JOTCS A
Journal of the Turkish Chemical Society
Section: A
Chemistry

TURKISH CHEMICAL SOCIETY

Intrinsically Disordered Proteins by Homology Modeling and Replica Exchange Molecular Dynamics Simulations: A Case Study of Amyloid-β42

Orkid Coskuner-Weber[1*]

[1]Turkish-German University, Molecular Biotechnology, Sahinkaya Caddesi, No. 106, Istanbul 34820 Turkey.

Abstract: Homology modeling emerges as a potent tool unveiling the structural enigma of intrinsically disordered proteins (IDPs), with recent advancements such as AlphaFold2 enhancing the precision of these analyses. The process usually involves identifying homologous proteins with known structures and utilizing their templates to predict the three-dimensional architecture of the target IDP. However, IDPs lack a well-defined three-dimensional structure, and their flexibility makes it difficult to predict their conformations accurately. On the other hand, special sampling molecular dynamics simulations have been shown to be useful in defining the distinct structural properties of IDPs. Here, the structural properties of the disordered amyloid-β42 peptide were predicted using various homology modeling tools, including C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2. In parallel, extensive replica exchange molecular dynamics simulations of Aβ42 were conducted. Results from homology modeling were compared to our replica exchange molecular dynamics simulations and experiments to gain insights into the accuracy of homology modeling tools for IDPs used in this work. Based on our findings, none of the homology modeling tools used in this work can fully capture the structural properties of Aβ42. However, C-I-TASSER yields a radius of gyration and tertiary structure properties that are more in accord with the simulations and experimental data rather than I-TASSER, Phyre2, SwissModel, and AlphaFold2.

Keywords: Intrinsically disordered proteins, Amyloid-β42, Homology modeling, Replica exchange molecular dynamics simulations.

Submitted: March 22, 2024. **Accepted:** June 22, 2024.

Cite this: Coskuner Weber O. Intrinsically Disordered Proteins by Homology Modeling and Replica Exchange Molecular Dynamics Simulations: A Case Study of Amyloid-β42. JOTCSA. 2024;11(3): 1151-64.

DOI: https://doi.org/10.18596/jotcsa.1457169

***Corresponding author's E-mail:** weber@tau.edu.tr

1. INTRODUCTION

The history of homology modeling of IDPs is relatively recent compared to the longer history of homology modeling for well-folded proteins (1,2). In fact, the field of IDPs has gained attention over the past two decades as the prevalence and functional importance of IDPs that lack a stable, well-defined three-dimensional structure have been recognized more and more (3-6) Homology modeling, initially developed for structured proteins, saw an expansion of its application to IDPs in the mid-2000s (7). However, we should note that the lack of well-defined templates and the dynamic nature of IDPs pose significant hurdles. Advances in computational methods, such as machine learning and deep learning, have played a crucial role in improving the accuracy of predicting the structures of IDPs (8). AlphaFold2, developed by DeepMind, made headlines with its success in predicting the structures of both structured and disordered proteins (9-14). Current research continues to focus on refining methods for homology modeling for IDPs. The incorporation of experimental data, such as NMR and cryo-electron microscopy, along with improved algorithms, contributes to the ongoing development of accurate predictive models.

Replica exchange molecular dynamics (REMD) simulations offer a valuable computational approach for studying the structures and dynamics of IDPs (see, for example 15-19). REMD simulations involve running multiple replicas of a system at different temperatures and periodically exchanging their conformations (20). This exchange allows for enhanced conformational sampling, particularly in regions of the energy landscape that may be challenging to explore with conventional molecular

dynamics simulations (21). IDPs often exist as dynamic ensembles, sampling a wide conformational space. REMD simulations facilitate the exploration of this space, enabling the observation of the distribution of different conformations and transitions between them (21).

Amyloid-β42 (Aβ42) is an intrinsically disordered protein with 42 amino acid residues and is at the center of Alzheimer's disease (5,22). It has been studied experimentally and computationally extensively (see, for example 23-28). However, a comparative study of Aβ42 using the most widely used homology modeling tools and REMD simulations is currently lacking in the literature. Such a comparative study is crucial for cross-validation of structural models related to IDPs. REMD simulations and experiments can be used to refine and validate the homology-based models for IDPs. Therefore, in this study, different homology modeling tools used widely for the structural prediction of IDPs (C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2 were utilized to obtain the three-dimensional structures of Aβ42. Furthermore, REMD simulations of Aβ42 were conducted. The obtained structural properties were compared to each other and experiments.

2. MATERIALS AND METHODS

The three-dimensional structures of Aβ42 were predicted using its amino acid sequence by C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2. C-I-TASSER (Contact-guided Iterative threading ASSEmbly Refinement) is a new method extended from I-TASSER (29). Starting from an amino acid sequence, C-I-TASSER generates inter-residue contact maps using various deep neural network predictors (29). Next, it identifies the structural templates from the PDB by multiple threading approaches with full-length atomic models assembled by contact map-guided replica exchange Monte Carlo simulations (29). The large-scale benchmark tests showed that C-I-TASSER generates significantly more accurate models than I-TASSER, especially for proteins that do not have homologous templates in the PDB (29).

I-TASSER (Iterative Threading ASSEmbly Refinement) is a homology modeling method that combines threading, ab initio modeling, and iterative refinement process to generate three-dimensional models for proteins (30). Threading is the initial step where I-TASSER searches for homologous protein structures in a structural database. It identifies template structures that are similar to the target protein sequence. Threading involves aligning the target sequence onto these templates to create a preliminary three-dimensional model. After threading, I-TASSER generates additional models using ab initio modeling techniques. This consists of predicting the protein's structure based on its amino acid sequence without relying on template structures. The threaded and ab initio models are then assembled into a pool of candidate models. These models are ranked based on their energy and structural compatibility with the input sequence and experimental constraints, if available. The algorithm employs an iterative optimization process to refine the models. This involves simulation and optimization steps to improve the accuracy of the models. The final step consists of selecting the best model from the refined pool of candidates. The model is chosen based on its energy score, structural quality, and compatibility with experimental data, if available (31).

Phyre2 (Protein Homology/analogy Recognition Engine) is a homology modeling tool that is used for protein structure prediction (32). It utilizes a combination of homology modeling, profile-profile matching, and ab initio methods to generate three-dimensional structural models for a given amino acid sequence. Phyre2 initially attempts to find homologous proteins with known structures in sequence databases. If a homologous protein is found, it uses the structural information from that template to predict the structure of the target protein. Homology modeling is effective when there is a significant similarity between the target sequence and the template. In cases where conventional homology modeling may not be applicable due to low sequence similarity, Phyre2 employs profile-profile matching techniques. This involves comparing the target sequence's profile with profiles of known protein structures. If homology modeling and profile-profile matching do not yield suitable templates, Phyre2 resorts to ab initio modeling. This involves predicting the protein structure without relying on known templates. The models undergo refinement to improve their accuracy and quality. This step may include energy minimization and optimization processes. The algorithm then evaluates the quality of generated models using different metrics, such as energy scores and structural consistency.

The SwissModel generates protein structures, and it operates on the principle of homology modeling, where the structure of a target protein is predicted based on the known structures of homologous proteins (33). It begins by searching a database of experimentally determined protein structures, such as PDB, to find homologous proteins with known structures that are similar to the target protein. Then, it performs a sequence alignment, which aids in establishing the correspondence between amino acids in the target sequence and the templates. It builds a three-dimensional model using the known three-dimensional structure of a homologous template as a starting point. This step involves aligning the target sequence onto the template structure and adjusting the backbone and side-chain conformations to fit the target sequence. The homology model is subjected to energy minimization and refinement steps to improve the geometry and overall quality of the model. It assesses the quality of generated models using various criteria, including stereochemical properties, bond lengths, and angles (33).

AlphaFold2 is an advanced artificial intelligence system designed for protein folding prediction (12,34,35). It gained significant attention for its

exceptional performance in the Critical Assessment of Structure Prediction (CASP) competition. It employs deep learning to predict protein structures. It involves a deep neural network that is trained on a large dataset of protein structures. This model is trained on a diverse set of known protein structures from PDB. The model learns to recognize patterns and relationships between amino acid sequences and corresponding three-dimensional structures. It uses an attention mechanism inspired by the Transformer architecture, which is a kind of neural network architecture usually used in natural language processing tasks. This attention mechanism enables the model to capture long-range interactions between amino acids in the sequence. It incorporates information from multiple sequence alignments, which considers evolutionary relationships between related proteins. This helps to improve the accuracy, especially for regions where the sequence similarity is low. It predicts inter-residue distances between pairs of amino acids in the sequence. This distance prediction provides important information about the spatial relationships between different parts of the protein. It assembles the three-dimensional structure of the protein and generates a probability distribution for each atom's position in the structure. The refinement process involves iterative optimization processes to enhance the geometry and overall quality of the predicted structures. It uses an ensemble approach by generating multiple models for a given protein.

These homology modeling and structure prediction tools were chosen because these are heavily used in the studies of IDPs in commercial and academic settings. See, for example, references (7,36-40).

All-atom REMD simulations of the Aβ42 monomer in an aqueous solution environment were performed with the AMBER22 software package (41). An extended structure was used initially for Aβ42. T-REMD simulations were conducted using the Amber ff99SB parameters for the protein, and the Onufriev-Bashford-Case generalized Born implicit water model (42). These models were chosen to avoid confined aqueous volume effects using an explicit water model, and we recently showed that these parameters yield results in excellent agreement with experiments for Aβ (42,43). Specifically, T-REMD simulations using the Amber ff99SB parameters for Aβ along with the Onufriev-Bashford-Case generalized Born implicit model for water were shown to yield structural properties in excellent agreement with NMR experiments (42,43). Langevin dynamics was used to control the temperature with a collision frequency of 2 ps^{-1} (21). The particle mesh Ewald method was used for treating the long-range interactions with a cutoff value of 25 Å (44). The temperatures of each replica for Aβ42 were exponentially distributed between 280 K and 400 K, yielding an exchange ratio of 0.74 for Aβ42. After energy minimization of the initial structure with the steepest descent method, the initial conformation was equilibrated for 200 ps for each replica (21). The peptide was then simulated using an integration timestep of 2 fs for each replica, and trajectories were saved every 500 steps. Exchanges between replicas were attempted every five ps. The system was simulated for 100 ns for each replica with a total simulation time of 2.4 μs. 60 ns are required to reach convergence, which is in agreement with earlier studies (45). The structural properties were then calculated for Aβ42 obtained after convergence from the replica closest to 310 K. The end-to-end distance values, the radius of gyration, secondary structure properties per residue, and intra-molecular contact map, as well as salt bridges, were calculated and compared to those obtained using I-TASSER, C-I-TASSER, Phyre2, SwissModel, and AlphaFold2 as well as to experiments. Following our earlier studies, intramolecular peptide interactions occur when the centers of mass of two residues are within 9.0 Å of each other (45). An interaction is considered to be a salt bridge when a hydrogen bond exists between the two residues and the hydrogen-bonded residues possess opposite electrostatic charges. If the distance between the donor and acceptor atoms of the hydrogen bond is ≤ 2.5 Å and the hydrogen bond angle is larger than 113°, then a hydrogen bond exists (45).

3. RESULTS AND DISCUSSION

The top-ranked three-dimensional structures of Aβ42 obtained from C-I-TASSER, I-TASSER. Phyre2, SwissModel, and AlphaFold2 are depicted in Figure 1. As seen in Figure 1, different homology modeling methods yield varying three-dimensional structures for Aβ42. Selected Aβ42 structures obtained from REMD simulations are illustrated in Figure 2.

To gain deeper insights into the structural properties of these three-dimensional model structures for Aβ42, the end-to-end (R_{EE}) and radius of gyration (R_g) values were calculated and compared to experiments (Table 1) (46-48).

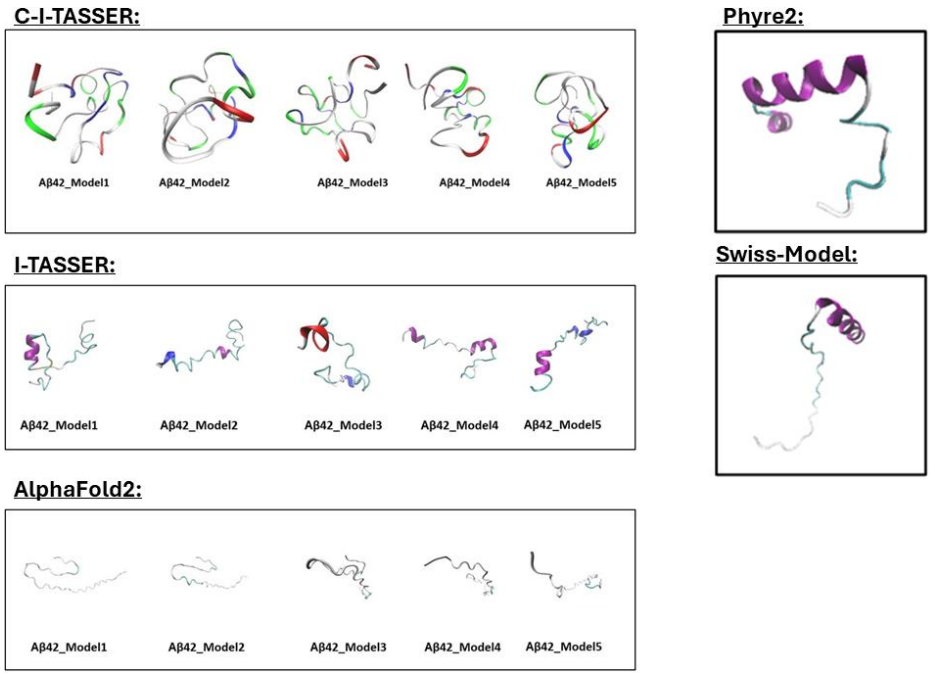

Figure 1: Top-ranked three-dimensional structures from C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2 for Aβ42.

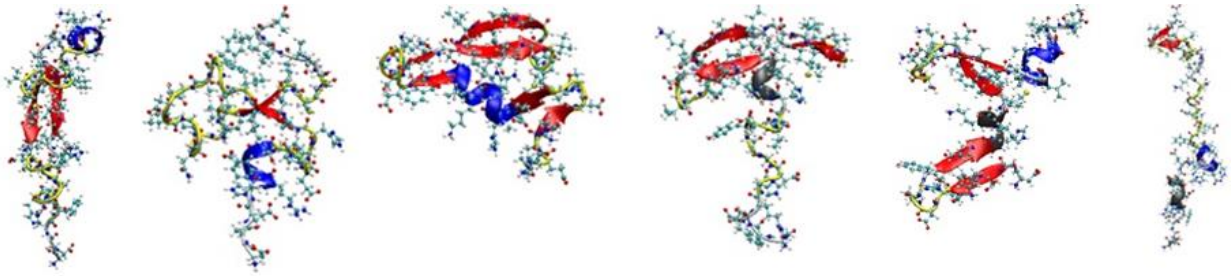

Figure 2: Selected three-dimensional structures for Aβ42 from REMD simulations.

Table 1: Calculated R_{EE} and R_g values for Aβ42 and their comparison to experiments.

	R_{EE} (Å)	R_g (Å)
C-I-TASSER		
Model 1	9.64	8.53
Model 2	8.17	8.41
Model 3	7.93	8.47
Model 4	7.85	8.34
Model 5	4.58	8.53
I-TASSER		
Model 1	33.57	12.53
Model 2	35.23	14.49
Model 3	18.91	10.74
Model 4	39.72	14.96
Model 5	37.43	14.58
Phyre2		
Model	39.91	14.96
SwissModel		
Model	47.47	17.96
AlphaFold2		
Model 1	50.77	19.82
Model 2	47.48	18.11
Model 3	41.84	16.84
Model 4	45.93	17.08
Model 5	39.20	17.03
REMD Simulations		
Average Values	37.73 ± 8.69	10.81 ± 1.83
Experiments (47,48)		
	35.66 ± 8.51	9.00 ± 1.00

REMD simulations yield R_{EE} and R_g values that are in agreement with the experiments. The R_g values obtained for the three-dimensional models for Aβ42 using C-I-TASSER show excellent agreement with the results obtained from REMD simulations and experiments. However, the R_{EE} values for Aβ42 using C-I-TASSER do not agree with those values obtained from REMD simulations or experiments. Instead, the R_{EE} values obtained using I-TASSER and Phyre2 show more agreement with REMD simulations and experiments. Interestingly, AlphaFold2, which is widely used for IDPs, cannot capture the experimental R_{EE} and R_g values for Aβ42. To gain even more insights, the potential of mean force (PMF) surfaces based on R_{EE} and R_g values were computed using the trajectories obtained from REMD simulations (Figure 3). Figure 3 shows that there are two most energetically stable basins for Aβ42 (basin IA and basin IB). These most preferred basins are located at R_g values ranging from 10.6 Å to 11.4 Å and 10.3 Å to 10.7 Å. The first basin has Aβ42 structures that have 10% α-helix and 6% β-sheet structures. The second most preferred basin has structures with about 22% α-helix and only 1% β-sheet structures. Energetically preferred structures are located in the basin with R_{EE} and R_g values closer to experimental data, as seen in Figure 3 (see also Table 1), indicating that most homology modeling methods used in this work yield for Aβ42 three-dimensional structures that are not preferred energetically. Incorporating constraints related to R_{EE} and R_g experimental values could help improve the outcome of homology modeling tools for IDPs.

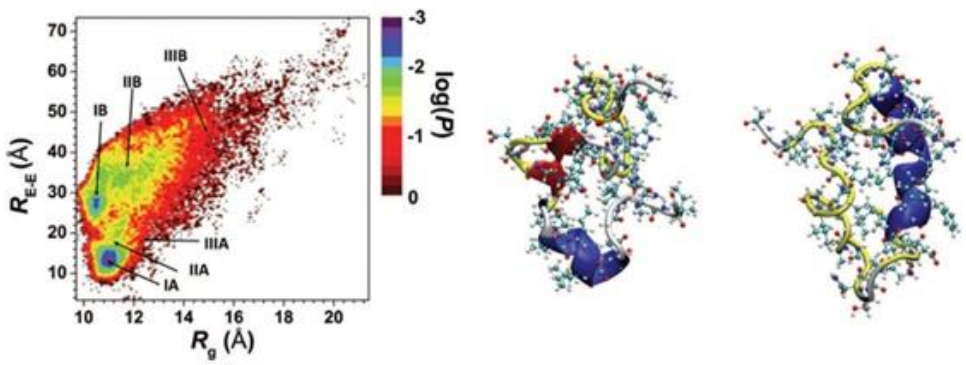

Figure 3: The potential of mean force surface area based on R_g and R_{EE} values of Aβ42 using the trajectories obtained from REMD simulations.

The secondary structure elements per residue, along with their abundances using the trajectories obtained from REMD simulations, were calculated. REMD simulations yield overall 11.0% α-helix, 10.9% 3_{10}-helix, 0.4% π-helix, 3.4% β-sheet, 25.1% turn and 49.2% random coil structure for Aβ42. To gain a deeper insight into the distinct structuring – given the crucial roles of α-helix and β-sheet structure adopting residues in Alzheimer's disease – the probabilities per residue of these secondary structure elements were calculated (Figure 4).

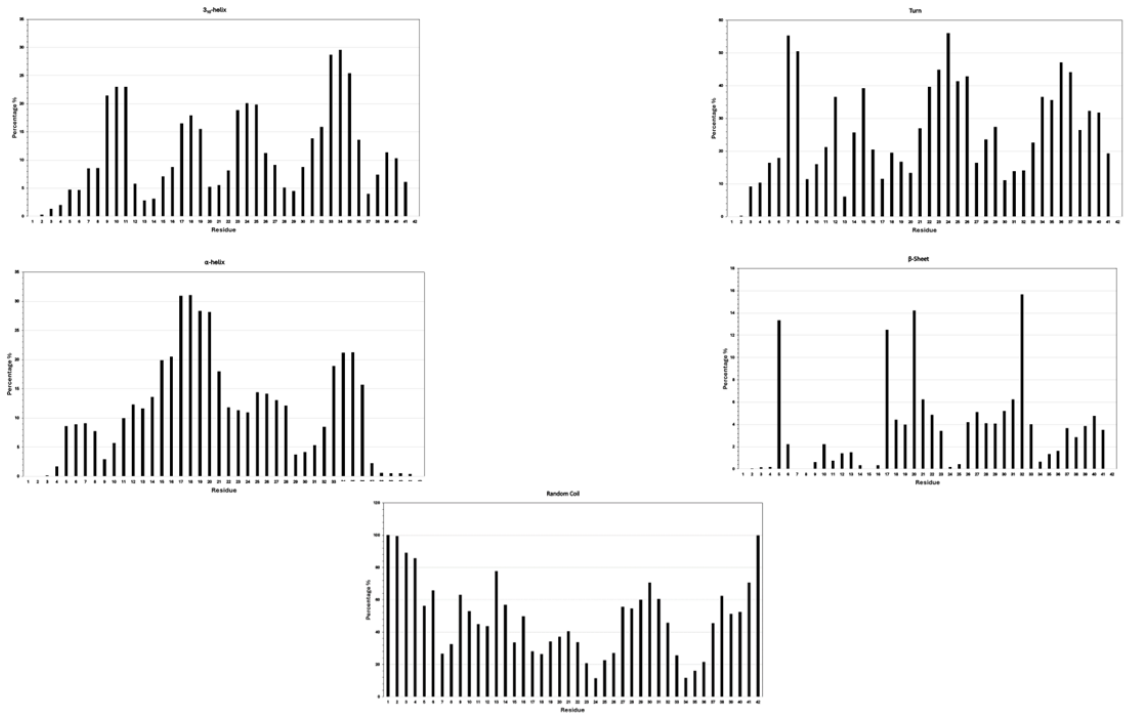

Figure 4: The calculated secondary structure elements, along with their probabilities for Aβ42, using the trajectories obtained from REMD simulations.

The most abundantly formed secondary structure elements are the turn and random coil structures. B-sheet structure formation in Aβ42 is crucial due to its involvement in oligomer, fibril, and aggregate formation processes in Alzheimer's disease. Therefore, it is crucial to adapt tools that can detect the residues forming β-sheet structure. Based on REMD simulations, β-sheet formation occurs at Arg5, Glu11, Val12, Phe19-Glu22, Gly25-Asn27, Gly33, Met35, Val36, Val40 and Ile41 with probabilities up to 23%. A-helix formation is detected at Arg5-Gln15, Lys16-Asp23, and Gly29-Val36. Turn structure formation is detected at Arg5-Val12 and Glu22-Lys28, with the largest abundance occurring in the Ala21-Ala30 region. These findings support the NMR measurements that reported bend-like and turn conformations for the Asp7-Glu11 and Phe20-Ser26 regions (48). Experiments also demonstrated turn structure formation for the Ala21-Ala30 region (49). A-helix formation occurs in the N-terminal, mid-domain, and C-terminal regions. Still, residues in the mid-domain region form a more abundant α-helix structure than those in the N-terminal and C-terminal regions. Experiments and computational studies also detected the abundant β-sheet structure formation in the C-terminal region and showed an agreement with the findings reported herein (48,50). The calculated secondary structure abundances of Aβ42 conformations located in different basins on the PMF surface (Figure 2) are listed in Table 2.

Table 2: The calculated secondary structures, along with their probabilities for Aβ42 structures located in each basin on the PMF surface.

	Basin IA	**Basin IIA**	**Basin IIIA**	**Basin IB**	**Basin IIB**	**Basin IIIB**
α-helix (%)	10.0	14.5	15.3	21.5	15.8	15.2
3_{10}-helix (%)	9.1	9.5	11.0	9.5	11.4	11.1
β-sheet (%)	6.2	3.6	2.4	1.1	2.6	2.9
Turn (%)	32.4	29.5	28.1	28.2	30.3	29.5
Coil (%)	41.6	42.5	42.7	39.6	39.7	41.0

Overall, REMD simulations show an in-depth representation of secondary structure properties of intrinsically disordered proteins and agree with experiments overall. Additionally, the secondary structure components of the three-dimensional models for Aβ42 obtained from C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2 were calculated. The obtained results are presented in Scheme 1. C-I-TASSER, I-TASSER, and AlphaFold2 yield bend and turn structures for the Ala21-Ala30 region of Aβ42 in agreement with experiments (49). None of the homology modeling tools studied in this work can capture the distinct β-sheet structuring in the C-terminal region of the peptide. We should note here that the detection of amino acid residues forming β-sheet structure is crucial because these residues play a role in the oligomerization, fibrillation, and aggregation processes of Aβ42 in Alzheimer's disease (5). In addition, the prominent distinct α-helix formation in the mid-domain region can be obtained using Phyre2 and I-TASSER homology modeling tools. All in all, these homology modeling tools cannot fully capture the distinct structuring of the intrinsically disordered Aβ42 peptide. The incorporation of dynamic effects in homology modeling methods may improve the outcome of homology modeling tools for inherently disordered proteins.

Sequence:
D1A2E3F4R5H6D7S8G9Y10E11V12H13H14Q15K16L17V18F19F20A21E22D23V24G25S26N27K28G29A30I31I32G33L34M35V36G37G38V39V40I41A42
C-I-TASSER:
Model 1: CCSCSGGGGTSSSCCCCTTTTTSCBTCTTCSBSSCSSSSCCC
Model 2: CCSSCSTTTTTSSCCCCSSSTTSCBTCTTCSBSSCSSSSCCC
Model 3: CTTBCCGGGTTCCCCCCSTTSSSCSSCTTTSCSCCTTTTCBC
Model 4: CCSCCCTTGGGTSCCBCSSSSTTCSCCTTCCCBCSSSSSCCC
Model 5: CCSSSCTTTTSSSCCCCSSSSSSSCCCSSCCCCCCSSSCCCC
I-TASSER:
Model 1: CCSTTTSBTTBCCCCCCBBTHHHHHHHTTTSCCBSCCSTTBCC
Model 2: CCTTTTTSSSSCTTCTTTTTHHHHTTSTTTSCSSBTBTBBCC
Model 3: CCIIIIITTTCCCSSCSCTTTTTTTTTGGGGCSSCBTTBSCC
Model 4: CHHHHHTBTTBCCCCCCCSHHHHHHHHHSSSCSCSCTTTTCCC
Model 5: CTTTTTHHHHSTTCSCCBTTBTTTTTTSCGGGCSTTSTTCCC
Pyhre2:
Model: CCCCCCSSSCSHHHHHHHHHHHHHHCCTTHHHHHHHHHHHHHC
Swiss-Model:
Model: CCCCCCCCCCCCCCCCCCCCCCCCCSHHHHHHHHHHHHHHHC
AlphaFold2:
Model 1: CCCCCCCCCCCCCCCCCCCCCCCCCSSSCCCCCCCTTSCCCC
Model 2: CCCCCCCCCCCCCTTCCCCCSSSCCSSCCCCCCCCSSSCCCC
Model 3: CCCSCSCCCCCCCCCCCCCCCCCCCSSSCCCCCCCSSSCCCC
Model 4: CCCCCCCCCCCCCCCCCCCCCCCCSCTTCCCCCCSSSSCCCC
Model 5: CCCCCCCSCCCCCCCCCCCCCCCCSSCCCCCCCCCTTSCCCC

Scheme 1: The calculated secondary structure elements per residue of Aβ42 using the three-dimensional structures obtained from C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2. H is for α-helix, B for isolated β-bridge, G for 3_{10}-helix, I for π-helix, T for turn, and S for bend structure.

The tertiary structure properties were studied by means of intra-molecular interactions. Figure 5 shows the calculated tertiary structure properties for Aβ42 using the trajectories obtained from REMD simulations. We note stark interactions between the central hydrophobic core (CHC) region and the N-terminal region, the N- and C-terminal regions, and the CHC region and the C-terminal region. These findings are in agreement with those of Yang and Teplow for Aβ42 (50).

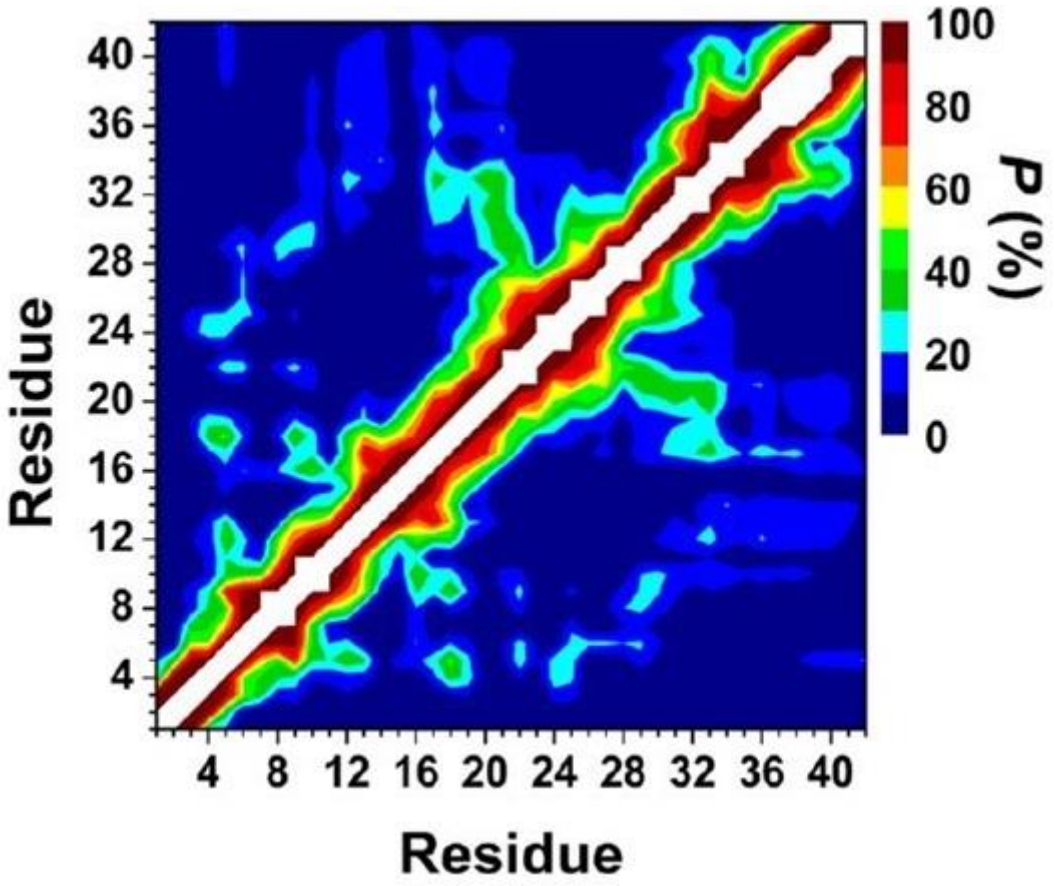

Figure 5: The calculated intra-molecular contact map for Aβ42 using the trajectories obtained from REMD simulations.

The intra-molecular contact maps for the three-dimensional Aβ42 models obtained from C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2 were calculated as well. Figures 6-10 present the findings from these calculations.

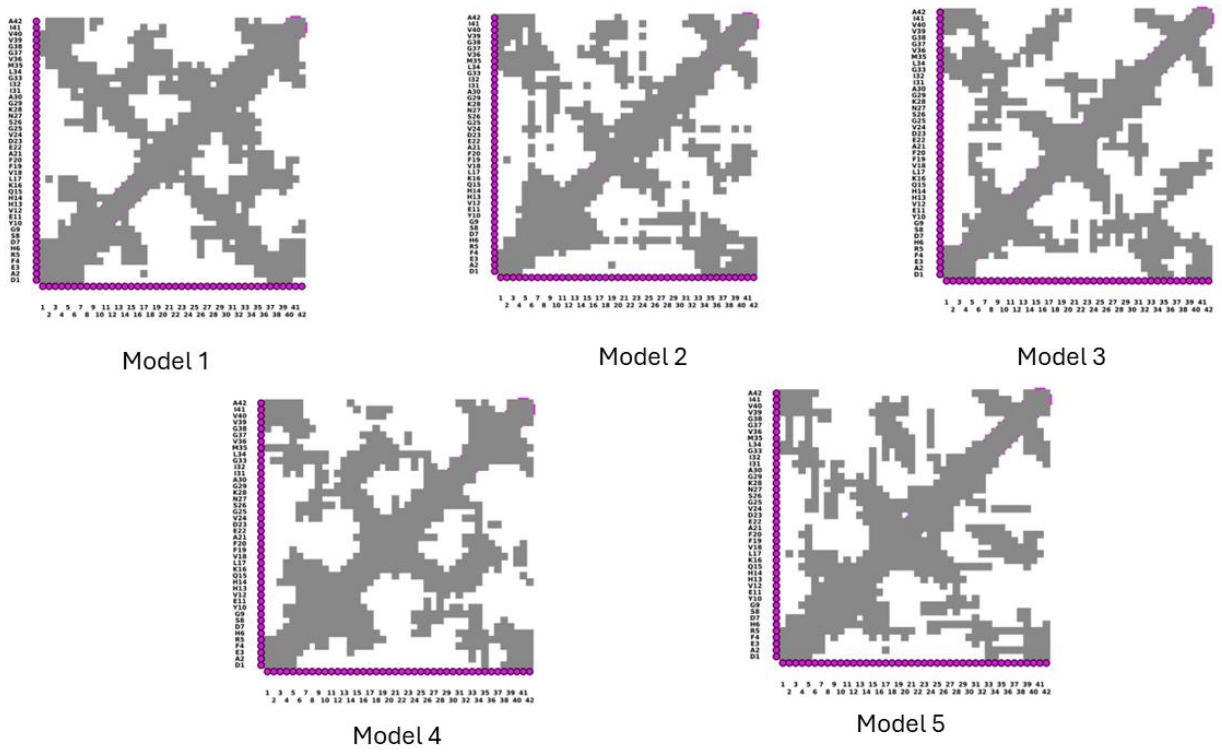

Figure 6: The calculated intra-molecular interaction maps for Aβ42 using the three-dimensional models obtained from C-I-TASSER.

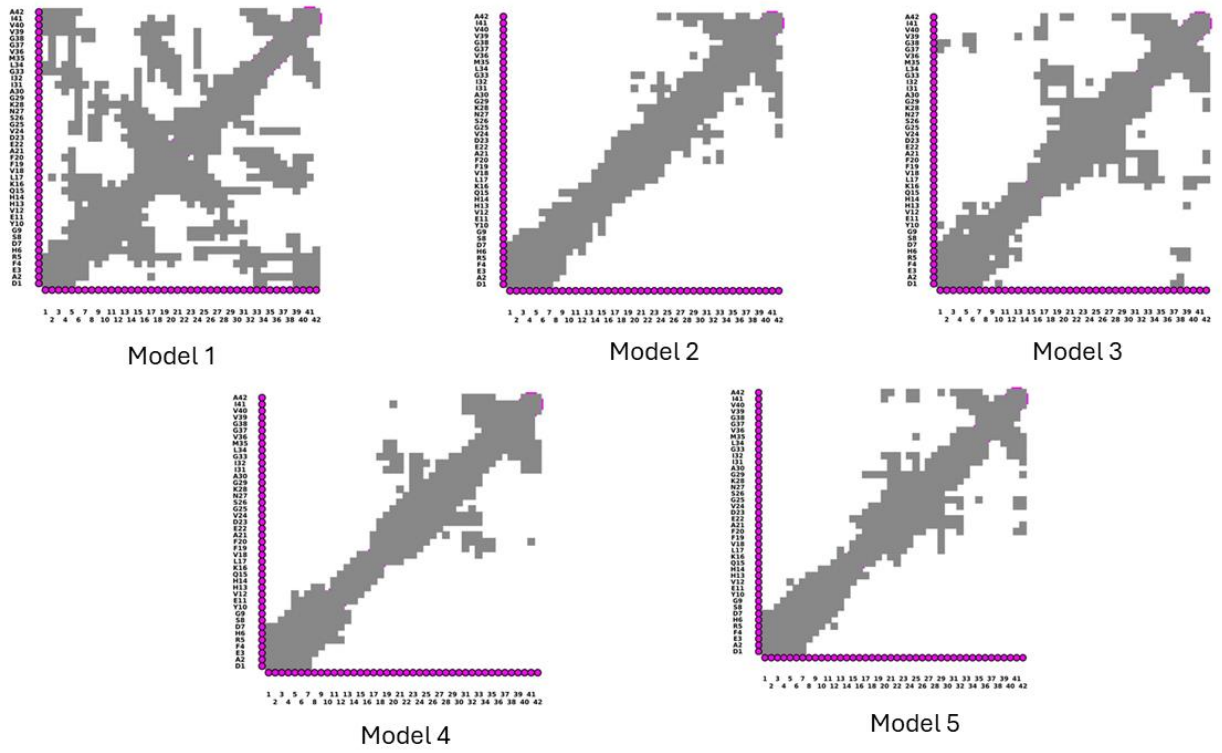

Figure 7: The calculated intra-molecular interaction maps for Aβ42 using the three-dimensional models obtained from I-TASSER.

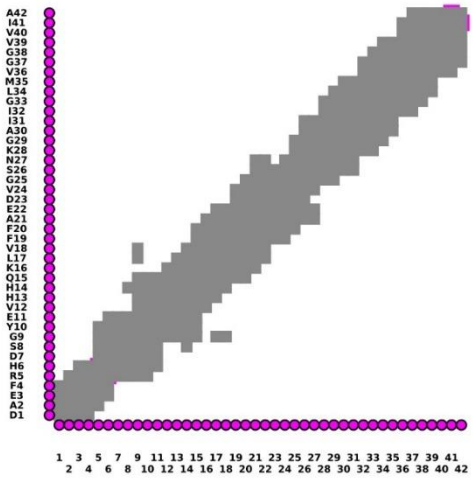

Figure 8: The calculated intra-molecular interactions map for Aβ42 using the three-dimensional models obtained from Phyre2.

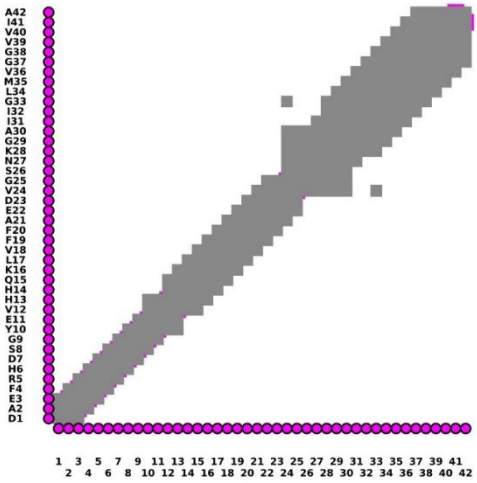

Figure 9: The calculated intra-molecular interactions map for Aβ42 using the three-dimensional models obtained from Swiss-Model.

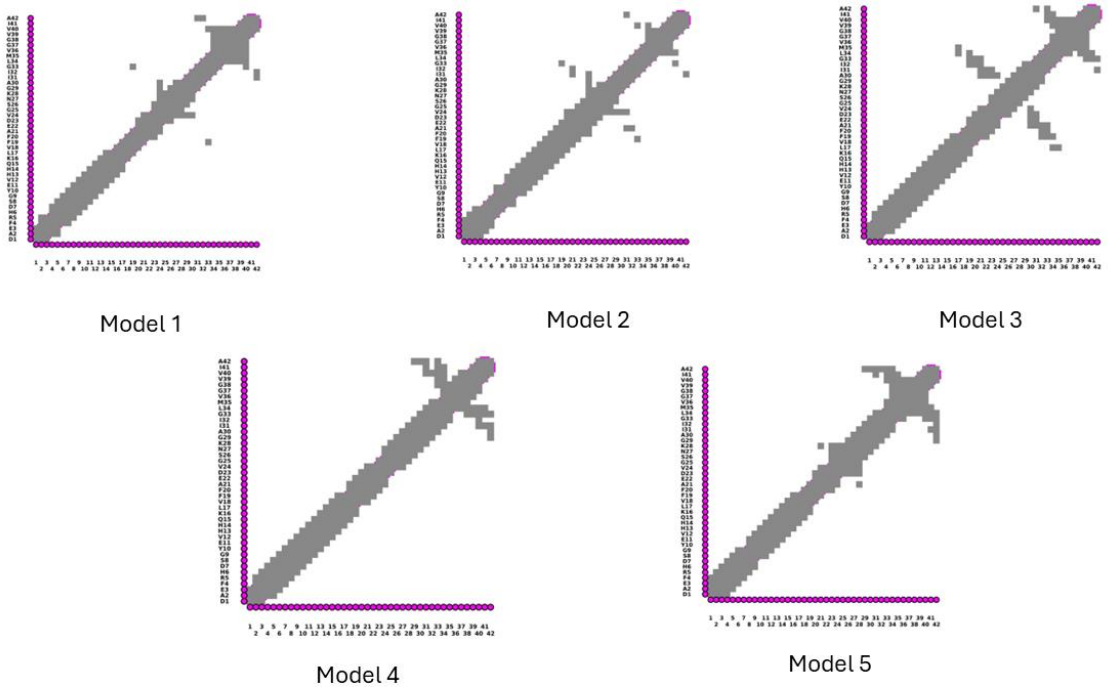

Figure 10: The calculated intra-molecular interaction maps for Aβ42 using the three-dimensional models obtained from AlphaFold2.

As seen in Figure 6, C-I-TASSER can capture even the weak interactions between the N-terminal and C-terminal regions, CHC region and C-terminal region, N-terminal and mid-domain regions, as well as mid-domain and C-terminal regions of Aβ42. On the other hand, such interactions could be obtained using Model 1 for Aβ42 from I-TASSER (Figure 7). Furthermore, Model 3 for Aβ42 from I-TASSER represents the CHC region and C-terminal interactions as well. Interestingly, intra-molecular interaction maps calculated using the three-dimensional models from Phyre2, Swiss-Model, and AlphaFold2 cannot reproduce the data obtained from REMD simulations. Given the wide usage of AlphaFold2 in the studies of Aβ and IDPs in general, including its implementation in databases of IDPs (51), these findings indicate that caution has to be given to such studies since AlphaFold2 is not capable of fully reproducing the structural properties of Aβ.

Table 4 lists the calculated salt bridges for Aβ42 using the trajectories obtained from REMD simulations.

Table 5 lists the salt bridges calculated using the three-dimensional models from C-I-TASSER, I-TASSER, Phyre2, Swiss-Model, and AlphaFold2.

Table 4: The salt bridges of Aβ42 from REMD simulations.

Residue	**Residue**	**Probability (%) R(C-N) ≤ 6.0 Å**
Arg5	Glu3	59.6
Arg5	Glu22	26.7
Arg5	Asp1	15.0
Arg5	Glu11	14.1
Arg5	Ala42	21.3
Lys16	Glu11	8.0
Lys28	Glu22	13.2
Lys28	Asp23	7.1
Asp1	Glu3	5.3
Arg5	Asp7	1.2
Lys16	Asp7	11.1
Lys16	Glu3	1.8

Table 5: The calculated salt bridges for Aβ42 using the three-dimensional structures obtained from C-I-TASSER.

C-I-TASSER		
Model 1		
Residue	Residue	Distance (Å)
Arg5	Asp1	2.9
Arg5	Asp7	4.9
Asp7	His14	5.0
Glu11	His14	5.8
Model 2		
Residue	Residue	Distance (Å)
Arg5	Glu11	2.6
Lys16	Asp23	2.5
Model 3		
Residue	Residue	Distance (Å)
Arg5	Glu3	3.5
Asp7	His6	5.6
Glu11	His13	3.8
Lys28	Glu11	2.4
Lys16	Asp23	2.4
Model 4		
Residue	Residue	Distance (Å)
Arg5	Asp1	5.9
Lys16	Asp7	5.6
Glu11	His13	5.8
Glu11	Lys28	2.5
Lys16	Asp23	2.5
Model 5		
Residue	Residue	Distance (Å)
Arg5	Asp7	2.8
Asp23	His14	5.6
Lys16	Asp23	3.2

Table 6: The calculated salt bridges for Aβ42 using the three-dimensional structures obtained from I-TASSER, Phyre2, and Swiss-Model.

I-TASSER		
Model 1		
Residue	Residue	Distance (Å)
Arg5	Glu3	2.7
Glu3	His6	5.5
Glu11	His14	4.7
Model 2		
Residue	Residue	Distance (Å)
Arg5	Asp1	2.7
Asp1	His6	5.3
Glu11	His14	5.8
Glu22	Lys28	2.5
Model 3		
Residue	Residue	Distance (Å)
Asp1	His6	5.0
Arg5	Glu11	5.7
Glu11	His13	3.7
Lys16	Glu22	2.5
Model 4		
Residue	Residue	Distance (Å)
Arg5	Asp1	3.8
Asp1	His6	5.6
Asp7	His13	5.3
Glu11	His13	5.7
Glu11	His14	2.8
Lys16	Glu22	3.3
Model 5		
Residue	Residue	Distance (Å)
Arg5	Asp1	2.7
Glu11	His14	5.0
Phyre2		
Residue	Residue	Distance (Å)
Arg5	Glu3	5.0
Swiss-Model		
Residue	Residue	Distance (Å)
Glu11	His13	4.9

Table 7: The calculated salt bridges for Aβ42 using the three-dimensional structures obtained from AlphaFold2.

AlphaFold2		
Model 1		
Residue	Residue	Distance (Å)
Glu11	His13	5.3
Model 2		
Residue	Residue	Distance (Å)
Glu11	His13	5.5
Model 3		
Residue	Residue	Distance (Å)
Arg5	Glu3	4.7
Glu11	His13	4.5
Model 4		
Residue	Residue	Distance (Å)
Glu11	His13	3.4
Model 5		
Residue	Residue	Distance (Å)
Glu11	His13	5.4

As seen in Tables 4-7, none of the homology modeling methods used in this work can fully capture the salt bridges obtained from REMD simulations. We should note that C-I-TASSER yields results closer to REMD simulations. However, the results are still not fully captured for the full spectrum of the salt bridges. AlphaFold2 performs poorly in detecting the salt bridges of Aβ42.

All in all, the structural properties for the intrinsically disordered Aβ42 peptide reported by experiments or REMD simulations cannot be fully captured by C-I-TASSER, I-TASSER, Phyre2, Swiss-Model, and AlphaFold2.

We should mention here that I-TASSER relies heavily on existing protein structures (templates) in the Protein Data Bank (PDB). If there are no suitable templates for the target protein, the predictions can be less accurate. On the other hand, the accuracy of C-I-TASSER heavily depends on the quality and correctness of the constraints provided. AlphaFold2's performance is influenced by the quality and quantity of existing protein structures in its training dataset. Proteins with novel folds or those not well-represented in the training data may have lower prediction accuracy. Furthermore, AlphaFold2 predicts static structures and does not inherently provide insights into protein function and dynamics, which are critical for understanding biological processes. Moreover, Phyre2 relies heavily on template-based modeling, which means it can only be as accurate as the template it uses. Like AlphaFold2, Phyre2 provides only static structures and does not account for the dynamic nature of proteins. Proteins often undergo conformational changes that are crucial for their functions, which a single static model cannot capture. Swiss-Model relies on the availability of suitable templates. Swiss-Model generates static structures as well, but these do not capture the dynamic nature of proteins. Errors or discrepancies in template annotation or structural data can propagate into the model, leading to inaccurate predictions. Despite this, REMD simulations are computationally expensive because they require running multiple replicas in parallel, each at different temperatures. The accuracy of the results depends on the quality of the force field parameters and the chosen boundary conditions. Inaccurate force fields can lead to erroneous interpretations of the conformational space.

We also should mention here that C-I-TASSER utilizes predicted residue-residue contact maps from deep learning algorithms. These contact maps provide additional structural constraints that can significantly improve the accuracy of the model based on our findings, especially for regions or proteins where threading templates are less reliable.

4. CONCLUSION

Homology modeling methods, including C-I-TASSER, I-TASSER, Phyre2, SwissModel, and AlphaFold2, are used in the studies of IDPs, including Aβ. In this work, C-I-TASSER, I-TASSER, Phyre2, Swiss-Model, and AlphaFold2 were used to generate the three-dimensional models for Aβ42, which is at the center of Alzheimer's disease. In parallel, extensive REMD simulations of Aβ42 were conducted. Results obtained from different homology modeling methods were compared to the results obtained from REMD simulations and available experimental data. None of the homology modeling methods used in this study can fully reproduce the REMD simulation results or available experimental data. Given the wide usage of these homology modeling methods in the studies of IDPs, these findings show that outermost care must be provided to such studies.

Specifically, C-I-TASSER performs better than I-TASSER, Phyre2, Swiss-Model, and AlphaFold2 in terms of the radius of gyration, parts of secondary structure, parts of salt bridges, and tertiary structure properties. However, the end-to-end distance and obtained full spectrum of salt bridges, as well as the full spectrum of the secondary structure properties, cannot be reproduced accurately using C-I-TASSER either. Surprisingly, AlphaFold2 performs poorly for Aβ42. Given the significant usage of AlphaFold2 in the studies of IDPs, our findings show that the incorporation of end-to-end distance and radius of gyration constraints in further development of AlphaFold2 could improve the outcomes for IDPs.

5. CONFLICT OF INTEREST

The author declares no conflict of interest.

6. REFERENCES

1. Hameduh T, Haddad Y, Adam V, Heger Z. Homology modeling in the time of collective and artificial intelligence. Comput Struct Biotechnol J [Internet]. 2020 Jan 1;18:3494–506. Available from: <URL>.

2. Coskuner-Weber O, Uversky VN. Current stage and future perspectives for homology modeling, molecular dynamics simulations, machine learning with molecular dynamics, and quantum computing for intrinsically disordered proteins and proteins with intrinsically disordered regions. Curr Protein Pept Sci [Internet]. 2024 Feb 17;25(2):163–71. Available from: <URL>.

3. Uversky VN. Introduction to Intrinsically disordered proteins (IDPs). Chem Rev [Internet]. 2014 Jul 9;114(13):6557–60. Available from: <URL>.

4. Uversky VN. Dancing protein clouds: The strange biology and chaotic physics of intrinsically disordered proteins. J Biol Chem [Internet]. 2016 Mar 25;291(13):6681–8. Available from: <URL>.

5. Coskuner-Weber O, Mirzanli O, Uversky VN. Intrinsically disordered proteins and proteins with intrinsically disordered regions in neurodegenerative diseases. Biophys Rev [Internet]. 2022 Jun 8;14(3):679–707. Available from: <URL>.

6. Dunker AK, Obradovic Z. The protein trinity—linking function and disorder. Nat Biotechnol [Internet]. 2001 Sep;19(9):805–6. Available from: <URL>.

7. Trivedi R, Nagarajaram HA. Intrinsically disordered proteins: An overview. Int J Mol Sci [Internet]. 2022 Nov 14;23(22):14050. Available from: <URL>.

8. Liu Y, Wang X, Liu B. A comprehensive review and comparison of existing computational methods for

intrinsically disordered protein and region prediction. Brief Bioinform [Internet]. 2019 Jan 18;20(1):330–46. Available from: <URL>.

9. Cramer P. AlphaFold2 and the future of structural biology. Nat Struct Mol Biol [Internet]. 2021 Sep 10;28(9):704–5. Available from: <URL>.

10. Bryant P, Pozzati G, Elofsson A. Improved prediction of protein-protein interactions using AlphaFold2. Nat Commun [Internet]. 2022 Mar 10;13(1):1265. Available from: <URL>.

11. Jones DT, Thornton JM. The impact of AlphaFold2 one year on. Nat Methods [Internet]. 2022 Jan 11;19(1):15–20. Available from: <URL>.

12. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature [Internet]. 2021 Aug 26;596(7873):583–9. Available from: <URL>.

13. Ruff KM, Pappu R V. AlphaFold and implications for intrinsically disordered proteins. J Mol Biol [Internet]. 2021 Oct 1;433(20):167208. Available from: <URL>.

14. Yang Z, Zeng X, Zhao Y, Chen R. AlphaFold2 and its applications in the fields of biology and medicine. Signal Transduct Target Ther [Internet]. 2023 Mar 14;8(1):115. Available from: <URL>.

15. Cecchini M, Rao F, Seeber M, Caflisch A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. J Chem Phys [Internet]. 2004 Dec 1;121(21):10748–56. Available from: <URL>.

16. Nguyen PH, Ramamoorthy A, Sahoo BR, Zheng J, Faller P, Straub JE, et al. Amyloid Oligomers: A joint experimental/computational perspective on Alzheimer's disease, Parkinson's disease, type II diabetes, and amyotrophic lateral sclerosis. Chem Rev [Internet]. 2021 Feb 24;121(4):2545–647. Available from: <URL>.

17. Nguyen P, Derreumaux P. Understanding amyloid fibril nucleation and Aβ oligomer/drug interactions from computer simulations. Acc Chem Res [Internet]. 2014 Feb 18;47(2):603–11. Available from: <URL>.

18. Coskuner-Weber O, Uversky V. Insights into the molecular mechanisms of Alzheimer's and Parkinson's diseases with molecular simulations: Understanding the roles of artificial and pathological missense mutations in intrinsically disordered proteins related to pathology. Int J Mol Sci [Internet]. 2018 Jan 24;19(2):336. Available from: <URL>.

19. Coskuner-Weber O, Uversky VN. Alanine scanning effects on the biochemical and biophysical properties of intrinsically disordered proteins: A case study of the histidine to alanine mutations in amyloid-β 42. J Chem Inf Model [Internet]. 2019 Feb 25;59(2):871–84. Available from: <URL>.

20. Zhou R. Replica exchange molecular dynamics method for protein folding simulation. In: Protein Folding Protocols [Internet]. New Jersey: Humana Press; 2007. p. 205–24. Available from: <URL>.

21. Allison TC, Coskuner O, Gonzalez CA. Metallic systems: A quantum chemist's perspective [Internet]. Allison TC, Coskuner O, Gonzalez CA, editors. CRC Press; 2011. Available from: <URL>.

22. Coskuner-Weber O, Habiboglu MG, Teplow D, Uversky VN. From quantum mechanics, classical mechanics, and bioinformatics to artificial intelligence studies in neurodegenerative diseases. In: Methods in Molecular Biology [Internet]. Humana, New York, NY; 2022. p. 139–73. Available from: <URL>.

23. Tycko R. Solid-State NMR studies of amyloid fibril structure. Annu Rev Phys Chem [Internet]. 2011 May 5;62(1):279–99. Available from: <URL>.

24. Karamanos TK, Kalverda AP, Thompson GS, Radford SE. Mechanisms of amyloid formation revealed by solution NMR. Prog Nucl Magn Reson Spectrosc [Internet]. 2015 Aug 1;88–89:86–104. Available from: <URL>.

25. Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM. Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. Nature [Internet]. 2011 Dec 8;480(7376):268–72. Available from: <URL>.

26. Ma B, Nussinov R. Simulations as analytical tools to understand protein aggregation and predict amyloid conformation. Curr Opin Chem Biol [Internet]. 2006 Oct 1;10(5):445–52. Available from: <URL>.

27. Buchete NV, Tycko R, Hummer G. Molecular dynamics simulations of Alzheimer's β-amyloid protofilaments. J Mol Biol [Internet]. 2005 Nov 4;353(4):804–21. Available from: <URL>.

28. Zhang M, Ren B, Chen H, Sun Y, Ma J, Jiang B, et al. Molecular simulations of amyloid structures, toxicity, and inhibition. Isr J Chem [Internet]. 2017 Jul 16;57(7–8):586–601. Available from: <URL>.

29. Zheng W, Zhang C, Li Y, Pearce R, Bell EW, Zhang Y. Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. Cell Reports Methods [Internet]. 2021 Jul 26;1(3):100014. Available from: <URL>.

30. Yang J, Zhang Y. I-TASSER server: New development for protein structure and function predictions. Nucleic Acids Res [Internet]. 2015 Jul 1;43(W1):W174–81. Available from: <URL>.

31. Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc [Internet]. 2010 Apr 25;5(4):725–38. Available from: <URL>.

32. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. Nat Protoc

[Internet]. 2015 Jun 7;10(6):845–58. Available from: <URL>.

33. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res [Internet]. 2018 Jul 2;46(W1):W296–303. Available from: <URL>.

34. Akdel M, Pires DE V., Pardo EP, Jänes J, Zalevsky AO, Mészáros B, et al. A structural biology community assessment of AlphaFold2 applications. Nat Struct Mol Biol [Internet]. 2022 Nov 7;29(11):1056–67. Available from: <URL>.

35. Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. Nat Methods [Internet]. 2022 Jun 30;19(6):679–82. Available from: <URL>.

36. Shafat Z, Ahmed A, Parvez MK, Parveen S. Intrinsic disorder in the open reading frame 2 of hepatitis E virus: A protein with multiple functions beyond viral capsid. J Genet Eng Biotechnol [Internet]. 2023 Dec 1;21(1):33. Available from: <URL>.

37. O'Brien DP, Hernandez B, Durand D, Hourdel V, Sotomayor-Pérez AC, Vachette P, et al. Structural models of intrinsically disordered and calcium-bound folded states of a protein adapted for secretion. Sci Rep [Internet]. 2015 Sep 16;5(1):14223. Available from: <URL>.

38. Kheirabadi M, Taghdir M. Is unphosphorylated Rex, as multifunctional protein of HTLV-1, a fully intrinsically disordered protein? An in silico study. Biochem Biophys Reports [Internet]. 2016 Dec 1;8:14–22. Available from: <URL>.

39. Yang J, Zhang Y. Protein structure and function prediction using I-TASSER. Curr Protoc Bioinforma [Internet]. 2015 Dec 17;52(1):5.8.1-5.8.15. Available from: <URL>.

40. Lee YT, Ayoub A, Park SH, Sha L, Xu J, Mao F, et al. Mechanism for DPY30 and ASH2L intrinsically disordered regions to modulate the MLL/SET1 activity on chromatin. Nat Commun [Internet]. 2021 May 19;12(1):2953. Available from: <URL>.

41. Case DA, Aktulga HM, Belfon K, Cerutti DS, Cisneros GA, Cruzeiro VWD, et al. AmberTools. J Chem Inf Model [Internet]. 2023 Oct 23;63(20):6183–91. Available from: <URL>.

42. Caliskan M, Mandaci SY, Uversky VN, Coskuner-Weber O. Secondary structure dependence of amyloid-β(1–40) on simulation techniques and force field parameters. Chem Biol Drug Des [Internet]. 2021 May 22;97(5):1100–8. Available from: <URL>.

43. Weber OC, Uversky VN. How accurate are your simulations? Effects of confined aqueous volume and AMBER FF99SB and CHARMM22/CMAP force field parameters on structural ensembles of intrinsically disordered proteins: Amyloid-β 42 in water. Intrinsically Disord Proteins [Internet]. 2017 Jan 30;5(1):e1377813. Available from: <URL>.

44. Darden T, York D, Pedersen L. Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. J Chem Phys [Internet]. 1993 Jun 15;98(12):10089–92. Available from: <URL>.

45. Wise-Scira O, Xu L, Kitahara T, Perry G, Coskuner O. Amyloid-β peptide structure in aqueous solution varies with fragment size. J Chem Phys [Internet]. 2011 Nov 28;135(20):1448–57. Available from: <URL>.

46. Sgourakis NG, Yan Y, McCallum SA, Wang C, Garcia AE. The Alzheimer's Peptides Aβ40 and 42 Adopt Distinct Conformations in Water: A Combined MD / NMR Study. J Mol Biol [Internet]. 2007 May 18;368(5):1448–57. Available from: <URL>.

47. Tomaselli S, Esposito V, Vangone P, van Nuland NAJ, Bonvin AMJJ, Guerrini R, et al. The α-to-β Conformational Transition of Alzheimer's Aβ-(1–42) Peptide in Aqueous Media is Reversible: A Step by Step Conformational Analysis Suggests the Location of β Conformation Seeding. ChemBioChem [Internet]. 2006 Feb 6;7(2):257–67. Available from: <URL>.

48. Nag S, Sarkar B, Banerjee A, Sahoo B, Varun KAS, Maiti S. The Nature of the Amyloid-β Monomer and the Monomer-Oligomer Equilibrium. Biophys J [Internet]. 2011 Feb 2;100(3):202a. Available from: <URL>.

49. Murray MM, Krone MG, Bernstein SL, Baumketner A, Condron MM, Lazo ND, et al. Amyloid β-protein: Experiment and theory on the 21−30 fragment. J Phys Chem B [Internet]. 2009 Apr 30;113(17):6041–6. Available from: <URL>.

50. Yang M, Teplow DB. Amyloid β-protein monomer folding: Free-energy surfaces reveal alloform-specific differences. J Mol Biol [Internet]. 2008 Dec 12;384(2):450–64. Available from: <URL>.

51. Piovesan D, Del Conte A, Clementel D, Monzon AM, Bevilacqua M, Aspromonte MC, et al. MobiDB: 10 years of intrinsically disordered proteins. Nucleic Acids Res [Internet]. 2023 Jan 6;51(D1):D438–44. Available from: <URL>.