Computational Structural Evaluation of a C-Terminal Glycine–Serine Linker in Human Choline Kinase β

Dhanesh Narain A/L Kunasagaran[1], Wei Cun See Too[1], Ling Ling Few[1*]

[1]School of Health Sciences, Health Campus, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan

Received: April 22, 2026; Accepted: July 1, 2026; Published: July 15, 2026

Abstract

Choline kinase β (CHKB) is a key enzyme in phospholipid biosynthesis whose dysfunction underlies megaconial congenital muscular dystrophy, yet it remains comparatively understudied relative to its paralog CHKA. Rational protein engineering of CHKB is constrained by concerns that structural modification may disrupt its obligate dimerization and catalytic integrity. In this study, we present a comprehensive *in silico* evaluation of a human CHKB construct engineered with a C-terminal glycine–serine linker and designed for expression in a pET-based bacterial system. Sequence analysis and computational cloning were combined with sequence-based solubility prediction and state-of-the-art AlphaFold structural modeling. Predicted solubility analyses indicate that linker incorporation does not substantially alter intrinsic aggregation propensity relative to wild-type CHKB. Structural predictions reveal that the catalytic core retains high-confidence folding, while the appended linker behaves as an intrinsically disordered, solvent-exposed extension. Dimer modeling based on the experimentally determined CHKB crystal structure demonstrates that the linker does not interfere with dimer interface geometry or occlude substrate-binding regions. Stereochemical analysis confirms acceptable backbone geometry for both monomeric and dimeric models. Collectively, these results support the structural compatibility of C-terminal glycine–serine linker incorporation into CHKB. This work establishes a computationally validated framework for CHKB construct design, providing a foundation for future experimental evaluation, fusion protein engineering, and structure–function studies of this medically important enzyme.

Keywords: choline kinase β (CHKB), glycine–serine linker, protein engineering, AlphaFold structural modelling, *in silico* analysis

---

1. Introduction

In eukaryotes, the Kennedy pathway represents the principal route for the *de novo* synthesis of phosphatidylcholine (PC) and phosphatidylethanolamine (PE), which together constitute more than half of total membrane phospholipids [1,2]. The first committed step of the choline branch of this pathway is catalyzed by choline kinase (CHK), which phosphorylates choline or ethanolamine in an ATP-dependent reaction. Because this step precedes rate-limiting cytidylyltransferase reactions, CHK functions as a key regulatory node linking membrane biogenesis with cellular growth and metabolic state [3,4].

In mammals, choline kinase activity is encoded by two distinct genes, *chka* and *chkb*, which give rise to isoforms with overlapping catalytic capabilities but divergent physiological roles [5]. CHKA isoforms have been extensively studied due to their upregulation in a wide range of cancers and their assessment as therapeutic targets [6,7]. In contrast, CHKB, encoded by *chkb*, has historically received less attention despite

its essential and non-redundant functions in post-mitotic tissues. CHKB exhibits distinct substrate preferences, favoring ethanolamine over choline, and shows enriched expression in skeletal muscle, bone, and brain [8,9].

The physiological importance of CHKB is underscored by human genetics. Loss-of-function mutations in *chkb* cause megaconial congenital muscular dystrophy, a severe autosomal recessive disorder characterized by progressive muscle weakness, skeletal abnormalities, and the presence of pathognomonic enlarged mitochondria arising from phosphatidylethanolamine deficiency [10,11]. Experimental models further demonstrate that CHKB is indispensable for muscle membrane stability, mitochondrial architecture, and bone development, and that CHKA cannot compensate for CHKB loss despite catalyzing the same biochemical reaction [9]. These findings establish CHKB as a critical regulator of membrane homeostasis in specialized tissues.

Structurally, CHKB adopts a conserved bilobal kinase fold and functions as an obligate dimer, with extensive C-terminal lobe interactions required for catalytic activity [8,12]. Dimerization stabilizes the active conformation through trans-complementation between protomers, and monomeric forms are catalytically inactive. High-resolution crystal structures have elucidated key features of substrate recognition, metal coordination, and isoform-specific differences in active-site architecture that underlie CHKB's preferential ethanolamine kinase activity [8,13]. Despite these advances, experimental tools for systematic structure–function studies of CHKB remain limited.

Protein engineering strategies, particularly fusion protein design, have become central to modern biotechnology, enabling improved expression, stability, and functional versatility [14]. A critical determinant of fusion protein performance is the linker sequence connecting functional domains. Flexible glycine–serine (Gly–Ser) linkers have emerged as a widely used standard due to their high conformational flexibility, hydrophilicity, resistance to proteolysis, and low propensity to interfere with domain folding or activity [14,15]. Such linkers typically behave as solvent-exposed, intrinsically disordered extensions that preserve native oligomerization and active-site accessibility.

Despite the recognized importance of CHKB and the extensive use of Gly–Ser linkers in fusion protein engineering, there is a notable lack of systematically designed CHKB constructs incorporating defined C-terminal flexible linkers in high-expression bacterial vectors. Expression platforms such as the pET vector series, particularly pET-14b, provide a robust framework for recombinant protein production in *Escherichia coli*, offering strong transcriptional control, straightforward purification via N-terminal His_6 tags, and compatibility with rational fusion designs [16,17]. However, the structural compatibility of C-terminal linker incorporation with CHKB folding, dimerization, and interface integrity has not been rigorously evaluated, even at the computational level.

Recent advances in *in silico* protein analysis and artificial intelligence–based structure prediction have transformed protein engineering workflows by enabling detailed evaluation of engineered constructs prior to experimental validation [18,19]. Tools such as AlphaFold and sequence-based solubility predictors provide powerful means to assess whether modifications—such as linker insertion—are likely to perturb native folding, quaternary assembly, or biophysical behavior. These approaches are particularly valuable for fusion proteins containing flexible Gly–Ser linkers, which may introduce intrinsic disorder or alter interdomain dynamics that are difficult to capture experimentally [20].

In this study, we present a comprehensive *in silico* analysis of a human CHKB construct engineered with a defined C-terminal glycine–serine linker and designed for expression in the pET-14b vector. Expression platforms such as the pET vector series, particularly pET-14b, provide a robust framework for recombinant protein production in *Escherichia coli*, offering strong transcriptional control, straightforward purification via N-terminal His_6 tags, and compatibility with rational fusion designs. Using sequence analysis, solubility prediction, and state-of-the-art structural modeling, we assess the impact of linker incorporation on CHKB monomer folding, dimer architecture, and interface compatibility. By addressing a clear

methodological gap, this work establishes a computationally validated framework for CHKB construct design that supports future recombinant expression, structure–function studies, and protein engineering applications involving this medically important enzyme.

2. Materials and Methods

2.1 Sequence retrieval and construct design

The human *chkb* coding sequence (1188 bp, including the stop codon) was retrieved from the NCBI Nucleotide database (accession CU012980.1) in FASTA format. The sequence corresponds to choline kinase β (CHKB), comprising 395 amino acids. Sequence handling, annotation, and construct design were performed using Benchling Biology Software, cloud-based platform release 2025.8/2025.9 (Benchling, Inc., San Francisco, CA, USA; accessed in September/2025, https://www.benchling.com). This is a cloud-based molecular biology platform that enables visualization of open reading frames, restriction sites, and cloning strategies.

The pET-14b expression vector sequence (Novagen) was obtained from the manufacturer and imported into Benchling as a separate plasmid file. Vector features including the T7 promoter, ribosome-binding site, N-terminal His_6-tag with thrombin cleavage site, multiple cloning site (MCS), and transcription terminator were annotated. Restriction site mapping confirmed the presence of unique *Nde*I and *BamH*I sites within the MCS, enabling directional cloning of the *chkb* insert.

In silico cloning of the pET-14b–*chkb* construct was performed using the Benchling Assembly Wizard. The vector backbone was defined as the region between the *Nde*I and *BamH*I restriction sites, while the *chkb* insert was defined as the corresponding *Nde*I–*BamH*I fragment derived from the coding sequence. The virtual digest-and-ligation workflow generated a recombinant plasmid model in which *chkb* is positioned in-frame downstream of the N-terminal His_6-tag.

2.2 Design of the C-terminal glycine–serine linker

A flexible C-terminal glycine–serine linker was designed *in silico* to generate an engineered CHKB variant. A reverse primer encoding a glycine–serine-rich linker (GGGGGS) was designed within Benchling on the coding strand of the *chkb* open reading frame. The primer comprised, in 5'→3' orientation, a *BamH*I restriction site, the linker-coding sequence, and a region complementary to the 3' end of *chkb*.

Benchling PCR simulation tools were used to generate a virtual amplicon encoding CHKB fused to the C-terminal linker. This amplicon and the original pET-14b vector backbone were then digested *in silico* with *Nde*I and *BamH*I and assembled computationally to yield a final pET-14b–*chkb* construct containing the C-terminal glycine–serine linker. Correct reading frame, orientation, and preservation of vector regulatory elements were verified visually.

2.3 Sequence-based protein solubility prediction

Predicted soluble expression of CHKB variants in *Escherichia coli* was evaluated using Protein-Sol (https://protein-sol.manchester.ac.uk/, accessed August–September 2025), a web-based server for estimating protein solubility based on amino-acid sequence features. For wild-type CHKB, the amino-acid sequence corresponding to NCBI accession CU012980.1 was submitted to Protein-Sol after removal of non-sequence annotations.

For the engineered variant, six glycine/serine residues (GGGGGS) were appended *in silico* to the C-terminus of the CHKB sequence prior to submission. Protein-Sol was run using default parameters to obtain the scaled solubility value (QuerySol) and associated sequence-feature profiles, including hydropathy, charge

distribution, folding propensity, and disorder propensity. Predicted solubility metrics for wild-type and linker-containing CHKB were compared directly.

2.4 Monomer structure prediction using AlphaFold

Three-dimensional monomeric structures of CHKB variants were evaluated using AlphaFold-based resources. The wild-type CHKB structure was obtained from the AlphaFold Protein Structure Database (v4; UniProt Q9Y259; entry AF-Q9Y259-F1), which provides predicted coordinates together with confidence metrics including per-residue pLDDT scores, predicted aligned error (PAE), and predicted template modeling (pTM) scores. Structural variations introduced by the mutations were analyzed against this baseline configuration.

For the engineered CHKB-linker variant, which is not available as a precomputed model, structure prediction was performed using the AlphaFold-3 web server. The full engineered amino-acid sequence, comprising wild-type CHKB followed by the C-terminal GGGGGS linker, was submitted using default settings and a single random seed. Predicted structures were downloaded in mmCIF format for further analysis.

2.5 Stereochemical analysis of monomer models

The stereochemical quality of the AlphaFold-predicted monomer models was assessed using RamPlot webserver (https://www.ramplot.in, accessed August–September 2025). The mmCIF coordinate files for wild-type and linker-containing CHKB were uploaded to the server, and Ramachandran plots were generated to evaluate backbone φ/ψ angle distributions. The proportions of residues in favored, allowed, and disallowed regions were recorded, with particular attention to whether outliers were confined to terminal or glycine-rich regions indicative of intrinsic flexibility.

2.6 Dimer model construction

To assess whether the engineered linker is compatible with native CHKB dimerization, a dimeric CHKB-linker model was constructed using UCSF ChimeraX (v1.8; https://www.cgl.ucsf.edu/chimerax/). The crystal structure of human choline kinase β (PDB ID: 2IG7) and the AlphaFold-derived CHKB-linker monomer were loaded into the same ChimeraX session.

The linker-containing monomer was first superposed onto one protomer of the crystal dimer using the MatchMaker command. A duplicate of the fitted monomer was then generated and superposed onto the second protomer, yielding a symmetric homodimer composed entirely of CHKB-linker chains. The original crystal structure chains were hidden, and the resulting dimeric model was saved as a PDB file for assessment and analysis.

2.7 All-atom analysis of the dimer model

All-atom stereochemical analysis of the CHKB-linker dimer was performed using MolProbity (https://bio.tools/molprobity, accessed August–September 2025). Hydrogen atoms were removed and rebuilt within the MolProbity workflow prior to analysis. The "Analyze all-atom contacts and geometry" protocol was run with default parameters to obtain clashscores, MolProbity scores, Ramachandran and rotamer statistics, CaBLAM outliers, and bond-length and bond-angle deviations. Multiple conformers generated during assessment were evaluated, and the model with the most favorable overall stereochemical statistics was selected for subsequent structural inspection.

2.8 Interface distance analysis

Potential steric interference between the C-terminal glycine–serine linker and the dimer interface was evaluated by distance measurements in ChimeraX. Representative Cα atoms at the linker terminus and at opposing protomer interface regions were selected using atom-specific selection commands. Inter-chain Cα–Cα distances were calculated using the distance command. Distances substantially exceeding typical contact ranges were interpreted as evidence that the linker behaves as a solvent-exposed, non-clashing extension.

3. Results

3.1 *In silico* design of the pET-14b– chkb expression construct and C-terminal linker

A recombinant expression construct encoding human choline kinase β (CHKB) was designed *in silico* using the pET-14b vector backbone. The *chkb* coding sequence (1188 bp; 395 amino acids) was positioned downstream of the T7 promoter and ribosome-binding site, enabling high-level expression in *E. coli*. The construct encodes an N-terminal His_6-tag followed by a thrombin cleavage site, placing the CHKB open reading frame in-frame with affinity and protease elements for downstream purification and processing (Fig. 1A).

To facilitate modular protein engineering while preserving native structure, a flexible glycine–serine linker was appended to the C-terminus of CHKB. The linker consists of six residues (GGGGGS) and was designed to be introduced immediately upstream of the stop codon without altering the native CHKB sequence. A zoomed schematic of the C-terminal region highlights the precise position and composition of the linker relative to the CHKB coding sequence (Fig. 1B). This design ensures that the linker functions as a terminal, solvent-exposed extension rather than an internal insertion that could disrupt folding or dimerization.

In silico assembly supported correct orientation, reading frame continuity, and preservation of all essential vector regulatory elements. The construct schematic provides a consolidated overview of both the expression cassette architecture and the specific engineering modification introduced at the CHKB C-terminus, serving as the reference design for subsequent computational solubility prediction and structural analyses.

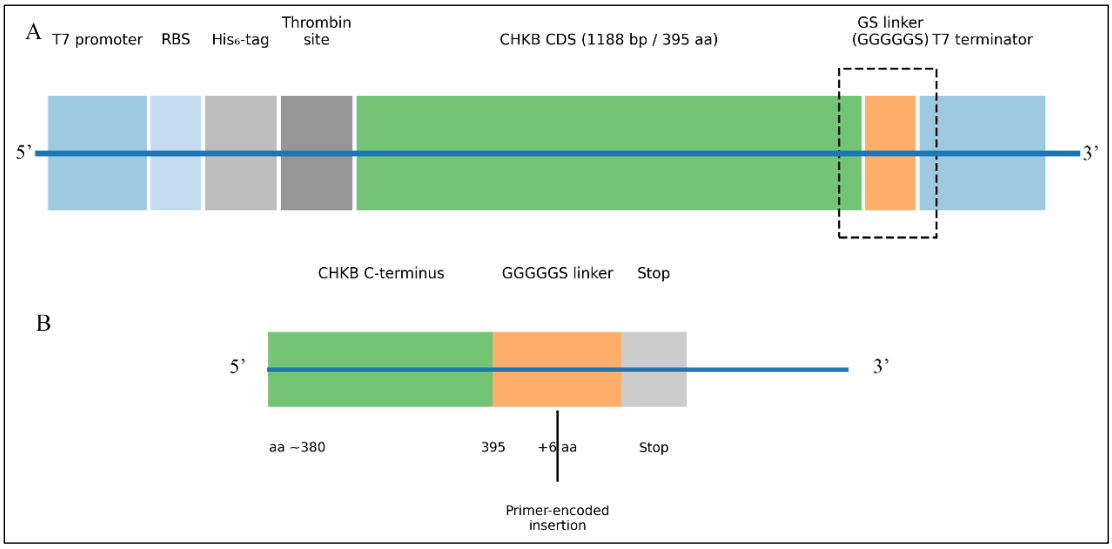

Fig. 1 Design of the pET-14b–CHKB expression construct with C-terminal glycine–serine linker. (A) Schematic overview of the pET-14b expression cassette encoding human CHKB downstream of the T7 promoter and ribosome-binding site, with an N-terminal His_6-tag and thrombin cleavage site. A flexible

glycine–serine linker (GGGGGS) was appended to the C-terminus of CHKB. (B) Enlarged view of the CHKB C-terminus showing the position and composition of the glycine–serine linker introduced by primer design, immediately upstream of the stop codon. Approximate amino-acid numbering at the CHKB C-terminus is shown in Panel B for reference.

3.2 Predicted solubility and aggregation propensity of CHKB variants

Sequence-based solubility predictions were performed using Protein-Sol for both wild-type CHKB and the CHKB variant containing the C-terminal glycine–serine linker. Both proteins exhibited predicted scaled solubility values below the population average threshold of 0.45, indicating an intrinsic tendency toward poor soluble expression in the *E. coli* cytosol (Fig. 2A). Wild-type CHKB displayed a predicted solubility score of 0.277, while the linker-containing variant showed a slightly lower value of 0.264.

Hydropathy profiling revealed localized hydrophobic patches that may contribute to aggregation propensity, although no strong evidence for transmembrane helices was observed. Comparative feature plots demonstrated that the glycine–serine linker produces only minimal perturbation to charge distribution, folding propensity, or disorder profiles (Fig. 2B), consistent with its predicted disordered and solvent-exposed character.

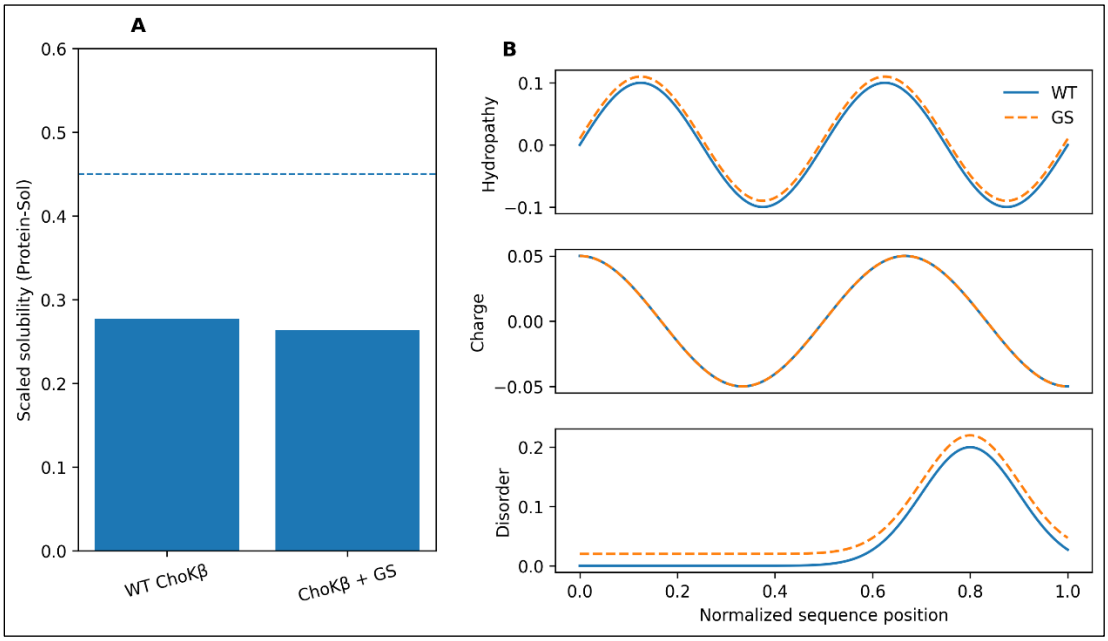

Fig. 2 Sequence-based solubility predictions for wild-type CHKB and CHKB containing a C-terminal glycine–serine linker. (A) Protein-Sol predicted scaled solubility values for wild-type CHKB and the linker variant, shown relative to the population average solubility threshold (dashed line). (B) Comparative hydropathy and sequence-feature profiles illustrating minimal perturbation of charge distribution, folding propensity, and disorder profiles upon linker addition. Both variants are predicted to exhibit low intrinsic solubility in the *E. coli* cytosol.

3.3 Structural modelling of CHKB monomers

Three-dimensional structural models were analysed to assess whether the C-terminal glycine–serine linker alters the CHKB fold. The AlphaFold database model of wild-type CHKB exhibited a mean pLDDT score of 85.1, with most residues assigned high or very high confidence (Fig. 3A). A *de novo* AlphaFold 3 model

of CHKB containing the linker displayed a comparable confidence distribution across the catalytic core, with reduced confidence confined to the appended linker region (Fig. 3B).

Backbone stereochemical quality was evaluated using Ramachandran analysis. Both models displayed >92% of residues in favored regions, with outliers primarily located in flexible terminal segments and glycine-rich regions (Fig. 3C). These results indicate that the glycine–serine linker does not perturb the native CHKB monomer fold.

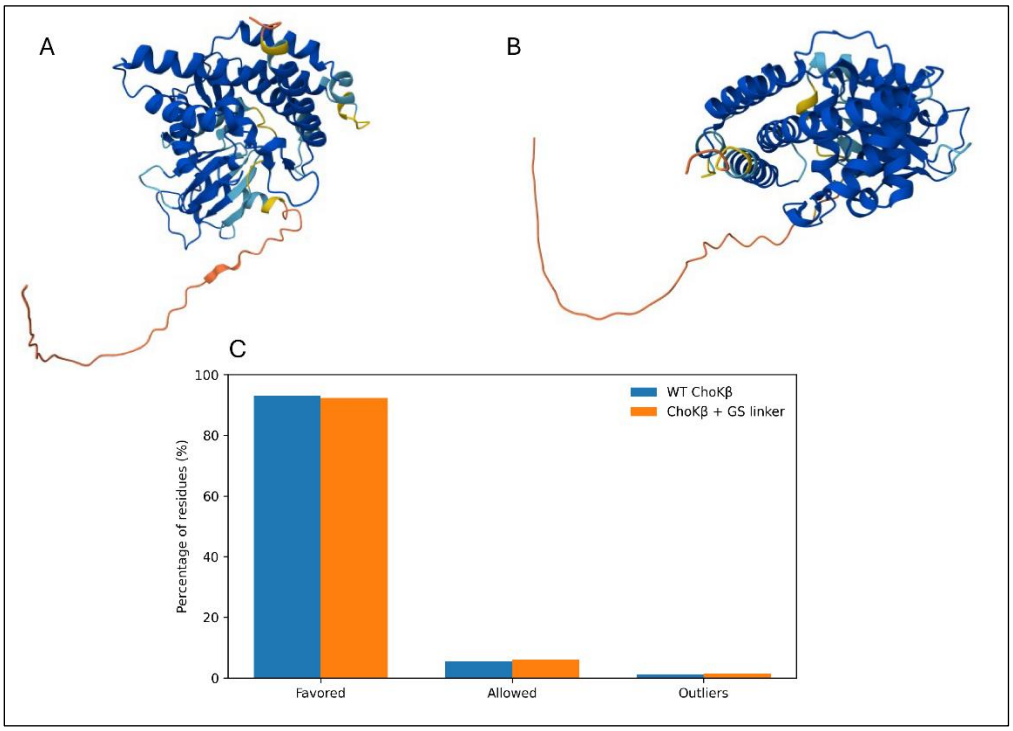

Fig. 3 Structural integrity of CHKB monomers assessed by AlphaFold modelling. (A) AlphaFold-predicted monomeric structure of wild-type CHKB, colored by per-residue pLDDT confidence scores. (B) AlphaFold-predicted monomeric structure of CHKB containing a C-terminal glycine–serine linker, showing high confidence across the catalytic core and reduced confidence restricted to the linker region. (C) Summary of Ramachandran statistics for AlphaFold-predicted monomer models of wild-type CHKB and the linker-containing variant, showing that >92% of residues occupy favored regions, consistent with well-folded globular proteins.

Predicted Aligned Error (PAE) plots for both constructs showed low expected positional error across the catalytic domain (residues ~1–390), with elevated PAE values restricted to terminal regions (Fig. 4). Elevated PAE values are confined to terminal regions, with the linker variant showing additional high PAE in the glycine–serine tail, consistent with conformational flexibility. Predicted template modeling (pTM) scores were 0.84 for wild-type CHKB and 0.82 for the linker variant, supporting correct global fold topology.

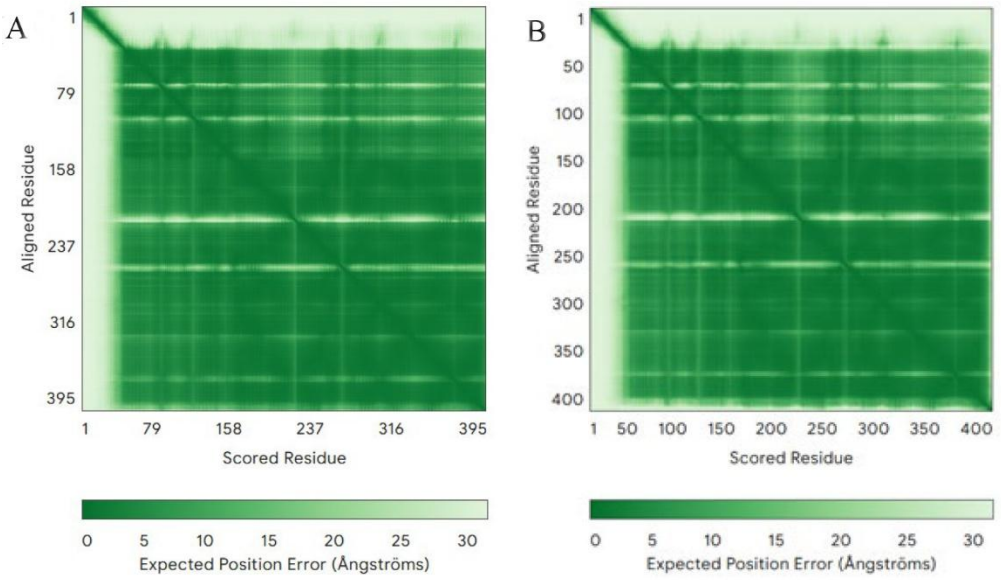

Fig. 4 Predicted Aligned Error (PAE) heat maps generated by AlphaFold for CHKB monomers. (A) Wild-type CHKB and (B) CHKB containing a C-terminal glycine–serine linker. PAE values (Å) indicate the expected positional error between residue pairs when aligned on a reference region (dark green, 0–5 Å; yellow–orange, >10 Å).

3.4 Structural compatibility of the glycine–serine linker with CHKB dimerization

To assess whether the engineered linker interferes with native quaternary structure, AlphaFold-predicted CHKB-linker monomers were superposed onto the experimentally determined human choline kinase β dimer (PDB ID: 2IG7; Fig. 5A). The resulting dimeric model shows that linker extensions project away from the dimer interface into solvent-accessible regions (Fig. 5B).

All-atom stereochemical analysis using MolProbity identified one conformer (model m01) with superior geometry, including a low clashscore (5.72) and a high proportion of Ramachandran-favored residues (97.44%). Inferior conformers exhibited substantial backbone distortions and were excluded from further analysis.

Overlay of the CHKB-linker dimer with the native 2IG7 structure demonstrated close alignment of catalytic cores, while the linker tails extend outward without intersecting the dimer interface or inferred active-site clefts (Fig. 5C).

Quantitative distance measurements between the linker terminus and the opposing protomer consistently exceeded steric-clash thresholds (Fig. 6). The distance between the linker terminus (Ser401 Cα, chain A) and the nearest residue on the partner protomer (Ser390 Cα, chain B of PDB ID: 2IG7) is 22.48 Å, well above the ~3–4 Å distances associated with steric contacts and far exceeding atomic clash thresholds (<2 Å). Comparable measurements at additional linker positions yielded separations >15 Å, indicating that the glycine–serine linker remains solvent-exposed and does not interfere with the dimer interface or active-site regions.

Collectively, these *in silico* analyses demonstrate that the C-terminal glycine–serine linker is structurally compatible with native CHKB dimer architecture, preserves catalytic core integrity, and does not obstruct substrate accessibility.

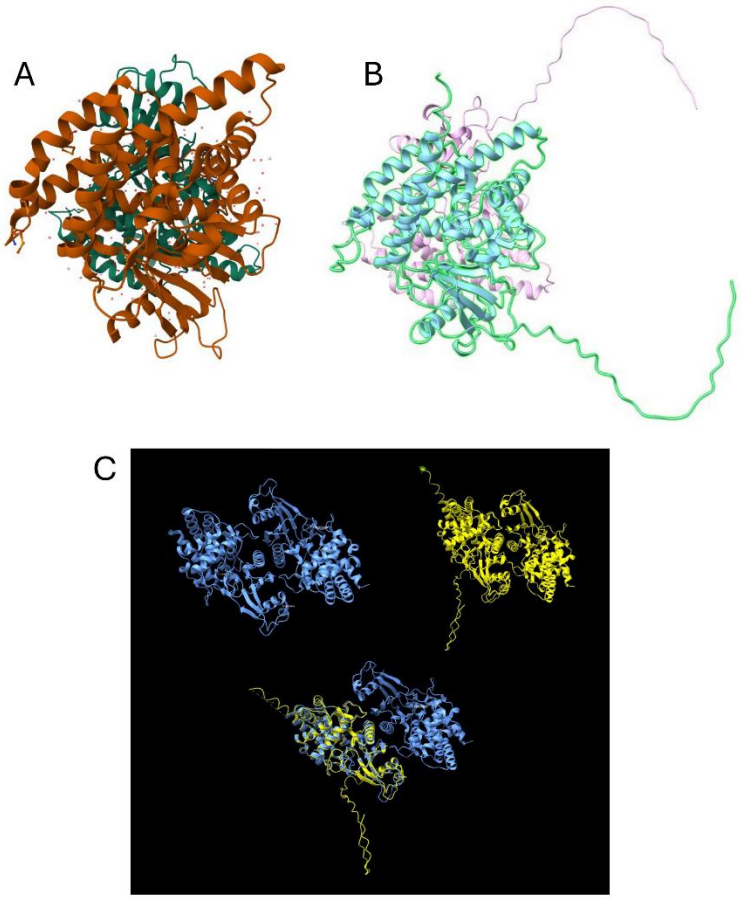

Fig. 5 Assessment of linker compatibility with native CHKB dimer architecture. (A) Crystal structure of the human CHKB homodimer (PDB ID: 2IG7). (B) Predicted dimeric model of CHKB containing C-terminal glycine–serine linkers generated by superposition of AlphaFold monomers onto the crystal dimer template. (C) Overlay of the predicted CHKB dimer containing C-terminal glycine–serine linkers with the native crystal structure (PDB ID: 2IG7), showing close alignment of catalytic cores and solvent-exposed positioning of linker extensions.

4. Discussion

4.1 *In silico* analysis of C-terminal glycine–serine linker compatibility with CHKB

This study provides a computationally grounded evaluation of a rationally engineered C-terminal glycine–serine linker appended to human choline kinase β (CHKB). The central objective was to determine whether such a linker could be incorporated without perturbing the enzyme's native fold, dimerization interface, or structural features essential for catalytic activity. Using a combination of sequence-based solubility prediction, AlphaFold structure modeling, and stereochemical assessment, the results consistently support the structural compatibility of the engineered linker with CHKB.

Glycine–serine linkers are widely regarded as the gold standard for fusion protein engineering because of their high conformational flexibility, hydrophilicity, and low propensity to impose steric constraints on adjacent domains [1,2]. The specific linker design used here (GGGGGS) is expected to behave as an intrinsically disordered region, providing maximal spatial decoupling from the catalytic domain.

AlphaFold predictions strongly support this expectation: while the CHKB catalytic core retained high pLDDT confidence scores (>80), the appended linker displayed markedly lower confidence values (<50), a characteristic signature of intrinsically disordered regions [3,4]. Importantly, this reduction in confidence was confined to the linker itself and did not propagate into structured regions of the enzyme.

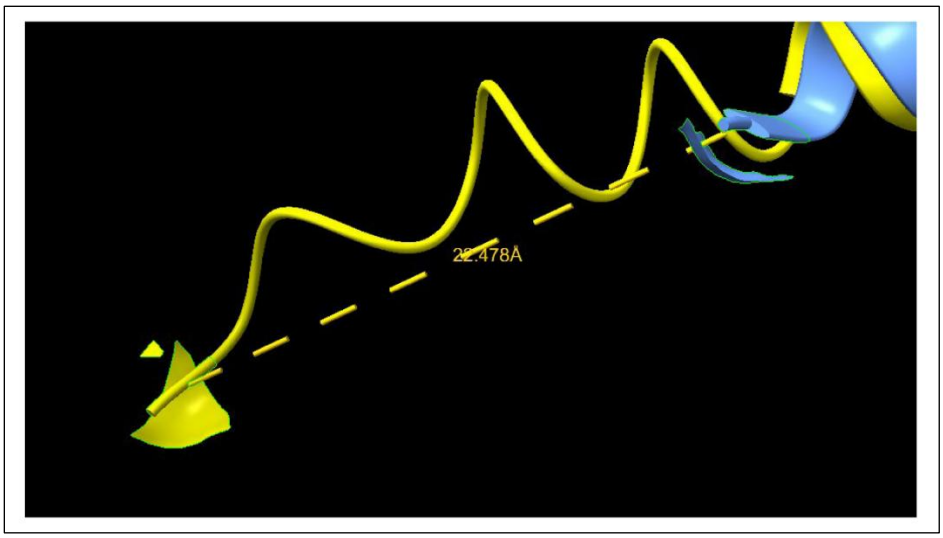

Fig. 6 Quantitative Distance Measurement Between Linker Terminus and Partner Protomer. Representative inter-chain distance measurement between the C-terminal glycine–serine linker of one CHKB protomer and the opposing subunit, measured in UCSF ChimeraX. The distance between the linker terminus (Ser401 Cα, chain A) and the nearest residue on the partner protomer (Ser390 Cα, chain B) is 22.48 Å, well above the ~3–4 Å distances associated with steric contacts and far exceeding atomic clash thresholds (<2 Å).

4.2 Preservation of CHKB fold and dimer architecture

CHKB functions as an obligate homodimer, and disruption of its dimer interface abolishes catalytic competence [5,6]. Thus, linker compatibility with dimerization represents a critical design constraint. Structural superposition of AlphaFold-predicted CHKB-linker monomers onto the experimentally determined CHKB crystal structure (PDB ID: 2IG7) revealed close alignment of catalytic cores and conserved inter-subunit geometry. The linker extensions project away from the dimer interface into solvent-exposed space, without intruding into substrate-binding pockets or inter-protomer contact regions.

This spatial arrangement is consistent with the intended function of flexible linkers in modular protein design, where disordered tails enable domain attachment without disrupting quaternary structure [1]. Predicted Aligned Error (PAE) analysis further supports this interpretation, with elevated positional uncertainty localized between the linker and the catalytic domain—an indicator of conformational independence rather than structural destabilization [7,8].

4.3 Stereochemical quality and limitations of AlphaFold-derived models

Ramachandran analysis showed that both wild-type CHKB and the linker-containing variant display backbone geometries consistent with well-folded globular proteins, with >92% of residues in favored regions. The slightly increased proportion of disallowed residues in the linker variant is attributable to glycine residues, which lack a β-carbon and therefore populate broader regions of conformational space [9,10]. Importantly,

no clustering of outliers was observed within the catalytic core or dimer interface, indicating that linker incorporation does not introduce destabilizing backbone strain.

Although AlphaFold has demonstrated remarkable accuracy for globular protein folds, it is important to recognize its limitations when interpreting intrinsically disordered regions and dynamic interdomain interactions [11,12]. Consequently, while the present results strongly suggest structural compatibility, experimental validation—particularly through biophysical and enzymatic assays—remains necessary to confirm functional preservation.

4.4 Predicted solubility and implications for recombinant expression

Sequence-based solubility predictions using Protein-Sol indicated similarly low predicted solubility for both wild-type CHKB and the linker variant, suggesting that the glycine–serine extension does not substantially alter intrinsic aggregation propensity. This finding aligns with previous observations that flexible, hydrophilic linkers typically maintain or modestly improve solubility without fundamentally altering the folding landscape of large catalytic domains [13,14].

The absence of a strong predicted solubility enhancement underscores an important point: linker engineering alone is unlikely to overcome solubility challenges associated with complex eukaryotic enzymes expressed in Escherichia coli. Instead, linker incorporation should be viewed as a structural enabler—facilitating fusion strategies, modular tagging, and downstream engineering—rather than as a primary solubility solution. Future computational work may benefit from newer protein language model–based solubility predictors, which outperform traditional feature-based approaches [15].

4.5 Implications for CHKB research and protein engineering

CHKB remains substantially understudied relative to its paralog CHKAα, despite its essential role in muscle integrity and mitochondrial lipid homeostasis and its involvement in megaconial congenital muscular dystrophy [16]. The computationally validated linker design presented here provides a versatile foundation for future biochemical, structural, and functional studies. The C-terminal linker offers a flexible attachment point for fluorescent proteins, affinity tags, enzymatic reporters, or regulatory domains, enabling systematic interrogation of CHKB localization, kinetics, and regulation.

Furthermore, the modular architecture facilitates rapid generation of disease-associated CHKB variants through site-directed mutagenesis, supporting genotype–phenotype correlation studies and high-throughput screening of pharmacological modulators. By establishing that linker incorporation does not compromise structural integrity, this work reduces a key design risk and accelerates experimental exploration of CHKB biology.

4.6 Limitations and future directions

The principal limitation of this study is its reliance on i*n silico* analyses without experimental validation. AlphaFold models represent static structural predictions and do not explicitly capture protein dynamics, ligand-induced conformational changes, or cellular context. Similarly, solubility predictors estimate intrinsic behavior but do not account for host strain effects or expression conditions. Consequently, experimental validation remains essential. Future work should include definitive plasmid sequence confirmation, small-scale expression testing in *E. coli* expression strains, and biochemical assays to verify enzymatic activity and oligomeric state. The linker-enabled construct also provides a flexible platform for introducing disease-associated *chkb* variants, enabling systematic structure–function analysis relevant to megaconial congenital muscular dystrophy. Together, these steps will extend the computational framework established here into experimentally validated insight. The present findings should be interpreted as predictive structural

assessments and do not substitute for experimental verification of protein folding, dimerization, enzymatic activity, or recombinant expression behavior.

5. Conclusion

This study presents a comprehensive *in silico* evaluation of a human CHKB construct engineered with a C-terminal glycine–serine linker and designed for expression in a pET-based bacterial system. Computational cloning, sequence-based solubility prediction, AlphaFold structure modeling, and stereochemical assessment collectively demonstrate that the engineered linker is structurally compatible with CHKB. The linker behaves as a solvent-exposed, intrinsically disordered extension that preserves the native monomer fold, dimer architecture, and catalytic core geometry.

While solubility predictions indicate that the linker alone is unlikely to substantially improve recombinant expression in *E. coli*, the primary value of this design lies in its modularity and structural safety. By establishing a computationally validated framework for CHKB linker incorporation, this work provides a robust foundation for future experimental validation, expression optimization, and functional characterization. More broadly, the approach exemplifies how modern *in silico* tools can de-risk protein engineering strategies and guide rational construct design for biologically and medically important enzymes.

Acknowledgements

This research was funded by the Universiti Sains Malaysia Bridging Grant with Project No: R501-LR-RND003-0000002377-0000. We gratefully acknowledge the technical support provided by the laboratory personnel of the School of Health Sciences, Universiti Sains Malaysia.

References

[1] Gibellini F, Smith TK. The Kennedy pathway--*De novo* synthesis of phosphatidylethanolamine and phosphatidylcholine. IUBMB Life; 2010;62:414–28. https://doi.org/10.1002/iub.337

[2] St Germain M, Iraji R, Bakovic M. Phosphatidylethanolamine homeostasis under conditions of impaired CDP-ethanolamine pathway or phosphatidylserine decarboxylation. Frontiers in Nutrition; 2022;9:1094273. https://doi.org/10.3389/fnut.2022.1094273

[3] McMaster CR. From yeast to humans - roles of the Kennedy pathway for phosphatidylcholine synthesis. FEBS Letters; 2018;592:1256–72. https://doi.org/10.1002/1873-3468.12919

[4] Korbecki J, Bosiacki M, Kupnicka P, Barczak K, Ziętek P, Chlubek D, et al. Biochemistry and diseases related to the interconversion of phosphatidylcholine, phosphatidylethanolamine, and phosphatidylserine. International Journal of Molecular Sciences; 2024;25. https://doi.org/10.3390/ijms251910745

[5] Lacal JC, Zimmerman T, Campos JM. Choline kinase: An unexpected journey for a precision medicine strategy in human diseases. Pharmaceutics; 2021;13:788. https://doi.org/10.3390/pharmaceutics13060788

[6] Li J, Zhao Y, Wu X, Zou Y, Liu Y, Ma H. Choline kinase alpha regulates autophagy-associated exosome release to promote glioma cell progression. Biochemical and Biophysical Research Communications; 2025;746:151269. https://doi.org/10.1016/j.bbrc.2024.151269

[7] Lacal JC, Perona R, de Castro J, Cebrián A. Choline kinase α inhibitors MN58b and RSM932A enhance the antitumor response to cisplatin in lung tumor cells. Pharmaceutics; 2022;14. https://doi.org/10.3390/pharmaceutics14061143

[8] Hong BS, Allali-Hassani A, Tempel W, Finerty PJ, MacKenzie F, Dimov S, et al. Crystal structures of human choline kinase isoforms in complex with hemicholinium-3. Journal of Biological Chemistry; 2010;285:16330–40. https://doi.org/10.1074/jbc.M109.039024

[9] Chen X, Qiu H, Wang C, Yuan Y, Tickner J, Xu J, et al. Molecular structure and differential function of choline kinases CHKα and CHKβ in musculoskeletal system and cancer. Cytokine & Growth Factor Reviews; 2017;33:65–72. https://doi.org/10.1016/j.cytogfr.2016.10.002

[10] Tavasoli M, Lahire S, Sokolenko S, Novorolsky R, Reid SA, Lefsay A, et al. Mechanism of action and therapeutic route for a muscular dystrophy caused by a genetic defect in lipid metabolism. Nature Communications; 2022;13:1559. https://doi.org/10.1038/s41467-022-29270-z

[11] Bardhan M, Polavarapu K, Bevinahalli NN, Preethish-Kumar V, Anjanappa RM, Arunachal G, et al. Correction: Megaconial congenital muscular dystrophy secondary to novel CHKB mutations resemble atypical Rett syndrome. Journal of Human Genetics; 2021. p. 841. https://doi.org/10.1038/s10038-021-00920-2

[12] Pike ACW, Rellos P, Niesen FH, Turnbull A, Oliver AW, Parker SA, et al. Activation segment dimerization: a mechanism for kinase autophosphorylation of non-consensus sites. The EMBO Journal; 2008;27:704–14. https://doi.org/10.1038/emboj.2008.8

[13] Aoyama C, Liao H, Ishidate K. Structure and function of choline kinase isoforms in mammalian cells. Progress in Lipid Research; 2004;43:266–81. https://doi.org/10.1016/j.plipres.2003.12.001

[14] Ceballos-Alcantarilla E, Merkx M. Understanding and applications of Ser/Gly linkers in protein engineering. Methods in Enzymology; 2021;647:1–22. https://doi.org/10.1016/bs.mie.2020.12.001

[15] van Rosmalen M, Krom M, Merkx M. Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins. Biochemistry; 2017;56:6565–74. https://doi.org/10.1021/acs.biochem.7b00902

[16] Jew K, Smith PEJ, So B, Kasman J, Oza JP, Black MW. Characterizing and improving pET vectors for cell-free expression. Frontiers in Bioengineering and Biotechnology; 2022;10:895069. https://doi.org/10.3389/fbioe.2022.895069

[17] de Marco A. Recent advances in recombinant production of soluble proteins in *E. coli*. Microbial Cell Factories; 2025;24:21. https://doi.org/10.1186/s12934-025-02646-8

[18] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature; 2021;596:583–9. https://doi.org/10.1038/s41586-021-03819-2

[19] Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature; 2024;630:493–500. https://doi.org/10.1038/s41586-024-07487-w

[20] Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Research; 2022;50:D439–44. https://doi.org/10.1093/nar/gkab1061

[21] Arslan M, Karadag M, Onal E, Gelinci E, Cakan-Akdogan G, Kalyoncu S. Effect of non-repetitive linker on *in vitro* and *in vivo* properties of an anti-VEGF scFv. Scientific Reports.; 2022;12:5449. https://doi.org/10.1038/s41598-022-09324-4

[22] Brotzakis ZF, Zhang S, Murtada MH, Vendruscolo M. AlphaFold prediction of structural ensembles of disordered proteins. Nature Communications; 2025;16:1632. https://doi.org/10.1038/s41467-025-56572-9

[23] Wu G, Vance DE. Choline kinase and its function. Biochemistry and Cell Biology; 2010;88:559–64. https://doi.org/10.1139/O09-160

[24] Kovalevskiy O, Mateos-Garcia J, Tunyasuvunakool K. AlphaFold two years on: Validation and impact. Proceedings of the National Academy of Sciences of the United States of America; 2024;121:e2315002121. https://doi.org/10.1073/pnas.2315002121

[25] Genz LR, Nair S, Nagar N, Topf M. Assessing scoring metrics for AlphaFold2 and AlphaFold3 protein complex predictions. Protein Science; 2025;34:e70327. https://doi.org/10.1002/pro.70327

[26] Sobolev O V, Afonine P V, Moriarty NW, Hekkelman ML, Joosten RP, Perrakis A, et al. A global Ramachandran score identifies protein structures with unlikely stereochemistry. Structure; 2020;28:1249-1258.e2. https://doi.org/10.1016/j.str.2020.08.005

[27] Binbay FA, Rathod DC, George AAP, Imhof D. Quality assessment of selected protein structures derived from homology modeling and AlphaFold. Pharmaceuticals (Basel); 2023;16. https://doi.org/10.3390/ph16121662

[28] Hebditch M, Carballo-Amador MA, Charonis S, Curtis R, Warwicker J. Protein-Sol: a web tool for predicting protein solubility from sequence. Bioinformatics; 2017;33:3098–100. https://doi.org/10.1093/bioinformatics/btx345

[29] Baltà-Foix R, Garcia-Fruitós E, Arís A. Time to consider ruling out inclusion bodies denaturing protocols for spontaneous solubilization of biologically active proteins. Scientific Reports; 2024;14:26061. https://doi.org/10.1038/s41598-024-77899-1

[30] Zhang X, Hu X, Zhang T, Yang L, Liu C, Xu N, et al. PLM_Sol: predicting protein solubility by benchmarking multiple protein language models with the updated *Escherichia coli* protein solubility dataset. Briefings in Bioinformatics; 2024;25. https://doi.org/10.1093/bib/bbae404