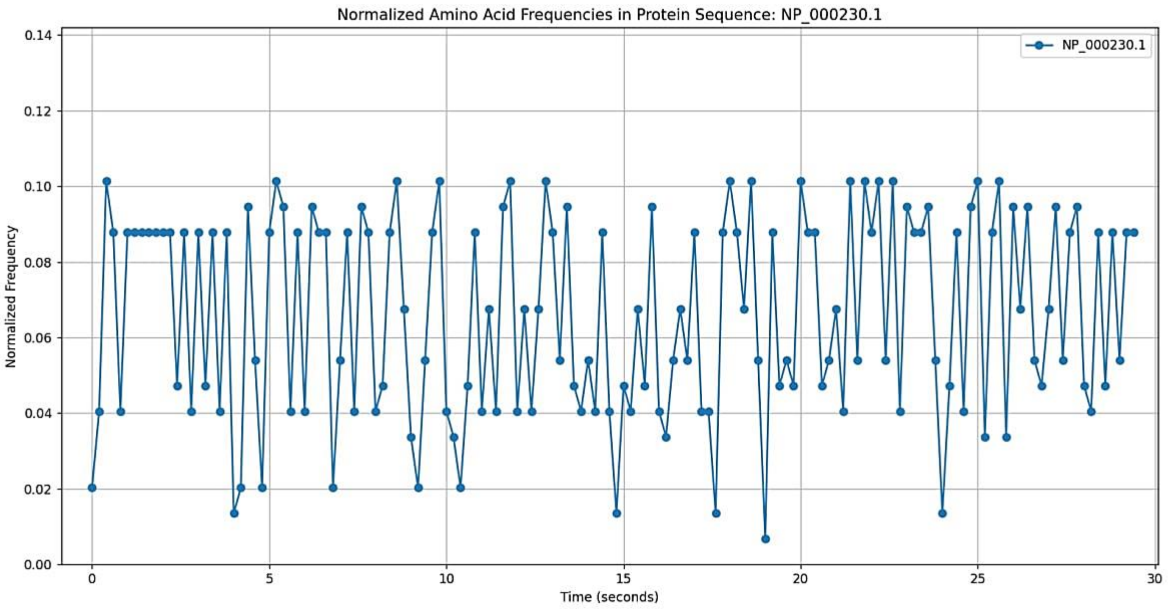
يوضح الشكل (٩) تحويل البروتين الثاني بيتا غلوبين إلى إشارة تشابهية.



الشكل(٩): الإشارة التشابهية الخاصة ببروتين بيتا غلوبين.

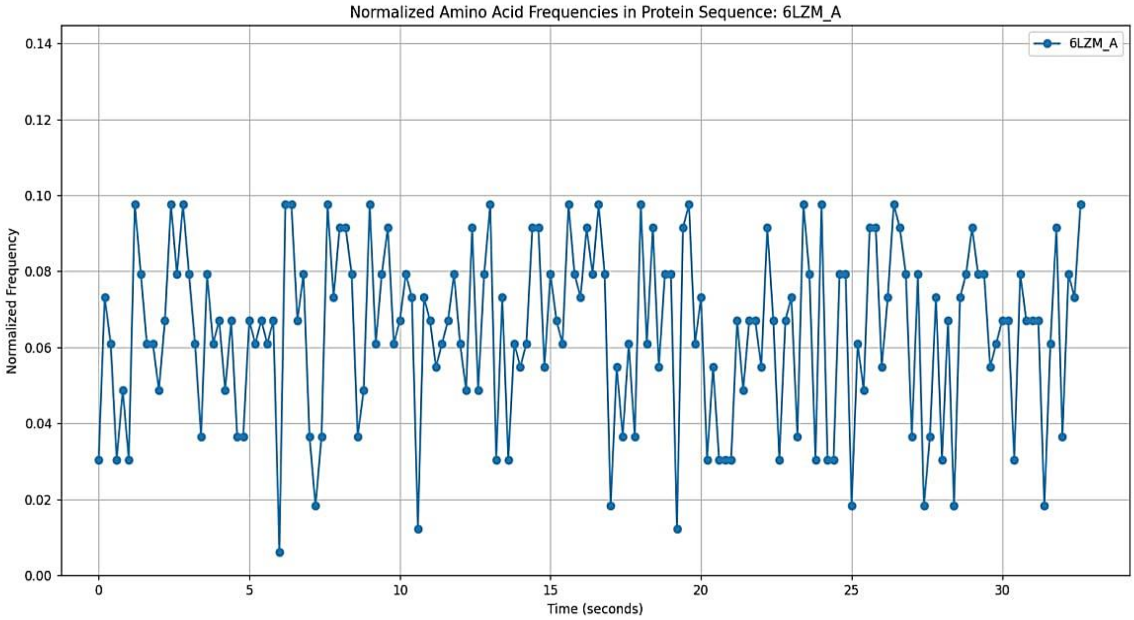
٤-١-٢ تحويل بروتينات الزوج الثاني إلى إشارات تشابهية

يوضح الشكل (١٠) تحويل البروتين الأول Lysozyme C precursor(Human Lysozyme) إلى إشارة تشابهية.



الشكل(10): الإشارة التشابهية الخاصة ببروتين (Lysozyme C precursor(Human Lysozyme

يوضح الشكل (١١) تحويل البروتين الثاني Lysozyme T إلى إشارة تشابهية.

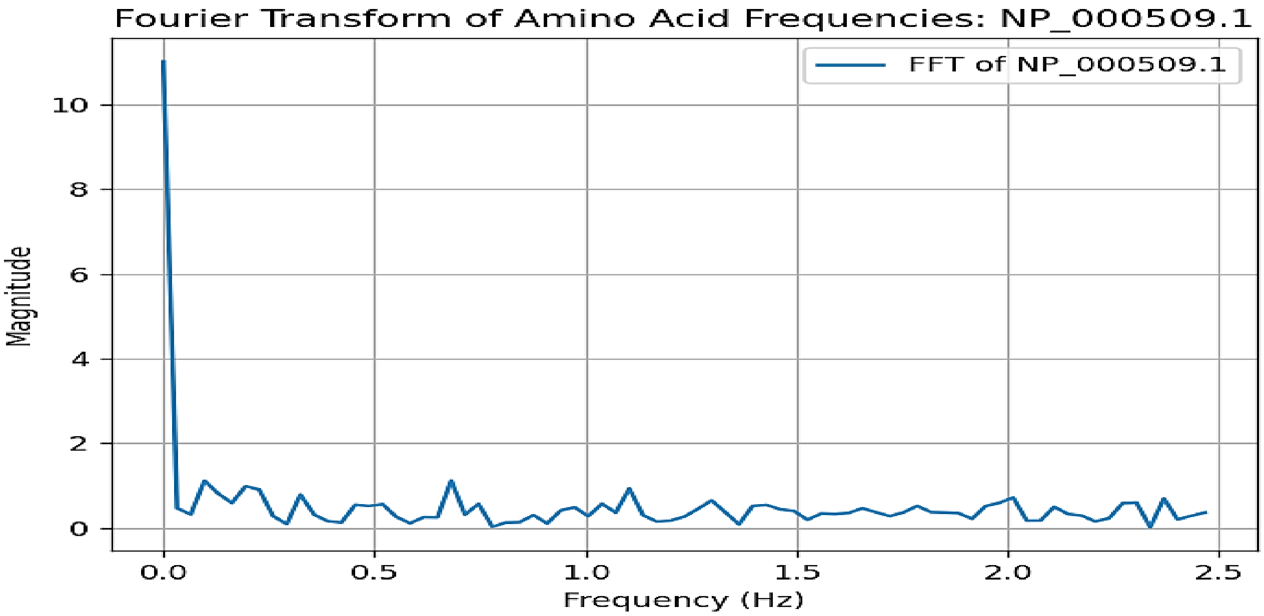


الشكل(11): الإشارة التشابهية الخاصة ببروتين Lysozyme T.

٤-٢ تحويل فورييه

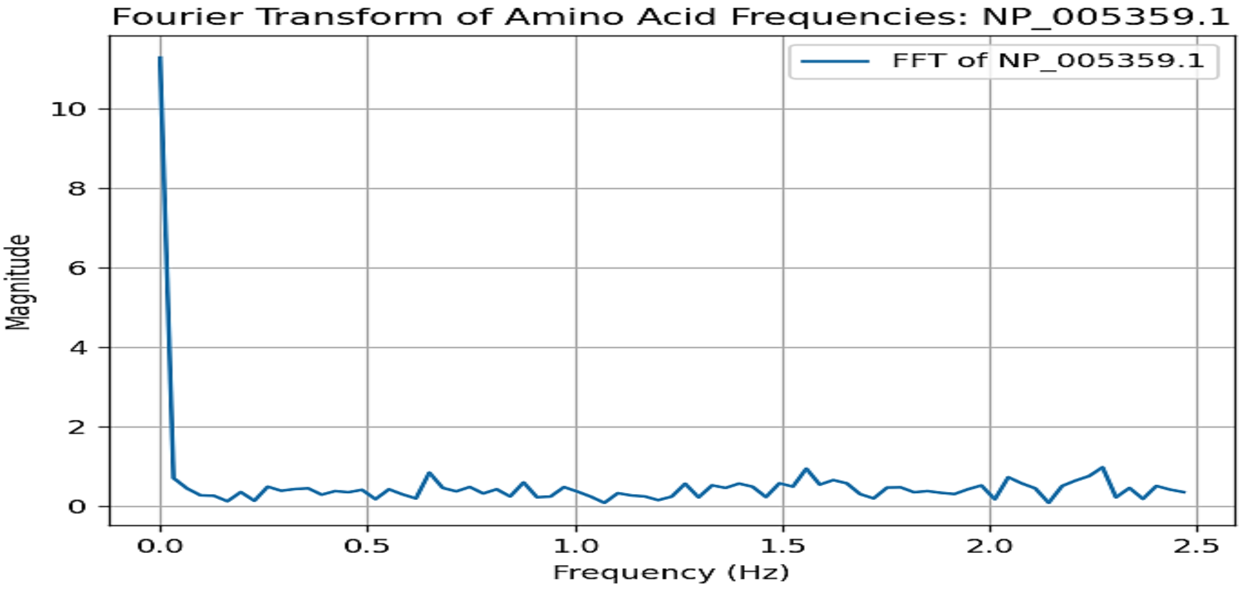
٤-٢-١ تحويل فورييه وحساب معامل الارتباط التبادلي للزوج الأول من البروتينات

تم في هذه الفقرة تطبيق تحويل فورييه باستخدام العلاقة (١) على الإشارات التشابهية التي حصلنا عليها، ويوضح الشكل (١٢) النتيجة التي حصلنا عليها بالنسبة للبروتين الأول Hemoglobin subunit beta.



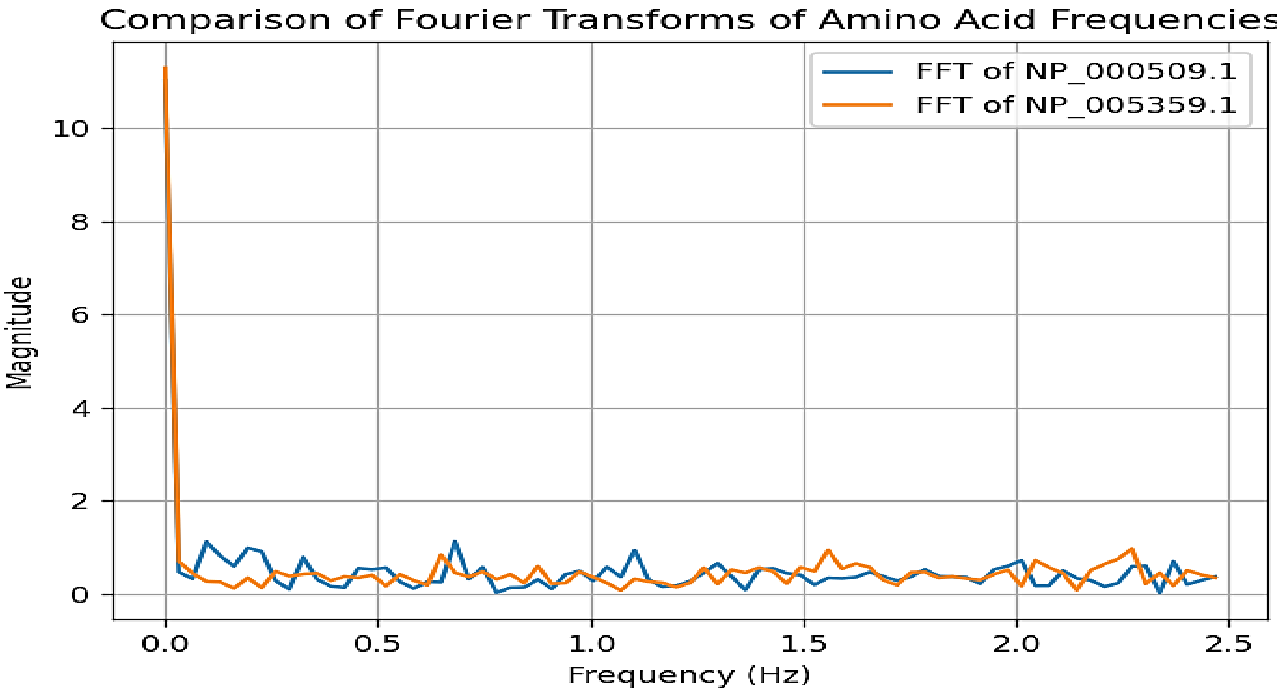
الشكل(١٢): تحويل فورييه للبروتين Hemoglobin subunit beta.

يوضح الشكل (١٣) تحويل فورييه للبروتين الثاني Myoglobin isoform 1.



الشكل(١٣): تحويل فورييه للبروتين Myoglobin isoform 1 [Homo sapiens].

يوضح الشكل (١٤) مقارنة تحويل فورييه للبروتينين Myoglobin isoform 1 و Hemoglobin subunit beta.



الشكل(14):مقارنة تحويل فورييه للبروتينين.

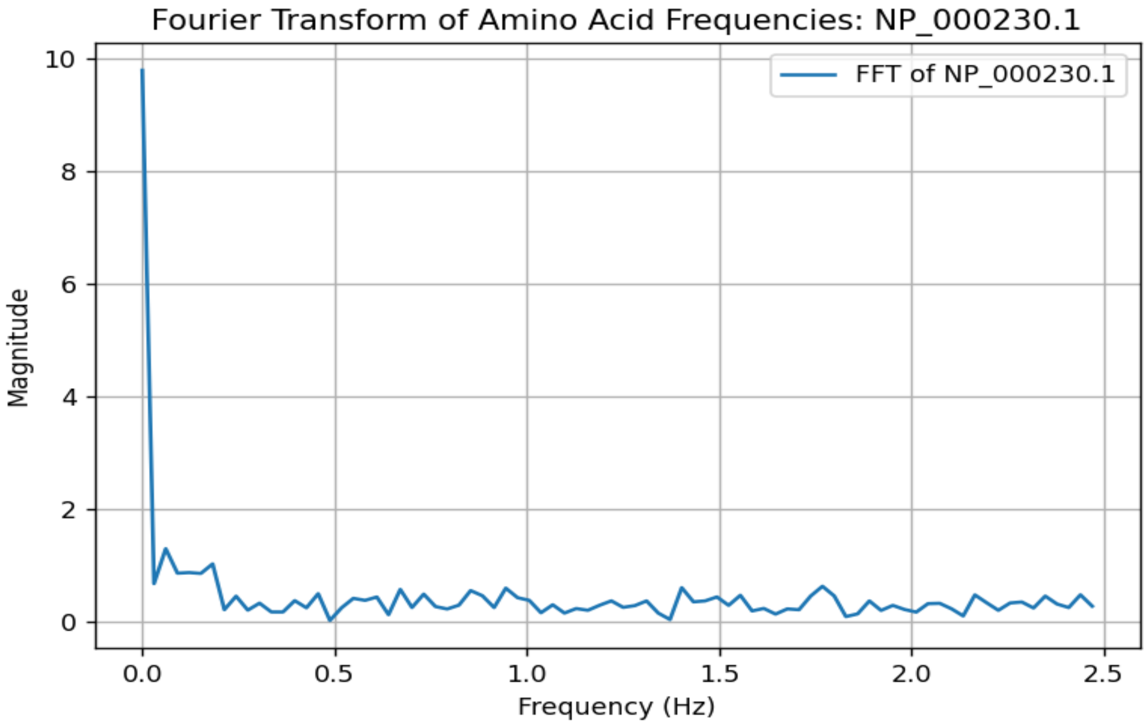
أظهرت الإشارات الناتجة عن تحويل فورييه لتتابعي الهيموغلوبين بيتا والميوغلوبين إيزوفورم١ نمطًا ترددياً متقارباً، وهذا التشابه الطيفي يُعزز بفعل معامل الارتباط التبادلي المرتفع (93%)، الذي تم حسابه بالعلاقة (٤) والذي يُشير إلى تطابق قوي بين الإشارتين في المجال الترددي، والتغيرات في الترددات بين البروتينين تحدث بشكل متزامن تقريباً، رغم أن المحاذاة التسلسلية تضع البروتينين ضمن منطقة الشفق.

على الرغم من أن نسبة التطابق التسلسلي بين البروتينين وفقًا لخوارزمية Needleman-Wunsch ومصفوفة PAM250 كانت منخفضة (23.2%)، فإن معامل الارتباط التبادلي المرتفع يُثبت أن هناك علاقة بنيوية مهمة بين البروتينين لا يمكن كشفها عبر المقارنة التسلسلية وحدها.

تشير هذه النتائج إلى أن التحليل الترددي للإشارات الناتجة يُمكن أن يكشف عن علاقات بنيوية غير واضحة عند محاذاة تسلسلات الأحماض الأمينية بشكل تقليدي، ويُعد مكملًا فعالًا لأدوات المحاذاة التقليدية، خاصة في حالات التطابق المنخفض كما أن معامل الارتباط التبادلي يُمكن اعتماده كمؤشر رياضي بديل لتقييم التشابه البنيوي بين البروتينات.

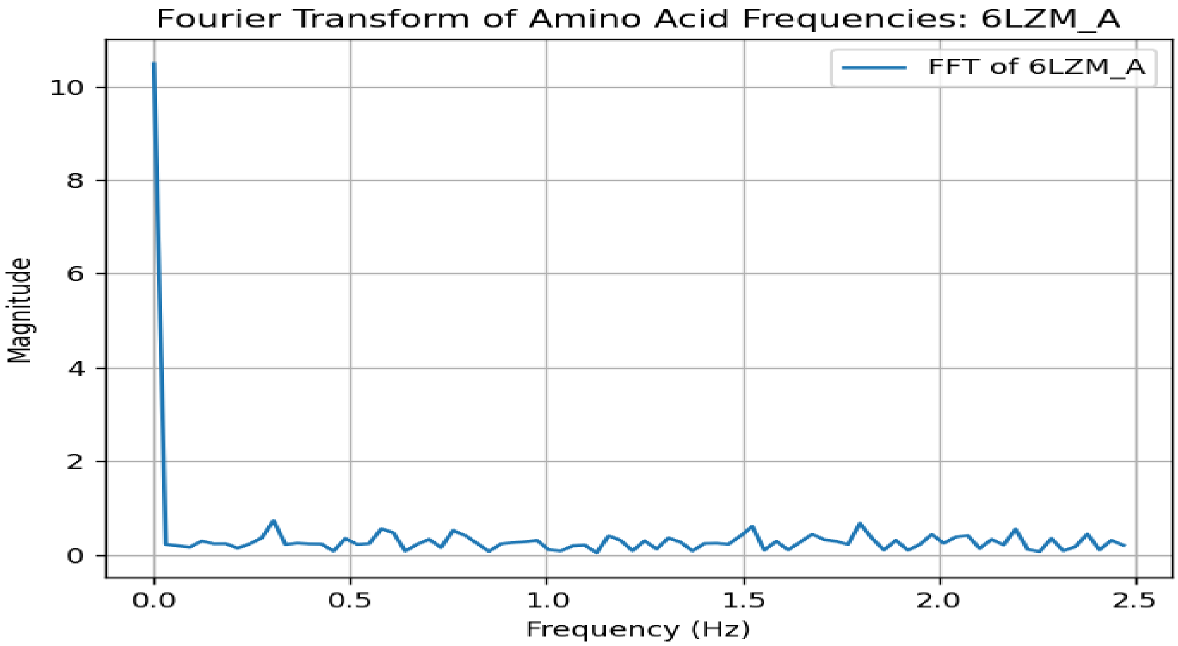
4-٢-2 تحويل فورييه وحساب معامل الارتباط التبادلي للزوج الثاني من البروتينات

تم في هذه الفقرة تطبيق تحويل فورييه باستخدام العلاقة (١) على الإشارات التشابهية التي حصلنا عليها ويوضح الشكل (١٥) النتيجة التي حصلنا عليها بالنسبة لبروتين Lysozyme C precursor(Human Lysozyme).



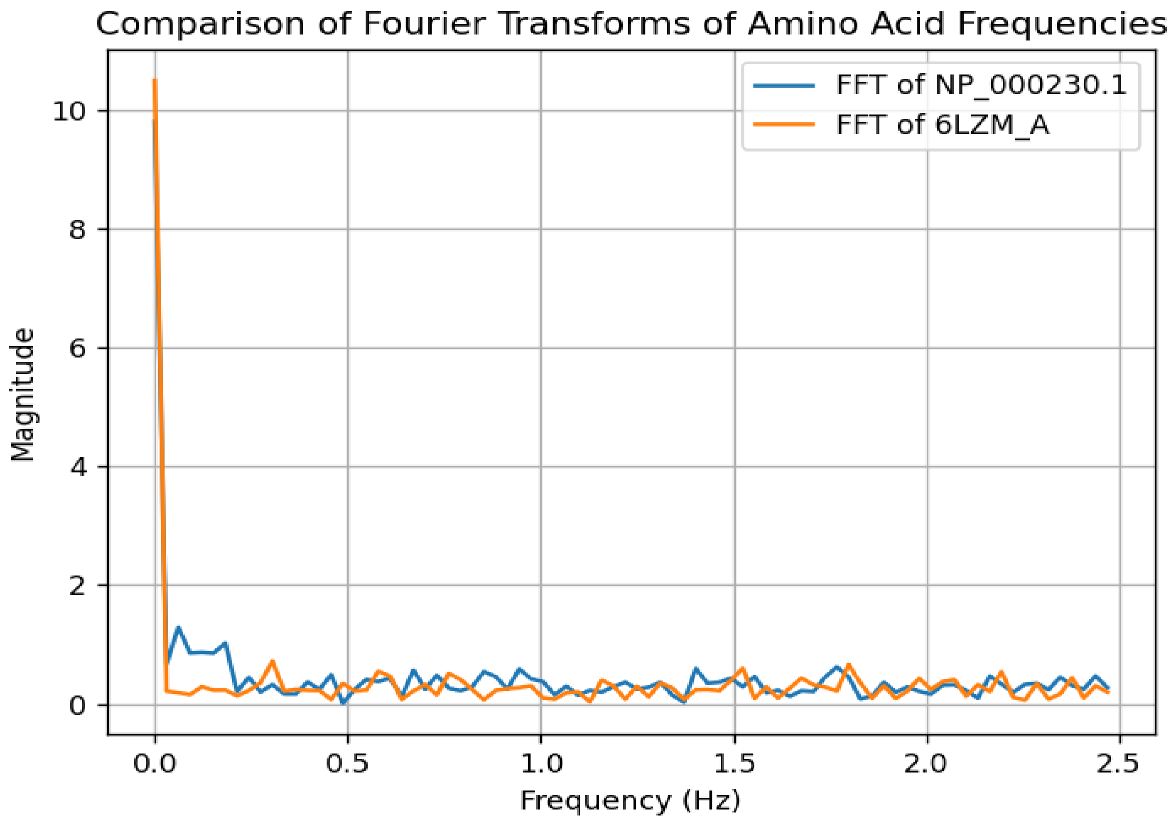
الشكل(١٥): تحويل فورييه السريع للبروتين Lysozyme C precursor(Human Lysozyme).

يوضح الشكل (١٦) تحويل فورييه للبروتين الثاني Lysozyme T



الشكل(١٦): تحويل فورييه السريع للبروتين Lysozyme T.

يوضح الشكل (١٧) مقارنة تحويل فورييه للبروتينين Lysozyme T و Lysozyme C precursor.



الشكل(17):مقارنة تحويل فورييه للبروتينين قيد الدراسة.

أظهر الزوج الثاني نمطًا ترددياً متقارباً للغاية انعكس في معامل ارتباط تبادلي مرتفع بلغ 94.3% عند تطبيق المعادلة (٤)، مما يدل على أن الإشارتين تتشاركان في توزيع دوري متشابه، ورغم أن الهوية التسلسلية تضعهما ضمن منطقة الشفق، فإن هذا التشابه في المركبات الترددية يُشير إلى وجود ضغط انتقائي حافظ على البنية الوظيفية المرتبطة بآلية التحليل البكتيري .

رغم المؤشرات الأولية المشجعة المستخلصة من تحليل المركبات الترددية ومعاملات الارتباط التبادلي، واجه هذا البحث تحدياً منهجياً يتمثل في غياب دراسات سابقة تعتمد تحويل فورييه لتحليل التطابق البروتيني في منطقة الشفق.

معظم الأدبيات العلمية تركز على المحاذاة الخطية أو النمذجة البنيوية [10,11]، مما يجعل المقارنة المباشرة غير ممكنة، ومع ذلك فإن التشابه الطيفي المتكرر الذي تم رصده في زوجين مختلفين يُشير إلى أن المركبات الترددية تُعبّر عن نمط وظيفي محفوظ، يُمكن أن يشكل أساساً لإطار معياري جديد في تحليل البروتينات المتباعدة.

٥-الاستنتاجات والتوصيات

تشير النتائج الواعدة التي تم الحصول عليها من تحويل فورييه وتحليل الارتباط التبادلي الى أن هذه الطريقة لديها القدرة على اكتشاف البروتينات المتباعدة التي تمتلك سلف تطوري مشترك.

ومع ذلك، ولإثبات متانتها وموثوقيتها، من الضروري إجراء دراسة إحصائية عامة مع مجموعة بيانات كبيرة.

وبعد الدراسة والتحليل نوصي بما يلي:

١. توسيع مجموعة البيانات:

•توسيع تطبيق التحليل الطيفي على أزواج بروتينية إضافية ضمن منطقة الشفق لاختبار موثوقية معامل الارتباط التبادلي كمؤشر للتشابه النمطي.

•تضمين بروتينات من كائنات مختلفة (بكتيريا، نباتات، فقاريات) لاختبار مدى عمومية النموذج.

٢. دمج التحليل الطيفي مع أدوات الذكاء الاصطناعي:

•تدريب نماذج تعلم آلي على الإشارات الترددية لتصنيف البروتينات أو التنبؤ بوظيفتها.

• استخدام الشبكات العصبية لتحديد أنماط ترددية مميزة لعائلات بروتينية معينة.

٣. الربط مع البيانات البنيوية ثلاثية الأبعاد:

•مقارنة نتائج تحويل فورييه مع بيانات البنية الفراغية (مثل PDB) لتحديد ما إذا كانت المركبات الترددية ترتبط فعلياً بطيّات بنيوية محفوظة.

• استخدام أدوات مثل DSSP أو STRIDE لتحديد العناصر الثانوية وربطها بالمركبات الترددية.

من خلال تنفيذ هذه التوصيات ودمج هذه الاقتراحات، يمكن تعميم هذه الطريقة وتحسينها وجعلها أكثر موثوقية ودقة وكفاءة، مما يعزز في نهاية المطاف فائدتها في دراسة تسلسلات البروتينات وتقديم رؤى ذات مغزى بيولوجي، مما يمهد الطريق لتطبيقها على نطاق أوسع في علم الأحياء الحسابي.

٦- المراجع:

[1] Alzubi, O. A., Alzubi, J. A., Alweshah, M.,Qiqieh, I., Al-Shami, S., & Ramachandran, M. (2020*). An optimal pruning algorithm of classifier ensembles: dynamic programming approach*. Neural Computing and Applications, 32, 16091-16107.

[2] Chao, J., Tang, F., & Xu, L. (2022). *Developments in algorithms for sequence alignment: A review. Biomolecule*s, 12(4), 546.

[**3**] szcześniak, M. W., Kubiak, M. R., Wanowska, E., & Makałowska, I. (2021). *Comparative genomics in the search for conserved long noncoding RNA*s. Essays in Biochemistry, 65(4), 741-749.

[**4**] Rost, B. (1999). *Twilight zone of protein sequence alignments*. Protein engineering, 12(2), 85-94.(twilight).

[**5**] Needleman, S. B., & Wunsch, C. D. (1970*). A general method applicable to the search for similarities in the amino acid sequence of two proteins*. Journal of Molecular Biology, 48(3), 443-453.

[**6**] Pevsner, J. (2015*). Bioinformatics and Functional Genomics (3rd ed.).* John Wiley & Sons Inc.

[**7**] Smith, T. F., & Waterman, M. S. (1981*). Identification of common molecular subsequences.* Journal of Molecular Biology, 147(1), 195-197.

[**8**] Dayhoff, M. O., Schwartz, R. M., & Orcutt, B. C. (1978). *A model of evolutionary change in proteins. In Atlas of Protein Sequence and Structure (Vol. 5,* Suppl. 3, pp. 345-352). National Biomedical Research Foundation.

[**9**] Rost, B. (1999). *Twilight zone of protein sequence alignments*. Protein Engineering, Design and Selection, *12(2), 85–94.*

[10] Iovino, B. G., & Ye, Y. (2024*). Protein embedding based alignment. BMC Bioinformatics, 25*, Article 85.

[11] Lin, Z., Akin, H., Rao, R., et al. (2023). *Language models of protein sequences at the scale of evolution enable accurate structure prediction*. Nature, 614*(7948), 343–350.*

[12] Laursen, L., Čalyševa, J., Gibson, T. J., & Jemth, P. (2021). *Divergent evolution of a protein–protein interaction revealed through ancestral sequence reconstruction and resurrection*. Molecular Biology and Evolution, 38(1), 152–167.

[13] Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2014*). Molecular biology of the cell* (6th ed.). Garland Science.

[14] Rajaby, E., & Sayedi, S. M. (2022). *A structured review of sparse fast Fourier transform algorithms. Digital Signal Processing*, 123, 103403.

[15] Schober, P., Boer, C., & Schwarte, L. A. (2018*). Correlation coefficients: appropriate use and interpretation. Anesthesia & analgesia*, 126(5), 1763-1768.