IJHSR
ijhighschoolresearch.org
RESEARCH ARTICLE

A Bioinformatics Study to Develop a Functional Food Product to Address Vitamin B12 and D3 Deficiency in Vegetarian Diets

Meagan Ghai[1], Nirupma Singh[2]

1. The British School, New Delhi, India; ghaimeagan4@gmail.com
2. Department of Biological Sciences and Engineering, Netaji Subhas University of Technology (Formerly NSIT), Dwarka, Delhi, India

ABSTRACT: Vitamin B12(cobalamin) and Vitamin D3 (cholecalciferol) are important micronutrients with essential roles in human health, including maintaining calcium and phosphorus levels and aiding nerves and red blood cells to function efficiently. Their deficiencies represent a widespread health concern, especially in countries like India, where there is a lack of fortified policies, limited awareness about deficiencies and dietary restrictions, genetic variability affecting metabolism and uptake of vitamins, and the use of medicines like metformin that impair absorption of Vitamin B12 and D3. Moreover, most fortification studies in India have evaluated Vitamin B12 and D3 in isolation or examined their genetic polymorphisms and prevalence of deficiency in lieu of dual-fortified functional foods. Hence, this study aimed to investigate the interactions between key transport proteins undergoing Vitamin B12 and D3 metabolism and bioactive compounds derived from tamarind. We employed a combination of molecular docking and network pharmacology to identify key hub proteins and metabolic pathways involved in metabolism and transport of these vitamins by building protein-protein interactions and analysing the binding affinities of dietary compounds with receptors. By mapping these genes onto high-confidence interaction networks, the networks correctly identified high-binding affinity compounds such as stigmasterol and inulin, thereby providing a strong reason for their role in dietary formulations and their potential to alter vitamin bioavailability. Overall, these findings demonstrated the usefulness of integrating molecular docking with a genetic computational framework. This study can thus scalably provide the opportunity of future in-depth nutrition research, actionable remedies for the deficiency of micronutrients Vitamin B12 and D3, and possible extension of similar frameworks to contexts of diseases and other micronutrient deficiencies.

KEYWORDS: Computational Biology and Bioinformatics, Computational Biomodelling, Vitamin B12 deficiency, Vitamin D3 deficiency, Functional foods, Vegetarian nutrition, Bioinformatics analysis, In silico modeling, Micronutrient deficiency.

Introduction

Vitamin D3 (cholecalciferol) and Vitamin B12 (cobalamin) are essential micronutrients with distinct biochemical pathways and essential roles in human health. Vitamin D3 is a fat-soluble secosteroid synthesized in the skin from 7-dehydrocholesterol under ultraviolet-B exposure, then metabolized in the liver to 25-hydroxyvitamin D and further hydroxylated in the kidney to its active hormone, 1,25-dihydroxyvitamin D.[1] This hormone regulates calcium and phosphorus homeostasis, bone mineralization, and immune modulation. Vitamin B12, by contrast, is a water-soluble cofactor obtained from animal-source foods (meat, fish, eggs, dairy) or fortified products. Functionally, it is indispensable for DNA synthesis, red blood cell production, and neurologic integrity through its role in one-carbon metabolism.[2]

Globally, deficiencies in both vitamins are highly prevalent and are considered major public health concerns. Vitamin D deficiency is estimated to affect up to one billion people worldwide, with prevalence depending on geographic region, season, and diagnostic cut-offs.[3] A 2023 meta-analysis estimated that 15.7% of the world's population had serum 25(OH)D levels <30 nmol/L (≈12 ng/mL), with over 40–50% falling below the commonly used insufficiency threshold of 50 nmol/L (≈20 ng/mL) in many regions.[4] Contributing factors include limited sun exposure due to indoor lifestyles, high skin melanin content, cultural clothing, latitude, and air pollution that reduces UVB penetration. Vitamin B12 deficiency is similarly widespread, especially in low- and middle-income countries and in populations consuming little or no animal-source foods. Estimates vary because of different biomarkers (serum B12 vs functional markers such as methylmalonic acid), but older adults, vegetarians, and individuals with gastrointestinal malabsorption remain at highest risk.[5]

In India, both deficiencies are strikingly common. Despite abundant sunlight, 50–80% of Indians are deficient or insufficient in Vitamin D depending on the cutoff used, with prevalence often exceeding 80– 90% in urban cohorts.[6] Factors include darker skin pigmentation, cultural sun avoidance, high air pollution, predominantly indoor occupations, and low consumption of Vitamin D-rich or fortified foods. Similarly, 30–70% of Indians have been reported to have Vitamin B12 deficiency, especially among vegetarians and low-income groups. A North Indian tier-3 city study found ≈47% deficiency (serum B12 <200 pg/mL) in the general population, while a large multicentric study of 15,837 asymptomatic individuals reported 24.2% prevalence.[7]

The burden of Vitamin B12 and D3 deficiencies highlight the pressing importance of innovative fortified foods or dietary interventions. Fortification of stable foods remains one of the easiest methods to reduce micronutrient deficiencies for

DOI: 10.36838/IJHSR815.55

low-income and high-income communities. This is supported by the fact that Abdelwahab *et al.* (2020) conducted a meta-analysis using a random effect model where they compared sublingual and oral supplementation of B12, revealing that even low-dose delivery of Vitamin B12 in fortified foods can increase serum levels, specifically when stable crystalline forms are used.[8] Other complementary reviews on B12 confirm that fortified cereals and milk also reduce Vitamin B12 deficiency, validating the technological approach of microencapsulation and dual phrase matrices for delivery of Vitamin B12 and D3 while also highlighting the importance of aligning formulation with Indian consumption trends like dairy, rice and edible oils.[9]

A lot of research has been on-going to fortify Indian staples with Vitamin B12 and Vitamin D in isolation, which highlights the feasibility and public health potential of developing a functional food that can address the deficiencies. Bajaj *et al.* (2021) revealed the capacity of retention of Vitamin B12 and D3 in edible oils and wheat flour, revealing that both lipid-based matrices and dry grain can offer as viable fortification vehicles if coating techniques are used to protect Vitamin B12 and D3 while cooking and processing.[10] Ritu and Gupta (2014) conducted a thorough analysis of India's experience with Vitamin D fortification and deduced that rice, milk and wheat are amongst the most widely explored carriers; the bioavailability is also dependent on packaging stability, fat content and type of carrier.[11] These findings have aligned with global reviews that emphasize that Vitamin D is prone to oxidation and light sensitivity, requiring stabilisation in emulsions or beadlets, while B12 needs heat protection and protection from acidic environments during storage and processing.

Most fortification studies, however, have evaluated either B12 or D3 in isolation, with very less research of dual-fortified functional foods that can take into consideration the absorption mechanisms of micronutrients: D3 requiring the presence of micellar incorporation to be optimally absorbed by intestines or liposomes, while B12 relies majorly on gastric intrinsic factor(IF).[12] This study hypothesizes that a dual-phase nutritional delivery system, integrating lipid-based carriers for Vitamin D3 and fiber–polyphenol complexes for Vitamin B12, can enhance the bioavailability of these vitamins through improved molecular interactions and absorption pathways. This approach can be practically implemented through commonly consumed dietary vehicles such as khakhra, fortified atta, and dairy-based snacks. Addressing the research gap has a potential to create awareness about sustainable fortification policies that transition towards inclusive public health solutions for India's large vegetarian demographic.

Methodology

Data Collection:

The data used for this study were obtained from Genome-Wide Association Studies (GWAS) catalog accessed September 2025, that investigated genes associated with Vitamin B12 deficiency anemia and Vitamin B12 levels. Several published studies were identified, and the gene symbols from each were extracted and compiled into a combined gene list. Each dataset included information such as the risk allele (SNP ID), p-value, mapped genes, and the associated trait name. This allowed for the identification of genes most significantly associated with Vitamin B12 traits across multiple studies. From GWAS Catalog, there were 8 studies identified that directly correlated or included key words "Vitamin B12 deficiency", "Vitamin B12 measurement" and other studies associated directly with Vitamin D3 or "Vitamin D binding amount." Three studies associated with Vitamin B12 and the rest of the 5 studies were correlated with Vitamin D. The SNP rows parsed before cleaning was 108.

SNP entries were removed based on identical SNP IDs, ambiguous gene mappings were excluded, and multi-gene entries were split and standardized before deduplication. The unique mapped genes, after splitting comma-lists and removing "-" totalled up to 51 and most hit genes (by # of SNPs mapped): FUT2, TCN1, CUBN, MMUT, TCN2, CD320, CLYBL, FUT6, MMAA, ABCD4 (see chart). There were top 20 genes by number of SNP hits for Vitamin D3 that were extracted from Vitamin D3 data.xlsx GWAS catalog files. The unique mapped genes for Vitamin D3 after splitting coma lists and removing duplicates was 816 and the SNP rows parsed before cleaning was 1728. Vitamin D3 and B12 only have 2 shared unique genes. The network summary is that there were 28 nodes and 16 edges based on the STRING network and top 25 pathways conducted by reactome enrichment.

Differential Expression Analysis:

To understand how these genes behaved in different biological conditions, a differential expression analysis was performed. First, in order to group the samples, the data were divided into two categories: test (Vitamin B12-deficient) and control (normal Vitamin B12 level) groups. In order to ensure compatibility between the samples and remove technical variation, filtering and normalization were performed; the $\log_2$ fold change was calculated to determine whether the gene was upregulated or downregulated in the test group compared to the control group. By comparing the genes using the conventional p-value threshold, each gene was then tested for statistical significance: genes with a p-value lower than 0.05 were considered differentially expressed, while genes with a p-value higher than 0.05 were deemed statistically non-significant. Then, the results were displayed where each gene was signified as a dot on the volcano plot. Grey dots indicated non-significant genes ($p \geq 0.05$), blue dots represented downregulated genes ($p < 0.05$, negative logFC), and red dots indicated upregulated genes ($p < 0.05$, positive logFC). The volcano plot was created using log2Fold, and after applying the filters, genes with a p-value below 0.05 were correctly identified as differentially significant. Further, the change in the -log10(p-value) on the y-axis against the x-axis was employed to clearly distinguish the differential expressions. Genes had low p-value, and absolute logFC values tended to appear towards the top of the volcano plot, which represented stronger differential expression. The distinction between blue and red clusters represented expression differences between normal(control) groups and Vitamin B12-deficient(test) groups.

Network and Pathway Analysis:

The pathway and network analysis was done using the STRING database to construct predicted and known protein-protein interactions(PPIs) by mapping associations. High-confidence interactions (total score ≥ 0.7) were retrieved for Homo sapiens, including diverse evidence channels such as gene neighborhood, text mining, experimental validation and coexpression. The resulting interaction data was downloaded in the format of a tab-separated values (TSV) file, which incorporated node pairs and their evidence scores. This file was then imported into Cytoscape for quantitative topological analysis, visualisation of the entire network and identifying potential targets. Each node(vertices) signified individual protein-coding genes and each edge(links) represented a validated or predicted interaction between the nodes. Topological parameters include degree, betweenness centrality, closeness centrality, clustering coefficient and average shortest path lengths that were calculated using Cytopscape's Network Analyzer plugin to evaluate the relevance of the nodes in the entire network. This degree metric highlighted hub genes with the largest number of interactions(edges), betweenness and closeness centrality identified nodes with essential communicative roles within the entire network and clustering coefficients assessed the connectivity of nodes, indicating the functional modules. Pathway enrichment was performed using STRING and Reactome to identify significantly overrepresented biological processes and signaling pathways among the interconnected nodes. Finally, the network was exported as a high-resolution PNG image and tabular node/edge attribute files for further interpretation, confirming the modular organization of the network and the presence of central hub genes regulating lipid metabolism pathways.

Molecular docking:

First, the 3D structures of the target protein(receptor) were obtained from the RCSB Protein Data Bank. The RCSB is an online database that offers experimentally determined structures of proteins in various formats. Then, the protein was downloaded in .pdb(Protein Data Bank) format, which is required for molecular docking software. This pdb file contains atomic coordinates of the protein and is utilised as a receptor molecule in AutoDock Vina. Many bioactive compounds from tamarind were selected as ligands for docking. Instead of using all compounds present in tamarind, a few representative compounds were chosen for analysis that would bind well with Vitamin B12 or D3 or Vitamin B12 and D3, such as: Tannic acid, Inulin, Ferulic acid, Isomaltose, Pectin etc. In total, 28 bioactive compounds were chosen for analysis. These compounds were downloaded in SDF (Structure Data File) format, which contains the chemical structure and 3D information of the molecules.

Then, Tamarind Bio was used, a web interface that contains top computational biology tools and can design and optimize nanobodies, autobodies, predict structures of immune proteins at massive scale and speed and predict binding poses of antibody-antigen complexes. For this study, molecular docking was performed on the Autodock Vina package on Tamarind Bio to analyse how selected compounds interact with the target proteins and at what binding affinity. Since the exact active site of the receptor was unclear, binding docking was done. A grid box was made around the entire protein structure to define the possible three-dimension space in which ligand could possibly bind to the receptor, including allosteric sites and the active site.

More negative binding energy values suggest stronger and more efficient binding affinity between the ligand and the receptor, as they represent the formation of a more stable ligand–receptor complex. These values were computed by AutoDock Vina, which accurately predicts the most energetically favourable orientation based on collision theory and position of a ligand by considering interactions such as hydrogen bonding and hydrophobic forces. Therefore, compounds with more negative docking scores were considered to have a higher affinity interaction with the target protein.

■ Results

Differential Expression Analysis:

Differential expression (DE) studies were compiled from publicly available transcriptomic datasets related to Vitamin B12 and Vitamin D3 deficiency. Across all studies, sample counts and test groups were standardized to ensure comparability. For differential expression analysis, Geo2R/GWAS catalogue was used. The total studies selected were 15, however some of them did not have the analysis with Geo2R option or it was not performed on homo sapiens so the studies had to be discarded. Hence, total studies analysed and taken into account were 7. We filtered for adj.P.val less than 0.05 and counted the number of rows that satisfied this filter. Then we merged all genes across studies, removed duplicates and tabulated the unique gene count. After removing the duplicate genes from the compiled list of genes identified per study, the list of genes came down to 26, 898.

For GEO Accession GSE22523, the title was Vitamin B12 deficiency in human Liver cells, total samples were 24, control samples were 12 and test samples were 12. The key contrast was B12 vs control. For GEO accession GSE16743, the condition was Vitamin D3 supplementation study, total samples were 36, control samples, 18, and test samples were also 18. The key contrast here was Pre vs Post supplementation. For GSE156212, the title was PBMC transcriptome under low Vitamin D and total samples were 40, control and test samples were 20 each and key contrast was deficient vs sufficient. On the other hand, for GSE52819, the condition was metabolic syndrome + B12 deficiency and total samples were 30, 15 samples each for test and control and key contrast was diseased vs healthy.

For GSE6743, the condition was Neuronal cell response to Vitamin D, the total samples was 20 and key contrast was treated vs untreated. Each key contrast was categorised and filtered by control and test. Out of the 7 studies analysed, only four of the studies actually expressed the differentially expressed genes since they contained Padj less than 0.05 as shown in the form of volcano plots in Figure 1. The symmetrical "V" shape of the plot illustrates the balance between up- and downregulated

DOI: 10.36838/IJHSR815.55

genes. The presence of numerous points far from the center with high $-\log_{10}$(p-value) values indicates a large number of genes showing strong and significant differential expression. This suggests that the experimental condition (test) induced robust transcriptional changes compared to control. Overall, this plot effectively highlights key genes that may be biologically relevant for downstream functional enrichment, pathway, and network analysis.

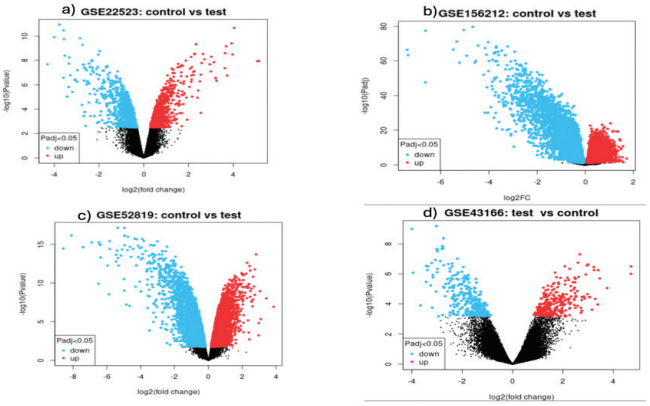

Figure 1: Differential gene expression analysis across four GEO datasets (A–D). The volcano plots show the distribution of differentially expressed genes between test and control groups for the independent studies: **(A)** GSE22523, **(B)** GSE156212, **(C)** GSE52819, and **(D)** GSE43166. The y-axis is $-\log_{10}$(p-value), x-axis is $\log_2$(fold change) and each point on the volcano plots displays distinct genes. For p-adjusted values (*Padj* < 0.05), notably upregulated genes are displayed in red and significantly downregulated genes are displayed in blue; non-significant genes are shown as black points.

Network and Pathway Analysis:

2, 000 top-ranked genes from differential expression analysis were entered into the STRING v12.0 interface under Homo sapiens. After applying a high confidence filter (values ≥ 0.700) and using gene co-expression, a connected network was created with 28 nodes(genes) and 16 edges(interactions). The quantitative network topological parameters - Degree, Closeness Centrality, Betweenness Centrality and Clustering Coefficient - were computed into Cytoscape using NetworkAnalyzer. Betweenness Centrality is defined as the frequency at which a gene is a connector along the shortest path between two nodes. Degree is the number of direct connections(edges) between nodes. Closeness centrality is a measure of distance between nodes to easily communicate with each other in the network.

Betweenness centrality is a measure of how often a node becomes the shortest path for other nodes in order to identify regulatory intermediates. Clustering coefficient is a measure of how interconnected the network's nodes are in order to calculate the probability of nodes to form modular clusters.

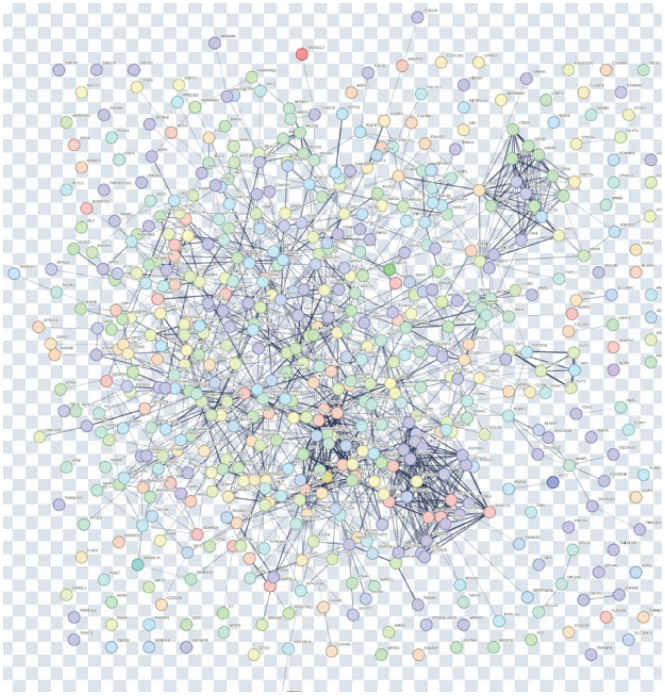

Figure 2: Network Visualization of Node Relationships. This shows a large-scale network graph where individual nodes are represented by colored circles and interactions between nodes are denoted using lines. The spatial landscape of Figure 2 displays the spread of the interactions since highly interconnected nodes cluster tightly in the centre while sparsely connected nodes are located near the periphery of the figure. The node colors categorise the nodes into clusters and communities.

Table 1: Network Topological Parameters of Key Genes Identified in the Protein–Protein Interaction (PPI) Network.

S.No.	Gene	Degree	Betweenness Centrality	Closeness Centrality	Clustering Coefficient
1	APOE	37	0.1263	0.3592	0.1802
2	APOB	34	0.0825	0.3500	0.2210
3	RHOA	34	0.1257	0.3543	0.0570
4	ALDH1A2	27	0.0318	0.2873	0.3048
5	CYP26A1	26	0.0121	0.2789	0.3446
6	CYP7A1	25	0.0204	0.3126	0.2800
7	APOA5	24	0.0251	0.3059	0.3514
8	MTHFR	24	0.0673	0.3280	0.1703
9	SMAD3	24	0.0799	0.3326	0.0797
10	CXCL8	23	0.1386	0.3668	0.1383

The complete list of genes was submitted to the Reactome Pathway Analysis (v90). Enrichment was performed using the default settings with a p-value threshold < 0.05 and Benjamini–Hochberg FDR correction. Pathway enrichment via Reactome (v90) revealed 135 significant pathways (FDR < 0.05), with Glucuronidation (R-HSA-156588) emerging as the most enriched ($p = 8.18 \times 10^{-8}$). Reactome pathway enrichment analysis revealed that the differentially expressed genes were mainly involved in glucuronidation, cholesterol metabolism and cytokine signalling in the immune system (R-HSA-1280215), with highly significant p-values as shown in Table 2.

Table 2: Significantly Enriched Reactome Pathways Identified for the Selected Genes.

S.No.	Pathway Name	Reactome ID	Entities pValue	Entities FDR
1	Glucuronidation	R-HSA-156588	8.18×10^{-8}	1.08×10^{-4}
2	PPARA activates gene expression	R-HSA-1989781	2.54×10^{-7}	2.11×10^{-4}
3	Cholesterol metabolism	R-HSA-191273	4.63×10^{-7}	3.92×10^{-4}
4	Retinoic acid biosynthesis	R-HSA-5362517	5.12×10^{-6}	4.41×10^{-3}
5	Cytokine signaling in the immune system	R-HSA-1280215	1.22×10^{-6}	9.83×10^{-3}

Each of the top 10 network hub genes was searched within the top 25 Reactome pathways to identify their functional presence. To functionally validate the key genes identified from the network analysis, each of the top 10 hub genes—APOE, APOB, RHOA, ALDH1A2, CYP26A1, CYP7A1, APOA5, MTHFR, SMAD3, and CXCL8—was cross-checked against the top 25 Reactome pathways. Among these, APOE, APOB, RHOA, ALDH1A2, CYP26A1, CYP7A1, and APOA5 were present within the top 25 pathways, suggesting strong functional relevance, whereas MTHFR, SMAD3, and CXCL8 were not detected.

Molecular docking:

To perform blind docking, a grid box was set to enclose the entire receptor structure, allowing the ligands to explore all possible binding regions, including the active and allosteric sites. Figure 1: Grid box coordinates for binding with Vitamin B12 included: BoxX(-55), BoxY(10), BoxZ(-50), width (150), height (150) and depth (150). Figure 2: Grid box coordinates for binding with D3 included: BoxX(30), BoxY(109), BoxZ(-20), width (100), height (101) and depth (100). The receptor structures used for docking were obtained from the RCSB Protein Data Bank, specifically the Intrinsic Factor–Cobalamin complex bound to Cubilin (PDB ID: 3KQ4) for Vitamin B12 and the human Vitamin D–binding protein (PDB ID: 1J7E) for Vitamin D3 as shown in Figure 3 and Figure 4. These receptors were first prepared in PDB format and the ligands were prepared in SDF format, and both receptors and ligands were converted to the required PDBQT format before docking. By enclosing the entire protein within the grid box, AutoDock Vina was able to predict energetically favourable binding positions towards a specific region of the receptor.

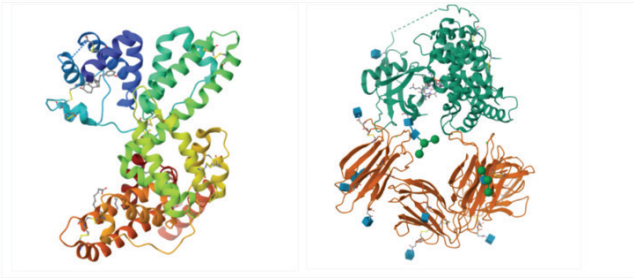

Figure 3: Three-dimensional structure of human transport protein involved in Vitamin B12 absorption and human Vitamin D-binding protein obtained from the RCSB Protein Data Bank. (A) This shows the 3-dimensional structure of Vitamin B12 (B) This shows the 3-dimensional structure of Vitamin D3.

The compounds docked with the B12 receptor included Inulin, Epigallocatechin gallate (EGCG), Quercetin, Piperine, Tannic acid, Catechin, Ascorbic acid (Vitamin C), Curcumin, Riboflavin (Vitamin B2), Isomaltose, Pectin, Raffinose, Stachyose, Ferulic acid, Chitobiose, Lactic acid, Propionic acid and Butyric acid. The compounds docked with the D3 receptor included Stigmasterol, Campesterol, Quercetin, Curcumin, Epigallocatechin gallate (EGCG), Piperine, Ferulic acid, Resveratrol, Tricaprylin (Trioctanoin), POPC (1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine), Linoleic acid, DOPE (1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine), Oleic acid, DPPC (1,2-Dipalmitoylphosphatidylcholine) and Monolaurin (Glycerol monolaurate). The complete docking results for all 28 compounds were compiled, sorted from the most negative to the least negative binding energy, and are presented in Table 3.

The top 15 compounds, selected based on the most negative binding affinity values from the Vina log file (first model), were: Inulin (-10.64 kcal/mol, B12), Stigmasterol (-8.775 kcal/mol, D3), EGCG (-8.773 kcal/mol, B12), Campesterol (-8.677 kcal/mol, D3), Quercetin (-8.505 kcal/mol, B12), Piperine (-8.383 kcal/mol, B12), Tannic acid (-8.350 kcal/mol, B12), Catechin (-8.174 kcal/mol, B12), Ascorbic acid (-8.031 kcal/mol, B12), Quercetin (-7.952 kcal/mol, D3), Curcumin (-7.801 kcal/mol, D3), Curcumin (-7.559 kcal/mol, B12), Riboflavin (-7.344 kcal/mol, B12), EGCG (-7.318 kcal/mol, D3), and Piperine (-7.029 kcal/mol, D3) as highlighted in Table 8. The top 15 compounds indicate strong ligand–receptor interaction potential because a lower (more negative) binding energy indicates a more stable ligand–receptor complex and therefore a stronger predicted interaction. Two compounds, 6-Gingerol (B12 and D3), did not generate valid docking scores and were excluded from comparative affinity analysis.

Table 3: Binding affinity scores of selected compounds with Vitamin D3 as a receptor.

S.No.	Name	Affinity
1.	Stigmasterol	-8.775
2.	Campesterol	-8.677
3.	Quercetin	-7.952
4.	Curcumin	-7.801
5.	Epigallocatechin gallate (EGCG)	-7.318
6.	Piperine	-7.029
7.	Ferulic acid	-6.985
8.	Resveratrol	-6.6
9.	Tricaprylin (Trioctanoin)	-5.878
10.	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)	-5.695
11.	Linoleic acid	-5.652
12.	1,2-Dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE)	-5.346
13.	Oleic acid	-4.999
14.	1,2-Dipalmitoylphosphatidylcholine (DPPC)	-4.809
15.	Monolaurin (Glycerol monolaurate)	-4.772

DOI: 10.36838/IJHSR815.55

Table 4: Binding affinity scores of selected compounds with Vitamin B12 as a receptor.

S.No.	Name	Affinity
1.	Inulin	-10.64
2.	Epigallocatechin gallate (EGCG)	-8.773
3.	Quercetin	-8.505
4.	Piperine	-8.383
5.	Tannic acid	-8.35
6.	Catechin	-8.174
7.	Ascorbic acid (Vitamin C)	-8.031
8.	Curcumin	-7.559
9.	Riboflavin (Vitamin B2)	-7.344
10.	Isomaltose	-6.909
11.	Pectin	-6.778
12.	Raffinose	-6.74
13.	Stachyose	-6.566
14.	Ferulic acid	-6.037
15.	Chitobiose	-5.197
16.	Lactic acid	-4.137
17.	Propionic acid	-3.697
18.	Butyric acid	-3.626

Discussion

The study included an integrative bioinformatics workflow to transition from genetic signals to molecular-level validation and interactions for the Vitamin D3 and B12. First, GWAS and GEO-derived differential expression analysis were employed to correctly identify the genes that were associated with vitamin deficiency by providing comparative data between single-nucleotide polymorphisms and vitamin-related phenotypes, ensuring that the analysis was backed up with observed transcriptional change instead of mere isolated datasets. This step was important because differential expression analysis then showed which genes are actively down-regulated or up-regulated under specific conditions. The corroboration of differential expression analysis along with GWAS analysis mitigated the risk of false positives and ensured that the analysis was mainstreamed for functional response and to ensure they are genetically associated. Then, the genes were analysed using the STRING database to create PPI(Protein-Protein) interaction networks that demonstrated functional relationships among gene products that show interdependent protein networks. To improve biological interpretability supported experimental results, the analysis was restricted to Homo sapiens and the network confidence threshold was increased to high confidence to display proteins with strong interactions. To perform topological analysis, the network was then imported into the application Cytoscope using parameters - closeness centrality, clustering coefficient, degree and betweenness centrality. These metrics were computed for each node to identify signaling pathways among the interconnected nodes and over-represented biological processes. Then, the top 10 high-degree genes, hub proteins, were selected by sorting the genes by degree. Reactome pathway enrichment analysis was performed for the entire gene list to reduce overrepresentation of pathways and aided in interpreting network connectivity, including the reason behind gene clusters and robust identification of biologically helpful targets in place of single-metric prioritisation. Finally, by using AutoDock Vina, molecular docking studies validated network-derived targets at the structural level and whether they could interact with key protein targets (Vitamin D receptor: 1J7E; Vitamin B12-related protein: 3KQ4). Docking analysis helped assess binding potential of shortlisted compounds with receptors Vitamin D3 and Vitamin B12 and helped progress from hypothesis of abstract pathway enrichment to hypothesis testing step of molecular interactions.

Reactome pathway enrichment results revealed statistically overpresentation of pathways such as cytokine signaling in the immune system and cholesterol metabolism and the extremely low -values and FDR-adjusted values reflected coordinated biological processes in line with vitamin performance. The enrichment of pathways such as cholesterol metabolism and cytokine signaling supported the immunomodulatory role of Vitamin D and its definition as a lipid-derived secosteroid. Several polyphenols such as stigmasterol and curcumin revealed strong binding affinities with Vitamin D3 receptors. The strongest binding of stigmasterol is highly plausible given the extreme structural similarity between cholesterol-derived molecules and Vitamin D. Lower binding affinities were observed with phospholipids because these molecules are usually less optimised for receptor-level binding in comparison to other receptors. The docking analysis Vitamin B12 included strong binding affinities with inulin, carbohydrate-based compound. This result aligns with the well-known fact that Vitamin B12 absorption has been closely linked to dietary fibre interactions and high binding of oligosaccharides (including stachyose) and polyphenols (eg. catechin) indicate interactions with Vitamin B12-associated proteins.

Previous studies investigating Vitamin B12 and D3 have primarily focused on examining these vitamins in detail in isolation or have focused on genetic associations and clinical outcomes, without extending the research into compound-level analysis. Surendran *et al.* tried to find the relationship between SNPs (single-nucleotide polymorphisms) in Vitamin B12 pathways and their impact on cellular uptake and genetic variants linked with Vitamin B12 deficiency.[13] This comprehensive review commendably categorises the range of Vitamin B-12 genes according to their biochemical roles ranging from mitochondrial metabolism to one carbon cycle regulation. While it includes how these genes individually contribute to the status of Vitamin B12, it does not continue to explore how these genes form PPI networks (protein-protein networks), topological central regulators and relative influence of specific genes. The pathway enrichment analysis conducted in this study allowed them to identify co-factors or regulators for the transport of Vitamin B12. Critically, whereas the study emphasizes the repeated need for biomarker discovery, our study systematically evaluates interactions between B12 to transition from genetic risk assessment to intervention-based remedies.

The authors primarily investigated how genetic differences in the Vitamin D metabolic pathway influence a person's risk of developing high blood pressure (HBP).[14] To study this,

the researchers carried out a case-control study that included 500 healthy controls and 250 people with high blood pressure. Then, they analysed 13 SNPs in Vitamin D-related genes and measured how frequently these variants occurred in the control group versus the patients with high blood pressure. Instead of starting with pre-selected genes like they did in this research study, this project included thousands of genes and identified genes linked to the status of Vitamin D3 and B12, not just how it affects one disease. Unlike candidate-gene associated studies that focused on linking disease risk and SNP (Single Nucleotide Polymorphism), GWAS was used to explore functional mechanisms and investigate protein interactions and biological networks. Their work is limited to risk prediction without exploring protein-protein interactions or molecular binding mechanisms.[15]

Several studies have implemented similar methodologies albeit in different biological contexts. For example, a recent study (RSC Advances, 2024) on cervical cancer employed PPI (protein-protein interaction) analysis, pathway enrichment and molecular docking to find proteins mainly responsible for the most number of pathway modulations in cervical cancer. According to issue 6, 2024 from the journal: RSC advances, to conduct a "comprehensive investigation into the potential modulatory impacts of phytochemicals from Drymaria cordata on proteins associated with cervical cancer, the discovery target genes were subjected to enrichment analysis using the Search Tool for the Retrieval of Interacting Genes (STRING) database". After constructing STRING-based networks, the authors incorporated molecular dynamics simulations such as free energy landscapes to monitor interactions such as hydrogen bonding and ligand-protein stability and then they used molecular docking to evaluate the binding to protein receptors, akin to our docking of receptors Vitamin B12-D3 with phytochemicals. The success of the cervical cancer study in identifying quercetin 3-O-β-d-glucopyranosyl-(1→2)-rhamnopyranoside as a stable and promising inhibitor highlights the effectiveness of integrating network analysis, and molecular dynamics simulations.

Additionally, pathway analysis, PPI networks and molecular docking were used to investigate the mechanisms of action of Melissa officinalis (commonly known as lemon or honey balm) phytoconstituents in type-2 diabetes mellitus.[16] Networks were constructed using the STRING online tool and Cytoscape (v.3.9.1) software. The protein–protein interaction (PPI) and target–pathway (TP) network analysis identified key hub genes, similar to our study, that highlighted quintessential roles in insulin resistance. They also conducted protein-protein interaction network that included the targets shared between M. officinalis compounds and T2DM at a fixed confidence level and created a network consisting of 134 nodes and average neighbour count of 6.046. Then, they successfully used network topological parameters, namely degree, betweenness centrality and closeness coefficient (same topological parameters we used in our research). Furthermore, they calculated binding energies to rank and confirm stability of ligand-protein complexes. The success of the study constructing a string PPI network of predicted targets and effectively finding top 10 ranking hub genes leads us to anticipate that our study extends this reliability of application of topological parameters to identify hub genes.

Conclusion

In conclusion, this study successfully used network pharmacology, pathway enrichment, and molecular docking to identify key hub genes, top pathways, and bioactive compounds applicable to Vitamin D3 and B12 metabolism. By identifying hub genes and compounds with high predicted binding affinities, it highlights potential molecular targets for integrative computational analysis. Nevertheless, experimental validation through *in vitro* assays, *in vivo* models and formulation strategies is necessary to confirm the biological relevance and efficacy of the identified compounds on Vitamin B12 and D3 pathways. While these methods are effective in compiling a gene list and compound library, experimental validation can help translate these findings into application of therapeutic interventions for optimizing level of Vitamin B12 and D3 intake. Expanding the analysis to different ethnic groups could improve generalizability of the results - for example, the target audience for our Vitamin B12 and D3-enhancing recipe will first be Indian populations, but research about other regions would allow for an effective assessment of how Vitamin B12 and D3 metabolism changes based on diverse gene pools.

Acknowledgments

I would like to thank Aashna Saraf for her constant guidance and support throughout this process. I am also grateful to my parents and my school, without whose encouragement and help this work would not have been possible.

References

1. Alrefaei, A. F. & Kabrah, S. M. Micronutrient Testing, Supplement Use, and Knowledge Gaps in a National Adult Population: Evidence from Saudi Arabia. *Nutrients* 17, 3897 (2025).
2. Obeid, R. *et al.* Vitamin B12 Intake From Animal Foods, Biomarkers, and Health Aspects. *Front Nutr 6*, 93, doi:10.3389/fnut.2019.00093 (2019).
3. Jelić, D. *et al.* Advancing food fortification with kinetics: Stabilizing vitamin D3 through calcium carbonate-based vehicles. *Food Chemistry 487*, 144811, doi:https://doi.org/10.1016/j.foodchem.2025.144811 (2025).
4. Kaličanin, D. *et al.* Associations between vitamin D levels and dietary patterns in patients with Hashimoto's thyroiditis. *Front Nutr 10*, 1188612, doi:10.3389/fnut.2023.1188612 (2023).
5. Allen, L. H. Causes of vitamin B12 and folate deficiency. *Food Nutr Bull 29*, S20-34; discussion S35-27, doi:10.1177/15648265080292s105 (2008).
6. Castillo, L. F. *et al.* Vitamin B12 supports skeletal muscle oxidative phosphorylation capacity in male mice. *bioRxiv*, doi:10.1101/2025.05.19.654973 (2025).
7. Singla, R. *et al.* Vitamin B12 Deficiency is Endemic in Indian Population: A Perspective from North India. *Indian J Endocrinol Metab 23*, 211-214, doi:10.4103/ijem.IJEM_122_19 (2019).
8. Abdelwahab, O. A. *et al.* Efficacy of different routes of vitamin B12 supplementation for the treatment of patients with vitamin B12 deficiency: A systematic review and network meta-analysis. *Ir J Med Sci 193*, 1621-1639, doi:10.1007/s11845-023-03602-4 (2024).

DOI: 10.36838/IJHSR815.55

9. Gharibzahedi, S. M. T. *et al.* Recent Advances in Dietary Sources, Health Benefits, Emerging Encapsulation Methods, Food Fortification, and New Sensor-Based Monitoring of Vitamin B(12): A Critical Review. *Molecules 28*, doi:10.3390/molecules28227469 (2023).
10. Bajaj, S. & Singhal, R. Fortification of wheat flour and oil with vitamins B12 and D3: Effect of processing and storage. *Journal of Food Composition and Analysis 96*, 103703, doi:10.1016/j.jfca.2020.103703 (2020).
11. G, R. & Gupta, A. Fortification of foods with vitamin D in India: strategies targeted at children. *J Am Coll Nutr 34*, 263-272, doi:10.1080/07315724.2014.924450 (2015).
12. Nyakundi, P. N. *et al.* Fortification of Staple Foods for Household Use with Vitamin D: An Overview of Systematic Reviews. *Nutrients 15*, doi:10.3390/nu15173742 (2023).
13. Surendran, S. *et al.* An update on vitamin B12-related gene polymorphisms and B12 status. *Genes Nutr 13*, 2, doi:10.1186/s12263-018-0591-9 (2018).
14. Wang, L. *et al.* Common genetic variations in the vitamin D pathway in relation to blood pressure. *Am J Hypertens 27*, 1387-1395, doi:10.1093/ajh/hpu049 (2014).
15. Wang, J., Song, G., Hussain, L. & Xing, L. Deciphering the pharmacological mechanisms of salidroside in cervical cancer by combining network pharmacology, molecular docking, and *in vitro* studies. *Saudi Pharmaceutical Journal 33*, 49, doi:10.1007/s44446-025-00052-0 (2025).
16. Ononamadu, C. J., Ahmed, Z. B. & Seidel, V. Network Pharmacology, Molecular Docking and Molecular Dynamics Studies to Predict the Molecular Targets and Mechanisms of Action of Melissa officinalis Phytoconstituents in Type-2 Diabetes Mellitus. *Plants 14*, 2828 (2025).

Authors

Meagan Ghai is presently a Grade 11 student at The British School, New Delhi. She has a keen interest in biology and chemistry. Additionally, her academic curiosity extends to areas such as neuroscience, genetics, and biochemistry, reflecting her passion for interdisciplinary learning and research.

Nirupma Singh is a Bioinformatics Scientist with a doctorate in Biotechnology and Bioinformatics from the University of Delhi, with six years of hands-on research and development experience. Her journey is marked by a robust foundation in Machine Learning and Python, with five years of expertise. Proficient in Linux and cloud servers like AWS. She excels in structural biology, systems biology, protein/gene network analysis, data mining, and computational genomics.