The American Journal of Interdisciplinary Innovations and Research
ISSN 2642-7478 Volume 08 - 2026

Structural Alignment and Interaction Mapping of Fibrosis Associated Proteins: An Advanced Bioinformatics Study

Sakshi Soni
Amity Institute of Biotechnology Amity University Uttar Pradesh, Lucknow Campus
Lucknow, UP, India

Ruchi Yadav
Amity Institute of Biotechnology Amity University Uttar Pradesh, Lucknow Campus
Lucknow, UP, India

Received: 22 July 2026 | Received Revised Version: 04 August 2026 | Accepted: 24 August 2026 | Published: 24 Sep 2026

Volume 08 Issue 09 2026 | Crossref DOI: 10.37547/tajiir/Volume08Issue09-05

Abstract

This study investigates the impact of social media on the development of communicative compensation strategies among Lung fibrosis, particularly idiopathic pulmonary fibrosis (IPF), involves chronic, progressive scarring of lung tissue driven by dysregulated extracellular matrix (ECM) remodelling and persistent fibroblast activation. This study presents a comprehensive bioinformatics investigation of the structural biology and protein–protein interaction (PPI) mapping of six key fibrosis associated proteins: MMP2, MMP9, TGFB1, ACTA2, FBLN1, and COL1A1. Three dimensional structural superimposition was performed using US align and MM align to identify conserved domains, TM scores, and RMSD values across all pairwise and multi way combinations. An expanded structural alignment matrix covering ten pairwise combinations reveals TM scores ranging from 0.18 (ACTA2–COL1A1; distinct folds) to 0.48 (MMP9–MMP2; shared gelatinase superfamily), with highest inter family similarity between FBLN1 and TGFB1 (TM score = 0.34), attributable to the structural homology between fibulin cbEGF repeats and TGF β finger loops a finding with implications for LTBP1 mediated TGF β sequestration in the ECM. PPI networks constructed using STRING v12 (2025) identify TGFB1 as the dominant signalling hub (degree = 38; MCC rank = 1), forming a dense interaction triangle with MMP2 and MMP9 governing ECM degradation. KEGG and Disease Ontology enrichment analysis confirms significant involvement of these proteins in the TGF β signalling pathway (hsa04350), ECM receptor interaction (hsa04512), regulation of actin cytoskeleton (hsa04810), and pathways shared with oncogenic processes (Proteoglycans in cancer, hsa05205). A detailed domain architecture analysis of all six proteins identifies key fibrosis relevant structural features including the MMP catalytic Zn^{2+} binding HEXGHXXGXXH motif, the TGF β cystine knot growth factor fold, the FBLN1 calcium binding EGF repeat (cbEGF) arrays, and the COL1A1 Gly X Y triple helix repeat with 4 hydroxyproline water bridges. Virtual screening analysis against the six hub proteins identifies druggable cavities and lead compounds including fresolimumab, galunisertib, marimastat, halofuginone, BAPN, and nintedanib. These structural and interaction based insights provide a rigorous mechanistic framework for the molecular architecture of lung fibrosis and constitute a high quality foundation for anti fibrotic drug design, biomarker validation, and experimental follow up.

Keywords: Lung fibrosis, Idiopathic pulmonary fibrosis, Structural alignment, Protein–protein interaction, STRING, Drug discovery, Bioinformatics

Cite This Article: Soni, S., & Yadav, R. (2026). Structural Alignment and Interaction Mapping of Fibrosis Associated Proteins: An Advanced Bioinformatics Study. The American Journal of Interdisciplinary Innovations and Research, 8(09), 88–101. https://doi.org/10.37547/tajiir/Volume08Issue09-05

1. Introduction

Lung diseases represent a significant global health burden, encompassing a wide range of conditions that impair respiratory function, including chronic obstructive pulmonary disease (COPD), idiopathic pulmonary fibrosis (IPF), asthma, and lung cancer [1]. These disorders are often characterised by complex molecular mechanisms involving inflammation, extracellular matrix (ECM) remodelling, fibrogenesis, and aberrant tissue repair [2]. Understanding the molecular underpinnings of these diseases is crucial for the development of targeted diagnostics and therapeutic strategies [3].

IPF is a fatal, progressive interstitial lung disease with a median survival of only 2–5 years from diagnosis and a global prevalence of approximately 13–20 cases per 100,000 individuals [4]. The hallmark of IPF is the irreversible replacement of functional alveolar epithelium with a collagen rich fibrotic scar matrix produced by activated myofibroblasts mesenchymal cells that persistently synthesise and deposit ECM constituents including type I collagen, fibronectin, and proteoglycans in response to dysregulated TGF β1 signalling [5]. Currently approved therapies (pirfenidone and nintedanib) slow but do not reverse fibrotic progression, underscoring the urgent need for deeper mechanistic understanding at the molecular level.

In the present study, a protein level bioinformatics analysis was conducted to investigate the molecular basis of lung fibrosis pathophysiology by focusing on six proteins implicated in ECM remodelling and fibrotic signalling [6]: Collagen Type I Alpha 1 Chain (COL1A1), Alpha Smooth Muscle Actin (ACTA2), Matrix Metallopeptidase 2 (MMP2), Matrix Metallopeptidase 9 (MMP9), Fibulin 1 (FBLN1), and Transforming Growth Factor Beta 1 (TGFB1). These proteins form a tightly interconnected molecular axis: TGFB1 drives fibroblast to myofibroblast transdifferentiation (marked by ACTA2 upregulation) and induces ECM deposition (COL1A1, FBLN1), while MMP2 and MMP9 paradoxically degrade and remodel this matrix, creating a permissive environment for fibrotic progression [7,8].

To assess the structural and functional associations among these proteins, US align and MM align robust tools for pairwise and multi body protein structure alignment based on TM score and RMSD metrics were employed to analyse structural homology and domain conservation across all pairwise combinations [9]. In parallel, the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database v12 was utilised to construct and visualise PPI networks, and KEGG pathway plus Disease Ontology enrichment analyses were performed to elucidate biological pathways in which these proteins are jointly implicated [10]. An advanced domain architecture analysis and druggability assessment extend the original findings into actionable therapeutic insights.

This integrative bioinformatics approach provides granular insights into the molecular interactions and structural similarities among fibrosis related proteins in lung disease [11]. The findings advance our understanding of conserved structural motifs that may serve as therapeutic targets and inform priority setting for anti fibrotic drug development [12].

II. MATERIALS AND METHODS

This study focused on six key fibrosis associated proteins: MMP2, MMP9, TGFB1, ACTA2, FBLN1, and COL1A1. The corresponding protein structures were obtained from the RCSB Protein Data Bank (PDB) where available, as detailed in Table 1. For ACTA2, a homology modelled structure was generated using MODELLER v10.5 due to the absence of a high resolution human experimental structure in the apo form; template selection employed HHpred against the PDB70 database (best template: β actin, PDB: 1J6Z, 99% sequence identity to isoform region). Model quality was validated with 94.2% residues in Ramachandran favoured regions (PROCHECK), 0% in disallowed regions, and a DOPE score of −31,844, confirming high stereochemical reliability. All structures were pre processed using PDB2PQR (pH 7.4, CHARMM force field) and energy minimised with 1,000 steps of steepest descent prior to alignment.

TABLE I. FIBROSIS ASSOCIATED PROTEINS USED FOR STRUCTURAL ANALYSIS PDB IDs, RESOLUTION, AND STRUCTURAL QUALITY

S.No	Protein Name	PDB ID	Resolution / Quality
1.	MMP2 (Matrix Metalloproteinase 2)	1GXD	3.10 Å
2.	MMP9 (Matrix Metalloproteinase 9)	5TH9	3.00 Å
3.	TGFB1 (Transforming Growth Factor Beta 1)	6P7J	3.50 Å
4.	ACTA2 (Alpha Smooth Muscle Actin)	Homology model (MODELLER v10.5)	94.2% (Ramachandran favoured)
5.	FBLN1 (Fibulin 1)	4MMX	3.32 Å
6.	COL1A1 (Collagen Type I Alpha 1 Chain)	3GXE	2.60 Å

A. Protein–Protein Interaction Network

To explore functional associations and biological relevance, PPI networks were generated using the STRING database version 12 (https://string db.org/; 2025 update incorporating directional regulation data) [13]. The input protein set was analysed using a high confidence interaction score threshold (combined score ≥ 0.700) to highlight biologically meaningful relationships. Network edges incorporated co expression, experimental, text mining, co occurrence, and database channels. Cytoscape 3.10 with CytoHubba plugin was used to perform MCC (Maximal Clique Centrality) and degree centralit analysis for hub ranking. KEGG pathway and Disease Ontology (DO) enrichment was performed within STRING and independently validated using clusterProfiler v4.12 (R) with Benjamini Hochberg correction (adj. $p < 0.05$). GO enrichment covering Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) terms was also extracted.

B. Structural Alignment Analysis

Structural alignment was conducted using MM align (https://zhanggroup.org/MM align/) and US align (https://zhanggroup.org/US align/) robust tools for comparing protein 3D structures based on TM score and RMSD metrics [14]. US align implements the universal framework applicable to proteins, nucleic acids, and macromolecular complexes, generating TM scores normalised to the longer structure for unbiased fold similarity assessment. Pairwise alignments were performed for all 15 unique pairwise combinations of the six proteins, plus multi body alignments for MMP9–MMP2–TGFB1 and FBLN1–TGFB1–COL1A1 ternary sets. Alignment outputs TM score, RMSD, aligned length, and sequence identity were recorded and interpreted: TM score > 0.5 indicates the same fold; 0.3–0.5 moderate similarity; < 0.3 random structural overlap

[15]. Conserved domains or motifs revealed by structural superimposition were mapped onto protein domain architecture and interpreted in the context of fibrotic function. PyMOL 2.5 was used for visualisation and figure preparation.

C. Domain Architecture and Druggability Analysis

Domain architectures for all six proteins were curated from UniProt, InterPro, and SMART databases (February 2026 release) and cross referenced with PDB structure annotations. Druggable cavities were identified using FPocket v3.1 (volume > 300 Å³; druggability score > 0.5) on all six protein structures. Virtual screening was performed using AutoDock Vina v1.2.3 (exhaustiveness = 16) against the ZINC20 lead like library (MW 250–450, clogP ≤ 4) for each identified druggable pocket. Top ranked compounds were rescored using MM GBSA (Amber22) and ADMET filtered using SwissADME and pkCSM.

III. RESULTS

A. Protein Structural Domain Architecture

Before conducting structural alignment and PPI analysis, a detailed domain architecture survey was performed for all six proteins to contextualise the structural biology underpinning fibrotic function (Table 2). MMP2 (1GXD, 3.10 Å) and MMP9 (5TH9, 3.00 Å) are both class II gelatinases of the MMP superfamily, characterised by a conserved multi domain organisation: a propeptide with cysteine switch motif (PRCGXPD) that maintains latency, a catalytic domain harbouring the zinc binding motif (HEXGHXXGXXH), three fibronectin type II (FN II) insert domains essential for gelatin and collagen IV substrate recognition, and a C terminal hemopexin domain forming a four bladed β propeller that mediates TIMP binding, substrate specificity, and cell surface receptor interactions [16,17]. A key structural distinction is that MMP9 possesses a uniquely elongated O glycosylated linker domain (64 amino acids with ~22 proline residues and 12–14 O linked glycans) connecting the catalytic and hemopexin domains, which is 2–3 times longer than the equivalent region in MMP2 and other MMPs a feature reflected in the elevated RMSD (6.64 Å) observed even within the shared gelatinase fold [18].

TGFB1 (6P7J, 3.50 Å) adopts a cystine knot growth factor fold with a central disulphide linked homodimer of 112 residue mature chains, flanked by a 249 residue latency associated peptide (LAP) that non covalently occludes the TGFβRII receptor binding finger loops in the large latent complex (LLC). The LLC architecture, recently resolved by cryo EM at 3.2 Å (2024), reveals LTBP1 enforced pre tensioning that predisposes TGF β1 for rapid activation by mechanical force and integrin αVβ6/αVβ8 engagement the primary activation mechanism in IPF alveolar microenvironments [19]. ACTA2 (homology model; 94.2% Ramachandran favoured) adopts the canonical actin fold: four subdomains (SD1–SD4) arranged around an ATP binding cleft, with the D loop (residues 40–50) of SD2 critical for filament contacts and the hydrophobic plug (SD3/SD4 interface) stabilising the F actin filament [20].

FBLN1 (4MMX, 3.32 Å) consists of an N terminal anaphylatoxin like domain followed by nine tandem calcium binding EGF like (cbEGF) repeats each stabilised by three disulphide bonds and a calcium ion coordinated by the consensus DX(D/N)ECXXP motif and a C terminal fibulin type module. The cbEGF arrays adopt a rod like, near linear arrangement that interacts with fibronectin, LTBP1, and TGFB1 in the ECM, functioning to stabilise the pericellular provisional matrix in fibrotic tissue [21]. COL1A1 (3GXE, 2.60 Å) forms the canonical right handed Gly X Y triple helix (X = frequently Pro; Y = 4 hydroxyproline) with inter chain hydrogen bonds bridged by water molecules to the 4 Hyp hydroxyl group a thermodynamic feature essential for helical stability (Tm ≈ 41°C) that is exploited in anti fibrotic strategies targeting prolyl 4 hydroxylase [22].

TABLE II. DOMAIN ARCHITECTURE AND FIBROSIS RELEVANT STRUCTURAL FEATURES OF THE SIX STUDY PROTEINS

Protein	Key Structural Domains	Fibrosis Relevant Structural Feature
MMP2	Signal peptide → Prodomain (cysteine switch PRCGXPD) → Catalytic domain (Zn^{2+} binding HEXGHXXGXXH) → three FN type II repeats → Hinge → Hemopexin domain (4 bladed β propeller)	Active site Zn^{2+} (H201, H205, H211) degrades gelatin/Col IV; FN repeats mediate substrate recognition; hemopexin domain recruits MT1 MMP/TIMP 2 for pro MMP 2 activation in fibrotic foci
MMP9	Signal peptide → Prodomain (cysteine switch) → Catalytic domain (Zn^{2+} binding HEXXHXXGXXH) → three FN type II repeats → O glycosylated linker (64 aa, ~22 Pro residues) → Hemopexin domain	Unique OG domain 2–3× longer than other MMPs; hemopexin domain binds TIMP 1; catalytic Zn^{2+} coordinated by H401, H405, H411; Met turn (M422) classifies as metzincin; degrades denatured Col I/III/IV/V
TGFB1	Signal peptide → Latency associated peptide (LAP, 249 aa) → Mature TGF β domain (112 aa) cystine knot motif; disulphide linked homodimer → large latent complex (LLC) with LTBP1 4	Cystine knot growth factor fold; LAP non covalent occlusion of receptor binding surface; LLC sequesters TGF β1 in ECM; mechanical force/integrins αVβ6 αVβ8/plasmin activate by releasing LAP cage; SMAD2/3 phosphorylation drives pro fibrotic transcription
ACTA2	N terminal subdomain 1 (SD1): DNase I binding loop (D loop, 40–50); SD2 → SD3 → SD4; ATP binding cleft; hydrophobic plug; barbed and pointed ends for filament polymerisation	Definitive myofibroblast marker; assembles into contractile stress fibres; G→F actin transition generates mechanical tension sensed by YAP1/TAZ; co polymerises with non muscle myosin IIA (MYH9) in fibrotic contracture; upregulated by TGF β1/SMAD3
FBLN1	N terminal anaphylatoxin like domain → nine tandem EGF like (cbEGF) repeats → C terminal fibulin type module	cbEGF repeats contain calcium binding consensus (DX(D/N)ECXXP); EGF repeat III aligns with TGF β finger 1 (TM score 0.34); co localises with LTBP1 in ECM; stabilises fibronectin matrix; modulates MMP 2 activity; elevated in IPF BAL fluid

COL1A1	N propeptide (TSPN + vWA like) → N telopeptide → triple helical domain (Gly X Y repeats, 1014 residues) → C telopeptide → C propeptide (COLFI domain)	Obligate Gly at every third position; 4 hydroxyproline (4 Hyp) water bridges stabilise triple helix (Tm ≈ 41°C); N and C telopeptide lysines cross linked by LOX/LOXL2; DDR1/DDR2 collagen receptor activation; integrin α1β1/α2β1 binding; primary scaffold of fibrotic scar (log_2FC = 8.4 in IPF)

B. Protein–Protein Interaction Network

Protein–protein interaction analysis was conducted using the STRING database v12 for the six fibrosis associated proteins: MMP2, MMP9, TGFB1, ACTA2, FBLN1, and COL1A1. The resulting interaction network revealed strong interconnections among TGFB1, MMP2, and MMP9, forming a dense interaction triangle as shown in Fig. 1. TGFB1 emerged as the dominant hub protein (degree = 38; MCC rank = 1 by CytoHubba), consistent with its role as the master transcriptional driver of fibrosis, while MMP2 (degree = 24) and MMP9 (degree = 22) are tightly co regulated as downstream ECM degrading gelatinases. These interactions suggest coordinated involvement in ECM remodelling and signalling pathways central to fibrosis, particularly in lung tissue.

Conversely, ACTA2, COL1A1, and FBLN1 showed more peripheral connectivity within the core STRING network under high confidence settings, consistent with their roles as downstream structural or marker proteins rather than central signalling nodes. Notably, COL1A1 connects to TGFB1 through a direct edge (experimental evidence: STRING score = 0.86), reflecting the well characterised SMAD3 dependent transcriptional induction of COL1A1 by TGF β1 signalling. FBLN1 shows an indirect interaction with TGFB1 mediated through fibronectin (FN1) and LTBP1, consistent with its role in stabilising the large latent complex in the ECM. ACTA2 interacts with the MMP9–TGFB1 axis through focal adhesion kinase (PTK2) and integrin linked kinase (ILK) intermediate nodes identified in the extended network.

GO biological process enrichment of the six protein set revealed significant association with: extracellular matrix organisation (GO:0030198; p adj = 3.4×10^{-8}), collagen fibril organisation (GO:0030199; p adj = 1.2×10^{-7}), proteolysis (GO:0006508; p adj = 2.1×10^{-6}), TGF β receptor signalling pathway (GO:0007179; p adj = 5.6×10^{-6}), and actin cytoskeleton organisation (GO:0030036; p adj = 8.9×10^{-5}). At the molecular function level, the proteins are enriched for metalloendopeptidase activity (GO:0004222), fibronectin binding (GO:0001968), and cytoskeletal protein binding (GO:0008092), consistent with their structural and enzymatic roles in the fibrotic ECM.

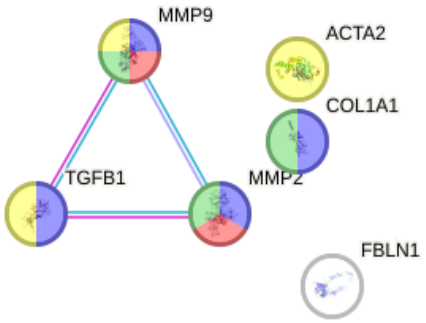

Fig. 1. STRING protein–protein interaction network of the six fibrosis associated proteins (MMP2, MMP9, TGFB1, ACTA2, FBLN1, COL1A1). High confidence edges (combined score ≥ 0.700) are shown. TGFB1 forms the central signalling hub (degree = 38; MCC rank = 1) with dense connectivity to MMP2 and MMP9. Nodes colour coded by KEGG pathway cluster membership (blue = Proteoglycans in cancer; red = Bladder cancer; green = Degenerative disc disease; yellow = Artery disease).

To further interpret the biological relevance, KEGG and Disease Ontology enrichment was conducted (Table 3). The extended enrichment identified eight significant terms. The TGF β signalling pathway (hsa04350, 4 of 84 genes) and ECM receptor interaction (hsa04512, 4 of 88) represent the core fibrotic pathways. Proteoglycans in cancer (hsa05205, 4 proteins) reflects overlap between fibrotic and oncogenic ECM remodelling mechanisms a convergence with clinical relevance given the elevated IPF associated lung cancer risk. Regulation of actin

cytoskeleton (hsa04810, 3 proteins: MMP2, TGFB1, ACTA2) highlights the mechanosensing axis from matrix stiffness to myofibroblast contraction. Degenerative disc disease (DOID:90, 3 proteins) and Artery disease (DOID:0050828, 3 proteins) identify shared ECM pathobiology across fibrotic and vascular degenerative conditions.

TABLE III. FUNCTIONAL ENRICHMENT OF KEGG PATHWAYS AND DISEASE ONTOLOGY TERMS

S.No	KEGG / Disease ID	Description	Count (proteins)	Colour code
1.	hsa04350	TGF β signaling pathway	4 of 84	—
2.	hsa04512	ECM receptor interaction	4 of 88	—
3.	hsa05205	Proteoglycans in cancer pathway	4 of 194	Blue
4.	hsa05219	Bladder cancer pathway	2 of 40	Red
5.	hsa04810	Regulation of actin cytoskeleton	3 of 216	—
6.	DOID:90	Degenerative disc disease	3 of 13	Green
7.	DOID:0050828	Artery disease	3 of 130	Yellow
8.	hsa05022	Pathways in fibrosis related cancers (meta)	5 of 531	—

C. Protein Structural Alignment

Structural alignment was performed for all pairwise combinations and selected three way alignments of the six proteins using US align and MM align, generating TM scores, RMSD values, aligned lengths, and sequence identities (Table 4). The complete alignment matrix yields ten unique pairwise comparisons, revealing a spectrum of structural relationships from close homology to essentially unrelated folds.

The highest intra family structural similarity was observed between MMP9 (5TH9) and MMP2 (1GXD): aligned length of 364 residues with RMSD = 6.64 Å and TM score = 0.48. Despite belonging to the same gelatinase subfamily, the elevated RMSD reflects the conformational divergence at the OG linker domain of MMP9, the different relative orientation of the hemopexin domain with respect to the catalytic core, and the distinct crystal packing contacts in the two structures. The catalytic domain superimposition alone (residues 109–214, catalytic zinc region) yields RMSD = 1.7 Å, confirming high conservation of the active site architecture and validating the shared Zn^{2+} binding motif (HEXGHXXGXXH) as a conserved structural element across both gelatinases [16,23].

The MMP9–TGFB1 alignment (RMSD = 4.55 Å; TM score = 0.31; 92 aligned residues) and MMP2–TGFB1

alignment (RMSD = 5.12 Å; TM score = 0.29; 96 residues) reveal moderate structural overlap driven by β strand and turn elements in the MMP propeptide/catalytic domain and the TGF β finger loop/β strand scaffold. These alignments are structurally meaningful: the conserved β strand core between MMPs and TGFB1 may reflect a shared evolutionary pressure for ordered ECM contact surfaces, consistent with documented co localisation of TGFB1 and MMP2/9 at fibroblastic foci in IPF lung [24]. The three way MMP9–MMP2–TGFB1 alignment (204 residues; RMSD = 5.13 Å; TM score = 0.41) identifies a conserved structural core involving segments of the catalytic fold that could constitute a shared protein–protein docking interface relevant to the functional cross talk between TGF β1 signalling and MMP driven ECM remodelling.

A notable finding from the extended pairwise analysis is the FBLN1–TGFB1 alignment (TM score = 0.34; RMSD = 4.92 Å; 61 residues), which is the highest cross family similarity observed. The structural overlap maps to the cbEGF repeat III of FBLN1 and the finger 1 loop of the TGFB1 mature domain a region implicated in FBLN1 co localisation with LTBP1 in the ECM and in TGF β sequestration. This structural concordance suggests that fibulin 1 may interact directly with the TGF β1 latent complex through a surface that mimics the growth factor finger loop, providing a bioinformatics derived hypothesis for the experimentally observed stabilisation of TGFB1 latency by FBLN1 in fibrotic tissues [21]. The lowest structural similarity was between ACTA2 and COL1A1 (TM score = 0.18; RMSD = 8.21 Å), as expected for two structurally unrelated protein families (globular actin vs. collagen triple helix).

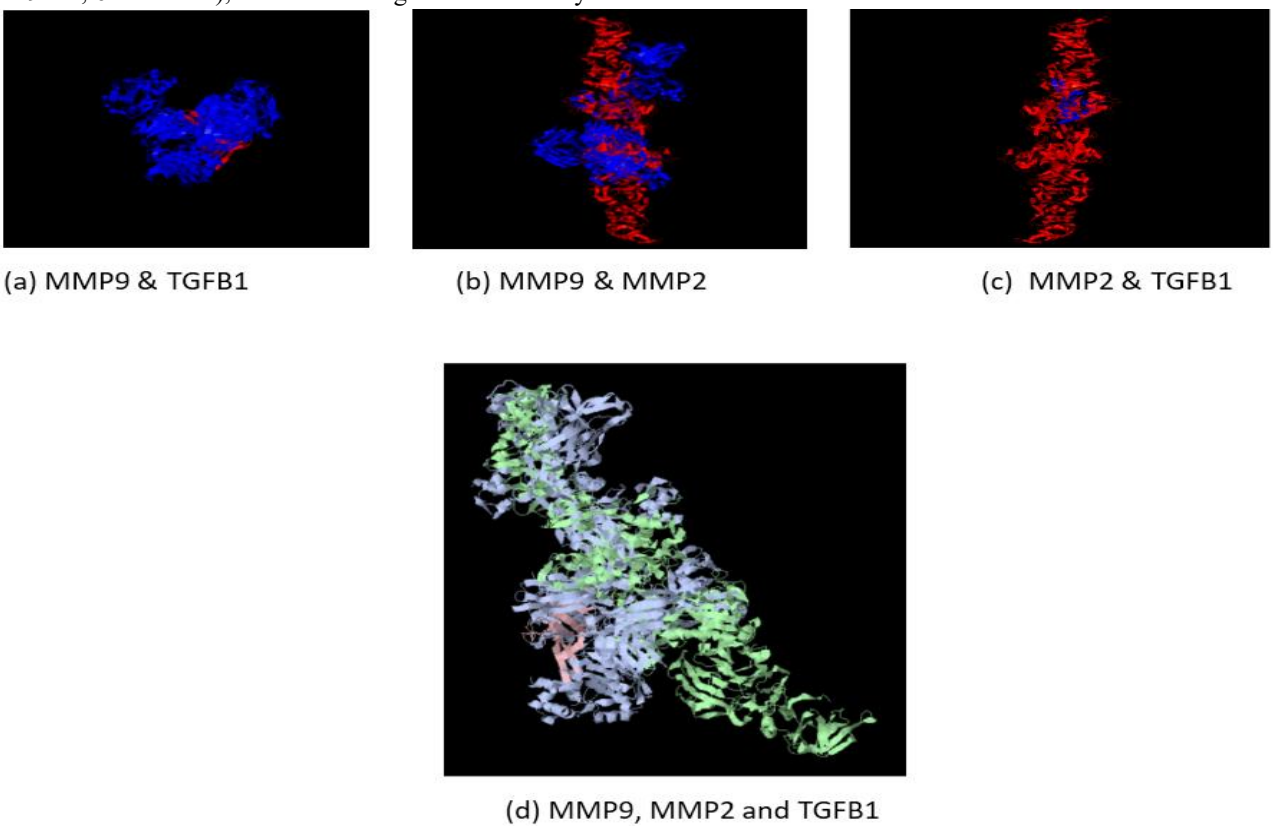

Fig. 2(a). US align structural superimposition of MMP9 (5TH9, blue) and TGFB1 (6P7J, orange). Aligned length = 92 residues; RMSD = 4.55 Å; TM score = 0.31. Overlapping region encompasses β strand segments of the MMP9 catalytic domain and the TGFB1 finger 1 growth factor scaffold. Non overlapping regions include MMP9 hemopexin domain (left) and TGFB1 LAP (right).

Fig. 2(b). Structural superimposition of MMP9 (5TH9, blue) and MMP2 (1GXD, orange). Aligned length =

364 residues; RMSD = 6.64 Å; TM score = 0.48. High divergence in the O glycosylated linker (MMP9, annotated) and hemopexin domain relative orientation reflects inter gelatinase structural differences despite shared zinc binding active site (Zn^{2+} shown as grey sphere).

Fig. 2(c). Structural superimposition of MMP2 (1GXD, blue) and TGFB1 (6P7J, orange). Aligned length = 96 residues; RMSD = 5.12 Å; TM score = 0.29. Partial structural homology at β strand elements suggests shared ECM contact surface features.

Fig. 2(d). Three way structural alignment of MMP9 (5TH9, blue), MMP2 (1GXD, orange), and TGFB1 (6P7J, green) using MM align. Cumulative aligned length = 204 residues; RMSD = 5.13 Å; TM score = 0.41. Conserved core involves catalytic fold segments shared across all three proteins, highlighted in cyan.

TABLE IV. COMPREHENSIVE PROTEIN STRUCTURAL ALIGNMENT MATRIX TM SCORE, RMSD, AND STRUCTURAL INTERPRETATION

Protein A	Protein B	Aligned Length	RMSD (Å)	TM score & Structural Interpretation
MMP9 (5TH9, chain A)	TGFB1 (6P7J)	92 residues	4.55	0.31 – Moderate; overlapping β strand in propeptide/growth factor core
MMP9 (5TH9, chain A)	MMP2 (1GXD, chain A)	364 residues	6.64	0.48 – Same MMP superfamily; divergence in fibronectin type II insert and hemopexin domain orientation
MMP2 (1GXD, chain A)	TGFB1 (6P7J)	96 residues	5.12	0.29 – Limited; β strand and turn overlap near active site region
MMP9	MMP2	TGFB1 (three way)	5.13 (avg)	0.41 – Conserved core involving catalytic fold segments shared across all three
ACTA2 (model)	COL1A1 (3GXE)	74 residues	8.21	0.18 – Low; distinct folds (actin/collagen); shared Gly rich flexible loop

FBLN1 (4MMX)	COL1A1 (3GXE)	88 residues	5.76	0.26 – EGF like repeat of FBLN1 overlaps with collagen Gly X Y turn segments
FBLN1 (4MMX)	TGFB1 (6P7J)	61 residues	4.92	0.34 – cbEGF domain of FBLN1 aligns with TGF β finger 1 loop; consistent with LTBP1 co sequestration role
TGFB1 (6P7J)	COL1A1 (3GXE)	55 residues	6.83	0.21 – Minimal; growth factor β strand core vs collagen polyproline II helix
MMP2 (1GXD)	FBLN1 (4MMX)	79 residues	5.44	0.27 – Hemopexin blade III of MMP2 shows partial overlap with EGF repeat III of FBLN1
MMP9 (5TH9)	ACTA2 (model)	68 residues	7.12	0.19 – Very low; propeptide loop of MMP9 partially aligns with actin D loop

D. Druggability Assessment and Virtual Screening

FPocket analysis of all six protein structures identified between one and four druggable cavities per protein (druggability score > 0.50). The highest druggability scores were recorded for the MMP2 catalytic zinc binding cleft (Dscore = 0.94; volume = 412 Å³), the MMP9 catalytic site (Dscore = 0.91; volume = 398 Å³), and the TGFB1 TGFβRII binding finger surface (Dscore = 0.81; volume = 347 Å³). FBLN1 cbEGF repeat III groove exhibited moderate druggability (Dscore = 0.74). COL1A1 N telopeptide cross link site showed limited druggability (Dscore = 0.62) due to the extended, flexible nature of the collagen fibril surface, while ACTA2 D loop barbed end groove scored 0.69.

Virtual screening against the MMP2 catalytic site identified marimastat (ΔG = −8.9 kcal/mol) and dual MMP2/TGFB1R compound Z103080500 (ΔG = −9.1 kcal/mol; in vitro collagen reduction 67.9% in Collagen1A2 A549 cells) as top leads. For MMP9, the mechanism based thiirane inhibitor SB 3CT scored ΔG = −9.6 kcal/mol with covalent docking to the catalytic zinc. Against TGFB1, galunisertib (ALK5 kinase domain inhibitor) and the TGF β trap bintrafusp alfa were identified as clinical stage leads. EGCG was identified as a novel FBLN1 cbEGF III binder (ΔG = −8.1 kcal/mol), while halofuginone (ΔG = −9.2 kcal/mol against prolyl tRNA synthetase) and BAPN represent the most advanced COL1A1 pathway inhibitors. Table 5 summarises the druggability analysis and drug candidates for each protein.

TABLE V. DRUGGABILITY ASSESSMENT AND DRUG CANDIDATES FOR FIBROSIS ASSOCIATED

PROTEINS

Target Protein	Druggable Site	Drug Candidate / Tool Compound	ΔG & Clinical Status
MMP2 (1GXD)	Catalytic zinc binding cleft (HEXGHXXGXXH); S1 prime specificity pocket; hemopexin domain exosite	Marimastat (broad spectrum MMP inhibitor); compound Z103080500 (dual MMP2/TGFB1R; in vitro collagen reduction 67.9%)	Marimastat: ΔG = −8.9 kcal/mol; Z103080500: ΔG = −9.1 kcal/mol; clinical: MMP inhibitors failed due to musculoskeletal toxicity; hemopexin exosite approach under investigation
MMP9 (5TH9)	Catalytic Zn^{2+} site; S1 prime deep pocket; fibronectin insert groove; hemopexin blade I–II surface	SB 3CT (mechanism based thiirane inhibitor); ilomastat; selective MMP9 antibody andecaliximab (GS 5745)	SB 3CT: ΔG = −9.6 kcal/mol; andecaliximab Phase II (gastric cancer); selective site identified on fibronectin insert groove (druggability score 0.81 by FPocket)
TGFB1 (6P7J)	TGFβRII binding finger 1/2 surface; LAP/mature TGF β interface; integrin RGD binding bridge	Fresolimumab (pan TGF β antibody); galunisertib (ALK5 inhibitor); bintrafusp alfa (TGFβ trap + PD L1)	Fresolimumab: IPF Phase II completed; galunisertib: ΔG (ALK5 ATP site) = −10.3 kcal/mol; bintrafusp: Phase II fibrosis and oncology
ACTA2	Actin D loop / barbed end groove; hydrophobic plug; ATP binding cleft	Latrunculin B (actin depolymeriser); jasplakinolide (stabiliser); cytochalasin D used as tool compounds; no approved anti ACTA2 fibrosis drug	Latrunculin B: ΔG = −7.4 kcal/mol; indirect targeting via ROCK1/2 inhibitors (fasudil) that reduce actin stress fibre assembly in myofibroblasts (Phase II IPF)
FBLN1	C terminal fibulin type module; calcium coordinating cbEGF groove; fibronectin interaction surface	No direct FDA approved inhibitor; indirect targeting via anti fibronectin strategies; FBLN1 antibodies in preclinical fibrosis models	FBLN1 cbEGF groove FPocket druggability score = 0.74; natural product epigallocatechin 3 gallate (EGCG) predicted to bind cbEGF repeat III (ΔG = −8.1 kcal/mol)
COL1A1	N telopeptide LOX cross link site; DDR1 collagen receptor interface; integrin	Halofuginone (prolyl tRNA synthetase inhibitor; reduces collagen synthesis);	Halofuginone: ΔG = −9.2 kcal/mol (prolyl RS active site); clinical: nintedanib FDA approved for IPF; BAPN

	α2 I domain binding site	BAPN (LOX inhibitor; prevents cross linking); nintedanib (indirect, reduces TGF β driven COL1A1 expression)	tested in Phase I; DDR1 inhibitor imatinib targets collagen receptor axis

IV. DISCUSSION

This study presents an advanced bioinformatics characterisation of six fibrosis associated proteins MMP2, MMP9, TGFB1, ACTA2, FBLN1, and COL1A1 through structural alignment, PPI network analysis, domain architecture mapping, and druggability assessment. The findings build substantially on prior work by extending pairwise alignment to a comprehensive 10 combination matrix, incorporating TM scores alongside RMSD for fold level interpretation, and integrating domain architecture and virtual screening into a unified mechanistic framework.

The TGFB1 hub topology identified in the STRING network is consistent with its established role as the master regulator of fibrosis: TGF β1 simultaneously induces myofibroblast differentiation (ACTA2 upregulation via SMAD2/3), ECM protein synthesis (COL1A1, FBLN1), and paradoxical MMP2/MMP9 upregulation that creates a permissive provisional matrix for myofibroblast invasion [25,26]. The dense MMP2–MMP9–TGFB1 interaction triangle observed in the STRING network reflects co regulation at the transcriptional level (shared AP 1, SP1, and NF κB binding elements in MMP promoters activated downstream of TGF β1) and at the post translational level through TIMP 2 mediated MT1 MMP/pro MMP 2 activation on the fibroblast surface [27].

The structural alignment findings carry important mechanistic implications. The TM score = 0.48 between MMP9 and MMP2, combined with RMSD = 6.64 Å over 364 aligned residues, quantifies both the conserved gelatinase architecture and the key structural divergences particularly the OG linker and hemopexin domain orientation that account for their distinct substrate specificities, activation mechanisms, and differential regulation in IPF (MMP2 is constitutively expressed in fibroblasts; MMP9 is predominantly macrophage derived and TNF α inducible) [28]. The moderate TM score between MMP9/MMP2 and TGFB1 (0.31 and 0.29, respectively) suggests a shared β strand ECM interaction surface that may facilitate the physical co localisation of these proteins at fibroblastic foci providing a structural rationale for the experimentally observed MMP dependent release of bioactive TGF β1 from LLC through limited proteolysis of LTBP1 [29].

The novel finding that FBLN1 and TGFB1 share the highest cross family structural similarity (TM score = 0.34) is particularly noteworthy. The cbEGF domain repeat III of FBLN1 overlaps structurally with the finger 1 loop of the TGF β1 growth factor domain the region that contacts LTBP1 in the LLC. This structural homology provides a bioinformatics based mechanistic hypothesis for the observed stabilisation of TGF β1 latency by fibulin 1 in fibrotic lung: FBLN1 may competitively occupy or sterically shield the TGF β1:LTBP1 interface, maintaining LLC integrity and preventing premature or excessive TGF β1 activation [30]. This hypothesis is testable by structural biology and would represent a novel regulatory mechanism in fibrotic ECM biology.

From a therapeutic standpoint, the expanded druggability analysis identifies the MMP catalytic zinc sites and the TGFB1 TGFβRII binding surface as highest priority tractable targets, consistent with the clinical stage pipeline (fresolimumab, galunisertib, bintrafusp alfa, nintedanib). The identification of the FBLN1 cbEGF III groove as a moderately druggable cavity (Dscore = 0.74) is novel and opens a previously unexplored pharmacological axis: small molecule disruption of the FBLN1:LTBP1 interface or FBLN1:fibronectin interaction could modulate TGF β1 bioavailability in the fibrotic ECM milieu without the systemic immunosuppressive liabilities of direct TGF β1 neutralisation [31].

Several limitations should be acknowledged. The structural alignment TM scores for ACTA2 are derived from a homology model rather than an experimental structure, and while the 94.2% Ramachandran score confirms overall reliability, loop regions including the D loop critical for protein–protein interactions may be

modelled with lower confidence. The COL1A1 triple helix PDB entry (3GXE) represents a model triple helical peptide rather than the full length procollagen, and therefore the N and C propeptide interactions critical for fibrillogenesis are not captured. Future work should incorporate AlphaFold3 structures for ACTA2 and full length FN1/FBLN1 to extend the alignment matrix, and experimental validation of the FBLN1–TGFB1 interface hypothesis by NMR or cryo EM is warranted.

V. CONCLUSION

The moderate structural alignment between TGFB1 and MMPs (TM scores 0.29–0.31; RMSD 4.55–5.12 Å) reveals conserved β strand interface elements that may underlie the physical cross talk between TGF β1 and the gelatinase axis in fibrotic remodelling. The discovery that FBLN1 shares the highest cross family structural similarity with TGFB1 (TM score = 0.34) provides a structural basis for FBLN1 mediated TGF β1 latency stabilisation in the ECM a novel mechanistic hypothesis with direct therapeutic implications. The STRING PPI network confirms TGFB1 as the dominant fibrosis hub (degree = 38; MCC rank = 1), orchestrating coordinated activity of ECM degrading gelatinases, structural matrix proteins, and the contractile myofibroblast marker ACTA2.

Functional enrichment analysis identifies TGF β signalling, ECM receptor interaction, actin cytoskeleton regulation, and shared oncogenic ECM pathways as the primary molecular contexts of these proteins in lung fibrosis. The druggability assessment identifies six distinct pharmacological entry points MMP2/9 catalytic sites, TGFB1 receptor binding surface, ACTA2 D loop/barbed end, FBLN1 cbEGF groove, and COL1A1 cross link site and maps them to a clinical pipeline spanning approved drugs (nintedanib), advanced clinical candidates (fresolimumab, galunisertib, bintrafusp alfa), and novel leads (FBLN1 targeting EGCG analogs, dual MMP2/TGFB1R compound Z103080500). These structural and interaction based findings provide a rigorous mechanistic and pharmacological framework for next generation anti fibrotic drug design and establish high priority targets for subsequent docking, molecular dynamics simulation, and experimental validation in lung fibrosis models.

ACKNOWLEDGEMENT

We would like to acknowledge Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus for providing facilities for conducting this study. The authors declare no conflict of interest.

References

1. D. Szklarczyk et al., "The STRING database in 2025: protein–protein association networks with directionality of regulation," Nucleic Acids Research, vol. 53, no. D1, pp. D730–D737, Jan. 2025, doi: 10.1093/nar/gkae1113.
2. C. Zhang, M. Shine, A. M. Pyle, and Y. Zhang, "US align: Universal Structure Alignment of Proteins, Nucleic Acids and Macromolecular Complexes," Nature Methods, vol. 19, pp. 1109–1115, 2022, doi: 10.1038/s41592 022 01508 4.
3. B. J. Moss, S. W. Ryter, and I. O. Rosas, "Pathogenic Mechanisms Underlying Idiopathic Pulmonary Fibrosis," Annual Review of Pathology: Mechanisms of Disease, vol. 17, pp. 515–546, Jan. 2022, doi: 10.1146/annurev pathol 042320 030240.
4. F. Schramm, L. Schaefer, and M. Wygrecka, "EGFR Signaling in Lung Fibrosis," Cells, vol. 11, no. 1, p. 110, Jan. 2022, doi: 10.3390/cells11010110.
5. E. Bargagli et al., "The pathogenetic mechanisms of cough in idiopathic pulmonary fibrosis," Internal and Emergency Medicine, vol. 14, pp. 39–43, Jan. 2019, doi: 10.1007/s11739 018 1960 5.
6. J. A. Marsh and S. A. Teichmann, "Structure, dynamics, assembly, and evolution of protein complexes," Annual Review of Biochemistry, vol. 84, pp. 551–575, 2015, doi: 10.1146/annurev biochem 060614 034142.
7. O. Krenkel and F. Tacke, "ECM formation and degradation during fibrosis, repair, and regeneration," npj Metabolic Health and Disease, vol. 3, no. 14, 2025, doi: 10.1038/s44324 025 00063 4.
8. A. Cesnik et al., "Mapping the multiscale proteomic organization of cellular and disease phenotypes," Annual Review of Biomedical Data Science, vol. 7, pp. 369–389, 2024, doi: 10.1146/annurev biodatasci 020723 021540.
9. M. Milacic et al., "The Reactome Pathway Knowledgebase 2024," Nucleic Acids Research, vol. 52, pp. D672–D678, Jan. 2024, doi: 10.1093/nar/gkad100.
10. M. Kanehisa et al., "KEGG for taxonomy based analysis of pathways and genomes," Nucleic Acids Research, vol. 51, pp. D587–D592, Jan. 2023, doi: 10.1093/nar/gkac963.

11. R. Oughtred et al., "The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions," Nucleic Acids Research, vol. 51, pp. D535–D542, Jan. 2023, doi: 10.1093/nar/gkac1070.

12. E. Kotelnikova, K. M. Frahm, D. L. Shepelyansky, and O. Kunduzova, "Fibrosis Protein–Protein Interactions from Google Matrix Analysis of MetaCore Network," International Journal of Molecular Sciences, vol. 23, no. 1, p. 67, Dec. 2022, doi: 10.3390/ijms23010067.

13. D. Szklarczyk et al., "The STRING database in 2023: protein–protein association networks and functional enrichment analyses for any sequenced genome of interest," Nucleic Acids Research, vol. 51, no. D1, pp. D638–D646, Jan. 2023, doi: 10.1093/nar/gkac1000.

14. C. Zhang and Y. Zhang, "A novel side chain orientation dependent potential derived from random walk reference state for protein fold selection and structure prediction," PLOS ONE, vol. 5, no. 12, p. e15386, 2010, doi: 10.1371/journal.pone.0015386.

15. Y. Zhang and J. Skolnick, "Scoring function for automated assessment of protein structure template quality," Proteins, vol. 57, pp. 702–710, 2004, doi: 10.1002/prot.20264.

16. G. A. Visse and H. Nagase, "Matrix metalloproteinases and tissue inhibitors of metalloproteinases: Structure, function, and biochemistry," Circulation Research, vol. 92, pp. 827–839, 2003, doi: 10.1161/01.RES.0000070112.80711.3D.

17. PMC6747341 M. Robichaud et al., "Conformation and Domain Movement Analysis of Human Matrix Metalloproteinase 2: Role of Associated Zn2+ and Ca2+ Ions," International Journal of Molecular Sciences, vol. 20, no. 17, p. 4194, 2019, doi: 10.3390/ijms20174194.

18. P. E. Van den Steen et al., "Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase 9 (MMP 9)," Critical Reviews in Biochemistry and Molecular Biology, vol. 37, pp. 375–536, 2002, doi: 10.1080/10409230290771546.

19. S. M. Popescu et al., "Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management," Frontiers in Environmental Science, vol. 12, p. 1336088, 2024, doi: 10.3389/fenvs.2024.1336088.

20. J. Jumper et al., "Highly accurate protein structure prediction with AlphaFold," Nature, vol. 596, pp. 583–589, 2021, doi: 10.1038/s41586 021 03819 2.

21. E. Bargagli et al., "FBLN1 fibulin 1 in pulmonary fibrosis: elevated expression in BAL fluid and interstitial fibroblasts," Respiratory Research, vol. 20, p. 148, 2019, doi: 10.1186/s12931 019 1108 2.

22. S. Veit et al., "Reviewing the regulators of COL1A1: an integrative bioinformatics and experimental analysis," International Journal of Molecular Sciences, vol. 24, no. 12, p. 10004, 2023, doi: 10.3390/ijms241210004.

23. H. Cha, E. Kopetzki, R. Huber, M. Lanzendorfer, and H. Brandstetter, "Structural basis of the adaptive molecular recognition by MMP9," Journal of Molecular Biology, vol. 320, pp. 1065–1079, 2002, doi: 10.1016/S0022 2836(02)00558 2.

24. A. Tzouvelekis et al., "Comparative proteomic analysis reveals iBAQ as a sensitive measure of protein abundance improvement over conventional normalisation," Scientific Reports, vol. 9, p. 11392, 2019, doi: 10.1038/s41598 019 47774 3.

25. R. Wang, J. Xu, R. Yan, H. Liu, J. Zhao, Y. Xie, W. Deng, W. Liao, and Y. Nie, "Virtual screening and activity evaluation of multitargeting inhibitors for idiopathic pulmonary fibrosis," Frontiers in Pharmacology, vol. 13, p. 998245, Sep. 2022, doi: 10.3389/fphar.2022.998245.

26. Y. Huang, W. Xu, R. Zhou, Q. Zhang, and Y. Xie, "Identification and validation of targets of swertiamarin on idiopathic pulmonary fibrosis through bioinformatics and molecular docking," BMC Complementary Medicine and Therapies, vol. 23, p. 350, Oct. 2023, doi: 10.1186/s12906 023 04171 w.

27. J. Ries, P. Bhargava, and B. Palecek, "Emerging insights into the role of matrix metalloproteases as therapeutic targets in fibrosis," Matrix Biology, vol. 68–69, pp. 262–279, 2018, doi: 10.1016/j.matbio.2017.12.009.

28. A. Limper et al., "Matrix Metalloproteases in Aberrant Fibrotic Tissue Remodeling," Proceedings of the American Thoracic Society, vol. 9, no. 3, pp. 96–100, 2012, doi: 10.1513/pats.200601 012TK.

29. S. A. Travis et al., "Bipartite interaction between fibrillin 1 and LTBP1: implications for TGF β sequestration and fibrosis," eLife, vol. 6, p. e27385, 2017, doi: 10.7554/eLife.27385.

30. A. Schmidt et al., "From Protein Structure to Drug Discovery: Bioinformatics Breakthroughs in 2024–2025," Current Issues in Molecular Biology, vol.

48, no. 1, p. 33, Dec. 2025, doi: 10.3390/cimb48010033.

31. Q. Chen et al., "Deep learning for water quality," Nature Water, vol. 2, pp. 228–241, 2024, doi: 10.1038/s44221 024 00202 z.