JBS
Research Article

Journal of Biological Studies

eISSN: 2209-2560

V. 9, N. 3: 180-194, 2026

https://doi.org/10.62400/jbs.v9i3.18555

Integrated phylogenetic and bioinformatic analysis of rbcl sequences in *Origanum* species

Gledjan Caka[1*], Rexhep Shkurti[1], Dhimitër Peçi[2], Aida Dervishi[1*]

[1]Department of Biotechnology, Faculty of Natural Sciences, University of Tirana, Tirana, Albania
[2] Research Center of Flora and Fauna, Faculty of Natural Sciences, University of Tirana, Tirana, Albania
*Correspondence author: gledjan.caka@fshn.edu.al; aida.dervishi@fshn.edu.al

Received 27 May 2026|Accepted 05 August 2026|Published 02 September 2026

Abstract

This study presents an integrated phylogenetic and bioinformatic analysis of the chloroplast *rbcL* gene in different species of the genus *Origanum*, *Lamiaceae* family. The aim of the study is to evaluate genetic diversity, evolutionary relationships, and structural characteristics of the corresponding protein. Homologous sequences retrieved from the NCBI database were aligned with MAFFT, trimmed and analyzed using the Maximum Likelihood method based on the Tamura–Nei model in MEGA 12 software. The phylogenetic tree reconstruction identified monophyletic *Origanum* group clearly separated from the outgroup, but provided limited resolution among the sampled species. Genetic distance analysis demonstrated low molecular divergence among species within the genus. A comprehensive bioinformatic analyses of the rbcL protein included physicochemical characterization, conserved motif identification, phosphorylation site prediction, secondary structure assessment, and three-dimensional modeling. Six conserved motifs associated with structural and functional roles were identified across most species. The analyzed proteins predominantly exhibited hydrophilic properties and high thermostability. Structural predictions and 3D models demonstrated a high degree of conformational conservation. Furthermore, phosphorylation analysis revealed species-specific differences in post-translational modification potential, particularly in serine-rich regions. The study findings provide valuable insights into the molecular evolution, phylogenetic relationships, and functional protein architecture of *Origanum* species, supporting future phylogenetic, pharmacological, and biotechnological investigations.

Keywords: rbcL, gene, phylogenetics, protein, bioinformatics

1. Introduction

The genus *Origanum*, belonging to the *Lamiaceae* family, comprises a diverse group of aromatic and medicinal plant species widely distributed throughout the Mediterranean region, as well as Western Asia (Chaachouay *et al.*, 2024; Nurzyńska-Wierdak and Walasek-Janusz, 2025). These species are considered of great importance because of their pharmacological and economic properties due to their extensive use in medicine, cosmetics and pharmaceutical industry. Several Origanum species have been recognized for their high content of essential oils and bioactive compounds which exhibit a variety of beneficial activities such as antioxidant, antimicrobial, anti-inflammatory and anticancer (de Torre *et al.*, 2020; Sarfaraz et al., 2023; Li Y *et al.*, 2025; Toso et al., 2025; Veljovic *et al.*, 2023; Nurzyńska-Wierdak and Walasek-Janusz, 2025; Solomou *et al.*, 2026). Despite the wide commercial and medicinal use, the evolutionary and taxonomic relationships and classification among the *Origanum* species are still complex because of their morphological similarities, hybridization and environmental influences (Lukas *et al.*, 2013; Buck and Flores-Renteria, 2022; Nazar *et al.*, 2022). Consequently, a molecular approach has become an essential tool for resolving phylogenetic relationships and understanding the genetic diversity within this genus.

One of the most widely utilized genes used in plant phylogenetics as a molecular marker is the chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (rbcL), due to its conserved nature and suitability for barcoding. Its high amplification efficiency and low mutability rate makes this gene an appealing choice for studies regarding phylogenetic diversity (Pere *et al.*, 2023). In addition to phylogenetic interference, advances in bioinformatics and computational biology allow a comprehensive analysis into protein functionality and structure.

Although rbcL is widely used as a chloroplast DNA barcode, relatively little is known about whether its sequence conservation is reflected at the protein structural level within the genus *Origanum*. Furthermore, comparative analysis integrating phylogenetic reconstruction with physicochemical characterization, conserved motif identification, post-translational modification prediction, and three-dimensional structural modeling have not been performed for this marker in *Origanum*. Therefore, this study aims to compare the evolutionary behavior of the marker at both the nucleotide and protein level using an integrated bioinformatics framework.

2. Materials and Methods

Sequence retrieval and phylogenetic analysis

Chloroplast rbcL and matK gene sequences representing *Origanum* species were retrieved from the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/) (O'Leary, 2024). Sequence alignments were performed using the Multiple Alignment using Fast Fourier Transform (MAFFT) algorithm available through the European Bioinformatics Institute (EMBL-EBI) web server (https://www.ebi.ac.uk/jdispatcher/msa/mafft) (Madeira *et al.*, 2024). The resulting multiple sequence alignments were manually inspected to ensure positional homology, and poorly aligned terminal regions containing missing data were removed to retain only the homologous overlapping coding region shared among all *Origanum* species included in this study.

Phylogenetic relationships were inferred using the Maximum Likelihood (ML) method implemented in MEGA version 12 (Kumar *et al.*, 2024). Evolutionary distances were estimated using the Tamura-Nei substitution model, with branch support assessed by 1000 bootstrap replicates. Sites were treated using partial deletion with 95% site coverage cutoff. Pairwise distances among taxa were also calculated in MEGA 12 using the same substitution model.

Protein sequence analysis

Protein sequences corresponding to each rbcL and matK coding sequences were obtained from the translated protein records linked to the respective GenBank entries. The physicochemical properties of each protein were determined using the Expasy ProtParam server (https://web.expasy.org/protparam/) (Gasteiger *et al.*, 2005). The evaluated parameters included amino acid composition, molecular weight, theoretical isoelectric point (pI), instability index, aliphatic index, the grand average of hydropathicity (GRAVY) and the extinction coefficient. Because only the homologous barcode region common to all available sequences was analyzed, the translated protein sequences represent partial proteins rather than complete rbcL and matK proteins. The observed amino acid lengths (174-184 aa for rbcL and 248-278 aa for matK) correspond to the analyzed barcode region.

Three-dimensional structure prediction

The 3D protein models were generated using SWISS-Model homology modeling server (https://swissmodel.expasy.org/interactive) (Waterhouse *et al.*, 2018). Model quality was assessed through the Ramachandran plots generated by the SWISS-Model evaluating stereochemical properties and structural reliability.

Conserved motif and domain analysis

Conserved protein motifs were analyzed using MEME suite (https://meme-suite.org/meme/tools/meme) (Bailey *et al.*, 2015). The identified motifs were compared among *Origanum* species to evaluate conservation and distribution of functionally sequence regions (six motifs) and motif width of 50.

Phosphorylation site prediction

Potential phosphorylation sites were predicted using NetPhos 3.1 (https://services.healthtech.dtu.dk/services/NetPhos-3.1/) (Blom *et al.*, 1999). This tool was used to identify putative phosphorylation residues on serine (Ser), threonine (Thr) and tyrosine (Tyr) amino acids within the predicted sequences with a threshold of 0.5

Structural feature prediction

Secondary structure, solvent accessibility, intrinsic disorder and residue geometry were predicted using NetSurfP 3.0 (https://services.healthtech.dtu.dk/services/NetSurfP-3.0/) (Høie *et al.*, 2022). The predictions were used to characterize the structural organization of the proteins and to assess the distribution of ordered and disordered regions, as well as the exposure of amino acid residues to the solvent environment.

3. Results and Discussion

The rbcL gene ranged from 523 (*Origanum* onites) to 555bp (*O. minutiflorum).* The rbcL DNA sequences of different species were obtained from NCBI GenBank. The final trimmed sequence of the gene comprised a 450 bp homologous rbcL alignment, in which every sequence (including the *T. vulgaris* used as an outgroup) has complete coverage and no gaps or missing bases. The maximum likelihood (ML) phylogenetic tree was constructed with this alignment (Figure 1).

Maximum likelihood analysis of the chloroplast rbcL sequences recovered a single monophyletic *Origanum* clade, with *T. vulgaris* forming the expected outgroup. Internal

branches within the *Origanum* clade were extremely short, indicating low sequence divergence among the species. This slow sequence variation observed among *Origanum* species reflects the slow evolutionary rate of the chloroplast rbcL gene.

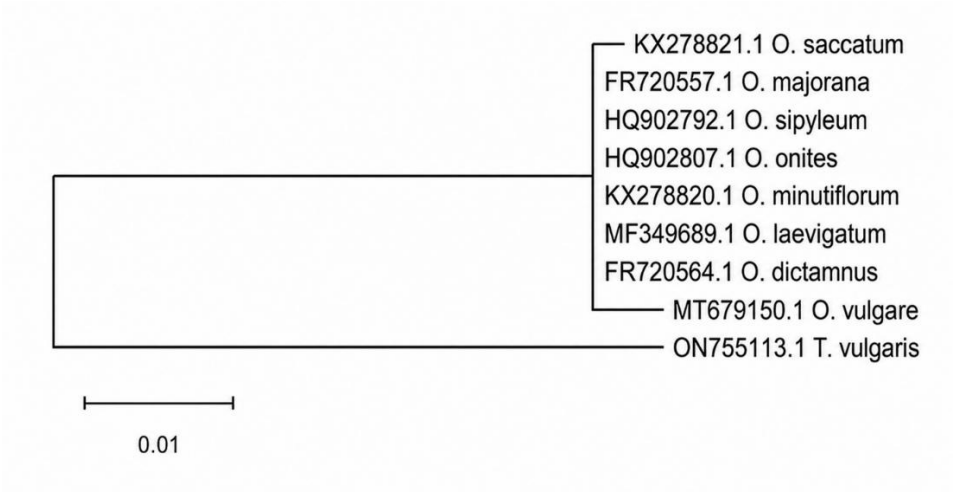

Figure 1. Phylogenetic tree of the rbcL gene in Origanum species

The high conservation of the rbcL gene among *Origanum* species is expected because chloroplast coding genes generally evolve slowly, and rbcL is one of the most conserved plastid genes widely used for plant phylogenetics and barcoding (Borsch and Quandt 2009; Daniell *et al.*, 2016). Although the rbcL gene effectively distinguished *Origanum* from the outgroup, it provided limited resolution among closely related species. This finding is consistent with previous studies showing that rbcL is highly conserved in angiosperms and is often insufficient as a sole marker for resolving recent species-level divergences (Kress and Erickson, 2007; Hollingsworth *et al.*, 2016; Kress, 2017).

In the RAPD analysis of *O. vulgare* subspecies, Katsiotis *et al.* (2009) grouped *O. onites*, *O. majorana*, *O. dictamnus* and *O. vulgare* in the same clades as our study, albeit not in the same subclades. A similar grouping was reported in the cpDNA ITS analysis study by Antaloudaki et al. in 2022. The study by Özgen *et al.* (2020) shows resemblance to the present study by grouping *O. onites* and *O. majorana* in the same subclade and *O. vulgare* in the same major clade. It should also be mentioned that in the same study, *O. elongatum* and *O. compactum* are also part of the same subclade. No data were reported for the other species of *Origanum*.

This pattern of limited divergence among species is also supported by the pairwise genetic distance analysis (Table 1), which revealed that most *Origanum* species exhibit identical sequences (genetic distance = 0.0000), or only minimal divergence (maximum = 0.0067). In contrast, genetic distances between *Origanum* species and *T. vulgaris* ranged from approximately 0.073 to 0.078, corresponding to the long branch separating the outgroup from the ingroup in the phylogenetic tree.

Table 1. Pairwise distance of rbcl gene in *Origanum* species

	O. sipyleum	O. onites	O. saccatum	O. minutiflorum	O.laevigatum	O. vulgare	T. Vulgaris	O. dictamnus	O. majorana
O. sipyleum									
O. onites	0.0000								
O. saccatum	0.0022	0.0022							
O. minutiflorum	0.0000	0.0000	0.0022						
O.laevigatum	0.0000	0.0000	0.0022	0.0000					
O. vulgare	0.0045	0.0045	0.0067	0.0045	0.0045				
T. Vulgaris	0.0732	0.0732	0.0758	0.0732	0.0732	0.0784			
O. dictamnus	0.0000	0.0000	0.0022	0.0000	0.0000	0.0045	0.0732		
O. majorana	0.0000	0.0000	0.0022	0.0000	0.0000	0.0045	0.0732	0.0000	

This agreement between the phylogeny and pairwise distance indicates that the observed topology accurately reflects the low level of sequence variation present in the rbcL dataset. These findings demonstrate that the chloroplast rbcL gene contains sufficient variation to distinguish *Origanum* from the outgroup, but provides limited phylogenetic resolution among closely related species because of its highly conserved nature. The absence of parsimony-informative sites indicates that the analysed chloroplast regions provide very limited resolution of relationship among the included *Origanum* species. Consequently, internal branching patterns should be interpreted cautiously, particularly when associated with low bootstrap support. Another factor taken into consideration in the study is the limited availability of homologous rbcL sequences for several recognized *Origanum* species in the NCBI GenBank database. Although *Origanum* comprises numerous taxonomically recognized species, only a subset is represented by comparable rbcL barcode sequences suitable for phylogenetic analysis. As a result, the present phylogeny includes only species for which homologous sequence data were available, potentially limiting the completeness of the inferred evolutionary relationships within the genus.

To address these limitations, maturase K was used as a marker for comparison. This marker was expected to provide greater phylogenetic resolution because it generally evolves more rapidly than rbcL and contains a higher proportion of informative nucleotide substitutions. The length of the sequences varied from 745 to 836bp and the final alignment length after trimming was 695 bp.

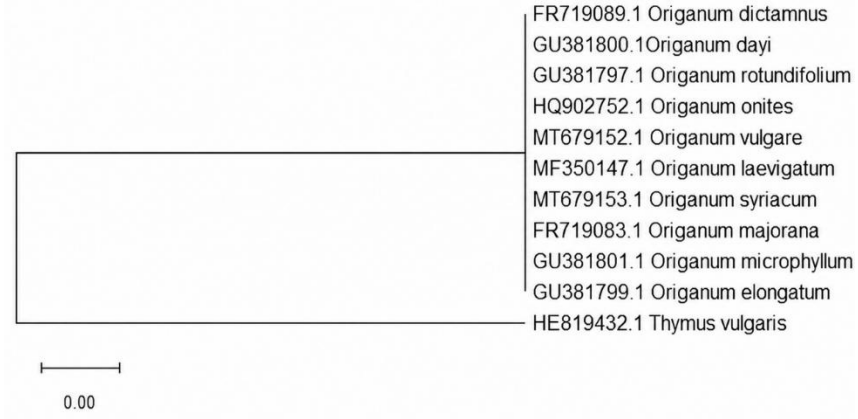

Figure 2. matK gene phylogenetic tree of *Origanum* species

The phylogenetic tree (Figure 2) for the matK marker showed two principal components: the *Origanum* taxa formed a single unresolved cluster and *T. vulgaris* is separated from the *Origanum* cluster since it serves as an outgroup belonging to a different genus. Within this *Origanum* cluster, there are no visible internal branch lengths separating the species, effectively resulting in a polytomy. This is also shown by the scale bar, which displays as 0.00 because

the branch lengths are extremely small. This result is also strongly supported by the pairwise distance analysis, which shows that the Tamura-Nei distance of the *Origanum* sequences is 0.0000, whereas between *Origanum* and *T. vulgaris* is approximately 0.0065 substitutions per site. The absence of matK sequence divergence among the sampled *Origanum* species indicates limited discriminatory power of the analyzed region at the species level. Even though the separation of *T. vulgaris* confirms that the dataset retains phylogenetic information at a broader taxonomic scale, the identical *Origanum* sequences resulted in an unresolved topology. Although matK is generally considered the more variable chloroplast marker, the homologous barcode regions available in GenBank showed complete sequence identity among the sampled *Origanum* species, whereas rbcL retained limited but detectable sequence variation.

Table 2. Pairwise distance of matK gene in *Origanum* species

	O. vulgare	O. laevigatum	O. syriacum	O. majorana	O.dictamnus	O. onites	T. Vulgaris	O. dayi	O. microphyllum	O. rotundifolium	O. elongatum
O. vulgare											
O. laevigatum	0.0000										
O. syriacum	0.0000	0.0000									
O. majorana	0.0000	0.0000	0.0000								
O. dictamnus	0.0000	0.0000	0.0000	0.0000							
O. onites	0.0000	0.0000	0.0000	0.0000	0.0000						
T. Vulgaris	0.0065	0.0065	0.0065	0.0065	0.0065	0.0065					
O. dayi	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0066				
O. microphyllum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0065	0.0000			
O. rotundifolium	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0065	0.0000	0.0000		
O. elongatum	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0065	0.0000	0.0000	0.0000	

Table 3. Physicochemical parameters of the rbcl gene in *Origanum* species

rbcL	Amino acid	GenBank	GenBank Prot	MW	pI	-R	+R	EC	II	AI	GRAVY
O. dictamnus	183	FR720564	CBX88643	20465.33	7.81	22	23	30160	36.31	78.31	-0.354
O. laevigatum	184	MF349689.1	ATB25978	19362.1	7.58	21	22	30620	27.92	80	-0.269
O. majorana	183	FR720557.1	CBX88636.1	19362.1	7.81	22	23	30160	36.31	78.31	-0.354
O. minutiflorum	184	KX278820.1	APP91298	19587.41	8.35	21	23	30495	24.28	78.42	-0.315
O. onites	174	HQ902807.1	AEX55465	19697.5	6.92	20	20	29130	31.57	80.69	-0.283
O. saccatum	177	KX278821.1	APP91299	20426.33	8.36	20	22	27515	26.39	79.32	-0.299
O. sipyleum	174	HQ902792.1	AEX55450	20443.38	6.92	20	20	29130	31.57	80.69	-0.283
O. vulgare	176	MT679150.1	QNJ99666	20465.33	6.92	21	21	30620	30.23	79.77	-0.324

Another evaluated parameter was the isoelectric point (pI) of the protein to establish the acidic or basic nature of the protein. This in turn makes it possible to determine protein stability, diversity, as well as solubility. The species *O. onites*, *O. sipyleum* and *O. vulgare. phyllum* were found to be neutral proteins (pI= 6,92), whereas the rest were found to be basic, with a pI ranging from 7.58 to 8.36. The matK *Origanum* species proteins all had a higher MW in contrast to the rbcL proteins (≥29637.66) which is expected because of their greater amino acid composition ranging from 248 aa – 278 aa. All matK species were considered to be basic in nature because of their pI ≥ 9.82 (Supplementary Table 1). The predicted isoelectric points reveal clear differences between the two chloroplast proteins. The rbcL proteins exhibit relatively conserved pI values, consistent with their strong evolutionary conservation, whereas matK pI values suggest a greater enrichment of positively charged amino acid residues. According to Tokmakov *et al.* (2021), proteins localized within specific intracellular compartments exhibit characteristic pI distribution that reflect adaptation to the biochemical environment and molecular function.

The next parameter evaluated was protein stability. The Instability Index (II) of the proteins is marked by a range of values with a cutoff mark of 40. All values above 40 show that the protein is unstable, and if this value is beneath the cutoff mark, the protein is considered stable. All rbcL proteins are considered stable, whereas all matK proteins show instability. This observation is consistent with the role of matK as a chloroplast RNA-binding maturase, where positively charged residues may facilitate interactions with negatively charged

RNA molecules during intron splicing. In contrast to the highly conserved rbcL proteins which show protein stability, matK proteins display instability indices greater than 40. According to the analysis, it should be mentioned that this index is an empirical in vitro predictor based solely on amino acid composition and does not reflect the overall protein condition and stability under physiological conditions. In addition, the analysis was based on homologous partial coding regions rather than complete matK proteins, which could influence the physicochemical parameters, including pI and II.

To further assess the stability of the rbcL protein, the aliphatic index was measured. This index shows protein thermostability by evaluating the aliphatic side chains of the protein, such as valine (V), leucine (L), isoleucine (I) and alanine (A) (Table 5).

These fractions were found to be within the 78.31 and 80.69 range. The higher the value of this index, the more thermally stable these proteins are considered to be. These values show that the Origanum species are thermostable as well as containing a high number of hydrophilic amino acids. This is in correlation with the nature of thermophilic proteins that exhibit higher values of instrinsic thermal stability and are characterized by low segregation of hydrophilic surfaces and the hydrophobic core (Zhou *et al.*, 2008; Sterpone and Melchionna, 2012; Bianco *et al.*, 2017). The thermostability of the proteins is also supported by the GRAVY values generated, which show that the values for the rbcL protein vary from -0.354 to -0.269 (Table 3). These values indicate that the nature of the protein is hydrophilic, soluble in water and appears to be globular. Amino acid composition analysis revealed that leucine was the most abundant residue in the rbcL protein (8.2-9.2%), despite the protein exhibiting an overall hydrophilic character. This finding is consistent with the structural organization of Rubisco, in which hydrophobic residues such as leucine are concentrated within the protein core to stabilize folding and maintain the conserved tertiary structure (Spreitzer and Salvucci., 2002; Anderson and Backlund., 2008). In contrast, polar and charged residues, including threonine (8.6%) and glutamine (6.9%) were also abundant and are predominantly exposed on the surface, conferring the overall hydrophilic character for the enzyme to function efficiently (Anai *et al.*, 2023; Tomita *et al.*, 2025). The lowest values were for asparagine, histidine, and tryptophan at 0%. The predominance of leucine suggests the presence of a stable hydrophobic core that maintains the 3D architecture of the enzyme, whereas the relatively high proportion of glycine (9.2%) provides conformational flexibility in loop and connecting regions.

Table 5. Aliphatic index assessment of rbcL and matK protein

Taxa	Ala (A)	Ile (I)	Leu (L)	Val (V)
Origanum dayi (matK)	2.2	7.9	12.2	5.4
Origanum dictamnus (rbcl)	7.1	3.8	8.7	7.7
Origanum dictamnus (matK)	2.6	9.6	13	7
Origanum elongatum (matK)	2.2	8.3	12.6	5.4
Origanum laevigatum (rbcl)	7.6	3.8	8.7	8.2
Origanum laevigatum (matK)	2.6	8.8	13.2	5.9
Origanum majorana (rbcl)	6.9	3.2	7.8	7.7
Origanum majorana (matK)	2.9	8.7	13	5.4
Origanum microphyllum (matK)	2.2	8.3	12.2	5.4
Origanum minutiflorum (rbcl)	8.2	3.8	8.2	8.2
Origanum onites (rbcl)	7.5	4	9.2	7.5
Origanum onites (matK)	2.7	9	12.9	7.1

Taxa	Ala (A)	Ile (I)	Leu (L)	Val (V)
Origanum rotundiflorum (matK)	2.2	8.3	12.9	5.4
Origanum saccatum (rbcl)	7.9	4	8.5	7.9
Origanum sipyleum (rbcl)	7.5	4	9.2	7.5
Origanum syriacum	2.7	9.3	12	5.4
Origanum vulgare (rbcl)	7.4	4	9.1	7.4
Origanum vulgare (matK)	2.8	9.3	11.7	5.2

Such a combination of stabilizing and flexible protein residues is characteristic of highly conserved enzymes such as Rubisco (Zhang et al., 2021). Although leucine was the most abundant individual amino acid in the aliphatic index, the negative GRAVY index indicates that the cumulative physicochemical properties of all residues produce an overall hydrophilic protein, consistent with the soluble stromal localization of Rubisco. This corresponds with the recent studies stating that a negative GRAVY score indicates a predominantly hydrophilic, globular protein likely to effectively interact with water and other polar molecules (Nene *et al.*, 2022; Komal *et al.*, 2025).

The identification of conserved sequence motifs in the rbcL protein was performed using MEME Suite. Six highly conserved motifs were identified across all species, with extremely significant E-values, indicating that these motifs represent evolutionary domains rather than occurring by chance. Although the precise evolutionary significance of the individual motifs in protein networking is unknown, their key role in protein-protein interaction is well documented (Saito *et al.*, 2007). Because post-translational modification can further regulate protein activity, phosphorylation sites of these proteins were predicted using NetPhos 3.1. Only prediction scores greater than 0.5 were considered as potential phosphorylation sites of serine, threonine and tyrosine. The analysis revealed a high degree of conservation among the species; each protein contained between 12 and 13 phosphorylation sites, with Thr being completely conserved across all species, whereas Ser and Tyr differed by only one residue. This result parallels the MEME results showing that the structural and regulatory features of the protein have been highly conserved in the genus *Origanum* (Figure 3a, b, c).

Figure 3a. MEME conserved motifs of the rbcl gene

In contrast, matK MEME analysis showed that the same proteins do not display the same degree of motif conservation. Although the motifs remain highly significant, their arrangement is more variable among the species, consistent with the greater evolutionary flexibility of matK (Supplementary Figure 1). Unlike rbcL, matK is known to accumulate substitutions at a faster rate (Supplementary Table 2). It should be noted that the rbcL protein comprised of 174-178 aa, whereas the translated matK proteins ranged from 248-278 aa. The difference in length

reflects the distinct biological functions and coding capabilities of the two chloroplast genes. Despite the shorter length, MEME analysis identified six highly conserved motifs shared by all species, accompanied by nearly identical phosphorylation profiles. The longer matK protein contained a greater number of phosphorylation sites and exhibited increased variability in motif organization. These findings suggest that protein length alone does not determine the degree of evolutionary conservation.

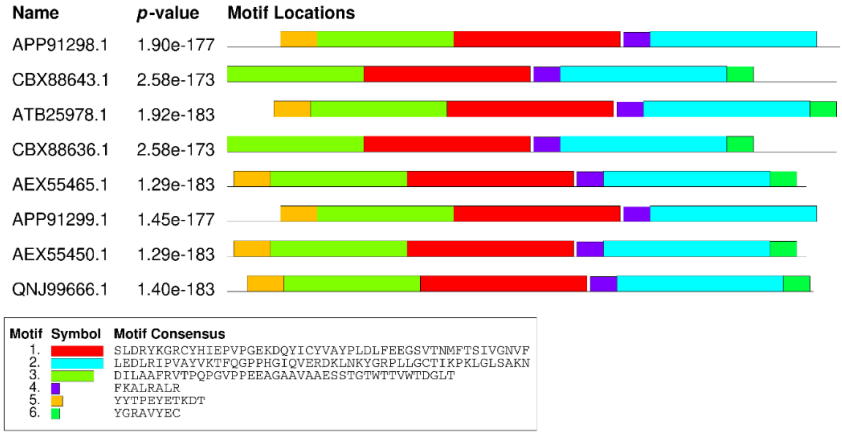

Figure 3b. MEME motif locations of the rbcl gene

	Ser (S)	Thre (T)	Tyr (Y)	Total
O. dictamnus	6	4	2	12
O. laevigatum	5	4	3	12
O. majorana	6	4	2	12
O. minutiflorum	6	4	3	13
O. onites	5	4	3	12
O. saccatum	6	4	3	13
O. sipyleum	5	4	3	12
O. vulgare	5	4	3	12

Figure 3c. Phosphorylation sites of rbcl gene

Additional support for these findings was provided by the predicted structural properties obtained using NetSurfP 3.0. NetSurfP 3.0 is a deep learning-based platform that predicts residue-specific secondary structure, solvent accessibility and intrinsic disorder from amino acid sequences (Høie *et al.*, 2022). The analysis indicated that the majority of amino acid residues were predicted to be buried (blue), suggesting that a large proportion of the protein contributes to the formation of a compact core, contributing to protein stability and proper folding (Johnson *et al.*, 2007; Jones and Gardner, 2015). Secondary structure prediction revealed that coil regions were predominant, interspersed with α-helices and β-strands (Figure 4). The abundance of coil regions is consistent with their role in connecting regular secondary structural elements and providing the local conformational flexibility required for catalytic function while preserving the overall tertiary structure of the enzyme. Furthermore, only a limited number of residues were predicted to exhibit intrinsic disorder, supporting the highly ordered and evolutionary architecture of rbcL. MatK secondary structure analysis displayed the same profile (Supplementary Figure 2).

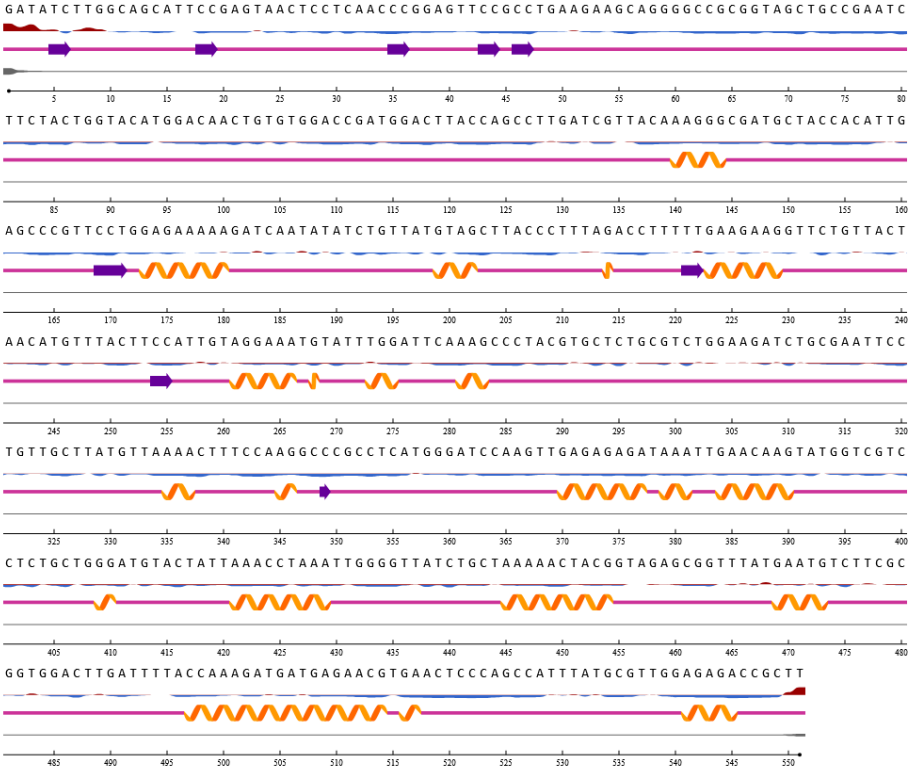

Figure 4. Secondary structure of the rbcL protein as evaluated by NetSurfP 3.0 (red is exposed aa, blue is buried aa; orange is helix structure, purple arrow is strand, purple line is coil)

The 3D structures from the analyzed *Origanum* species were predicted using the SWISS-MODEL tool to investigate structural conservation beyond primary sequence variation (Figure 5). This modeling provides insight into protein function, stability and interactions, contributing to our understanding of proteins of pharmacological and biological relevance (Mahdally *et al.*, 2024; Ravi *et al.*, 2026).

Despite minor sequence differences among the species, the predicted tertiary structures exhibited a highly conserved overall fold, indicating that sequence length did not substantially affect protein structure. This high degree of structural conservation is expected of rbcL, since its catalytic role in photosynthesis makes it subject to strong evolutionary constraints (Leung *et al.*, 2026). The reliability of the predicted models was also supported by structural validation. More than 95% of the amino acid residues were located within the favored regions of the Ramachandran plots, whereas outliers were either absent or represented less than 0.3% of residues (Figure 6). MolProbity scores ranging from 0.5 to 1.4, along with the low clashscores, confirm the excellent stereochemical quality of the models. According to current standards for protein structure assessment, the parameters are characteristic of highly reliable tertiary structure predictions (Waterhouse *et al.*, 2018; Waterhouse *et al.*, 2024). Analysis of the amino acid composition also revealed that glycine was among the most abundant residues within the predicted proteins. These glycine-rich regions have been shown to contribute to

structural flexibility and are involved in cellular metabolism and gene regulation (Lorković 2009; Cheng *et al.*, 2023).

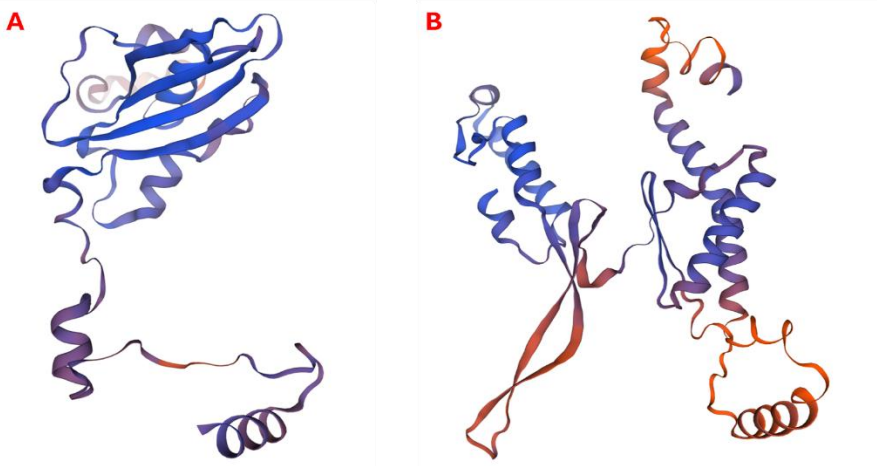

Figure 5. 3D-modeling of O. vulgare rbcl (A) and matK (B)

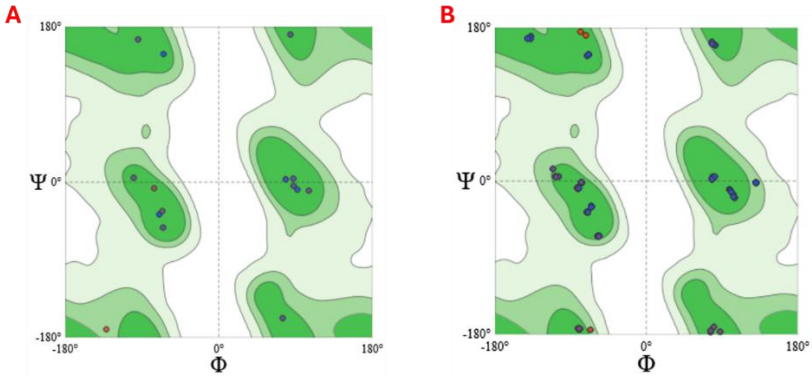

Figure 6. Ramachandran plots of two Origanum rbcL proteins (A: O. vulgare; B: O. majorana)

Overall, the structural analysis demonstrated that the rbcL protein of *Origanum* retains a conserved tertiary organization despite sequence variation among species. These findings agree with previous studies which similarly reported, despite sequence divergence among plant taxa, the overall motif organization and 3D structure of rbcL remain highly conserved because of strong functional and evolutionary constraints (Valegard et al., 2018; Sevindik *et al.*, 2024; Leung *et al.*, 2026). Rather than evaluating chloroplast markers solely by phylogenetic reconstruction, the present study integrates evolutionary, physicochemical, structural and functional analysis, providing a more comprehensive assessment of marker conservation.

5. Conclusions

This study demonstrates that integrating phylogenetic reconstruction with comparative protein characterization provides a more comprehensive evaluation of chloroplast barcode markers than sequence analysis alone. Although both rbcL and matk exhibited limited discriminatory power within the available *Origanum* dataset, their proteins remained remarkably conserved at the physiochemical and structural levels, indicating strong evolutionary constraints. These findings establish an integrated analytical framework for future studies of barcode evolution in *Origanum* and other *Lamiaceae* taxa.

Supplementary Materials

The following supporting information can be downloaded at:
https://onlinejbs.com/index.php/jbs/libraryFiles/downloadPublic/404

Funding

This study is funded by the National Agency for Scientific Research and Innovation (NASRI) in Albania within the framwork of the project "*Integrated analysis of genetic, histocytologic variability, the chemical profile, and bioactive compounds of Origanum vulgare L grown in distinct ecological conditions*".

Conflict of interests

The authors declare they have no conflicting interests to disclose.

References

Anai, T., Abe, S., Shobu, K., & Kadokawa, J.-i. (2023). Synthesis of hydrophobic poly(γ-glutamic acid) derivatives by enzymatic grafting of partially 2-deoxygenated amyloses. Applied Sciences, *13*(1), 489. https://doi.org/10.3390/app13010489

Andersson, I., & Backlund, A. (2008). Structure and function of Rubisco. Plant Physiology and Biochemistry, *46*, 275–291.

Antaloudaki, E., Mylonas, M., Kypriotakis, Z., & Poulakakis, N. (2022). Phylogenetic relationships of *Origanum* taxa (Lamiaceae) from Greece: Initial insights from molecular and morphological data. Botanica Serbica, *46*(1), 71–83. https://doi.org/10.2298/BOTSERB2201071A

Bailey, T. L., Johnson, J., Grant, C. E., & Noble, W. S. (2015). The MEME Suite. Nucleic Acids Research, *43*(W1), W39–W49.

Bianco, V., Franzese, G., Dellago, C., & Coluzza, I. (2017). Role of water in the selection of stable proteins at ambient and extreme thermodynamic conditions. Physical Review X, *7*(2), 021047. https://doi.org/10.1103/PhysRevX.7.021047

Blom, N., Gammeltoft, S., & Brunak, S. (1999). Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. Journal of Molecular Biology, *294*(5), 1351–1362. https://doi.org/10.1006/jmbi.1999.3310

Borsch, T., & Quandt, D. (2009). Mutational dynamics and phylogenetic utility of noncoding chloroplast DNA.

Buck, R., & Flores-Rentería, L. (2022). The syngameon enigma. Plants, *11*(7), 895. https://doi.org/10.3390/plants11070895

Chaachouay, N., Elachouri, M., Bussmann, R. W., & Khojimatov, O. K. (2024). *Origanum compactum* Benth., *Origanum majorana* L., *Origanum syriacum* L., *Origanum vulgare* subsp. *gracile* (K. Koch) Letsw. Lamiaceae. In R. W. Bussmann, M. Elachouri, & Z. Kikvidze (Eds.), Ethnobotany of Northern Africa and Levant. Springer. https://doi.org/10.1007/978-3-031-43105-0_155

Cheng, K., Zhang, C., Lu, Y., et al. (2023). The glycine-rich RNA-binding protein is a vital post-transcriptional regulator in crops. Plants, *12*(19), 3504. https://doi.org/10.3390/plants12193504

Daniell, H., Lin, C.-S., Yu, M., & Chang, W.-J. (2016). Chloroplast genomes: Diversity, evolution, and applications in genetic engineering. *Genome Biology, 17*, 134.

de Torre, M. P., Vizmanos, J. L., Cavero, R. Y., & Calvo, M. I. (2020). Improvement of antioxidant activity of oregano (*Origanum vulgare* L.) with an oral pharmaceutical form. Biomedicine & Pharmacotherapy, *129*, 110424. https://doi.org/10.1016/j.biopha.2020.110424

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., & Bairoch, A. (2005). Protein identification and analysis tools on the ExPASy server. In J. M. Walker (Ed.), *The proteomics protocols handbook* (pp. 571–607). Humana Press.

Høie, M. H., Kiehl, E. N., Petersen, B., Nielsen, M., Winther, O., Nielsen, H., Hallgren, J., & Marcatili, P. (2022). NetSurfP-3.0: Accurate and fast prediction of protein structural features by protein language models and deep learning. Nucleic Acids Research, *50*(W1), W510–W515. https://doi.org/10.1093/nar/gkac439

Hollingsworth, P. M., Li, D. Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: From barcodes to genomes. Philosophical Transactions of the Royal Society B: Biological Sciences, *371*, 20150338.

Johnson, R. J., Lin, S. R., & Raines, R. T. (2007). Genetic selection reveals the role of a buried, conserved polar residue. Protein Science, *16*(8), 1609–1616. https://doi.org/10.1110/ps.072938907

Jones, R. D., & Gardner, R. G. (2015). Digging for buried amino acids unearths new protein quality control treasure. Structure, *23*(7), 1151–1152. https://doi.org/10.1016/j.str.2015.06.008

Katsiotis, A., Nikoloudakis, N., Linos, A., Drossou, A., & Constantinidis, T. (2009). Phylogenetic relationships in *Origanum* spp. based on rDNA sequences and intra-genetic variation of Greek *O. vulgare* subsp. *hirtum* revealed by RAPD. Scientia Horticulturae, *121*(1), 103–108. https://doi.org/10.1016/j.scienta.2009.01.015

Komal, L., Kamran, A., Jahan, S., et al. (2025). Molecular insights into drought tolerance in wheat through in-silico genome-wide analysis of DREB1 transcription factor and peroxidase interactions. BMC Plant Biology, *25*, 1158. https://doi.org/10.1186/s12870-025-06938-4

Kress, W. J. (2017). Plant DNA barcodes: Applications today and in the future. Journal of Systematics and Evolution, *55*, 291–307.

Kress, W. J., & Erickson, D. L. (2007). A two-locus global DNA barcode for land plants: The coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. PLoS ONE, *2*(6), e508. https://doi.org/10.1371/journal.pone.0000508

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). Molecular Evolutionary Genetics Analysis version 12 for adaptive and green computing. Molecular Biology and Evolution, *41*, 1–9.

Leung, A., Chang, B. S. W., & Sage, R. F. (2026). Convergence, stability, and thermal adaptation of the Rubisco large subunit in plants. Evolution, qpag093. https://doi.org/10.1093/evolut/qpag093

Leung, A., Chang, B. S. W., & Sage, R. F. (2026). Convergence, stability, and thermal adaptation of the Rubisco large subunit in plants. Evolution. https://doi.org/10.1093/evolut/qpag093

Li, Y., He, L., Xue, Z., et al. (2025). Antimicrobial activity of *Origanum vulgare* L. essential oil and its effects on fungal gene expression. PLoS ONE, *20*(8), e0329548. https://doi.org/10.1371/journal.pone.0329548

Lorković, Z. J. (2009). Role of plant RNA-binding proteins in development, stress response and genome organization. Trends in Plant Science, *14*(4), 229–236. https://doi.org/10.1016/j.tplants.2009.01.007

Lukas, B., Samuel, R., Mader, E., Baser, K. H. C., Duman, H., & Novak, J. (2013). Complex evolutionary relationships in the *Origanum* L. section *Majorana* Benth. (Lamiaceae). Botanical Journal of the Linnean Society, *171*, 667–686. https://doi.org/10.1111/boj.12022

Madeira, F., Madhusoodanan, N., Lee, J., et al. (2024). The EMBL-EBI Job Dispatcher sequence analysis tools framework in 2024. Nucleic Acids Research, *52*(W1), W521–W525. https://doi.org/10.1093/nar/gkae241

Mahdally, N. H., Salih, A. E. M., El-Shiekh, R. A., et al. (2024). Exploring the antimicrobial activity of *Origanum majorana* L. against the highly virulent multidrug-resistant *Acinetobacter baumannii* AB5075: UPLC-HRMS profiling with in vitro and in silico studies. Future Journal of Pharmaceutical Sciences, *10*, 71. https://doi.org/10.1186/s43094-024-00641-1

Mirarab, S., Bayzid, M. S., & Warnow, T. (2016). Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. Systematic Biology, *65*(3), 366–380.

Nazar, N., Howard, C., Slater, A., & Sgamma, T. (2022). Challenges in medicinal and aromatic plants DNA barcoding—Lessons from the Lamiaceae. Plants, *11*(1), 137. https://doi.org/10.3390/plants11010137

Nene, T., et al. (2022). Plant catalase in silico characterization and phylogenetic analysis. Journal of Genetic Engineering and Biotechnology, *20*.

Nurzyńska-Wierdak, R., & Walasek-Janusz, M. (2025). Chemical composition, biological activity, and potential uses of oregano (*Origanum vulgare* L.) and oregano essential oil. Pharmaceuticals, *18*(2), 267. https://doi.org/10.3390/ph18020267

O'Leary, N. A., Cox, E., Holmes, J. B., Anderson, W. R., Falk, R., Hem, V., Tsuchiya, M. T. N., Schuler, G. D., Zhang, X., Torcivia, J., Ketter, A., Breen, L., Cothran, J., Bajwa, H., Tinne, J., Meric, P. A., Hlavina, W., & Schneider, V. A. (2024). Exploring and retrieving sequence and metadata for species across the tree of life with NCBI Datasets. Scientific Data, *11*(1), 732. https://doi.org/10.1038/s41597-024-03571-y

Özgen, A., Tan, N., Taştan, Ö., & Pehlevan, F. (2021). Phylogenetic analysis of *Origanum vulgare* and its antioxidant and antimicrobial activity. Gazi University Journal of Science, *34*(2), 311–325. https://doi.org/10.35378/gujs.751660

Pere, K., Mburu, K., Muge, E. K., Wagacha, J. M., & Nyaboga, E. N. (2023). Molecular discrimination and phylogenetic relationships of *Physalis* species based on ITS2 and *rbcL* DNA barcode sequences. Crops, *3*(4), 302–319. https://doi.org/10.3390/crops3040027

Ravi, P., Jasuja, H., Sarkar, D., Katti, D. R., Shetty, K., & Katti, K. S. (2026). Plant extracts from *Origanum vulgare* and *Vaccinium macrocarpon* induce apoptosis of bone metastasized breast cancer cells in a 3D bone-mimetic testbed of bone metastasis. International Journal of Molecular Sciences, *27*(5), 2355. https://doi.org/10.3390/ijms27052355

Saito, H., Kashida, S., Inoue, T., & Shiba, K. (2007). The role of peptide motifs in the evolution of a protein network. Nucleic Acids Research, *35*(19), 6357–6366. https://doi.org/10.1093/nar/gkm692

Sarfaraz, D., Rahimmalek, M., Sabzalian, M. R., Gharibi, S., Matkowski, A., & Szumny, A. (2023). Essential oil composition and antioxidant activity of oregano and marjoram as affected by different light-emitting diodes. Molecules, *28*(9), 3714. https://doi.org/10.3390/molecules28093714

Sevindik, E., Can, A. E., Özkara, Z., Düşgün, Y., Toksöz, Ş., & Özdemir, H. İ. (2024). Phylogenetic relationships among Lamiaceae species from Aydin (Türkiye), based on *rbcL* sequences. Biologica Nyssana, *15*(2), 53–60. https://doi.org/10.5281/zenodo.13902177

Solomou, A. D., Molla, A., & Skoufogianni, E. (2026). Medicinal and aromatic plant diversity of Greece: Biodiversity knowledge, ethnobotany and sustainable use—A short review. Diversity, *18*(1), 56. https://doi.org/10.3390/d18010056

Spreitzer, R. J., & Salvucci, M. E. (2002). Rubisco: Structure, regulatory interactions, and possibilities for a better enzyme. Annual Review of Plant Biology, *53*, 449–475.

Sterpone, F., & Melchionna, S. (2012). Thermophilic proteins: Insight and perspective from in silico experiments. *Chemical Society Reviews, 41*(5), 1665–1676. https://doi.org/10.1039/c1cs15199a

Tokmakov, A. A., Kurotani, A., & Sato, K. I. (2021). Protein pI and intracellular localization. Frontiers in Molecular Biosciences, *8*, 775736. https://doi.org/10.3389/fmolb.2021.775736

Tomita, N., Onoda, H., Chavas, L. M. G., & Chikenji, G. (2025). Exploring hydrophilic sequence space to search for uncharted foldable proteins by AlphaFold2. Biophysics and Physicobiology, *22*(1), e220005. https://doi.org/10.2142/biophysico.bppb-v22.0005

Toso, F., Buldain, D., Retta, D., Di Leo Lira, P., Marchetti, M. L., & Mestorino, N. (2025). Antimicrobial activity of *Origanum vulgare* L. and *Salvia rosmarinus* Spenn (syn. *Rosmarinus officinalis* L.) essential oil combinations against *Escherichia coli* and *Salmonella typhimurium* isolated from poultry. Processes, *13*(9), 2856. https://doi.org/10.3390/pr13092856

Valegård, K., Hasse, D., Andersson, I., & Gunn, L. H. (2018). Structure of Rubisco from *Arabidopsis thaliana* in complex with 2-carboxyarabinitol-1,5-bisphosphate. Acta Crystallographica Section D: Structural Biology, *74*(1), 1–9. https://doi.org/10.1107/S2059798317017132

Veljovic, K., Tesevic, V., Mitrovic, H., & Stankovic, M. (2023). Essential oil of *Origanum minutiflorum* exhibits anti-inflammatory and antioxidative effects in human bronchial cells and antimicrobial activity on lung pathogens. Journal of Herbal Medicine, *39*, 100651. https://doi.org/10.1016/j.hermed.2023.100651

Waterhouse, A. M., Studer, G., Robin, X., Bienert, S., Tauriello, G., & Schwede, T. (2024). The structure assessment web server: For proteins, complexes and more. Nucleic Acids Research, *52*(W1), W318–W323. https://doi.org/10.1093/nar/gkae270

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T. (2018). SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Research, *46*(W1), W296–W303. https://doi.org/10.1093/nar/gky427

Zhang, Z., Ryoo, D., Balusek, C., et al. (2021). Inward-facing glycine residues create sharp turns in β-barrel membrane proteins. Biochimica et Biophysica Acta (BBA) - Biomembranes, *1863*(10), 183662. https://doi.org/10.1016/j.bbamem.2021.183662

Zhou, X. X., Wang, Y. B., Pan, Y. J., et al. (2008). Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. Amino Acids, *34*, 25–33. https://doi.org/10.1007/s00726-007-0589-x