Integrated structural analysis of sex hormone binding globulin reveals allosteric modulation by distant mutations

Dwipanjan Sanyal, Deeptanshu Pandey, Arthur McLelland, Vladimir N. Uversky, Sourav Chowdhury, Shalender Bhasin, Ravi Jasuja

PII: S0141-8130(25)04602-1

DOI: https://doi.org/10.1016/j.ijbiomac.2025.144050

Reference: BIOMAC 144050

To appear in: *International Journal of Biological Macromolecules*

Received date: 27 December 2024

Revised date: 5 May 2025

Accepted date: 7 May 2025

Please cite this article as: D. Sanyal, D. Pandey, A. McLelland, et al., Integrated structural analysis of sex hormone binding globulin reveals allosteric modulation by distant mutations, *International Journal of Biological Macromolecules* (2024), https://doi.org/10.1016/j.ijbiomac.2025.144050

Integrated Structural Analysis of Sex Hormone Binding Globulin Reveals Allosteric Modulation by Distant Mutations

Dwipanjan Sanyal[1], Deeptanshu Pandey[2], Arthur McLelland[3], Vladimir N. Uversky[4], Sourav Chowdhury[2,5*], Shalender Bhasin[6*], and Ravi Jasuja[1,6*]

Affiliation
[1] Xyone Therapeutics, Boston, Massachusetts, USA
[2] Department of Biological Sciences, Birla Institute of Technology and Science-Pilani, Hyderabad, India.
[3] Centre for Nanoscale Systems, Harvard University, Cambridge, Massachusetts, USA
[4] USF Health Byrd Alzheimer's Research Institute, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA
[5] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA
[6] Research Program in Men's Health: Aging and Metabolism, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

[*]**Corresponding Authors**

Sourav Chowdhury
Department of Biological Sciences,
Office: J Block-109
Birla Institute of Technology and Science, BITS-Pilani Hyderabad Campus,
Hyderabad, Telangana, India, 500078
Email: sourav.chowdhury@hyderabad.bits-pilani.ac.in

Shalender Bhasin
Research Program in Men's Health,
Aging and Metabolism, Brigham and Women's Hospital,
Harvard Medical School, Boston, Massachusetts, USA

Ravi Jasuja
Ravi Jasuja, PhD
Director, Translational Research and Discovery
RPMH: Aging and Metabolism
Brigham and Women's Hospital
Harvard Medical School
Boston, MA 02118
Email: rjasuja@bwh.harvard.edu

Keywords: Sex hormone binding globulin (SHBG), Binding affinity, Conformational fluctuations, Allosteric perturbations, Raman spectroscopy, Structural dynamics, Network modeling

Abstract:

Sex hormone-binding globulin (SHBG), a glycoprotein in circulation, binds testosterone, dihydrotestosterone, and estradiol with high specificity, regulating their transport and bioavailability. This function relies on long-range conformational interactions between its N-terminal (NTD) and C-terminal (CTD) domains. Variations in SHBG levels or binding affinities alter free hormone concentrations, influencing reproductive and metabolic health. Despite its significance, the full-length SHBG structure and the conformational dynamics influencing hormone binding remain unclear. Deploying in-silico structural analysis, Raman spectroscopy, and network modeling, we investigated the intramolecular structural dynamics of the full length SHBG to understand how allosteric perturbations caused by natural mutations affect hormone binding and inter-residue interactions. Raman spectroscopy and in-silico analyses show that majority of the residues in SHBG (308 residues) constitute loop regions, whereas only 21% constitute beta sheet. Mutations in SHBG that alter its binding affinity, though distant from the ligand-binding pocket (LBP), induce long-range conformational changes. These mutations are clustered in flexible regions but maintain structural order through dense local interactions. Our in-silico analyses identified key substructures regulating allosteric interactions between mutation sites and ligand-binding residues. This study provides a template for further structural analyses of clinically reported mutations and their effect on hormone binding and action.

1. Introduction:

Sex hormone binding globulin (SHBG) is a 45KDa glycoprotein produced primarily in the liver[1,2]. It associates with sex hormones - testosterone, dihydrotestosterone (DHT), and estradiol - with high affinity in the bloodstream[3] and plays an important role in regulating their distribution, transport, and tissue bioavailability [4]. These hormones share a common binding pocket on SHBG. SHBG comprises a homodimeric assembly[5], made up of two identical monomers, each consisting of 401 amino acids. SHBG's distinct structural arrangement allows it to interact selectively with sex hormones[6]. Understanding the intricate interplay between SHBG and sex hormones is essential in comprehending the regulatory mechanisms governing hormonal actions in the body. Alterations in SHBG levels or its affinity for sex hormones are associated with alterations in the circulating concentrations of the unbound or the free hormone and can impact an individual's reproductive and metabolic health[7]. Together, SHBG's structure and function play an important role in modulating the bioavailability and actions of sex hormones, orchestrating a delicate hormonal balance. Monomeric structure of SHBG involves a tandem repeat of laminin G-like (LG) domains[8,9], which play a crucial role in its functionality and interactions with sex hormones. The LG domain is mainly responsible for its specific binding to sex hormones, particularly testosterone, dihydrotestosterone (DHT), and estradiol[10,11]. This domain within SHBG create a binding groove that accommodates and interacts with the hydrophobic regions of the sex hormones. Additionally, these LG domains contribute to SHBG's stability and structural integrity[11]. Mutations or alterations within these LG domains can affect SHBG's binding affinity for sex hormones, potentially leading to changes in free hormone levels and activity[9]. Moreover, the tandem arrangement of LG domains in SHBG allows for a range of interactions with multiple hormones within a steroid class[2].

Previous studies have revealed how SHBG distinguishes between androgens (such as testosterone and dihydrotestosterone) and estrogens (like estradiol), preferentially binding them in opposite orientations within the single steroid-binding site present in each SHBG monomer[12]. Androgens and estrogens possess distinct molecular structures and hydrophobic regions[13]. SHBG's steroid-binding site within the LG domains, is adaptable, enabling it to selectively recognize and bind androgens and estrogens through specific interactions within its steroid-binding pocket[2,14]. Several mutations in the gene encoding SHBG have been reported to impact

its binding affinity for its ligands, testosterone, dihydrotestosterone (DHT), and estradiol[9]. These mutations in the SHBG gene could influence its binding affinity for these ligands by altering its structure, stability, or functional domains. The extant studies on dynamics of hormone bioavailability built upon findings from LBD crystal structure[8,10] of truncated ligand binding domain may not accurately capture the solution dynamics and conformation flexibility across the full length SHBG. Despite previous research into SHBG's function and ligand interactions, the full-length structural organization and conformational dynamics that govern its ligand-binding properties remains unavailable to date. A particularly unresolved and clinically significant question is how naturally occurring mutations, often located far from the ligand-binding pocket (LBP), alter SHBG's hormone-binding affinity. The spatial separation of these mutations from the LBP suggests an allosteric mechanism, but the structural and dynamic pathways underlying such long-range communication are poorly characterized. Understanding these allosteric effects is essential to elucidate how SHBG regulates hormonal bioavailability and to interpret the functional consequences of disease-associated SHBG variants.

The present work provides a comprehensive insight on how the mutations which are distant from ligand-binding domain (LBD) could alter ligand specificity. These mutations, although spatially separated from the binding pocket, are hypothesized to induce conformational changes that propagate through the protein's structure, ultimately modulating its hormone-binding affinity. To elucidate this mechanism, we characterized the structural arrangement of SHBG monomer by which mutations impact its ligand binding. To generate a full-length model of SHBG, we performed homology modeling and validated the structure through AlphaFold comparison and Raman spectroscopy. To characterize regions of intrinsic disorder and flexibility, and to assess how mutation sites align with these dynamic regions, we conducted a comprehensive disorder propensity analysis using consensus-based predictive tools. We were particularly interested in understanding whether allosteric effects play a role in modulating the effects of mutations that are distant from the ligand binding pocket on SHBG's interactions with its ligands. In order to determine how mutation sites influence ligand-binding residues, we carried out structure network analysis, identifying key structural blocks (SBs) that mediate allosteric communication. Changes in distant regions of the protein due to mutations could potentially propagate conformational changes that indirectly affect the ligand-binding site, altering its affinity for sex hormones. Therefore, to evaluate the internal flexibility and dynamic coupling of different regions within

SHBG, we employed Monte Carlo simulations. Understanding the structural arrangement of LG domain, its interaction with rest part of the protein as well as internal dynamics provides valuable insights into the molecular mechanisms governing its function as a carrier protein for sex hormones selectivity and binding preferences within the single steroid-binding site of the SHBG monomer. Collectively, this integrative approach provides a mechanistic framework to investigate how distant mutations reshape the conformational landscape of SHBG and modulate its hormone-binding behavior.

2. Results:

As shown in Fig. 1, to accomplish our objectives, we deployed a combination of *in-silico* methods and spectroscopic techniques to characterize key regions in the SHBG protein structure which regulate the internal dynamics as well as ligand binding. We performed homology modelling and generated the model structure of the SHBG monomer. We verified the accuracy of our model by comparing it with the monomeric structure predicted by AlphaFold. Also, to complement the *in-silico* model and capture dynamic structural features, we performed Raman Spectroscopy, which provided experimental insights into the secondary structural organization and compared it with the *in-silico* model structure. This comparison helped us to bridge the resolution limitations of purely computational models while capturing dynamic features inaccessible to static structures. Subsequently, we analysed the structures of the DHT and estradiol-bound SHBG to investigate the residues involved in hormone binding and their spatial configuration within the protein structure. To determine the internal arrangement and mutual interactions between residues, we performed a comprehensive structure network analysis by performing normal mode analysis (NMA). The NMA included generating both all-residue structure network and community cluster network and a correlation network between amino acid residues. This approach allowed us to understand potential allosteric pathways linking distant regions to the ligand-binding site. We further studied the flexibility of SHBG monomer using MC simulation to identify key residues that regulate the structural dynamics of the protein. This integrative approach has provided important insights into critical aspects governing protein behaviour and its binding interactions with ligands. Fig. 1 represents the workflow outlined in the manuscript.

Together, this integrative framework addresses the limitations of earlier truncated or static SHBG models by offering a full-length, dynamic perspective essential for uncovering allosteric regulation and mutation-driven effects on ligand binding.

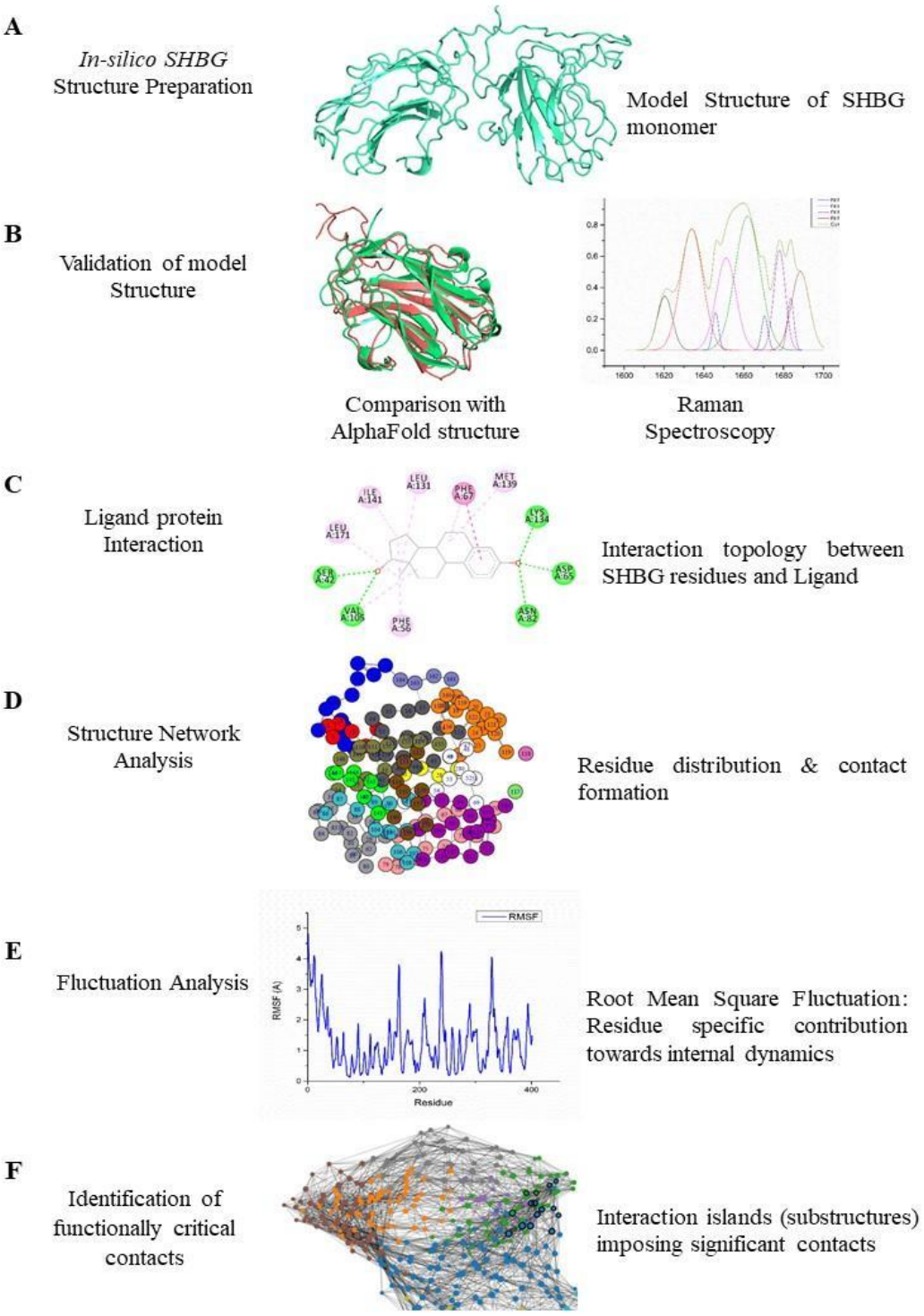

Fig. 1. Ensemble Structure Analysis Workflow to probe into SHBG structural dynamics. (A) Homology modelling was performed using the human SHBG sequence to generate a model structure of the SHBG monomer. (B) To validate our model structure, we compared it with the monomeric structure obtained from AlphaFold and further corroborated it by elucidation of secondary structural organization.using Raman Spectroscopy. (C) We analyzed ligand-bound SHBG structures to identify the protein residues interacting with DHT and estradiol and their specific interaction topology. (D) Structure network analysis was pwerformed to reveal the distribution and interactions of residues within the 3D protein structure. (E) We used Monte Carlo simulation to investigate backbone flexibility of the SHBG monomer, revealing coupled motions of different regions/residues. (F) We identified important residues that regulate intramolecular structural dynamics in SHBG and influence ligand affinity. By integrating experimental and *in-silico* data, this cumulative workflow has offered a mechanistic understanding of SHBG's structural dynamics and regions important in hormone binding.

2.1 Structure modeling of SHBG: Contact Map and Fluctuation Analysis

Since no full-length structure of the SHBG protein was available at the time of this study, we generated an *in-silico* structure for the monomeric SHBG protein. Within this structure, the LG domain houses nearly all of the reported mutations, which have been consistently associated with alterations in both estradiol and DHT binding affinities of SHBG (Fig. 2A, Table 1). This highlights the significance of the LG domain in mediating the functional impact of these mutations on the protein's interactions with these ligands.

The contact map generated shows inter-residue correlations within the SHBG model structure (Fig. 2B). The intensity of the red color is proportional to a higher extent of correlations/correlated motions among residues, while negative scores, indicated by green colors, represent residue-residue cross-correlation. White color represents uncorrelated residues. The 2D contact map indicates that residues in the LG domain mostly maintain positive correlation among themselves, while exhibiting anti-correlated motions with the rest of the protein.

Interestingly, we observed that residue 152 exhibited high positive correlation with the DHT binding site at residue 112 (140 with respect to the full-length SHBG structure), indicating strong

interaction between these two residues. Mutations at residue 152 have been reported to reduce DHT binding affinity of the protein. The majority of the mutation sites displayed anti-correlation with ligand binding sites, suggesting their allosteric effects on ligand interaction.

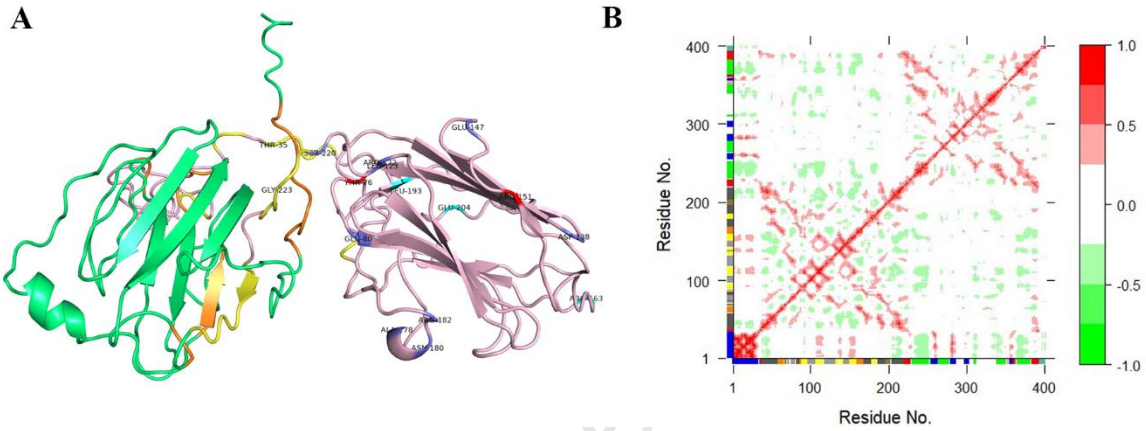

Fig. 2. Structural modelling and residue interaction analysis of SHBG reveals intra-domain communication pathways potentially regulating ligand binding and conformational dynamics. (A) The modelled full-length SHBG monomer structure (402 amino acids) is shown with domain-specific colour coding. The N-terminal Laminin G-like (LG) domain, containing the steroid-binding site, is shown in light pink, while the rest of the structure is depicted in lime green. Experimentally reported mutation sites are marked in blue. Mutations known to reduce ligand-binding affinity (e.g., T77I, R152C, R152H, G224E) are highlighted in red, and those associated with enhanced binding are shown in cyan. The residues forming the structural bridge between the LG domain and the rest of the monomer are coloured in yellow and light orange, indicating possible communication interfaces. (B) A residue–residue contact map was generated using a 5 Å distance threshold to identify spatial proximities in the modelled structure. Each dot represents a residue pair with Cα–Cα distance below the cutoff, and the colour gradient indicates interaction strength. The map reveals strong intradomain interactions within the LG domain, especially in the ligand-binding pocket, and highlights the presence of long-range contacts that may mediate allosteric communication between distant regions of the protein.

2.2 Model validation of the *in-silico* SHBG structure and insights into SHBG conformation

We compared the homology model of the SHBG monomer, generated using I-TASSER with default settings, to the AlphaFold-predicted structure (obtained via AlphaFold version 2.3.1 using the AlphaFold Protein Structure Database) to assess structural similarity. The overall root mean square deviation (RMSD) between the two models was 2.416 Å, indicating a moderate degree of structural deviation (Fig. 3A). To gain deeper insight into the distribution of these differences, we conducted domain-specific comparisons by separating the LG domain (residues 1–238) and the N-terminal domain (residues 239–401). As depicted in Fig. 3B, the N-terminal domain showed greater structural variance (RMSD = 3.813 Å) compared to the LG domain (RMSD = 2.175 Å). The primary deviation in the LG domain was traced to the final 40 residues at the C-terminal, which formed a helical segment and extended loop in the AlphaFold model but appeared only as an extended loop in the I-TASSER-generated structure. Similarly, the elevated RMSD in the N-terminal domain is attributed to differences in secondary structure predictions, where AlphaFold predicted helical elements that were absent in the homology model. These differences likely reflect the challenges of accurately modeling structurally dynamic or disordered regions. Notably, AlphaFold has been reported to overpredict ordered structures, particularly α-helices, in regions of intrinsic flexibility or disorder, particularly at the N- and C-terminal ends of proteins[15]. It is well-known that *in-silico* modelling approaches, including AlphaFold, can sometimes overestimate the disorder in terminal regions, leading to deviations in predicted structures[16]. Additionally, AlphaFold's tendency to overestimate certain secondary structures, such as α-helices, particularly in flexible or disordered regions, may have contributed to the observed differences[17]. This limitation is particularly relevant for SHBG, a loop-dominant protein where the dynamic nature of loops plays a critical role in its structure. Moreover, AlphaFold's performance, while remarkable in predicting monomeric structures, can be less reliable when dealing with proteins that exhibit significant flexibility or exist in multi-meric forms, as SHBG potentially does[18]. This highlights the importance of complementary structural modelling techniques, such as homology modelling, which may provide alternative structural insights where AlphaFold predictions diverge. Furthermore, experimental validation remains essential to refine these *in-silico* predictions and accurately capture the nuanced features of protein structures like SHBG. The integration of techniques such as Raman spectroscopy

provides crucial insights into secondary structure elements, offering validation and revealing potential limitations in computational models.

Raman spectroscopy analysis provided significant insights into the secondary structure of the SHBG monomer (Fig. 3C and 3D). The analysis underscored the prominence of loop regions within the protein's structure. Our *in-silico* analysis revealed that 21% of the SHBG monomer consists of β-sheet structures, while a remarkable 77% of the residues (equivalent to 308 amino acids) are involved in loop regions. These loop regions encompass both ordered and disordered segments of the protein. The Raman spectroscopy data corroborated these findings, indicating a comparable secondary structure distribution with approximately 23% β-sheet content and about 77% loop regions (Fig. 3D). The loop-rich nature of SHBG suggests that these segments are far more than passive linkers between secondary structure elements. Loops are often associated with enhanced conformational flexibility, which can allow proteins to adapt dynamically to various molecular interactions. In the case of SHBG, the extensive loop network may contribute significantly to the protein's ability to undergo subtle yet functionally important conformational changes in response to ligand binding or mutations. This structural adaptability is particularly relevant for mediating allosteric effects, where perturbations such as mutations at distant sites propagate through the protein structure to influence the behavior of the ligand-binding pocket.

Moreover, loops can serve as hubs for intramolecular communication, especially in multi-domain proteins like SHBG, where coordinated dynamics between the laminin G-like domains and surrounding flexible regions are critical for functional integrity. Given that many of these loops likely bridge both ordered and disordered segments, they may play a unique regulatory role in stabilizing transient conformations required for selective hormone recognition and binding. Despite this, loop regions remain underexplored in the context of full-length SHBG models, which have traditionally focused on static representations of structured domains. Our analysis shows the importance of these flexible elements, and we hypothesize that SHBG's loop-dominated architecture is integral to its function as a dynamic hormone carrier. Elucidating how these regions contribute to internal dynamics and allosteric signalling is therefore essential to fully understand the molecular basis of hormone binding and specificity in SHBG.

This spectroscopic validation highlights that despite the inherent flexibility and often unstructured nature of loop regions, they significantly contribute to the ordered structural

framework of the SHBG monomer. The substantial representation of loop regions suggests their critical role in maintaining the protein's structural integrity and functionality, bridging both ordered and disordered domains within the overall architecture of SHBG.

We conducted detailed mathematical analyses on the Raman spectrum of the wild-type (WT) SHBG to complement the secondary structure evaluation. The analytical approach involved baseline correction, derivative spectroscopy, spectral moment calculation, entropy analysis, and peak asymmetry determination to provide comprehensive quantitative insights. The first and second derivatives of the Raman spectrum were computed to enhance the detection of subtle spectral features. The first derivative plot highlighted minor intensity fluctuations and spectral shifts, while the second derivative was particularly effective in resolving overlapping bands, offering precise identification of peak positions indicative of specific structural elements or molecular interactions. Baseline fitting using a fifth-order polynomial provided a robust method to correct spectral data, ensuring the accurate quantification of peak intensities and positions. The corrected spectrum differentiated between genuine spectral features and baseline artifacts, significantly improving the reliability of complex structural analyses. Quantitative spectral metrics included Shannon entropy of 6.34, indicating moderate spectral complexity and suggesting an ordered yet diverse molecular environment; a spectral centroid located at 1656.15 cm^{-1}, typical of the Amide I region, reflecting predominant secondary structural features; variance of 315.89, indicative of moderate bandwidth reflecting a distinct yet diverse vibrational environment; and skewness of 1.31, emphasizing the asymmetric distribution of spectral intensity toward higher wavenumbers, possibly indicating specific hydrogen bonding or environmental effects. Peak detection identified eight distinct peaks, confirming the presence of multiple structural and molecular features as observed with the SHBG model.

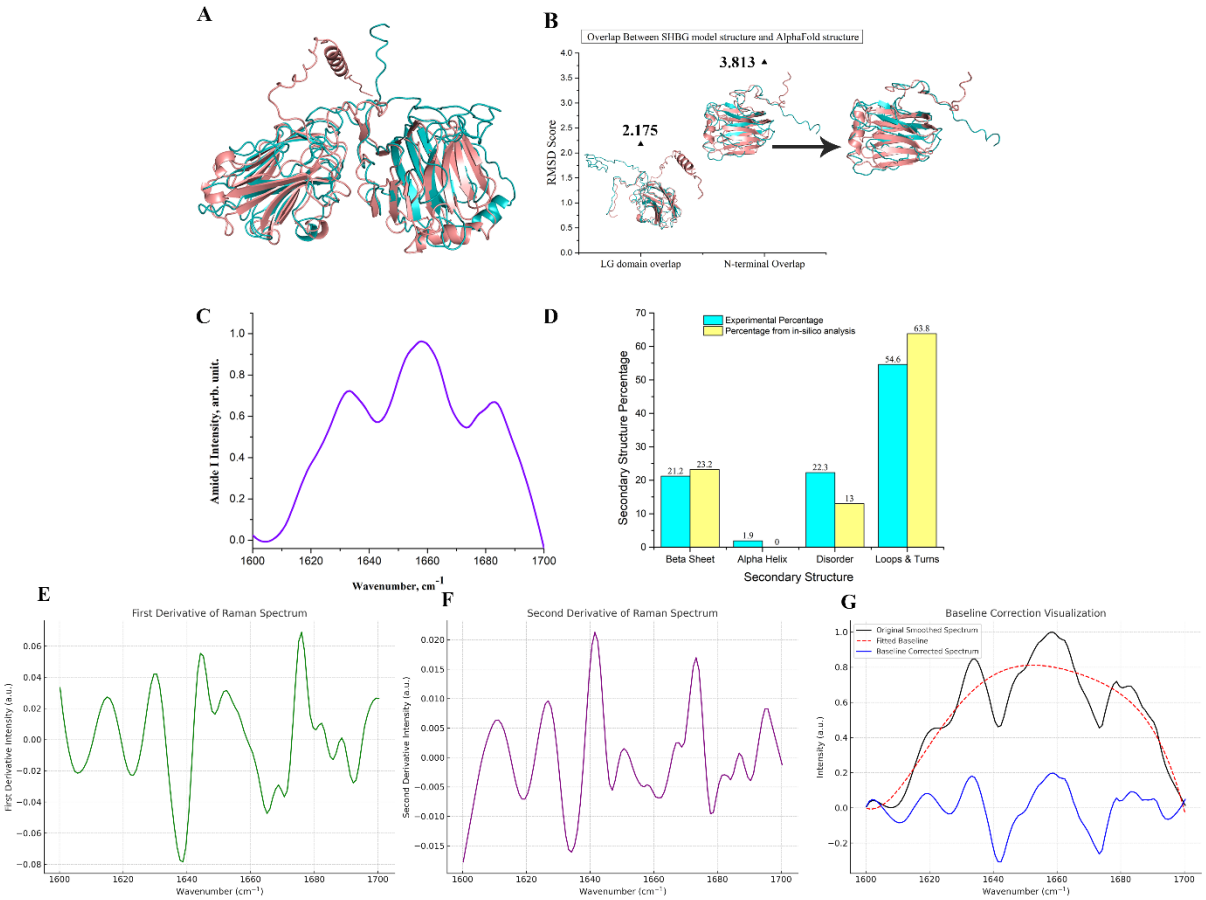

Fig. 3. Validation of SHBG Homology Model via *In-Silico* Comparison with AlphaFold Predictions and Raman Spectroscopic Analysis. (A) Overlap between *in-silico* model structure (Green) and AlphaFold structure (Red) shows an RMSD score of 2.416 (293 to 293 atoms). (B) The N-terminal domain exhibits higher deviation (RMSD score = 3.813) between the homology model (Green) and AlphaFold structure (Red) compared to the LG domain (RMSD score = 2.175). In the inset, we have shown the zoomed in view of the divergent N-terminal region (residues 239–401). Here *in-silico* model structure is represented as Green and AlphaFold structure as Red. (C) Raman Spectroscopy was performed as a control check (Amide I), showing the raw amide I spectra ranging of the SHBG monomer from 1600 to 1700 cm^{-1}. (D) The *in-silico* model structure shares similar secondary structure content with the experimental results obtained from the Raman Spectroscopy. This clearly indicates that although SHBG is a loop dominant protein, the extended loop regions are not intrinsically disordered. (E) The Fig. illustrates the first derivative spectrum, highlighting minor intensity fluctuations and spectral shifts. Panel (F) shows the second derivative spectrum, resolving overlapping bands to accurately

pinpoint peak positions. Panel (G) displays the baseline correction process, demonstrating the original smoothed spectrum (black), the polynomial-fitted baseline (red dashed line), and the baseline-corrected spectrum (blue), ensuring precise identification and quantification of spectral features.

2.3 Intrinsic disorder analysis revealed mutations were positioned in flexible regions of the protein

To understand the regions having flexibility and intrinsic disorderness within the SHBG protein sequence, we performed an analysis employing prevalent intrinsic disorder predictors, yielding a mean intrinsic disorder profile (Fig. 4A). The predictions were integrated into a consensus disorder profile, calculated as the unweighted average of individual predictor scores for each residue. Table 1 provides the Mean Predicted Disorder Scores (PDSs) for individual residues. We categorized the regions and residues surpassing the 0.5 PDS threshold as intrinsically disordered, while those within the 0.2 to 0.5 range are deemed flexible based on standard thresholds in the literature[19,20]. This distinction of the residue patches is visually represented with varied colors, highlighted in the protein structure (Fig. 4A and 4B). Despite the predominance of loop residues, only few residue patches showed intrinsically disordered nature having mean PDS values higher than 0.5. Here, specific stretches from residue 1 to 8, residue 39 to 51, residue 314 to 329, and residue 391 to 402 exhibited intrinsic disorder nature. Notably, most of the residues of SHBG displayed flexibility, with mean PDS values ranging from 0.2 to 0.5 (Table 1). Our analysis of mutation sites that affect ligand binding revealed that out of the 16 reported mutation sites (Supplementary Table 1), 12 demonstrated flexible characteristics (Table 1). Specifically, mutation sites 36, 139, 148, 152, 221, and 224 exhibited higher flexibility (PDS within the 0.4 to 0.5 range) compared to other mutation sites. Among the mutations that reduce DHT binding affinity to SHBG (T77I, R152C, R152H, and G224E), residues 152 and 224 displayed higher flexibility having elevated mean PDS values (0.4-0.5) in contrast to residue 77 (mean PDS: 0.2-0.3). Similarly, substitutions affecting estradiol's binding affinity (R164C, L194M, and E205K) showed mean PDS values within the 0.3 to 0.4 range.

This intrinsic disorder calculation aligns with the experimental data observed from Raman Spectroscopy. The disorder observed from Raman Spectroscopy was approximately 22%, while

the *in-silico* structure calculated about 13% disorder (Fig. 4A). Although these values are not identical, they are reasonably consistent given the inherent differences between experimental and computational methods. Moreover, this disorder calculation not only aligns with but also validates the earlier Raman Spectroscopy analysis conducted for secondary structure prediction (Fig. 3C). This convergence emphasizes the consistency and reliability of our computational predictions in elucidating the propensity for disorder within the intricate framework of the protein structure.

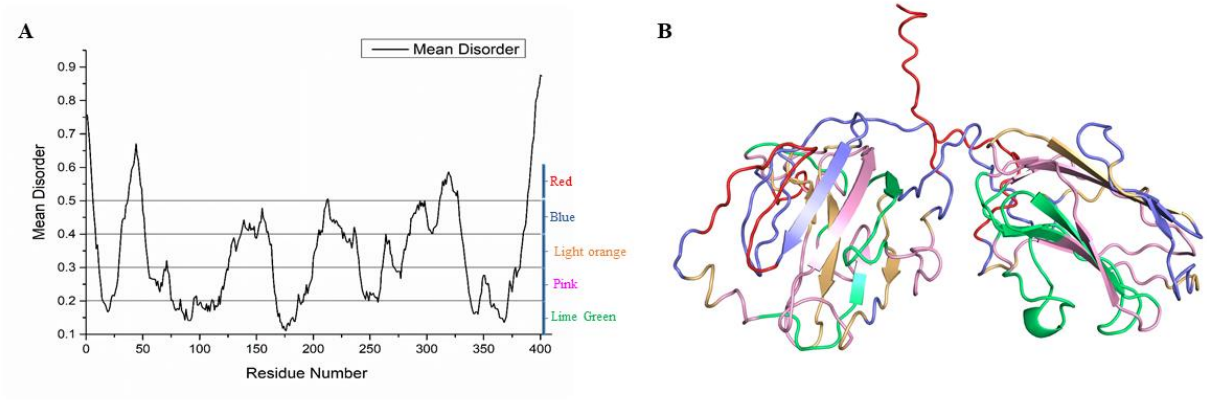

Fig. 4. Disorder propensity and mutation localization in SHBG protein reveal that the sequence naturally maintains an ordered structure, with a substantial number of observed mutations predominantly located within its flexible regions. (A) A multiparametric analysis elucidates the intrinsic disorder predisposition of SHBG. The plot demonstrates the mean disorder propensity, determined by averaging the disorder profiles from individual predictors. Residues and regions with Mean Predicted Disorder Scores (PDS) surpassing a threshold of 0.5 are classified as disordered, as indicated by the dark red background. Additionally, those with PDS values between 0.2 and 0.5 are earmarked as inherently flexible or presumed flexible. Varied line styles and shades of red boxes convey differing degrees of flexibility. (B) Visual representation of SHBG highlights intrinsically disordered regions in red, flexible domains in green, and documented mutations as light red sticks. It is noteworthy that a significant proportion of the reported mutations reside within SHBG's flexible zones.

2.4 Effect of mutations on local intrinsic disorder propensity of SHBG

Next, we analysed how mutations affect local disorder propensity of SHBG. To this end, we calculated per-residue disorder profiles for all the mutants and overlaid them with the disorder profile of the wild type protein. Fig. 5A represents the results of this analysis and shows that the local disorder propensity of SHBG was differently affected by mutations. It is also clear that some mutations increased the local disorder propensity, whereas others decreased it. To better see changes induced by mutations, we utilized the "difference disorder spectrum" approach, where a per-residue disorder profile generated for the wild type protein was subtracted from the disorder profiles of individual mutants. To better visualize changes induced by mutations, we utilized the "difference disorder spectrum" approach, where a per-residue disorder profile generated for the wild-type protein was subtracted from the disorder profiles of individual mutants. Fig. 5B represents the overlaid "difference disorder spectra" for all 16 mutants. While these trends suggest mutation-specific impacts on local disorder, it is important to note that the "difference disorder spectra" reflect correlations rather than causative relationships. Changes in local disorder propensity do not necessarily translate to functional effects such as altered binding affinity. Other molecular factors such as electrostatic changes, steric hindrance, or alterations in hydrophobicity may also contribute to the observed functional outcomes and should be considered alongside disorder shifts**.** Finally, Fig. 5C correlates local disorder propensity in a site of a given mutant with the disorder propensity of the wild type at the same position. Here, symbols located above the diagonal correspond to the mutations that caused the local increase in the disorder propensity, whereas mutations leading to the local decrease in disorder propensity are located below the diagonal.

Analysis of these data shows that based on their absolute effects on the local disorder propensity, mutations can be arranged as D139N < N181K < R152H < T36N < R123Q < A179P < E205K < L194M < L124M < E148D < R183W < E81I< G224E < T77I < R152C < S221L < R164C, with the smallest effects being introduced by mutations T36N, R123Q, D139N, R152H, and N181K. The strongest positive effects (i.e., largest increase in local disorder propensity) were introduced by G224E, L124M, and L194M, whereas the strongest negative effects (i.e., largest decrease in local disorder propensity) were promoted by R164C, S221L, R152C, T77I, E81I, R183W, and E148D. These changes were mostly agreed with the expected changes based on the known classification of residues as order- (C, W, Y, I, F, V, L, H, T, and N) and disorder-promoting (A, G, D, M, K, R, S, Q, P, and E). In fact, mutations causing increase in the local disorder

propensity were substitutions of order-promoting to disorder-promoting residues, whilst changes of disorder-promoting residues to order-promoting ones were associated with the decrease in local disorder propensity.

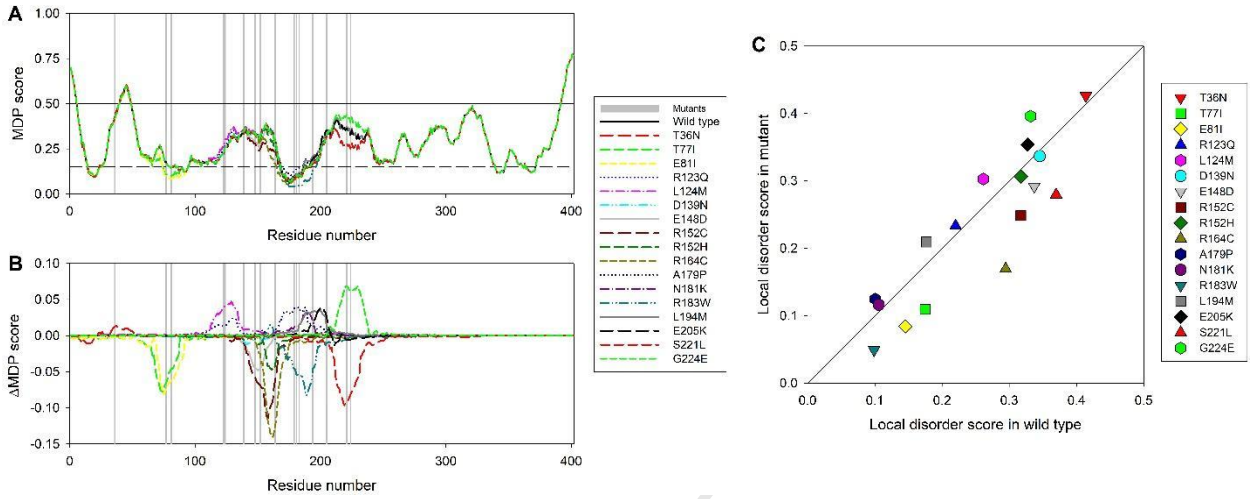

Fig. 5. **Analysis of the effects of mutations on local disorder propensity of human SHBG. A**. Overlaid per-residue disorder profiles generated for the wild type protein and 16 mutants. Vertical grey vertical bars show positions of mutations. **B**. Overlaid "different disorder spectra" calculated by subtracting the disorder profile of the wild type SHBG from the disorder profile of the mutant. Positions of mutations are shown by vertical grey bars. **C**. Local disorder score in mutant versus local disorder score in wild type plot.

2.5 SHBG Ligand Binding Sites were located in the Beta Barrel Region

To locate the ligand binding pockets within SHBG, we analysed the available DHT and estradiol-bound SHBG structures (1KDM and 1LHU, respectively) (Fig. 6A and 6B respectively). The reported mutation sites were located distantly from both ligands in the respective bound structures (Supplementary Table 1). In the DHT-bound structure, only one DHT binding site was approximately 3Å away from residue 152, a mutation site known to diminish DHT binding affinity (Distance between residue 152 and 112 is ~3Å) (Fig. 6A). Similarly, within the estradiol-bound structure, the closest distance observed between estradiol and the reported mutation sites was 8Å, whereas the minimum distance from estradiol binding residues was 7.5Å (Fig. 6B). These observations suggest that mutations distant from the ligand

binding sites influence the protein's functional behaviour, possibly through allosteric modulations.

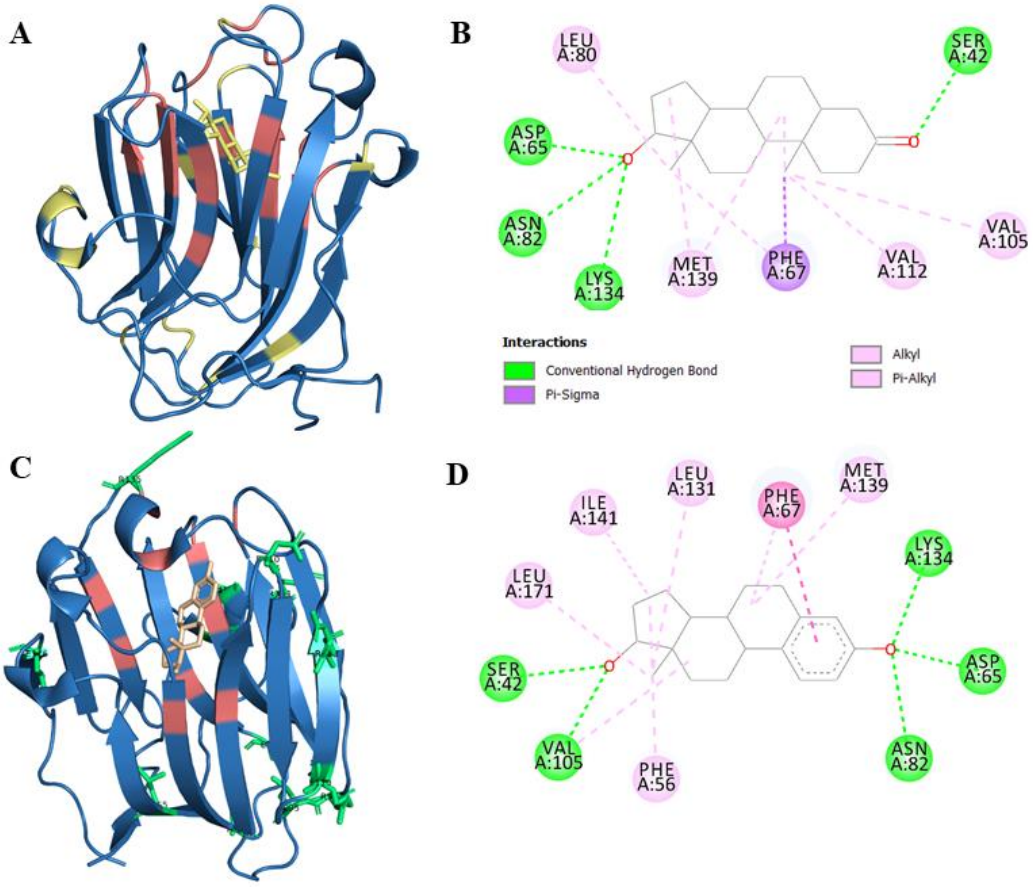

Fig. 6. Ligand interaction residues are located in the β-barrel region of SHBG. DHT-bound SHBG structure (PDB ID: 1KDM) is shown in (A), where DHT and the reported mutation sites are depicted as pale-yellow sticks, and the ligand-binding pocket is highlighted in red. The minimum distance between DHT and the mutation sites is approximately 8 Å; notably, residue R152 lies ~5 Å from DHT-binding residue 112, while no DHT-binding residues are located near mutation site T77. Panel (B) presents a 2D schematic of DHT-binding residues in the SHBG monomer. In (C), the estradiol-bound SHBG structure (PDB ID: 1LHU) shows estradiol as pale yellow sticks, ligand-binding residues in red, and mutation sites in green. The minimum distance between estradiol and the reported mutation sites is ~8 Å, with specific measurements including

E205–141 (~7.5 Å), L194–56 (~7.5 Å), and R164–65 (~8 Å); additionally, R164 (position 135 in the full-length SHBG) lies adjacent to residue 163 (134), a known estradiol-binding site. Panel (D) illustrates a 2D schematic of estradiol-binding residues. These observations suggest that the mutations, although spatially distant from the ligand-binding sites, may exert their influence through long-range allosteric effects A table summarizing distances between key mutation sites (e.g., R152C, L194M, R164C) and DHT/estradiol is provided in Supplementary Table 1 for quick reference.

2.6 Structure network analysis to determine how ligand binding residues and mutation sites are orchestrated in SHBG and their association with internal dynamics of the protein

To investigate the internal arrangement and mutual relationships between residues, we conducted a comprehensive structure network analysis[19,21] and generated both an all-residue structure network (Supplementary Fig. 1) and a community cluster network (Fig. 7A.) The community cluster network revealed the organization of SHBG residues into 11 distinct structural blocks (SBs) as illustrated in Fig. 6A and outlined in Supplementary Table 2. Fig. 7B shows the residues in the SHBG structure by different colors based on their positions in specific SBs. Notably, six of these structural blocks (SBs 1, 3, 5, 7, 8, and 9) contained a substantial number of residues, emphasizing their significance within the structural network (Fig. 7A). Additionally, SBs 2, 4, and 6 exhibited dense connections with other SBs, indicating their importance in terms of stability of the protein structure.

This study represents the first application of residue-level structure network analysis on the full-length monomeric SHBG protein. Previous structural analyses have largely been restricted to truncated forms, primarily focusing on the ligand-binding domain, leaving the role of distant regions and their long-range interactions unexplored. By analysing the entire protein, our approach captures a comprehensive map of inter-residue connectivity, allowing for the identification of structurally encoded allosteric communication pathways between distant mutation sites and the hormone-binding pocket. This network-based framework provides a novel lens through which to understand the molecular basis of allosteric regulation in SHBG.

We found that the hormone (DHT and estradiol) binding sites were mainly housed in SB 3, 5 and 7. Only one DHT binding site (Valine 65) was in SB 4. The reported mutation sites predominantly clustered within four interconnected SBs; SB 2, 3, 5, and 7, indicating their

crucial involvement in structural dynamics (Supplementary Table 2). It's noteworthy that, only one mutation site, R164C, resided within SB4.

The correlation network between amino acid residues suggested that majority of the residues in LG domain were highly correlated, whereas some residues showed correlated motions with the rest part of the structure (Fig. 7C). Positive correlations have been shown as red colour. Interestingly, most of the inter-domain contact forming residues having high positive correlation were housed within SB 3, 5 and 7 (Fig. 7A and 7C). We also observed that residues surrounding the binding cleft imposed anti-correlated motions indicating their allosteric effects (Fig. 7D). Interestingly, we found that some ligand binding residues and naturally occurring mutation sites displayed cross-correlation. Residues with higher correlated motions formed contact and clustered into same or highly interacting SBs.

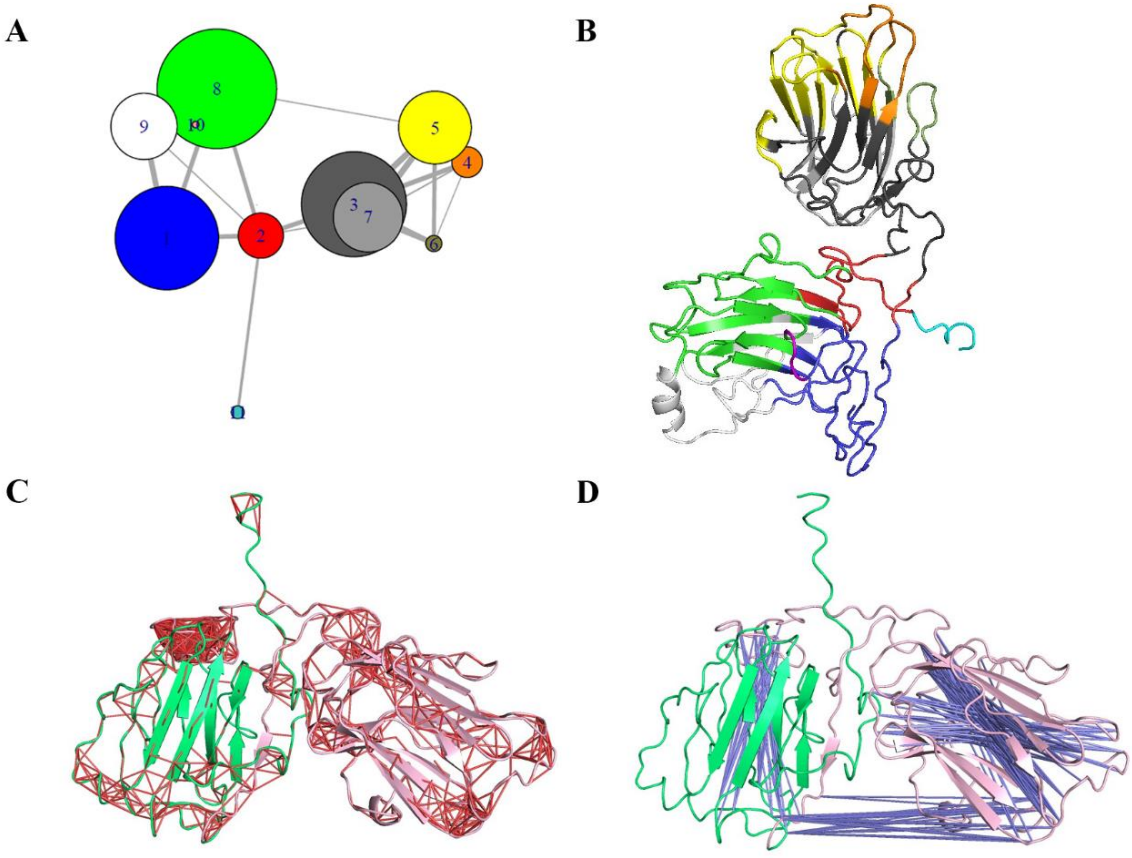

Fig. 7. A network, analysis of SHBG's structure reveals the distribution of key interacting residues and structural blocks in its 3D conformation. (A) Within the Structure Network of the SHBG protein, 11 distinct Structure Blocks (SBs) are present. The size of these SBs is dictated by the count of their constituent residues. The magnitude of interactions between these SBs is conveyed through the varying edge widths. Here SB 1, 2, 3, 7, 5, and 8 are comprised of large number of residues, whereas SB 2, 3, 4, 5,6, and 8 have a large number of connections in comparison with others. (B) A rendered model of SHBG is depicted, wherein the residue stretches have been marked with same color as their corresponding SBs. For example, SB 1 is colored as Blue in Fig. A. accordingly, the Blue colored region in this Fig. corresponds to the residues constituting SB 1. (C) SHBG monomer shows the residues having higher positive correlations among themselves. Red coloured lines indicate positive correlations, i.e. those residues form contacts among themselves. (D) SHBG structure indicating cross-correlations among residues. Blue lines indicate cross-correlations. Here inter domain residues imposed cross-correlation among themselves. Also, some mutation sites in the LG domain and ligand binding residues showed cross-correlation.

2.7 Conformational fluctuation analyses identified significant sub-structures/clusters in SHBG and their effect on overall structural dynamics

We deployed the course-grained Monte Carlo simulation by sampling 2*10^5 conformations, which provided a comprehensive exploration of collective motions and residue fluctuations within the SHBG structure. The Root Mean Square Fluctuation (RMSF) profile derived from this simulation revealed a distinct pattern: a notable lower extent of fluctuation among residues located within the LG domain (residues 1-239) in comparison to the remaining segments of the protein (Fig. 8A). The scatter plot depicted in Fig. 8A effectively showcased the comparison between the calculated RMSF values (illustrated as solid spheres) and the RMSF values derived from the predicted model (represented by solid lines). This visualization demonstrated the alignment between model performance and the actual RMSF values across the protein sequence. Notably, an intriguing observation emerged that the reported mutation sites exhibited relatively lower fluctuations compared to other residues. Moreover, the contact map derived from the RMSF profile (Fig. 8B) unveiled several regions characterized by high-frequency contacts,

forming island-like clusters within the protein structure. Among these, the most substantial cluster encompassed residues 1 to 259, forming a large interconnected region. Additionally, distinct dense islands comprised residues spanning from 113 to 188, 235 to 238, 275 to 264, continuing with 293 to 307, and extending to 355 to 385. These regions signify the most significant clusters, indicating potential sites of dense interaction within the protein structure. Furthermore, maximal clique analysis identified highly significant residues or patches within the SHBG structure (Fig. 8C and Supplementary Table 3). We found that many of the maximal clique residues participated in inter-domain interaction within the monomeric structure (Fig. 2). The top three identified maximal cliques are highlighted with thick black borders, among which most of them were housed in the island formed of residues 1 to 259. Remarkably, these top three cliques predominantly consisted of residues adjacent to mutation sites, suggesting their importance towards the structural integrity. Additionally, some of these residues exhibited close proximity to the ligand binding sites, further emphasizing their potential functional significance in ligand interactions. Again, the interesting observation of seven clique residues housed within SB3 further emphasizes the significance of this particular SB. Such a concentration of maximal clique residues within SB3 potentially denotes its role in facilitating essential structural and functional dynamics within the protein.

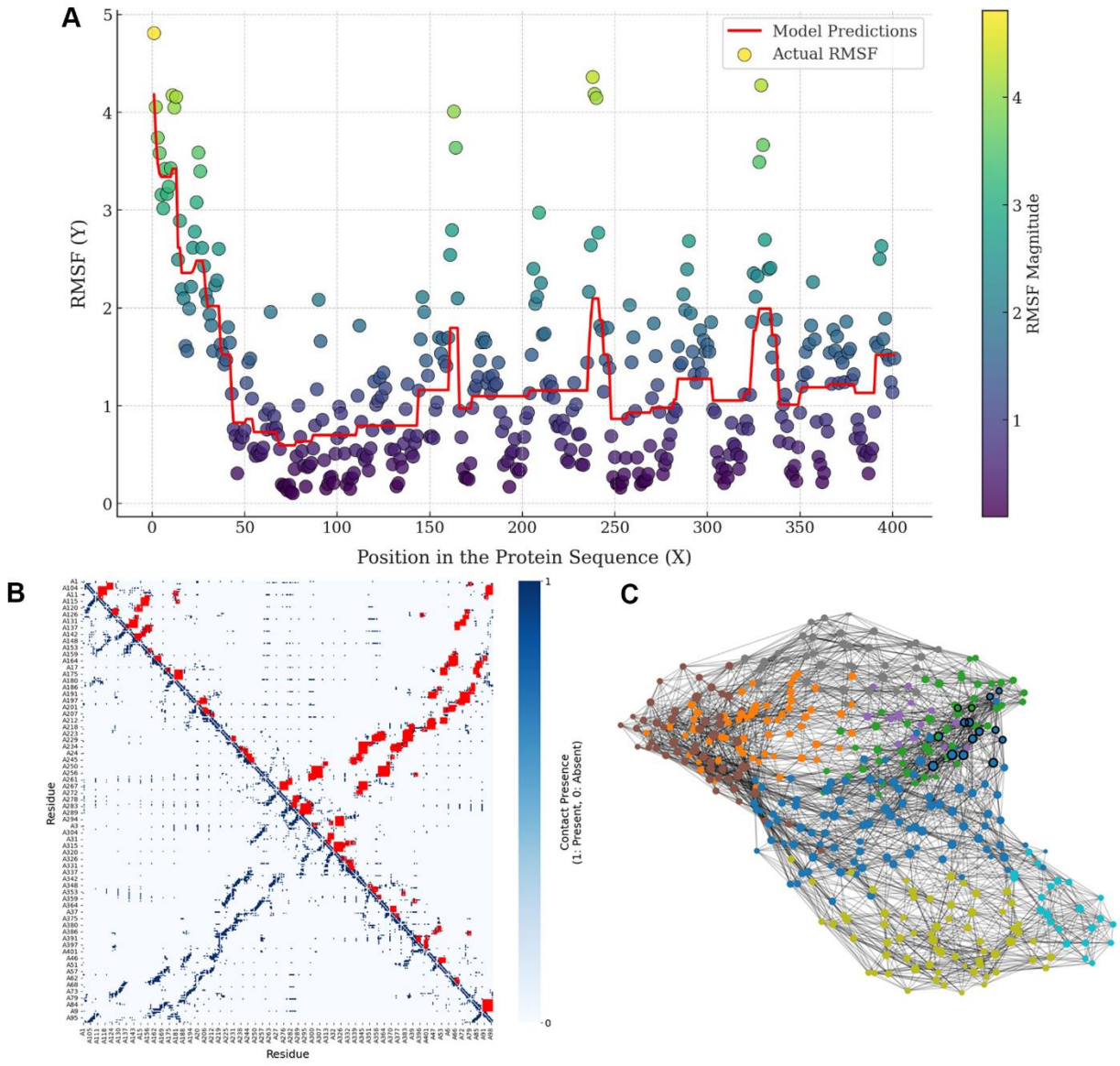

Fig. 8. Fluctuation analysis reveals the dense contact islands in SHBG. (A)The scatter plot represents the actual RMSF values with solid sphere markers, and each point is color-coded based on its RMSF magnitude, allowing for a more intuitive grasp of the variations in RMSF values across different amino acid positions. The line plot in red represents the model's predictions and shows how model performance aligns with the actual RMSF values across the protein sequence. (B) The dense regions with significant contacts in the binary contact map have been identified and marked with red rectangles on the contact map. These regions are indicative of stable, well-defined structures within the protein. (C) The Asynchronous Label Propagation Algorithm detected 8 communities (discussed in the results) within the protein graph. The

visualization displays these communities, and within this graph, the nodes are color-coded by the community and sized by degree of centrality. The top 3 cliques (discussed in the results) have been highlighted with thick black borders.

3. Discussion and Conclusion:

Our integrated approach reveals that loop-rich structure of SHBG facilitate extensive allosteric communication between mutation sites and the ligand-binding pockets. By generating the *in-silico* full-length SHBG structure and validating it through Raman spectroscopy, we investigated previously unresolved structural regions. Our findings showed that around 77% of the SHBG structure comprises loop regions which, despite their disordered tendencies, exhibit structural organization essential for ligand binding, inter-domain communication, and dimerization. Structure network and contact map analyses identified strong interconnections between these loop regions and the β-barrel binding site, suggesting potential allosteric regulation pathways. Notably, mutations were predominantly located within these flexible loops and, although spatially distant from the binding sites, were found to influence hormone affinity supporting the presence of long-range allosteric effects. Together, these findings highlight a structurally coordinated mechanism where loop-mediated allosteric communication modulates SHBG's ligand-binding property.

The complex interactions between proteins and their ligands influence the clinical manifestation of diverse physiological processes[22]. While hormone binding proteins (HBPs) and regular proteins participate in this intriguing interplay, their mechanisms and biological contexts vary distinctly. Binding pockets in HBPs have evolved to accommodate their specific hormone partners[23,24]. This high affinity ensures that the signal reaches the intended recipient, orchestrating specific cellular responses. HBPs, such as Sex Hormone-Binding Globulin (SHBG)[14] and Corticosteroid-Binding Globulin (CBG)[25,26], thyroxine-binding globulin (TBG)[27,28] have evolved to bind and transport specific hormones. HBPs are classified into two main types: steroid-binding and non-steroid-binding proteins. Non-steroid-binding proteins, including various enzymes, transport proteins, and receptors, interact with a diverse range of ligands such as neurotransmitters, enzymatic substrates, and signalling molecules, playing critical roles in biological processes[29-31]. In contrast, the steroid-binding proteins, like SHBG and

CBG, exhibit high specificity for steroid molecules such as testosterone, estrogen, and cortisol, facilitating their transport and modulating their bioavailability [32-34]. This specificity enables them to effectively regulate the concentration and distribution of steroid hormones in the bloodstream, influencing physiological processes like reproduction, metabolism, and stress responses.

SHBG's binding affinity towards specific hormones influences their distribution and accessibility, ensuring a delicate hormonal equilibrium crucial for various physiological functions. These associations highlight SHBG as a potential biomarker for diagnosing and monitoring hormonal disorders and related pathologies[35-37]. The loop-dominant SHBG structure is critical for interactions with specific ligands. The different orientations of androgens and estrogens within the SHBG steroid-binding site hold functional significance[12]. This alteration in orientation influences the molecular interactions and accessibility within the binding site.

Mutations at different sites of SHBG disrupt its hormone-binding affinity, leading to an array of health conditions[9,38]. Therefore, understanding the structural orchestration of SHBG, importance of its different regions from the perspective of its ligand binding as well as function is critical. But due to the unavailability of full length SHBG monomeric structure, clear understanding of the conformational dynamics associated with ligand binding and their disruptions under different mutation conditions have remained elusive and not fully understood. Unlike truncated structures, the full-length SHBG can reveal how the LBD communicates with other regions of the protein, allowing for a more accurate representation of its conformational dynamics and flexibility in solution. Additionally, full-length SHBG is vital for investigating allosteric mechanisms and how distant mutations can influence hormone bioavailability and regulatory functions. In this study, we generated *in-silico* SHBG model and investigated its structural arrangement and possible allosteric interactions.

Loop regions in proteins are essential components that contribute significantly to their structural dynamics, function, and versatility. SHBG, despite being comprised of 77% unstructured loop residues, shows a propensity for forming ordered substructures as observed from the disorder calculation (Fig. 4). Our Raman Spectroscopic analysis also showed that SHBG monomer contains only 13% disordered regions which validates the *in-silico* predictions. In monomeric SHBG, the flexible loop residues connect the N-terminal and C-terminal LG domains, facilitating dimerization[2,39], inter domain interactions, and ligand binding (Fig. 2). Contact map

and network analyses revealed a strong correlation between loop residues of the two domains, highlighting their role in structural communication (Fig. 1B and 7C). Also, most of the reported mutations which influence the ligand binding affinity of the protein are positioned in the flexible loop regions (Fig. 4). Notably, over 50% of reported mutations affecting ligand binding are in these flexible loops, suggesting potential allosteric crosstalk with the ligand-binding pocket in the β-barrel region, despite the mutation sites being distant from it.

The structure network topology of the SHBG monomer revealed the interplay among residues based on pairwise interactions and interaction strength, highlighting a cross-correlation between mutation sites and ligand-binding sites. This suggests that the effects of mutations are transmitted to the hormone binding site through allosteric crosstalk. Notably, naturally occurring mutations impacting DHT and EST binding clustered in specific SBs (Supplementary Table 1), which showed dense interconnections, emphasizing their role in maintaining structural integrity. Key ligand-binding residues were positioned in SBs 3, 4, 5, 6, 7, and 10 (Table 3), with SB 5 housing 9 out of 24 pocket-forming residues and SB 6 only one. Mutations reducing DHT binding were found in SBs 2, 5, and 7, while those enhancing EST binding were in SBs 3 and 4. These SBs (2, 3, 4, 5, 6, and 7) exhibited strong interconnections, with mutation sites positioned at least 7.5 Å from ligand molecules, suggesting indirect interactions through correlated motions. This allosteric crosstalk reveals the influence of distant mutations on ligand-binding dynamics.

The RMSF profile revealed variations in fluctuation intensity, highlighting distinct dynamics and stability of the LG domain compared to other regions within the protein structure (Fig. 8A). Contact map derived from RMSF identified residue islands formed through strong interactions, influencing the protein's structural dynamics. Residues 1 to 259, forming an island with maximal clique residues, demonstrated significant centrality and inter-domain contact, showing their structural importance. Most of the significant sites identified through maximal clique analysis were in SBs 2, 3, 5, 6, 7, and 8 (Supplementary Table 3), which contains the LG domain, ligand-binding sites, and mutation sites. These maximal clique residues strongly interconnect with ligand-binding and mutation sites, indicating their potential role in allosteric communication and the impact of mutations on ligand binding. This interconnected cluster plays a crucial role in mediating interactions crucial for the protein's functional properties.

Our study provides a comprehensive picture on the crucial roles of loop regions, mutation sites, and SBs in defining SHBG's structural dynamics and ligand recognition. Mutation sites in flexible loop regions indirectly influence ligand-binding pockets, revealing an intricate allosteric network governing protein-ligand interactions. The full-length SHBG model presented here serves as a foundational template for future structural studies, with mechanistic insights on the implications of reported mutations and their potential downstream effects.

Our study introduces an integrative framework that brings together experimental and computational strategies to unravel the structure-function dynamics of SHBG. This approach offers several distinct advantages. First, we present, for the first time, an *in-silico* model of the full-length monomeric SHBG protein, enabling detailed analysis of inter-domain interactions and structural communication that are absent in truncated versions. Second, Raman spectroscopic validation of secondary structure complements disorder predictions, enhancing confidence in model accuracy and structural interpretations. Third, by means of structure network and correlation analyses, we mapped previously uncharacterized allosteric pathways connecting distant mutation clusters with the hormone-binding pocket. Finally, the integration of RMSF profiling, contact maps, and maximal clique analysis offers a comprehensive view of SHBG's conformational landscape, identifying key residue islands involved in stability and ligand responsiveness. Together, these strengths provide a foundational basis for understanding the SHBG function in health and disease, with implications for targeted drug design and biomarker discovery.

Absence of functional experimental assays probing into conformational dynamics associated with the SHBG and its mutants is a limitation in the current study. Future work would aim at studying the mutant structures that effect ligand binding and investigating how the mutations alter critical contacts in the ligand binding pocket. Further, using *in-silico* approaches we would study different mutant structures of SHBG to understand how the loop movements modulate the pocket formation for specific hormones. Additionally, combining experimental and computational strategies using both ligand-bound and unbound structures would help to identify intermediate protein conformations and provide a deeper understanding of SHBG's conformational landscape. Such studies would provide insights into the allosteric regulation

mechanisms and might inform the development of targeted therapeutics that modulate SHBG function.

Acknowledgement:

D.S acknowledges the Xyone Theraputics. The studies were supported in part by NIH grants: 5P30AG031679(SB), 5 R44 AG045011(RJ) and R33 AG068234 (RJ). S.C and DP acknowledges BITS Pilani-Hyderabad for the generous funding (RGrant Reference: N4/24/1032)

Author Contributions:

R.J, S.C. and S.B conceptualized the overall project outline. D.S and S.C designed the work plans, S.C. and A.M carried out the experiments. D.S and D.P performed the computational calculations, analyses and wrote the paper. D.S performed all the analyses and prepared the Fig.s. A.M performed Spectroscopic analysis, and S.C analysed the spectroscopic data during his stay at Harvard University. V.U performed the disorder propensity analysis. R.J, S.C, S.B and V.U reviewed the manuscript.

Declaration:

The authors do not have any conflicts of interest.

Table 1. Mean Predicted Disorder Score (PDS).

Mean Disorder Score	Residue Number
0.809311	1
0.757418	2
0.720336	3
0.674606	4
0.620881	5
0.566172	6

0.502125	7
0.507669	38
0.548752	39
0.556564	40
0.592245	41
0.60112	42
0.609302	43
0.63396	44
0.67039	45
0.633811	46
0.62255	47
0.583331	48
0.546652	49
0.535958	50
0.501491	213
0.505582	214
0.49511	312
0.515839	313
0.520528	314
0.55882	315
0.544666	316
0.564331	317
0.54962	318
0.575563	319
0.586136	320
0.577758	321
0.571685	322
0.549809	323
0.529854	324
0.527038	325

0.523057	326
0.514384	327
0.529279	328
0.521117	390
0.542641	391
0.582573	392
0.623887	393
0.668668	394
0.686604	395
0.74365	396
0.800271	397
0.80994	398
0.834895	399
0.843206	400
0.875417	401
0.873391	402

4. Materials and Methods:

4.1 Intrinsic Disorder Analysis

Due to the absence of an experimentally determined full-length structure for SHBG at the time of this study, we began by retrieving its protein sequence from UniProt (ID: P04278). To evaluate the intrinsic disorder characteristics of the protein, we applied a combination of well-established per-residue disorder prediction tools, including PONDR® VLXT[40], PONDR® VSL2[41], PONDR® VL3[42], PONDR® FIT[43]. Also in our analysis, we deployed two types of IUPred algorithms for predicting short as well as long intrinsically disordered regions, IUPred_short and IUPred_long[43], respectively. We also computed the mean disorder propensity of the SHBG protein by averaging the disorder profiles provided by the individual predictors. We employed a consensus approach to evaluate intrinsic disorder, as empirical observations have shown that this

approach usually improves predictive performance compared to using a single predictor. In these analyses, we considered predicted disorder scores above 0.5 as indicating disordered residues and regions, while considered scores ranging from 0.2 to 0.5 as the fluctuating regions within the protein[19].

4.2 Structure-Based Analysis

Model Structure Preparation

The SHBG protein sequence was obtained from UniProt (ID: P04278), and it's in silico structural model was generated using I-TASSER (I-TASSER Suite version 5.2)[44]. Subsequently, we performed an *in-vacuo* energy minimization process on this model structure to mitigate unfavorable interactions and steric clashes, following a previously established protocol[45].

4.3 Raman Spectroscopy and Data Processing

To investigate the conformational dynamics and solution-phase secondary structure of Sex Hormone-Binding Globulin (SHBG), Raman spectroscopy was employed as a structural validation tool for our *in silico* model. Raman spectroscopy is a powerful vibrational technique that is both non-destructive and highly sensitive to molecular structure, particularly capable of probing the backbone conformation and side-chain environment of proteins. The technique's sensitivity to vibrational modes such as C=O stretching in the amide backbone (primarily the Amide I band, 1600–1700 cm^{-1}) is informative for assessing secondary structural elements, including α-helices, β-sheets, and turns[53-55].

4.3.1 Instrumentation and Experimental Setup

All Raman measurements were conducted using the Horiba XploRa PLUS confocal Raman microscope equipped with a thermoelectrically cooled CCD detector maintained at -70°C. Spectral acquisition utilized a 1200 gr/mm diffraction grating blazed at 750 nm, providing a spectral resolution of 1 cm^{-1}. The excitation source was a 785 nm near-infrared solid-state diode laser, selected to minimize fluorescence interference and sample degradation; power at the sample was maintained at approximately 41 mW. To ensure spatial resolution and spectral

clarity, the spectrometer slit width was set to 200 µm and the confocal pinhole aperture to 500 µm. Prior to each batch of measurements, the system was calibrated using a silicon wafer, aligning the 520.7 cm^{-1} reference peak for spectral accuracy. SHBG samples (20 µM in sodium phosphate buffer, pH 7.4) were measured in a total volume of 20 µL, and each spectrum was normalized to the sharp, reproducible band at 330 cm^{-1}, which served as an internal standard.

Each spectrum was acquired with an integration time of 180 seconds, and four consecutive spectra were recorded per sample to enhance signal-to-noise ratio and to enable robust cosmic ray rejection.

4.3.2 Baseline Correction and Denoising

Raw spectra inherently contain noise components and fluorescence background that must be carefully removed prior to quantitative analysis. These steps were performed using Labspec 6 software (Horiba Scientific).

Baseline-Subtraction: Fluorescence background was removed using a polynomial baseline correction method. Specifically, a 7th-order polynomial fitting algorithm was applied to the entire spectral window. This polynomial order was selected empirically to best model the complex curvature typically seen in protein spectra, particularly in the Amide I region. The polynomial was iteratively fit to non-peak regions of the spectrum using a moving anchor point algorithm, and then subtracted from the raw signal to flatten the baseline across the entire range.

Spectral-Denoising: After baseline correction, Horiba's built-in denoising algorithm (standard mode) was applied. This method utilizes a Savitzky-Golay filter, conFig.d with a window width of 11 points and a second-order polynomial smoothing function[56]. This setting balances retention of spectral features with effective noise suppression, minimizing distortion of narrow peaks, particularly in the Amide I band.

Cosmic-Ray-Removal: An automatic cosmic ray rejection filter integrated into the acquisition routine was engaged during multi-scan accumulation. This filter uses a threshold-based pixel comparison across successive scans to identify and eliminate sharp, non-reproducible spikes.

4.3.3 Secondary Structure Analysis from Raman Spectra

Raman-based structural analysis was carried out to analyze the Amide I region (1600–1700 cm^{-1}), which is sensitive to secondary structure due to C=O stretching vibrations in the protein backbone[57].

Spectral-Deconvolution: Deconvolution was performed using Gaussian-Lorentzian mixed curve fitting within the Labspec 6 software. The Amide I envelope was decomposed into its component bands corresponding to distinct secondary structure types. Initial peak positions and widths were selected based on prior literature for protein spectra.

Peak fitting employed a Levenberg–Marquardt least-squares optimization, with constraints placed on full-width at half-maximum (FWHM) values (typically 8–25 cm^{-1}) to prevent overfitting. The final fits were evaluated using reduced χ^2 minimization and visual inspection to ensure model fidelity.

Secondary Structure Quantification: The area under each fitted peak was integrated, and relative percentages of secondary structures were calculated by normalizing the peak areas to the total Amide I envelope area. This approach assumes a linear proportionality between peak area and structural population.

4.3.4 Structure Network Analysis

The structural network representation of a protein offers a topological analysis of its 3D structure, independent of its secondary structure and folding characteristics[46]. Understanding the internal motions and structural dynamics of proteins is crucial for elucidating their functions and activities. Therefore, we employed normal mode analysis (NMA) to predict functional motions within the protein segment[21]. Subsequent to NMA, we conducted correlation analysis to construct a cross-correlation matrix. Utilizing correlation network analysis, we created an all-residue network using *in-silico* model structure of SHBG. This all-residue network was further dissected into a network of highly correlated, coarse-grained community clusters using the Girvan-Newman clustering method. In these clusters, residues with strong interactions were grouped together, forming what we referred to as "structure blocks" (SBs) in our study[47].

4.4 Monte-Carlo simulation

To investigate the dynamic behavior of the SHBG monomer, we applied the CABS (C-alpha, beta, and side chain) coarse-grained protein model (version 0.9.18) using a Monte Carlo simulation approach. Key simulation parameters were customized for this study: the random seed was set to 3864 to ensure reproducibility, and the number of simulation cycles (Ncycle) was defined as 100, resulting in a trajectory of 2,000 models (20 × 100). To control the frequency of model output, the 'cycles between trajectory frames' ($N_{skipped}$) was set to 100, producing a total of 200,000 models (20 × 100 × 100) throughout the simulation. The simulation temperature was maintained at T = 1.2, representing conditions close to the protein's native state. All other simulation parameters followed the default settings specified by the CABS model.

4.5 Analysis of the Contact Map for Critical Subnetwork Identification

The interaction between residue pairs and the strength of their interaction were depicted through a contact map. This map provided a visual representation of the interacting residues and the magnitude of their interactions. In order to differentiate notable contacts from the negligible ones, we considered a threshold for the 'Contact' values. Analyzing the distribution of these values allowed for the determination of a threshold based on the median contact value. This threshold effectively segregated high-affinity contacts from lower ones.

A graph-based representation of the protein contact map was constructed using the NetworkX library (version 3.4.2) in Python. In this graph, individual residues were modeled as nodes, while edges represented significant residue-residue contacts that exceeded a predefined interaction threshold. By analyzing the connected components within the graph, we identified discrete structural subnetworks or 'islands', which may correspond to functionally or structurally critical regions within the SHBG protein.

We prioritized connected components by size, selecting the largest component for further analysis as it represented the most significant subnetwork in terms of involved residues.

4.5.1 Graphical Representation of the Main Island

We positioned nodes using a spring layout, ensuring a balanced spatial distribution. Node size corresponded to their degree, aiding in the identification of highly connected residues (hubs) within the network. Here the nodes were color-coded to represent their sequential position within the protein, forming a gradient that visually mapped the progression from the N-terminus to the C-terminus. Additionally, nodes representing residues with a high degree were labeled, emphasizing their potential significance in the protein structure. We adjusted edges in the graph to increase transparency, reducing visual clutter arising from the dense network of connections.

The analysis employed Python, leveraging critical libraries like Pandas for data manipulation, Matplotlib and Seaborn for data visualization, and NetworkX for graph-theoretical analyses. Handling protein structure data involved the use of the BioPython library for parsing and management.

4.5.2 Asynchronous Label Propagation for Community Detection

We employed the Asynchronous Label Propagation algorithm to analyze the contact map derived from the Root-Mean-Square Fluctuation (RMSF) distribution of protein simulations. We considered this algorithm due to its efficiency in managing/dealing large-scale networks and for revealing the inherent modular structure within the contact map.

Here are the descriptions of the Algorithm used:

Initialization: Each node (i) representing amino acid residues was initialized with a unique label L_i.

Convergence: The algorithm iterated until a steady state was achieved, with no label changes or upon reaching the maximum iterations.

Gradient Boosting Regressor for RMSF Modeling:

We utilized the Gradient Boosting Regressor to model RMSF data of protein amino acid residues, aiming to predict RMSF values based on residue properties. The regressor constructed an ensemble of weak prediction models, minimizing a loss function through gradient descent.

Loss Function Minimization: The ensemble model prediction was optimized by minimizing the loss function.

Distinguishing Clique Communities: We identified clique communities within the protein contact map to unveil tightly interconnected residue clusters. Maximal cliques were detected, ensuring that every pair of nodes within a clique was interconnected.

4.6 Protein Purification

In order to overexpress the SHBG protein, we transiently transfected HEK293 cells with plasmids mixed with transfection reagents at an optimal ratio. We cultured the cells in a serum-free medium, maintaining them in Erlenmeyer flasks or a bioreactor under stirring conditions at 37°C for 6 days. After harvesting, we centrifuged the cell culture broth to separate the supernatant, which we then loaded onto an affinity purification column at an appropriate flow rate. Finally, we analyzed the purified proteins by SDS-PAGE to assess their purity and molecular weight.

References

1 Hammond, G. L. & Bocchinfuso, W. P. Sex hormone-binding globulin: gene organization and structure/function analyses. *Hormone Research in Paediatrics* **45**, 197-201 (1996).

2 Avvakumov, G. V., Cherkasov, A., Muller, Y. A. & Hammond, G. L. Structural analyses of sex hormone-binding globulin reveal novel ligands and function. *Molecular and cellular endocrinology* **316**, 13-23 (2010).

3 Guadarrama-García, C., Bello, M. & Soriano-Ursúa, M. Molecular insights into how SHBG dimerization exerts changes on ligand molecular recognition. *The Journal of Steroid Biochemistry and Molecular Biology* **197**, 105502 (2020).

4 Hammond, G. L. Diverse roles for sex hormone-binding globulin in reproduction. *Biology of reproduction* **85**, 431-441 (2011).

5 Hammond, G. L. Potential functions of plasma steroid-binding proteins. *Trends in Endocrinology & Metabolism* **6**, 298-304 (1995).

6 Baker, M. Beyond carrier proteins: albumin, steroid hormones and the origin of vertebrates. *J Endocrinol* **175**, 121-127 (2002).

7 Hammond, G. L. Access of reproductive steroids to target tissues. *Obstetrics and Gynecology Clinics* **29**, 411-423 (2002).

8 Avvakumov, G. V., Grishkovskaya, I., Muller, Y. A. & Hammond, G. L. Resolution of the human sex hormone-binding globulin dimer interface and evidence for two steroid-binding sites per homodimer. *Journal of Biological Chemistry* **276**, 34453-34457 (2001).

9 Wu, T.-S. & Hammond, G. L. Naturally occurring mutants inform SHBG structure and function. *Molecular endocrinology* **28**, 1026-1038 (2014).

10 Bocchinfuso, W. P. & Hammond, G. L. Steroid-binding and dimerization domains of human sex hormone-binding globulin partially overlap: steroids and Ca2+ stabilize dimer formation. *Biochemistry* **33**, 10622-10629 (1994).

11 Hildebrand, C., Bocchinfuso, W. P., Dales, D. & Hammond, G. L. Resolution of the steroid-binding and dimerization domains of human sex hormone-binding globulin by expression in Escherichia coli. *Biochemistry* **34**, 3231-3238 (1995).

12 Grishkovskaya, I., Avvakumov, G. V., Hammond, G. L., Catalano, M. G. & Muller, Y. A. Steroid ligands bind human sex hormone-binding globulin in specific orientations and produce distinct changes in protein conformation. *Journal of Biological Chemistry* **277**, 32086-32093 (2002).

13 Gao, W., Bohl, C. E. & Dalton, J. T. Chemistry and structural biology of androgen receptor. *Chemical reviews* **105**, 3352-3370 (2005).

14 Hammond, G. L. & Bocchinfuso, W. P. Sex hormone-binding globulin/androgen-binding protein: steroid-binding and dimerization domains. *The Journal of Steroid Biochemistry and Molecular Biology* **53**, 543-552 (1995).

15 Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *nature* **596**, 583-589 (2021).

16 Gomes, P. S., Gomes, D. E. & Bernardi, R. C. Protein structure prediction in the era of AI: Challenges and limitations when applying to in silico force spectroscopy. *Frontiers in Bioinformatics* **2**, 983306 (2022).

17 Perrakis, A. & Sixma, T. K. AI revolutions in biology: The joys and perils of AlphaFold. *EMBO reports* **22**, e54046 (2021).

18 Evans, R. *et al.* (2022).

19 Chowdhury, S. *et al.* Evolutionary analyses of sequence and structure space unravel the structural facets of SOD1. *Biomolecules* **9**, 826 (2019).

20 Varricchio, L. *et al.* Calreticulin: challenges posed by the intrinsically disordered nature of calreticulin to the study of its function. *Frontiers in cell and developmental biology* **5**, 96 (2017).

21 Bahar, I. & Rader, A. Coarse-grained normal mode analysis in structural biology. *Current opinion in structural biology* **15**, 586-592 (2005).

22 May, L. T., Leach, K., Sexton, P. M. & Christopoulos, A. Allosteric modulation of G protein–coupled receptors. *Annu. Rev. Pharmacol. Toxicol.* **47**, 1-51 (2007).

23 Caltabiano, G. *et al.* The specificity of binding of glycoprotein hormones to their receptors. *Cellular and molecular life sciences* **65**, 2484-2492 (2008).

24 Herington, A., Ymer, S. & Stevenson, J. Identification and characterization of specific binding proteins for growth hormone in normal human sera. *The Journal of clinical investigation* **77**, 1817-1823 (1986).

25 Hammond, G., Smith, C. & Underhill, D. Molecular studies of corticosteroid binding globulin structure, biosynthesis and function. *The Journal of steroid biochemistry and molecular biology* **40**, 755-762 (1991).

26 Lin, H.-Y., Muller, Y. A. & Hammond, G. L. Molecular and structural basis of steroid hormone binding and release from corticosteroid-binding globulin. *Molecular and cellular endocrinology* **316**, 3-12 (2010).

27 Hocman, G. Human thyroxine binding globulin (TBG). *Reviews of Physiology, Biochemistry and Pharmacology, Volume 91: Volume: 91*, 45-89 (2005).

28 Zhou, A., Wei, Z., Read, R. J. & Carrell, R. W. Structural mechanism for the carriage and release of thyroxine in the blood. *Proceedings of the National Academy of Sciences* **103**, 13321-13326 (2006).

29 Raymoure, W. J., McNaught, R. W. & Smith, R. G. Reversible activation of non-steroid binding oestrogen receptor. *Nature* **314**, 745-747 (1985).

30 Hadcock, J. R. & Maibon, C. C. Regulation of receptor expression by agonists: transcriptional and post-transcriptional controls. *Trends in neurosciences* **14**, 242-247 (1991).

31 Lee, K. & Cohen, P. Beyond carrier proteins. Nuclear effects: Unexpected intracellular actions of insulin-like growth factor binding protein-3. *J Endocrinol* **175**, 33-40 (2002).

32 Hammond, G. L. Plasma steroid-binding proteins: primary gatekeepers of steroid hormone action. *The Journal of endocrinology* **230**, R13 (2016).

33 Hammond, G. L. Molecular properties of corticosteroid binding globulin and the sex-steroid binding proteins. *Endocrine reviews* **11**, 65-79 (1990).

34 Grasa, M. d. M. *et al.* Modulation of SHBG binding to testosterone and estradiol by sex and morbid obesity. *European journal of endocrinology* **176**, 393-404 (2017).

35 Qu, X. & Donnelly, R. Sex hormone-binding globulin (SHBG) as an early biomarker and therapeutic target in polycystic ovary syndrome. *International journal of molecular sciences* **21**, 8191 (2020).

36 Ramsey, J. M., Cooper, J. D., Penninx, B. W. & Bahn, S. Variation in serum biomarkers with sex and female hormonal status: implications for clinical tests. *Scientific reports* **6**, 26947 (2016).

37 Toljan, K., Grgić, F., Pavičić Baldani, D., Jurković, I. & Šprem Goldštajn, M. Sex hormone binding globulin (SHBG) as a marker of clinical disorders. *Collegium antropologicum* **40**, 199-209 (2016).

38 Hogeveen, K. N. *et al.* Human sex hormone–binding globulin variants associated with hyperandrogenism and ovarian dysfunction. *The Journal of clinical investigation* **109**, 973-981 (2002).

39 Grishkovskaya, I. *et al.* Crystal structure of human sex hormone-binding globulin: steroid transport by a laminin G-like domain. *The EMBO journal* (2000).

40 Romero, P. *et al.* Sequence complexity of disordered protein. *Proteins* **42**, 38-48 (2001).

41 Peng, K., Radivojac, P., Vucetic, S., Dunker, A. K. & Obradovic, Z. Length-dependent prediction of protein intrinsic disorder. *BMC Bioinformatics* **7**, 208, doi:1471-2105-7-208 [pii]10.1186/1471-2105-7-208 (2006).

42 Xue, B., Dunbrack, R. L., Williams, R. W., Dunker, A. K. & Uversky, V. N. PONDR-FIT: A meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* **1804**, 996-1010, doi:S1570-9639(10)00013-0 [pii]10.1016/j.bbapap.2010.01.011 (2010).

43 Dosztanyi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433-3434, doi:bti541 [pii]10.1093/bioinformatics/bti541 (2005).

44 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725 (2010).

45 Sanyal, D. *et al.* An integrated understanding of the evolutionary and structural features of the SARS-CoV-2 spike receptor binding domain (RBD). *International Journal of Biological Macromolecules* **217**, 492-505 (2022).

46 Grant, B. J., Rodrigues, A. P., ElSawy, K. M., McCammon, J. A. & Caves, L. S. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **22**, 2695-2696 (2006).

47 Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proceedings of the national academy of sciences* **99**, 7821-7826 (2002).

48 Kurcinski, M. *et al.* CABS-flex standalone: a simulation environment for fast modeling of protein flexibility. *Bioinformatics* **35**, 694-695, doi:10.1093/bioinformatics/bty685 %J Bioinformatics (2018).

49 Tuma, R. Raman spectroscopy of proteins: from peptides to large assemblies. *Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering* **36**, 307-319 (2005).

50 Rygula, A. *et al.* Raman spectroscopy of proteins: a review. *Journal of Raman Spectroscopy* **44**, 1061-1076 (2013).

51 Benevides, J. M., Overman, S. A. & Thomas Jr, G. J. Raman spectroscopy of proteins. *Current protocols in protein science* **33**, 17.18. 11-17.18. 35 (2003).

52 Lippert, J., Tyminski, D. & Desmeules, P. Determination of the secondary structure of proteins by laser Raman spectroscopy. *Journal of the American Chemical Society* **98**, 7075-7080 (1976).

53. Tuma, R., Raman spectroscopy of proteins: from peptides to large assemblies. *Journal of Raman Spectroscopy: An International Journal for Original Work in all Aspects of Raman Spectroscopy, Including Higher Order Processes, and also Brillouin and Rayleigh Scattering* **2005,** *36* (4), 307-319.

54. Rygula, A.; Majzner, K.; Marzec, K. M.; Kaczor, A.; Pilarczyk, M.; Baranska, M., Raman spectroscopy of proteins: a review. *Journal of Raman Spectroscopy* **2013,** *44* (8), 1061-1076.

55. Punzalan, J. M.; Hartono, P.; Fraser-Miller, S. J.; Leong, S. Y.; Sutton, K.; Moggre, G. J.; Oey, I.; Gordon, K. C., Fingerprinting of Semi-Refined Flaxseed Protein Using Raman Spectroscopy and Multivariate Analysis: Pulsed Electric Field (PEF)-Assisted Alkali and Aqueous Extraction Methods Alter Composition and Protein Conformation. *Journal of Raman Spectroscopy* **2025**.

56. Huang, J.; Shi, T.; Gong, B.; Li, X.; Liao, G.; Tang, Z., Fitting an optical fiber background with a weighted Savitzky–Golay smoothing filter for Raman spectroscopy. *Applied Spectroscopy* **2018,** *72* (11), 1632-1644.

57. Chowdhury, S.; Sen, S.; Banerjee, A.; Uversky, V. N.; Maulik, U.; Chattopadhyay, K., Network mapping of the conformational heterogeneity of SOD1 by deploying statistical cluster analysis of FTIR spectra. *Cellular and Molecular Life Sciences* **2019,** *76* (20), 4145-4154.