ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

Aggregation of Intrinsically Fluorescent Proteins: Combining Coarse-Grained Molecular Dynamics with Alphafold Predictions

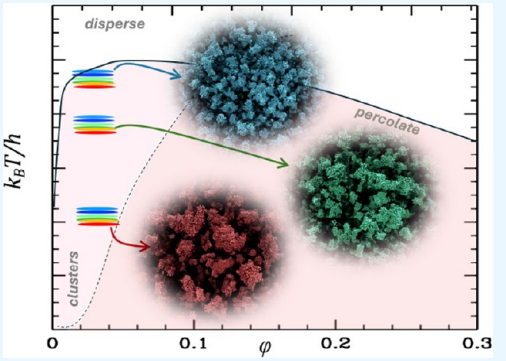
Margherita Bini,* Giorgia Brancolini, and Valentina Tozzini

Cite This: *ACS Omega* 2025, 10, 54341−54356

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Aggregation is a pervasive phenomenon affecting both disordered and folded proteins, modulated by the balance between hydrophobic attraction and electrostatics, and generally explored by means of molecular dynamics simulations, often performed with low-resolution models. On the other hand, the AI based structure prediction AlphaFold software, extremely accurate for single chains, has been recently provided with the capability of predicting the structure of protein complexes. In this work, we compare coarse-grained molecular dynamics simulations to assess the relative performance of different force fields and those of AlphaFold3 (AF3) in predicting the aggregation of four variants of intrinsically fluorescent proteins. Our results are useful for guiding the choice or design of new potentials for systems containing both disordered and folded proteins and outline a possible scheme to integrate AlphaFold with molecular dynamics simulations.



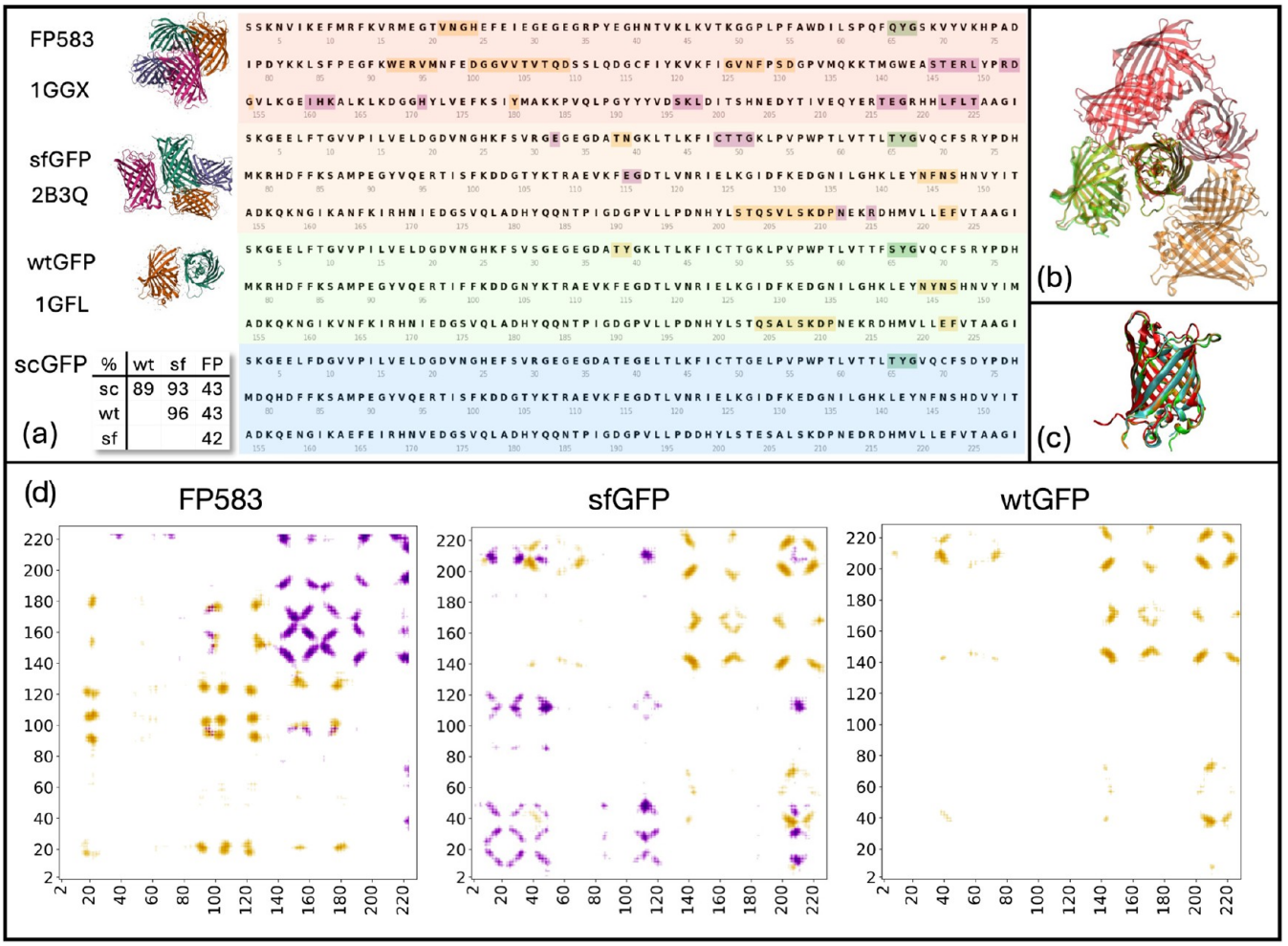
INTRODUCTION

The aggregation of proteins is a pervasive phenomenon, crucial in fundamental biological processes[1] and a cause of a large number of degenerative diseases.[2,3] Besides, it is exploited in the design of molecular sensors, often combining nanoparticles[4] with signaling biomolecules, e.g., intrinsically fluorescent proteins (IFPs).[5,6] In fact, aggregation affects not only disordered or unstructured proteins, such as typical amyloidogenic agents, but also well-structured proteins like IFPs. Although their interactions are generally driven by hydrophobic forces, when they are charged or strongly polar, their interaction is modulated by competition with electrostatics,[7] and their phase diagram can be enriched with new features characterized by the coexistence of diluted domains with solute aggregates or percolates, depending on the concentration and ionic strength.[8,9] The experimental characterization of aggregates is complex and therefore often supported by computer modeling.[10] While atomistic classical molecular dynamics (MD) can accurately describe aggregation interfaces,[5] the exploration of the phase diagram requires lower-resolution representations, i.e., coarse-grained (CG) models.[3,11] Several CG models with different resolutions have been used to study protein aggregation. For instance, OPEP[12] combines atomistic backbones with single-bead side chains and was used to study amyloid-β aggregation within the *E. coli* cytoplasm.[13,14] Martini-3[15] uses up to five side-chain beads plus one for the backbone, capturing the conformations of α-synuclein,[16] while SIRAH—a multibead model tailored for disordered proteins—was used to study its aggregation.[16]

In this work, we aim to keep the complexity at a minimum and use "minimalist" single-bead-per-amino-acid models,[17] which are particularly inexpensive but still account for the molecule flexibility,[18,19] often triggering aggregation, and residue-level intermolecular interactions responsible for interface formation and structure.[20] Inter-residue potentials, therefore, assume a crucial role and must account for both short-range and electrostatic contributions, which are parameterized in different ways. For instance, Dignon et al.[21] use a Lennard−Jones (LJ) plus Debye−Hückel (DH) representation, with the former parameterized based on hydrophobicity. Lindorff-Larsen et al.[22] improved this approach by including more hydrophobicity scales and statistically extensive data-driven analyses, further optimizing against SAXS and NMR data of intrinsically disordered proteins (IDPs).[22] Finally, COCOMO, developed by Valdes et al.,[23] more specifically targets the experimental protein saturation concentration, with good performance in describing phase separation.

Although intended to be general, these FFs have mainly been tested on IDPs, where aggregation is influenced by intramolecular dynamics. Conversely, in this work, we focus on an almost opposite situation, that is, the aggregation of IFPs,[24] with a rather rigid β-barrel fold, widely used in engineered molecular sensors.[6] Indeed, many IFPs show a spontaneous aggregation tendency in solution and are crystallized as multimers[25,26]

Received: July 14, 2025
Revised: September 19, 2025
Accepted: September 25, 2025
Published: November 9, 2025



%	wt	sf	FP
sc	89	93	43
wt		96	43
sf			42

Figure 1. (a) IFP variants with sequences and crystallographic structures, when available (pdb code reported). The superfolder (sfGFP) is the F64L/S65T/F99S/M153T/V163A/S30R/Y39N/N105T/Y145F/I171 V/A206 V mutant of wtGFP, while the supercharged (scGFP) is the T9D/K26E/Y39E/K41E/K52E/F64L/S65T/R73D/K79D/N149D/K158E/N164E/K166E/N198D/Q204E/K214D mutant of the wtGFP (pdb structure not available). A table of the sequence identities of the variants is reported as inset. The chromophore residues are highlighted in green, the interface residues in orange or purple. (b) Asymmetric crystallographic units of the variants, aligned to the chain A, are colored as the sequences in (a). (c) Aligned chains A only (also aligned to the reference 1EMG[37] in cyan): all folds are very similar, with RMSD (Root Mean Squared Deviation) of coordinates between wtGFP and sfGFP ∼0.3 Å and that of FP583 ∼7 Å. (d) Contact maps of different chains. Orange and purple represent the two different interfaces of the tetramers (primary in orange, secondary in purple; involved amino acids are colored accordingly in the sequences in (a)).

revealing multiple potential interfaces. Because aggregation reduces the performance of molecular sensors, great effort is dedicated to engineering non-aggregating IFPs, generally by mutating solvent-exposed residues to like-charged amino acids.[27,28] "Super-charging" has been shown to alter intermolecular interactions, improve solubility, and confer resistance to aggregation without significantly compromising the native fold and function. One motivation of this work is to study the aggregation of different variants of IFPs and their sequence dependence. This allows one to investigate the fundamental balance between hydrophobic and electrostatic interactions, structural rigidity, and environmental factors such as ionic strength.

Due to the fold rigidity, in IFPs aggregation is mainly driven by the inter-residue potential terms. Consequently, the simulation of aggregation and interface structures of different FP variants is also an ideal case to assess the performance of the different FFs. In particular, we study the wtGFP,[29] the red variant FP583,[30] the superfolder sfGFP[25] with a larger aggregation tendency, and the supercharged variant (scGFP).[28] On these variants, we perform extensive simulations with the mentioned FFs, and with the additional one by Trovato et al., parametrized for folded proteins using a data set of IFPs and proteins with similar fold.[20] We characterize the aggregation, the structure and size of aggregates and oligomers, and compare them with experimentally known interfaces.[31]

The third aim of this work is the comparison of CG-MD simulation results with the predictions of the AI-based algorithm AlphaFold.[32,33] The integration of machine learning algorithms has recently become customary within many aspects of molecular modeling,[34] especially the prediction of structures.[35] The current version Alphafold 3[36] (AF3), includes the capability of identifying aggregate-prone regions and returning macromolecular complexes. Its performance on interface prediction, however, is still less assessed. Even for this aspect, the study of IFP aggregation is a convenient performance indicator: AF predicts the single IFP fold with extremely high accuracy, which allows us to focus on its reproduction of the interfaces and comparison with the MD-CG simulations. Besides directly comparing the aggregates structure output from AF3 to those

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

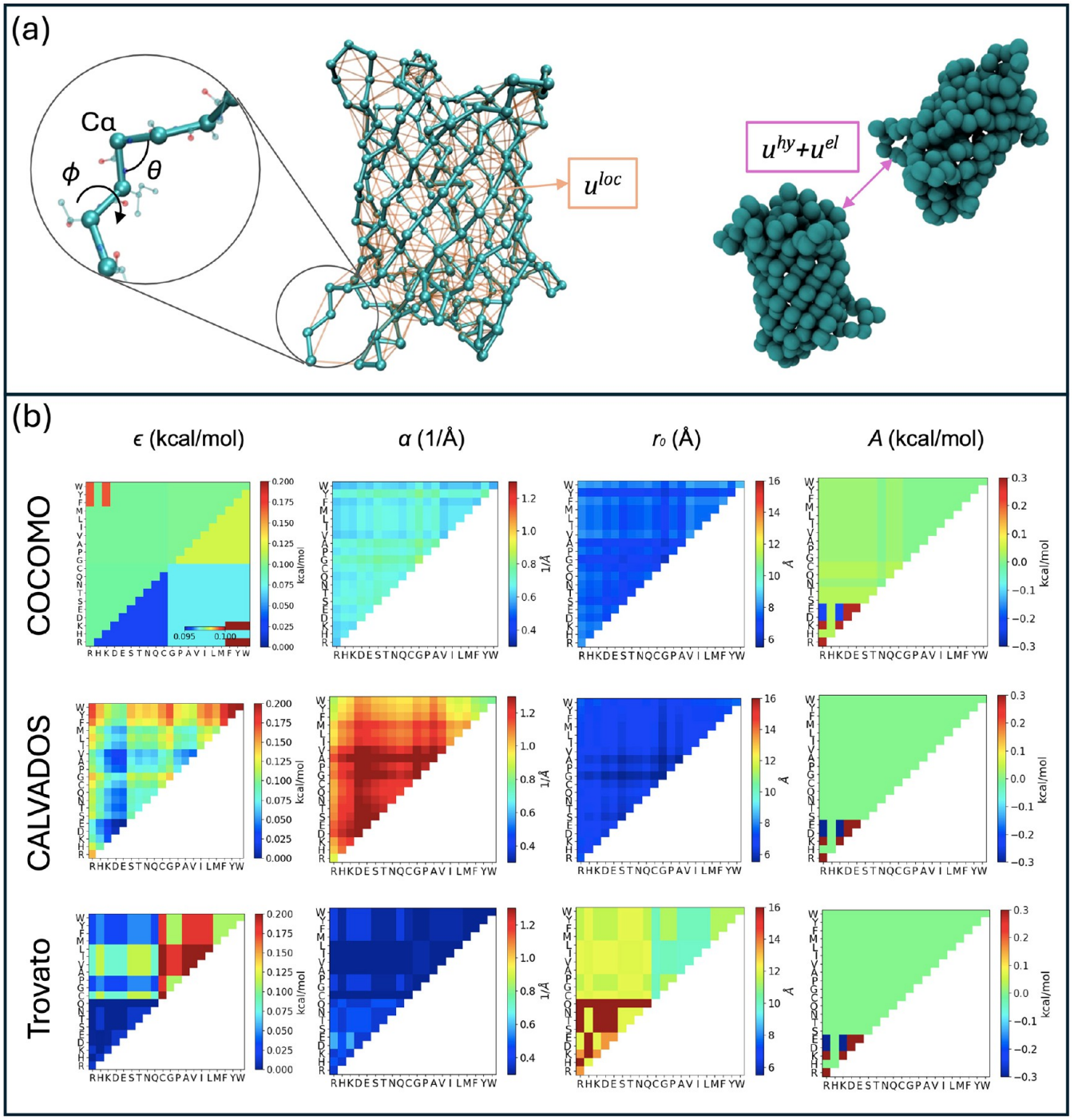


Figure 2. (a) Coarse-graining scheme and illustration of the potential terms of the model. The beads placed on the C_αs interact with angular (θ) and dihedral (ϕ) terms along the chain, and with intramolecular local interactions (represented as the network in orange); electrostatic and hydrophobic interactions act between beads of different molecules. (b) Matrix illustration of the parameters of the u_{hy} and u_{el} terms of the FFs COCOMO,[23] CALVADOS,[43] and Trovato.[20] Each of the parameters of the potentials (ε, α, r_0, and A) is a 20×20 symmetric matrix, whose elements are represented with color coding according to the scale reported alongside. (For COCOMO, ϵ is also reported in the lower triangle on a reduced scale to highlight variations.)

from MD, we also analyze the pair contact probability output by AF3 and compare it with that from simulations with the different FFs.

This work is organized into a methodological section, already including several elements of novelty, and a results section, followed by conclusions and perspectives.

SYSTEMS AND METHODS

In this section, we describe the proteins that are the subject of this study, the building of their models, the simulation and modeling methods, and the analysis methods.

IFP Reference Structures. The variants of the IFPs that we consider in this work are listed in Figure 1. In panel (a), the aligned sequences are reported with the representation of the

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

crystallographic structures on the left (different monomers in different colors). Although the crystallographic form may not reflect the oligomer structure in solution, it is often considered an indicator of aggregation tendency and preferred interfaces. In panel (b), all the crystallographic structures are represented with the first monomer superimposed to compare the interfaces (in this case, all the monomers of a single variant multimer are in the same color, corresponding to the one of the sequences in panel (a)). While the wtGFP (green) is crystallized as a dimer with a unique, well-defined interface (represented in the corresponding contact map of panel (d) in orange), the superfolder sfGFP (orange in panel (b))[38] appears as a tetramer, with one interface which is the same as that of wtGFP, but dysplaying a second interface (in purple in the map in panel (d)). Interestingly, the red variant FP583 (red in panel (b)) also appear as a tetramer (and oligomer in solution),[30] but with different interfaces (interface residues are highlighted in the sequences in panel (a)). This is coherent with the lower sequence identity of FP583 compared to the others (see the table reported as inset in panel (a)), which are all mutants of wtGFP. In particular, sfGFP was engineered to stabilize the fold,[25] but aggregation in oligomers was observed in solution at the concentration of ∼2 mg/mL.[39] On the other hand, the scGFP mutations of neutral to charged residues lead to a total net charge of −30e.[27] [1]

The reported contact maps use a "fuzzy" definition of a pair contact matrix $p_{ij}^{exp} = w(d_{ij})/\sum_{k,l} w(d_{kl})$, based on a distance-dependent weight function $w(d)$ which assigns high weight within an internal cutoff radius and decreases to 0 at a second cutoff radius (set at 8.2 Å and 18.1 Å, respectively; see also Section S4 for more details). The definition is such that p^{exp} is normalized to 1.

Coarse-Grained Model and Force Fields. We use a minimalist model[17,40] with the protein represented as an unbranched chain of beads placed on the C_α connected by rigid rods, and the backbone conformation defined by the angles θ_i and the dihedrals ϕ_i (see Figure 2a). We decompose the interaction energy as $U = U_{intra} + U_{inter}$, where the intramolecular contribution $U_{intra} = U_\theta + U_\phi + U_{local}$ includes, besides the conformational terms, a network of local interactions (in orange in Figure 2a) describing, e.g., hydrogen bonding and steric hindrance. These terms incorporate a partial bias toward the reference structure by setting their equilibrium values (bond angles, dihedrals, and local interaction distances) to match those of the reference conformation, as detailed in Section S1.[19,41,42]

More crucial to this work is U_{inter}, which we decompose into hydrophobic and screened electrostatics, taking the form of Morse and Debye−Hückel (DH), respectively

$$U_{inter} = \sum_{I>J} u_{hy}(r_{IJ}) + u_{el}(r_{IJ})$$

$$= \sum_{I>J} \varepsilon_{IJ}[(e^{-\alpha(r-r_{IJ}^0)} - 1)^2 - 1] + A_{IJ}\frac{e^{-kr}}{r} \quad (1)$$

(I, J runs over all of the pairs not already accounted for in U_{intra}). The parameters of u_{el} are the inverse screening length $k = 1/\lambda = \sqrt{4\pi e^2(n_+ + n_-)/\epsilon_\infty kT}$ related to the ionic strength, and $A_{IJ} = Q_I Q_J e^{kR_{IJ}^0}/\epsilon_\infty(1 + kR_{IJ}^0) \sim Q_I Q_J/\epsilon_\infty$ related to the effective charges and bead radii $R^0 = r^0/2$.[7] The well depth ε, the inverse range α, and the interbead distance r_0 describe the hydrophobic interaction. The four intrinsic parameters ε, α, r_0, and A depend on the pair I−J, i.e., each corresponds to a 20 × 20 symmetric matrix.

In order to compare the different FFs, we preliminarily mapped their different functional forms onto eq 1, as detailed in section S1. Briefly, except for an additional term accounting for electrostics of polar residues, the case of COCOMO is straightforward since it uses DH for electrostatics and 10−5 LJ for short-range interactions, which is mappable into Morse via a proper α definition. Similarly, CALVADOS[2] uses DH and 12−6 LJ tuned by a switching function, taken into account in the mapping. Trovato et al. merge electrostatics and hydrophobicity into a single functional form. Therefore, the DH term was first subtracted, and the residual part was used to fit the hydrophobicity parameters.

The result of this procedure is summarized in Figure 2b, reporting the parameter matrices for each FF. In CALVADOS and Trovato, the A matrix is null except for pairs of charged amino acids (AAs), while COCOMO has small contributions also for the polar AAs. The parameters of u_{hy} display larger differences: while the parameters ε in Trovato and CALVADOS are in a similar range (0−0.3 kcal/mol), COCOMO ε is flattened around the value 0.1 kcal/mol. In Trovato, both r_0 and the range are larger on average, as an effect of the fit of a complex potential onto the Morse form.

The aggregation tendency depends on the strength and range of both u_{hy} and u_{el}. We therefore defined a parameter $h \propto (\varepsilon/\alpha r_0 - Ae^{kr}/kr_0^2)$, roughly describing the aggregation enthalpy per pair (more details on the definition are reported in the Supporting Information S1). The average value of h in units of k_BT is 0.3, 0.6, and 1.1 for CALVADOS, COCOMO, and Trovato, respectively, which provides a preliminary indication of the order of aggregation tendency of the three FFs, although with rather different distributions and variances (see Figure S3), confirming that the three FFs treat the interactions in rather different ways.

Simulation Setup and Diffusivity Calculation. We performed simulations of lengths ranging from 80 ns to 2 μs, combining the four protein variants with the three FFs, at two different ionic concentrations of ∼10 mM and ∼100 mM (k ∼0.04 Å^{-1} and k ∼0.1 Å^{-1}), each case with two different starting random configurations (seeds), at both high and low protein concentrations, for a total of ∼22 μs of trajectories (list reported in Supporting Information S2). The 50 nm-side cubic box, including 200 copies of the IFPs, corresponds to a packing fraction ϕ ∼ 0.03 and concentration ∼70 mg/mL (see Supporting Information S3), which is larger than those typically achieved in solution but closer to that achieved in the cytoplasm, chosen to lie within the disperse-aggregate coexistence region.[8,44] The low concentration case is obtained with a single protein in the cell. The NVT conditions (T = 300 K) were maintained using the Langevin thermostat with the time step set at 10 fs and γ = 1.2 ps^{-1}, optimized to reproduce the experimental GFP diffusion coefficient in dilute solution[20] (D ∼ 87 μ m^2/s^2). The simulations were performed with LAMMPS,[45] using in-house coded tools to implement the different FFs[46] (details in Supporting Information S2).

The center of mass mean squared displacement $MSD(t) = \langle(\mathbf{r}_{cm}(t+t_0) - \mathbf{r}_{cm}(t_0))^2\rangle$ was calculated by averaging over t_0, the seeds, and all the proteins. The translational diffusion coefficient was then evaluated using the relation $MSD(t) = 6Dt$. We corrected D for the systematic underestimation due to periodic boundary conditions using the empirical formula $D = D_{sim}/(1-$

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

$\xi R/L$),[47] where L is the box size and $\xi = 2.837$ is a constant depending on the box shape. Furthermore, the viscosity η is related to D by $D = k_B T/6\pi R\eta$, and both depend on the concentration c of proteins. A commonly used empirical formula to account for this effect is[31]

$$\frac{\eta}{\eta_0} = \frac{D_0}{D} = \exp\left(\frac{c \times \mu_p}{1 - c \times \beta_C \mu_p}\right) \quad (2)$$

with η_0 and D_0 evaluated at very low c, μ_p is the volume per mass of a single protein, and β_C a crowding parameter representing the maximum packing fraction, also accounting forshape; $c = 1/\mu_p\beta_C$ is the maximum concentration corresponding to closely packed proteins. These parameters were also experimentally evaluated for the sfGFP by fitting viscosity data, resulting in $\beta_C \sim 0.5$ and $\mu_p \sim 5.5$ mL/g.[31]

Clusters and Interfaces Analysis. The cluster analysis on simulations trajectories is carried out using OVITO[48] on the equilibrated portion of the trajectories. For each given snapshot of the trajectory, two IFPs are assigned to the same cluster if at least on pair of C_α belonging to different IFPs are "in contact" i.e., nearer than r_{cut} (set at 8.2 Å)[3].

To characterize the interfaces, we define the contact probability matrix

$$p_{ij}^{b} = \frac{\langle N_{ij} \rangle}{\left\langle \sum_{k,l} N_{kl} \right\rangle} \quad (3)$$

with N_{ij} the number of contacts between residues in positions i and j along the chains of two different proteins (averaged over all pairs of proteins in a snapshot and over the trajectory). p^b monitors the bare frequencies of formation of contacts between pairs of proteins, including all contacts, i.e., not only those necessarily forming due to the contact propensity of i and j but also those formed indirectly due to the effect of nearby contacts. An alternative definition is

$$p_{ij}^{d} = \frac{\left\langle \sum_k N_{ik} \sum_l N_{lj} \right\rangle}{\left\langle \left(\sum_{k,l} N_{kl} \right)^2 \right\rangle} \quad (4)$$

monitoring the number of contacts of residue i with any residue of the other protein, i.e., its propensity to be at the interface, multiplied by the same quantity evaluated for j in the second protein (and averaged over any pair of proteins and throughout the trajectory). This is a measure of the joint probability of residues i and j both being at the interface, which includes specific i–j contacts and returns the reduced values over incidental indirect contacts. An example of a comparison between p^b and p^d is reported in Figure 3: apart from the background effect due to the fact that in the simulations all possible contacts are explored, the p^d evaluation appears more regular and more easily comparable with the experimental interfaces reported in Figure 2. The calculations are performed using dedicated in-house coded Python scripts.[46]

Alphafold Setup and Output Analysis. We used AF3 through its web interface.[49] We generated m-mers with m up to 20, to compare with the cluster structures of homologous sizes from MD. Additionally, AF3 returns, among its accessory outputs, the contact probability matrix p_{ij}^m, with a similar meaning to those defined in the previous section, except that it has the global dimension $(m \times n_{res}) \times (m \times n_{res})$, i.e., it is composed of $m \times m$ blocks of dimension $n_{res} \times n_{res}$ (an example for $m = 5$ is shown in Figure 4a).

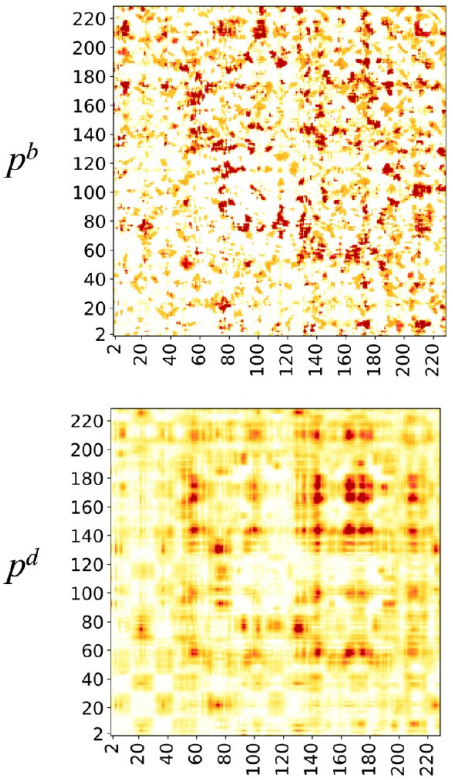


Figure 3. p^b and p^d are evaluated for the simulation of the protein wtGFP with the CALVADOS FF.

While the m blocks on the diagonal (shaded in blue) are the intramolecular contact probabilities, returning information on the secondary and tertiary structure of the single IFP, the $m \times (m - 1)$ out-of-diagonal blocks describe the intermolecular contact probabilities. The near-diagonal blocks (in green) generally give more weight to the primary interface, while other possible interfaces and contacts are explored in the others. We therefore averaged (and renormalized) all the out-of-diagonal blocks

$$p_{ij}^{AF} = \frac{\sum_{k,l=1,m}^{m} p_{i+k(m-1),j+l(m-1)}^{m}}{\sum_{ij}^{n_{res}} \sum_{k,l=1,m}^{m} p_{i+k(m-1),j+l(m-1)}^{m}} \quad (5)$$

obtaining a $n_{res} \times n_{res}$ matrix p_{ij}^{AF} (reported in panel (b)). This procedure summarizes all of the possible protein–protein interfaces explored by AF3 in multimers into a single two-protein contact probability matrix (further details in Section S4).

The comparison between p^{AF} and p from simulations is quantified using two indices. The Pearson correlation index

$$r = \frac{\sum_{ij} (p_{ij}^{AF} - p_{\mu}^{AF})(p_{ij} - p_{\mu})}{\sqrt{\sum_{ij} (p_{ij}^{AF} - p_{\mu}^{AF})^2} \sqrt{\sum_{ij} (p_{ij} - p_{\mu})^2}} \quad (6)$$

(with p_{ij} (and p_μ) the element i,j (and their average) from AF (p^{AF}) or from the simulation ($p = p^d$ or $p = p^b$) indicates the relative compatibility of the two matrices. The Jaccard index is

$$J = \frac{|P^{AF} \cap P|}{|P^{AF} \cup P|} \quad (7)$$

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

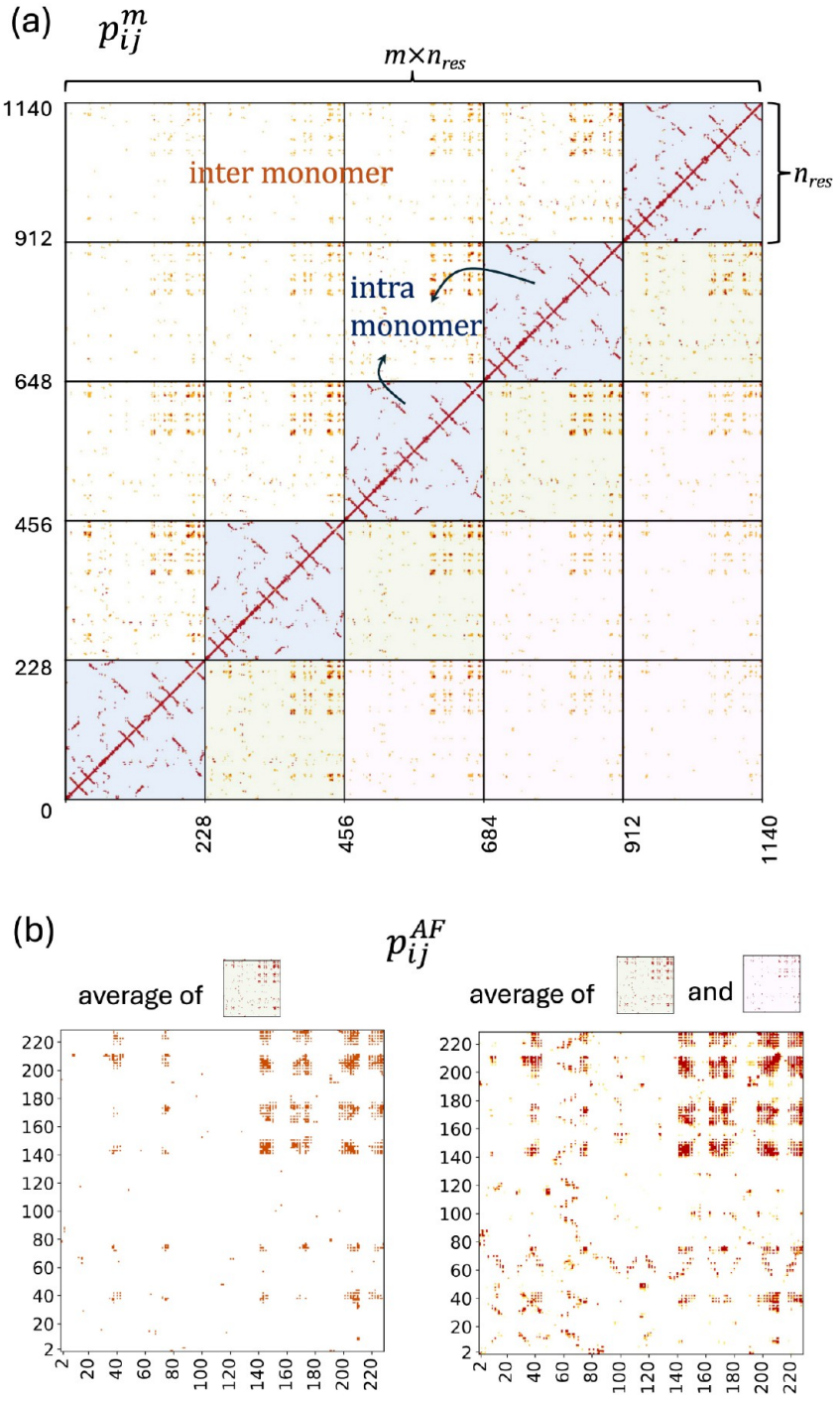


Figure 4. Panel (a): the p_{ij}^{m} from AF3 for the pentamer of scGFP (shaded colors as in the previous figure), with diagonal and off-diagonal blocks in different shades. (b) Result of the off diagonal averaging procedure described in the text.

where P^{AF} and P represent the sets of matrix elements that exceed a given interaction threshold, and || is the size of the set. Higher values of J indicate a greater overlap between high-valued regions of the matrices, which identify interfaces. J ranges between 0 and 1 and can be used to analyze the interfaces. While r can give some excessive weight to similarities in the out-of-interface regions, J gives more importance to interfaces and patterns of contacts. The same indexes are used to compare p with the normalized crystallographic contact matrix p^{exp}.

RESULTS AND DISCUSSION

We report, in the following, the results of simulations and analyses, with additional data reported in Supporting Information.

The Different FFs Return Different Viscosities. Figure 5 reports the MSD (mean squared deviation) and diffusion coefficients D from the simulation trajectories (energies and other details are reported in Supporting Information S5). Ds (PBC-corrected) and the corresponding viscosities, η, are also reported in Table 1 (uncorrected values are in Figure S6).

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

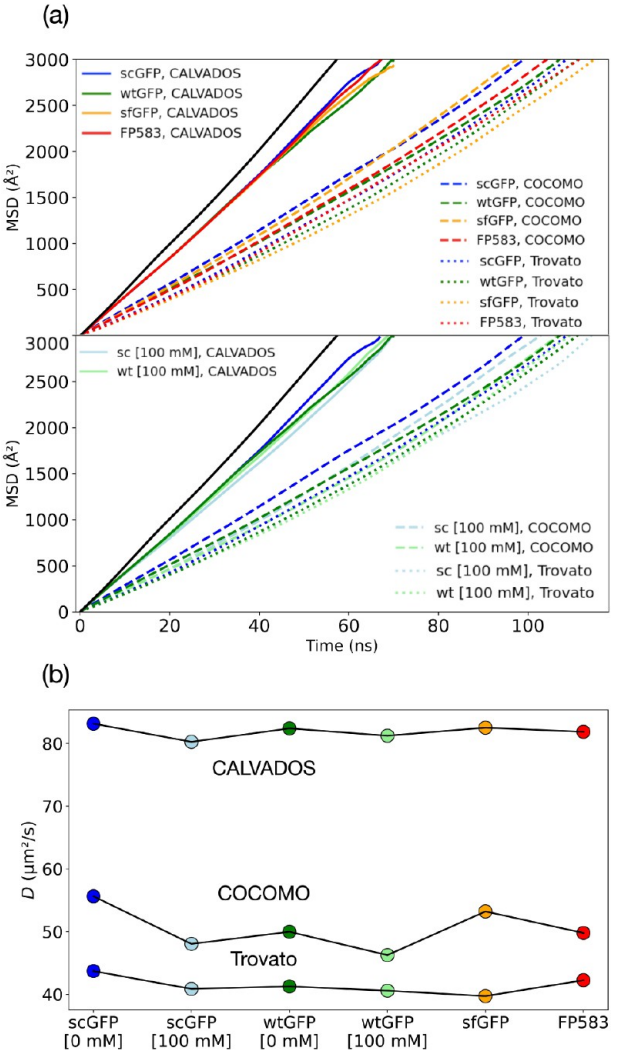


Figure 5. (a) MSD from simulations with the different FP variants and FFs as per the legend in the plots. The upper part reports the simulations at null ionic strength for all the variants, while the lower part reports the simulations at 100 mM of ion concentration only for wtGFP and scGFP. The black line represents the MSD from simulations at low protein concentration, evaluated by averaging over simulations with different variants and FFs. (b) The (PBC-corrected) diffusion coefficients for the different FFs, FP variants, and ionic strength.

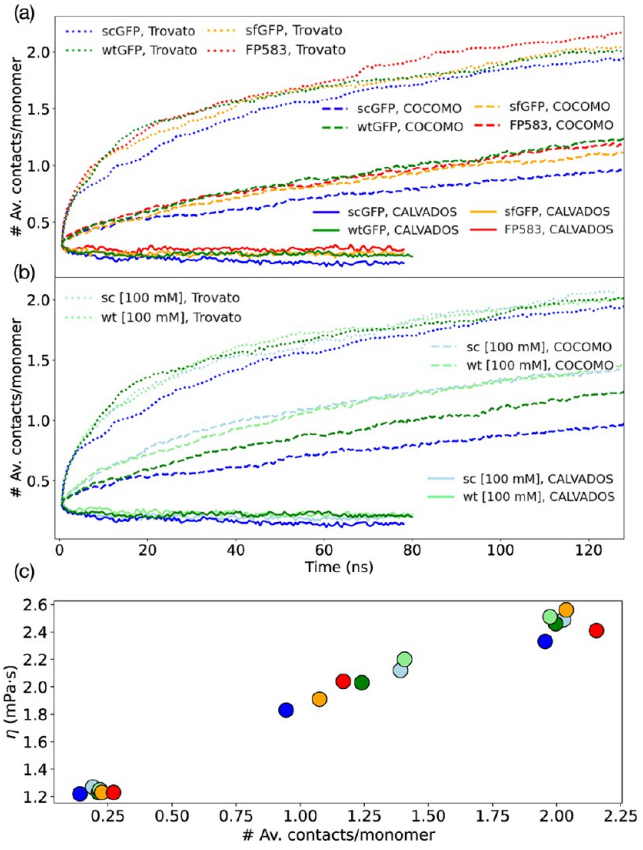


Figure 6. Contact formation dynamics. The average number of protein−protein contacts per protein is reported and counted with the same r_{cut} used for the cluster definition. Multiple AA contacts at the same interface are considered a single contact. (a) The three FFs and four IFP variants as per legend, at 0 mM of ionic strength. (b) High and low ionic strength (dotted lines = 0 mM, solid lines = 100 mM) for the scGFP and wtGFP. (c) Correlation plot of number of contacts vs viscosity.

The D at low concentration is reported as a black line in panel (a). We obtain an average value of the corresponding viscosity η_0 ∼0.0011 Pa·s for GFP variants and η_0 ∼0.0012 Pa·s for FP583, in good agreement with the experimental value in dilute solution. Interestingly, we observe a small but significant difference in the diffusive behavior of FP583 with respect to GFP mutants, which we attribute to its different mass and especially mass

Table 1. Diffusion Coefficients (Corrected by the PBC Effect) and Viscosities. The Values of η_0 Used to Calculate the Ratios η/η_0 are Extracted from the Single FPs Simulations (First Three Rows).[a]

	D (μm²/s)			η (10^{-3} Pa · s)			η/η_0			n. contacts		
	CA	CO	T	CA	CO	T	CA	CO	T	CA	CO	T
Single scGFP	95.42	91.91	94.33	1.06	1.11	1.08						
Single wtGFP	95.40	94.66	94.33	1.07	1.07	1.08						
Single FP583	89.85	88.68	90.42	1.13	1.15	1.12						
scGFP	83.16	55.64	43.73	1.22	1.83	2.33	1.13	1.69	2.16	0.141	0.945	1.955
wtGFP	82.41	49.99	41.27	1.23	2.03	2.46	1.14	1.88	2.28	0.213	1.240	1.996
sfGFP	82.52	53.22	39.72	1.23	1.91	2.56	1.14	1.77	2.37	0.226	1.075	2.037
FP583	82.86	49.80	42.26	1.23	2.04	2.41	1.07	1.77	2.10	0.272	1.168	2.155
sc 100 mM	80.25	48.06	40.87	1.27	2.12	2.49	1.18	1.96	2.31	0.191	1.391	2.027
wt 100 mM	81.23	46.27	40.59	1.25	2.20	2.51	1.16	2.03	2.32	0.219	1.407	1.974

[a]For FP583 we considered the average value over the three FFs $\eta_0^{FP583} = 1.15 \times 10^{-3}$Pa S, while for the GFP mutants we also averaged on variants, getting $\eta_0^{GFP} = 1.08 \times 10^{-3}$Pa S. The experimental value is η_0∼1.2×10^{-3} Pa·s. CO = COCOMO, CA = CALVADOS, and T = Trovato

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

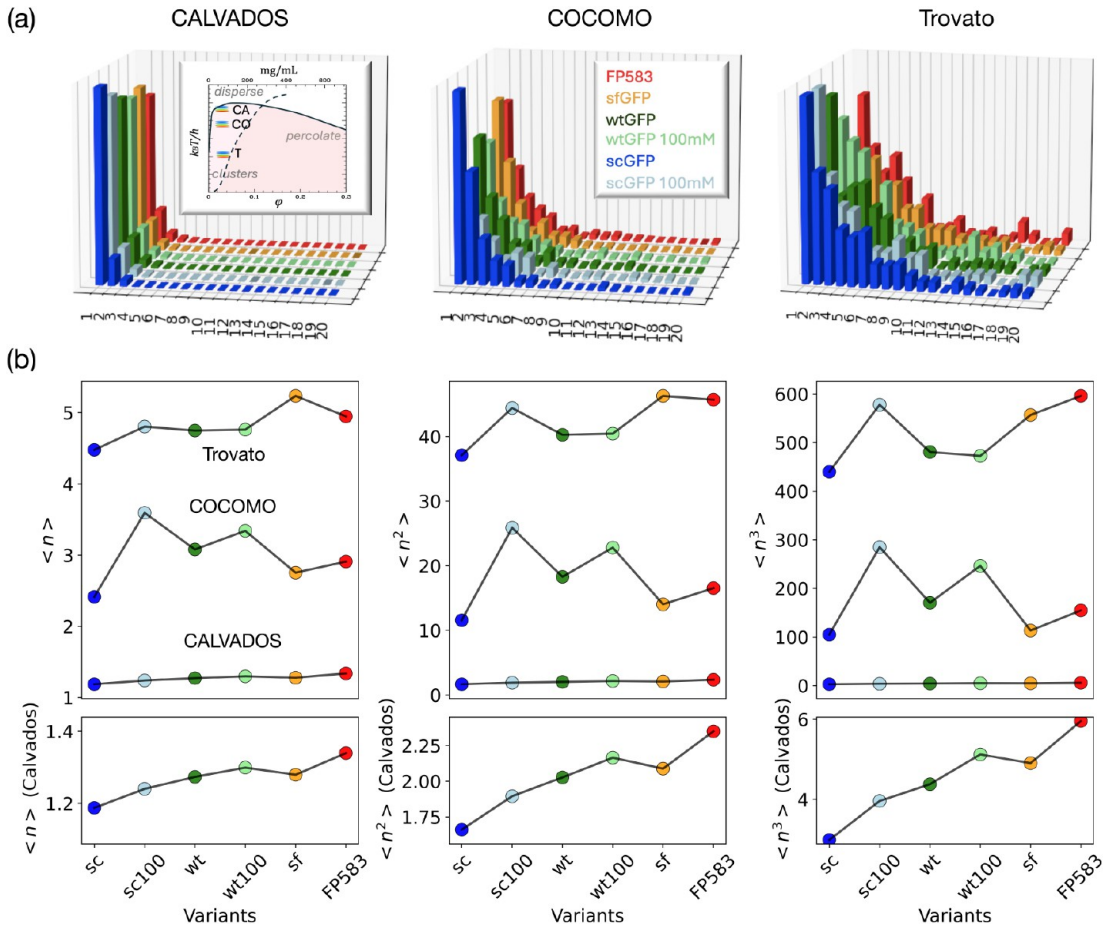


Figure 7. Cluster size statistics. (a) Histograms of the cluster size evaluated on the equilibrated part of the trajectories, for all the variants and ionic strengths, for the three FFs, as per legends. In the inset, the qualitative location of the simulations in the aggregation phase diagram. (b) The first, second, and third moments of the distributions of the different FFs and variants, with the same color coding as (a) (In the bottom part, the CALVADOS data are zoomed in. More details are in the SI Supporting Information S3).

distribution, producing a different relation between translational and rotational dynamics (details provided in Section S3).

The diffusivity in simulations at high protein concentration is clearly reduced in all cases (compare colored lines in panel (a) of Figure 5 with the black line). However, the different FFs return substantially different diffusivity. The upper part of panel (a) reports the comparison between CALVADOS (solid lines), COCOMO (dashed lines), and Trovato FF (dotted lines), which display decreasing diffusivity in the mentioned order (see also panel (b), where *D*s are reported in synthetic form). Since lower diffusivity usually depends on the formation of larger aggregates, this indirectly confirms the order of aggregation Trovato > COCOMO > CALVADOS implied by the values of the parameter h measuring the relative attraction strength evaluated in the previous sections.

Using the relation (2) with the experimental values of the parameters μ_p and β_C one gets $\eta/\eta_0 \sim 1.6$ for our working concentration (see Figure S8). The average values η/η_0 for the three FFs are ∼1.14, ∼1.8, and ∼ 2.3 for CALVADOS, COCOMO, and Trovato, respectively, indicating that COCOMO is the one better reproducing the experimental dependence of viscosity on the protein concentration. When looking at the differences among the IFP variants, however, the relative diffusivity differences are not very large. The scGFP is consistently the most diffusive across all FFs, but the other variants display very similar *D*. A deeper inspection reveals that CALVADOS, although with very small differences, returns the correct order of aggregation at null strength (scFP < wtGFP < sfGFP < FP583), while COCOMO inverts wtGFP and sfGFP, and Trovato inverts sfGFP and FP583.

Concerning the dependence on the ionic strength, evaluated only on the variants with larger net charge (scGFP and wtGFP), all the FFs generally reproduce the expected decrease of diffusivity as the ionic strength increase.[4] To this respect, the COCOMO FF results in the larger effect, as expected, since in this FF the electrostatic component plays a main role with respect to the hydrophobic one.

Diffusivity Correlates with Contact Formation Dynamics. The monitoring of protein−protein contact formation reported in Figure 6 reflects the viscosity behavior: panel (a) clearly shows that Trovato FF induces, on average, faster contact formation, reaching a value of ∼2.0 contacts per protein, followed by COCOMO, with ∼1.1.

Conversely, the CALVADOS FF shows a rather constant and low number of contacts (∼0.2). This indicates that, for this FF, the thermodynamic conditions are near the boundary of the condensate-disperse transition. Panel (b) confirms that increasing ionic strength is to increase the number of contact per protein and accelerate the dynamics of their formation, and this is particularly evident in COCOMO FFs, where the effects of electrostatics are amplified, and in scGFP, as expected due to its larger net charge.

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

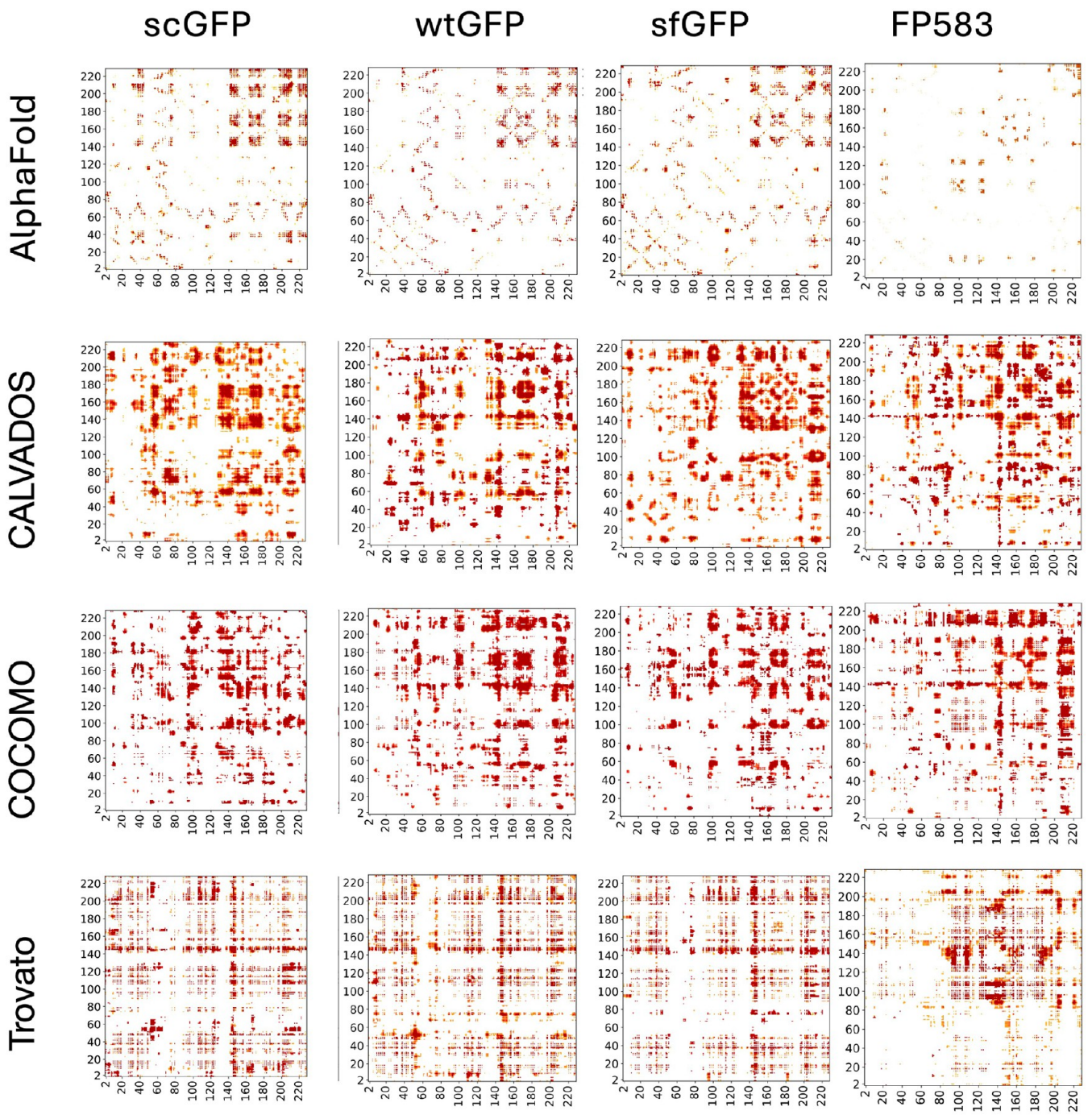


Figure 8. Contact probability matrices p^{d} evaluated for the four IFPs with AF3 (average over 10-mers) and in the simulations with the three FFs, as indicated (at 0 mM ionic strenght). The color range is from white (0) to yellow to red (0.0018). To facilitate the comparison with AF3, a cutoff on the low values of probability was subtracted from the matrices obtained from the simulations. The original versions with background are reported in SI, section S4.

As shown in panel (c), the correlation between the average number of contacts per monomer and the viscosity is direct and quantitative: as the average number of protein contacts increases, their motion is hindered and the diffusivity decreases. Overall, the average number of contacts per protein indicates the usual order of aggregation for the four variants, namely scGFP < wtGFP < sfGFP < FP583, and, in contrast to viscosity, it is the same for all FFs, making this indicator more reliable than the former for evaluating aggregation.

The Cluster Size Distribution Identifies the Location in the Phase Coexistence Diagram. Overall, our analyses indicate that, at given T and ϕ, the three FFs locate in different points of the phase diagram. In particular, CALVADOS appears to be near the aggregate-dispersed transition. We further explore this aspect by analyzing the cluster size distribution, as reported in Figure 7. Panel (a) shows a 3D view of the distribution of the different FFs and variants/ionic strengths (as per the legend), with the first moments of the distributions reported in panel (b). Data are coherent with the previous analyses of diffusion and

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

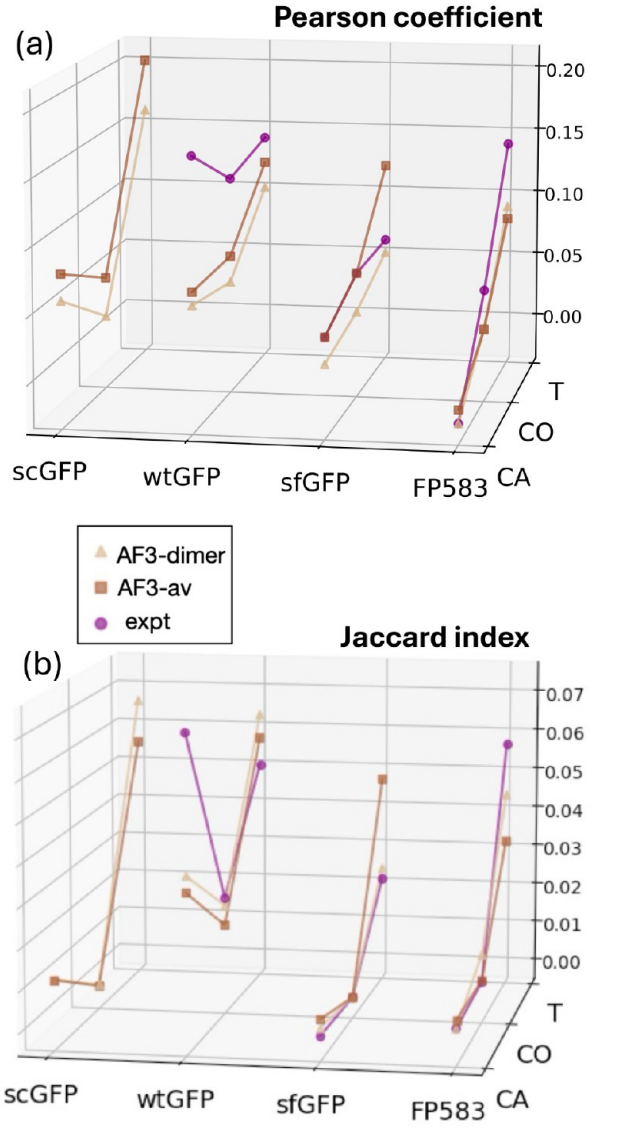


Figure 9. Graphical representation of the Pearson coefficient r (a) and Jaccard index J (b) in the vertical axes, for the four variants and three FFs (reported on the front axes and depth axes respectively) (CA = CALVADOS, CO = COCOMO, T = Trovato). The comparison of contact matrices from simulations with experiment or AF (average and obtained from dimers) is encoded into colors as per legend.

contacts formation: the distribution of cluster size is increasingly spread toward large sizes in the order CALVADOS < COCOMO < Trovato, and all of its first three moments follow the same order, confirming the already observed aggregation tendency of the three FFs.

Concerning the relative aggregation tendency of the variants, as shown in panel (b), CALVADOS displays a clear ordering of the average cluster size, indicating the scGFP as the least aggregating and FP583 as the most, with wtGFP and sfGFP having similar intermediate values. The order is maintained overall in COCOMO, which also displays the expected enhanced effect of ionic strength in increasing the cluster size, while Trovato inverts the order of sfGFP and FP583.

The size of clusters in CALVADOS near to 1 in all cases is coherent with the fact that the average number of contacts basically does not increase from the starting randomly generated structure (see Figure 6), indicating that the systems simulated with this FF are very near to the disperse state. Considering the meaning of h, we can use the ϕ,k_bT/h phase diagram in place of the usual ϕ,k_bT/ε, and locate the different FFs and variants according to their value of h in the bell-shaped coexistence region (see, e.g. ref 8). This is shown in the inset of panel (a) of Figure 7, which reports the location of CALVADOS-simulated systems just below the aggregate−disperse transition line, while the others are located below, with Trovato rather near the clusters-percolate separation line[8] (more details in Supporting Information S3).

The Contact Probability Matrices Identify the Protein−Protein Interfaces. Figure 8 reports the contact probability matrices p^{d} from simulations of different variants and with different FFs, compared to the p^{AF} from AF3.

The first difference between AF and the simulations is the already observed low but non-null probability for any i, j pair (appearing as a yellow background in Figure 3) due to the statistical interface sampling during dynamics. This is subtracted in Figure 8 to make the comparison easier (original versions are in Figure S10). This is absent in p^{AF}, since it results from learning over experimental structures where only the stable interfaces are revealed. Even with the background subtracted, the contact probability maps from simulations appear more diffuse compared to the more localized and sharply defined contacts predicted by AF. Nevertheless, both AF and all the FFs are able to identify the primary and secondary interfaces, where probability is higher. For GFP mutants, the primary is roughly located between positions 130 and 220, and the secondary one is located around position 60 (and their crossings), in agreement with the crystallographic interfaces reported in Figure 1. A similar agreement is found for the two interfaces of the FP583.

The analysis of Pearson's and Jaccard indexes allows the quantification of the agreement. The Pearson's coefficient r is reported in Figure 9a for the different variants and different FFs, and compared to experimental and AF matrices as per the legend reported in the figure. In general, the values of the coefficient are rather low, due to the already noted presence of a background of low probability in the matrices from simulations, which is negligible in p^{AF} and p^{exp}. Nevertheless, the relative comparison of indexes is significant. We first observe that the agreement with p^{exp} is generally slightly better than with p^{AF}, and a generally better match is obtained for wtGFP than for others, which is explained by considering that these are the most represented in the statistical data set over which the FF are parametrized and AF is trained. Concerning the performances of the FFs, we observe systematically better performances for Trovato, followed by COCOMO and CALVADOS. There are, however, exceptions: CALVADOS performs best of all on the wtGFP, while Trovato performs the best, especially in the less structurally represented variants, such as FP583 and scGFP. Considering the specialization of the Trovato FF on the IFPs, a higher accuracy in the interface prediction was somehow expected.

This analysis is confirmed by the Jaccard index reported in panel (b), displaying similar trends as r. Table S3 reports the numerical values of the indexes and the graphical analysis performed with J calculated and restricted to primary and secondary interfaces. Trovato confirms the best in predicting the primary interface for GFPs (residues 140−228), and with enhanced accuracy for both interfaces of FP583 (residues 90−130 and 120−220, respectively). Furthermore, Trovato is the only FF capable of identifying the primary interface of FP583, whereas all the other FFs exhibit lower J values for this variant.

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

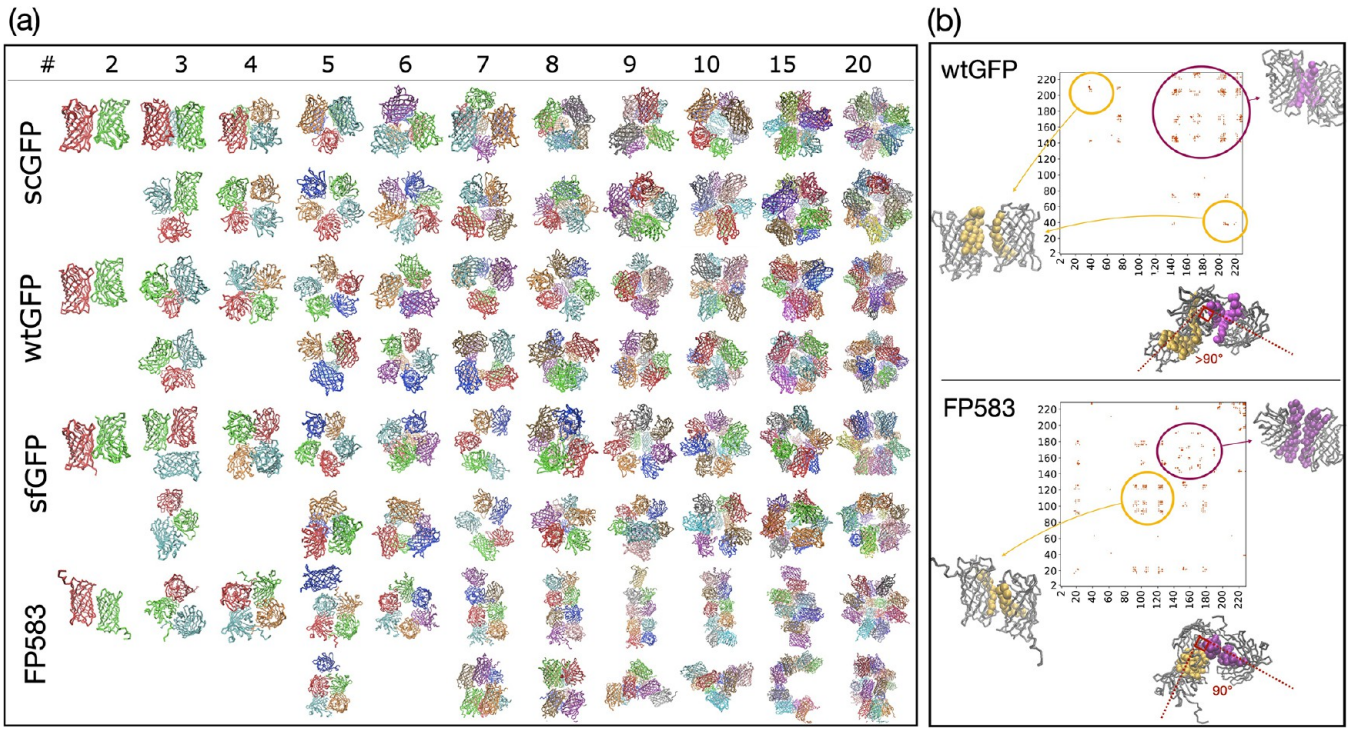


Figure 10. (a) Clusters generated by AF3. For each IFP variant, clusters up to 20-mer are reported, choosing two structures that are the most diverse. (b) Illustration of primary and secondary interfaces of wtGFP and FP583, and of the different orientations of trimers formed by the two kinds of interfaces.

The largest aggregation tendency produced by the Trovato FF is probably related to the highly specific stabilization of this interface. Conversely, CALVADOS appears to be more accurate in reproducing the secondary interface (residues 40−80) for both the wtGFP and scGFP variants (Table S3). The primary interface of all variants is mainly stabilized by hydrophobic contacts and side-chain hydrogen bonds, while secondary interfaces are mainly stabilized by polar interactions and side-chain hydrogen bonds.

In general, we note that the different behavior of Trovato FF, more specifically revealing the IFP interfaces, and CALVADOS-COCOMO, more general and less specific, can be understood primarily by considering the used data set for parametrization: in Trovato, it is composed of structured proteins with folds similar to the IFP, while for the others, it is composed of disordered proteins. In the latter, general hydrophobic interactions are weaker due to solvent exposure and are therefore enthalpically disfavored. Besides, specific interactions are less represented due to disorder and are therefore entropically unfavored.

Protein Interfaces Determine the Global Structure of Clusters. We now turn to compare the cluster structures generated during simulation to the multimers generated by AF3. We start from the latter, reported in Figure 10.

Coherent with their similar interfaces, the three mutants of GFP have similar multimer structures, while the FP583 multimers are different. This can be understood by examining the local geometry of the two interfaces in each cases. In GFP, the two interfaces are located so that, in a trimer, the three proteins form an obtuse angle when viewed from the top (see panel (b)). Therefore, in large multimers where proteins are organized with parallel axes, they can form a "ring". Small torsion between protein axes and the possibility of forming side-to-top contacts also favor the formation of hollow globular structures in larger multimers. FP583, conversely, has its two interfaces locate at about 90°; therefore, it preferentially forms compact tetramers, which can in turn organize into regular arrays that can also be elongated. These are preferred over hollow rings or globular structures, which can still, however, form by top-side contacts. Consistently, the radius of gyration predicted by AF3 for multimers of FP583 is slightly larger than that of GFP variants, e.g., for the 20 mer ∼50 Å and ∼60 Å, respectively (see also Figure 11).

The clusters obtained from simulations are generally more disordered. A classification of multimers into groups according to the number of monomers for the different FFs/IFPs is reported in Figure 12, while the size is also quantified by the radius of gyration reported in Figure 11. Again, we observe that the average size of clusters in simulations with CALVADOS, COCOMO, and Trovato increases in the mentioned order, in agreement with the results shown in the previous section on viscosity, contact formation dynamics, h value, and increasing aggregation tendency.

More specifically, CALVADOS primarily produces very small clusters whose aggregation-disaggregation dynamics are fast. Although some of the ring and globular conformations produced by AF3 are recognizable in the clusters from simulations, the few aggregated structures are generally more expanded and somewhat branched. In spite of the very small clusters formed, an increase in their average size is observed when transitioning from scGFP or wtGFP to sfGFP or FP583, for which clusters of up to 4−8 monomers are observed.

COCOMO produces clusters of 5−20 monomers on average, with a similar trend of cluster size increasing as one moves from GFPs to FP583. Interestingly, when the comparison is possible, the relationship between the number of monomers and the radius of gyration is similar to that of CALVADOS (see Figure 11a−d and (e) for the exponent of the fit). This indicates that although the aggregation propensity is different, the structure of

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

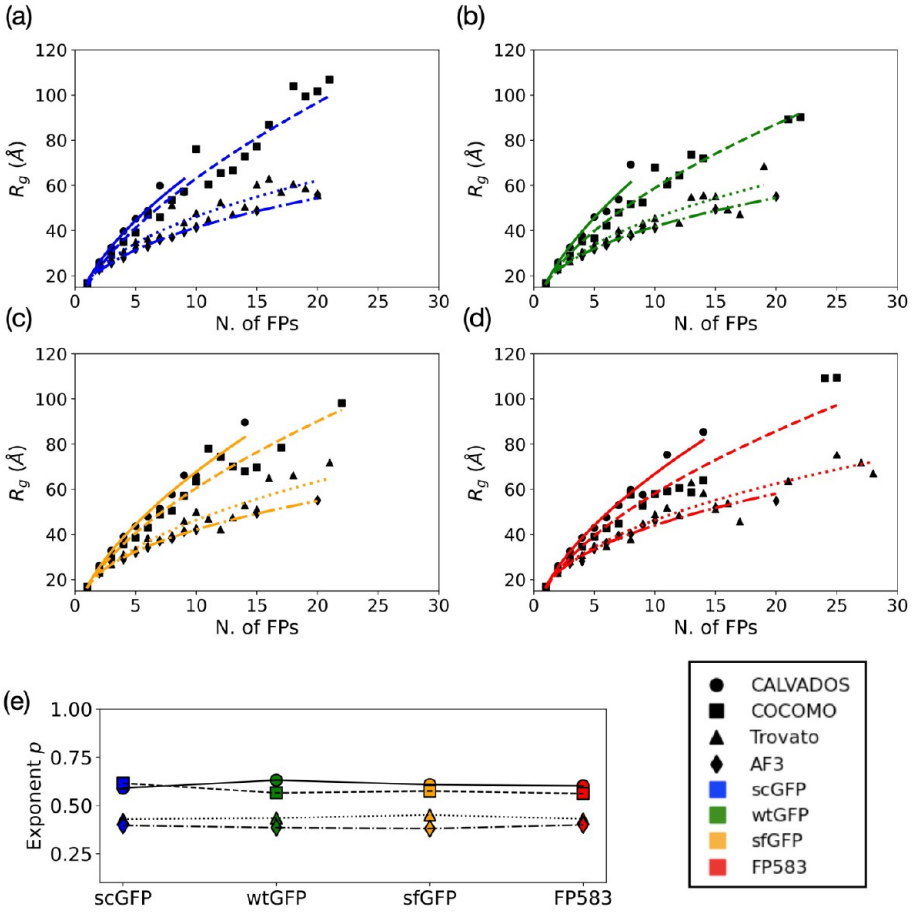


Figure 11. Cluster radius of gyration as a function of the number of proteins in the cluster, averaged over trajectories for different force fields (symbols reported in the legend) and different variants: (a) scGFP, (b) wtGFP, (c) sfGFP, and (d) FP583. Data are fitted using exponential scaling law $R_g \propto N^p$ (N = number of monomers in the cluster), with the exponent p reported in subplot (e). (For the sake of comparison, the exponent p is 1/3 for spherical clusters, $p \sim 0.6-1$ for linear clusters, and $p \sim 0.4-0.5$ for branched and fractal aggregates.) Line coding: solid = CALVADOS, dashed = COCOMO, dotted = Trovato, dash-dotted = AF3.

the formed clusters is similar in the two FFs. Indeed, we observe a considerable tendency to form branched and elongated structures. These are mainly present in FP583 but also in the other mutants, although for them, globular clusters are also present. Interestingly, fibrillar structures were experimentally observed in sfGFP.[39]

Trovato produces the largest clusters in terms of the number of proteins included and is more compact and generally more similar to AF3 prediction, with more elongated forms in FP583 and more globular ones in wtGFP. Coherently, the gyration radius data of Trovato superimpose fairly well with those of AF3, reproducing the relation between radius and the number of monomers (see Figure 11 panel (e)), consistent with the previously observed higher similarity of Trovato probability matrices to those of AF3.

SUMMARY AND CONCLUSIONS

In summary, in this work we have analyzed and compared the performances of three different inter-residue potentials in predicting the aggregation of four IFP variants with different aggregation tendencies and compared them with the predictions of AlphaFold3. Overall, the three FFs return different aggregation tendencies for all the IFP variants; specifically, the increasing order of predicted aggregation is CALVADOS < COCOMO < Trovato. This is coherently confirmed by the analysis of diffusion/viscosity, interprotein contact formation dynamics, and cluster size statistics, which we find to be quantitatively correlated, and is coherent with an *a priori* evaluation based on the definition of a parameter h representing an average attraction energy. The order of aggregation is in agreement with what is known from experiments, indicating sfGFP and FP583 are more aggregating than the wtGFP, and scGFP is less aggregating. Increasing the ionic strength is correctly observed to enhance aggregation, especially in charged variants.

The simulations allow one to locate the IFPs in the aggregation phase diagram. Specifically, IFP-concentrated solutions simulated with CALVADOS are located just below the disperse−aggregate transition line, with scGFP, wtGFP, sfGFP, and FP583 increasingly moving deeper into the aggregate region. COCOMO and Trovato simulations are located deeper in the aggregate region and increasingly nearer to the cluster-percolate region.

However, assessing which of these FFs is best in predicting aggregation is not straightforward. COCOMO is apparently the one that better reproduces the translational diffusivity, and in particular, the dependence of viscosity on concentration, as experimentally measured in aqueous solution. Moreover, although all FFs generally reproduce the differences in aggregation among the different variants—and all reproduce the same order of aggregation (scGFP < wtGFP < sfGFP <

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

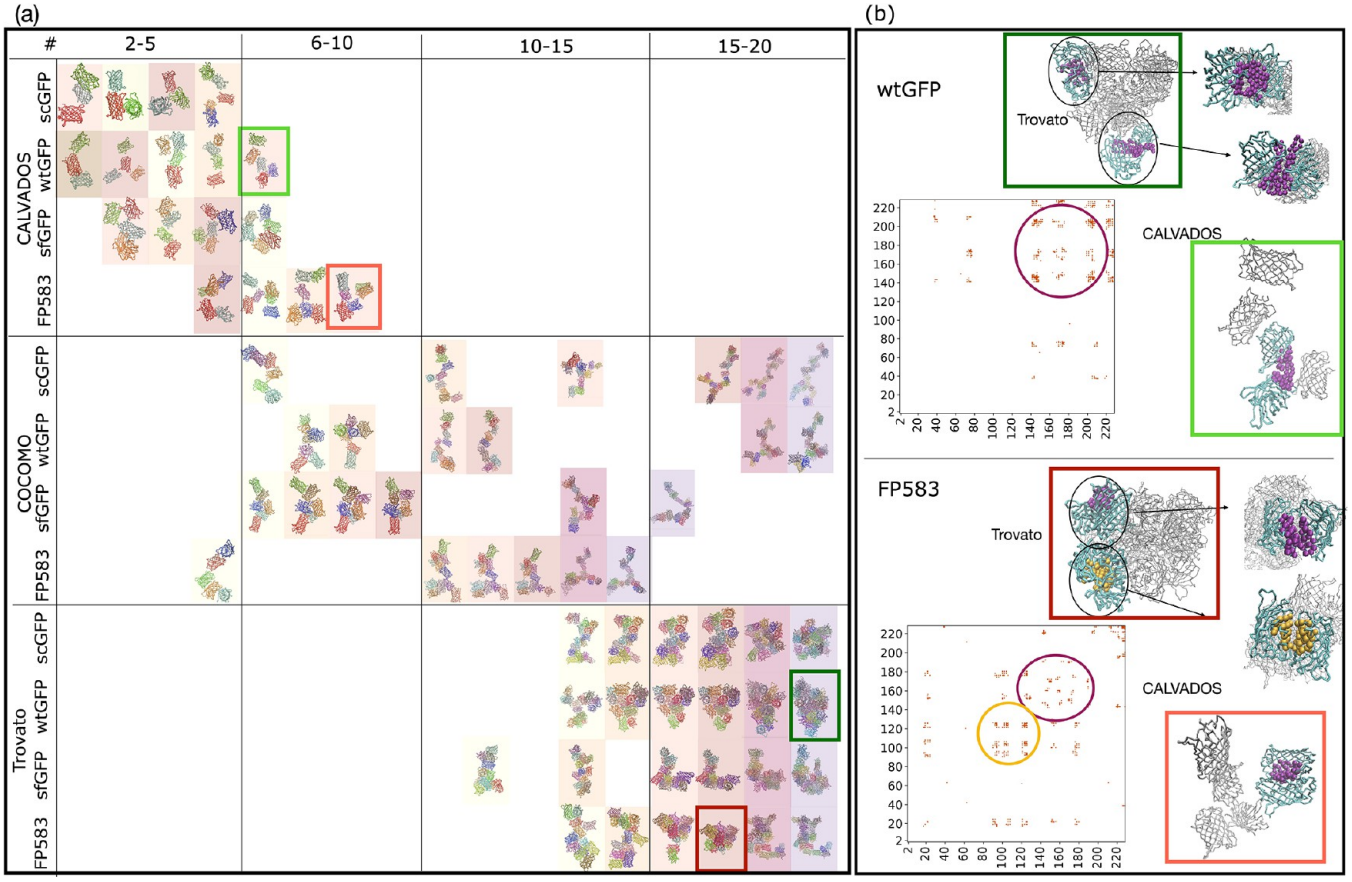


Figure 12. (a) Structure of the largest equilibrated clusters from the thermalized part of the trajectories. (Time evolution is encoded in the background color, from yellow to purple, with frames separated by 20 ns). (b) Selected largest clusters of wtGFP (green square in (a)) and FP583 (red squares in (a)) in simulations with the CALVADOS and Trovato FFs. Proteins interacting through the primary (purple) and secondary (orange) interfaces are highlighted, and the corresponding contacts are indicated in the reported contact matrix.

FP583) as well as the increase in aggregation with ionic strength—COCOMO is the one that most strongly enhances these effects, due to the stronger weight given to the electrostatic component in this FF.

On the other hand, CALVADOS predicts a very low number of protein−protein contacts, reflecting its tendency toward monomeric or weakly interacting states. This is due to its parameterization, which is optimized to capture intramolecular properties, such as the radius of gyration, rather than strong intermolecular interactions. As a consequence, CALVADOS often places structured proteins slightly below the aggregation threshold in phase diagrams and tends to underestimate aggregation, although it still correctly predicts the sequence-dependent aggregation tendency. We also note that hydrodynamic interactions, which may modify diffusivity in differential way on short and long time scales,[13] were not included in the simulations. These aspects represent an important direction for future work.

When looking at the structure of interfaces and clusters, AF3 data can also be included in the comparison, and new considerations emerge. Aside from a direct comparison between the clusters structure of similar sizes, we calculated inter-IFP contact probability matrices in the simulations and compare them with the contact probability matrices obtained by averaging those returned by AF3 among different monomers in multimers. The completely different methodologies used to calculate these inter-monomer contact probabilities produce different features in the matrices, such as a global background of non-null probability for any contact explored in simulations, which is absent in AF3.

In spite of this, both methodologies reveal primary and secondary interfaces, and therefore, a quantitative comparison is possible. It emerges that Trovato is the FF that displays the largest similarity to AF3, and more accurately captures the different features of the experimental interfaces, particularly those of the red variant FP583. This, in turn, produces some differences in the cluster structures, which again place AF3 and Trovato in a class different from CALVADOS and COCOMO: Trovato (and AF3) predicts more globular structures, especially for wtGFP, while allowing branching and elongation for other variants, especially sfGFP and FP583, whereas COCOMO tends to present less compact and more branched structures in general. Indeed, this is a good comparison feature, since fibrillar structures have been observed so far in sfGFP.

From the methodological point of view, the similarity between AF3 and the Trovato FF is interesting and can be understood by considering the derivation of this FF, which is based on the potential of mean force calculated on a statistical set including proteins with a fold of the IFP family. The AF3 algorithm is trained on a wide data set of proteins, mainly folded, and is specifically designed to predict the correct fold, which it does with high accuracy, especially for well-represented families such as that of IFPs. In both procedures, therefore, there is a

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

specificity toward folded proteins and toward the specific family, either included a priori or inferred from the sequence.

Trovato predicts a larger viscosity than that of COCOMO in simulations, as an effect of the stronger stabilization of specific interfaces. This indicates a larger capability of COCOMO to explore different and less specific interfaces, which is an added value if extensive explorations of the phases are needed. On the other way around, the accuracy in the interface prediction of COCOMO is lower, also with respect to CALVADOS, which performs comparably better on the variants of the GFP mutants. From this point of view, our results on the relative performances of COCOMO/CALVADOS, parametrized for disordered proteins, and Trovato on the IFPs are particularly useful in view of simulating systems that include both ordered and disordered proteins, such as, for instance, sensors based on the combination of GFPs with peptides.

On the methodological level, we have also shown here a possible way of combining, comparing, and integrating AlphaFold into molecular dynamics, which is somehow an alternative to the current trends. Neural network-based algorithms are recently mostly used in coarse-grained models, either directly to predict forces or even dynamics, or in advanced fitting procedures of CG FFs,[34] with excellent results. Here, we take a different perspective, considering that AF and FF-CG-based simulations can align on common concepts such as the contact probability matrix, which is the primary output of AF, and can also be calculated from simulations. This opens the way to using these two methods in deeper combination to take reciprocal advantage: on one side, to optimize the FF parametrization, and on the other, to enhance the explainability of the AF algorithm, especially in the prediction of interfaces, where its reliability is still under test.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.5c06850.

A pdf file named "SI.pdf" including some additional data and details of simulations; three text files with the nonbonded remapped parameters of the three FFs, ready to use for LAMMPS; Movies from simulations, specifically: FP583 with Trovato FF, wtGFP with COCOMO FF, scGFP with CALVADOS FF (PDF)

Non-bonded interaction parameters for CALVADOS coarse-grained potential in LAMMPS (TXT)

Non-bonded interaction parameters for Trovato coarse-grained potential in LAMMPS (TXT)

Non-bonded interaction parameters for COCOMO coarse-grained potential in LAMMPS (TXT)

Dynamics movie of wtGFP with COCOMO FF (MOV)

Dynamics movie of scGFP with CALVADOS FF (MOV)

Dynamics movie of FP583 with Trovato FF (MOV)

■ AUTHOR INFORMATION

Corresponding Author

Margherita Bini − *Scuola Normale Superiore, Pisa 56127, Italy; Istituto Nanoscienze, National Research Council (CNR-NANO), Pisa 56127, Italy;* orcid.org/0000-0002-7829-3507; Email: Email:, margherita.bini@nano.cnr.it

Authors

Giorgia Brancolini − *Istituto Nanoscienze, National Research Council (CNR-NANO), Modena 41100, Italy;* orcid.org/0000-0002-1083-9198

Valentina Tozzini − *Istituto Nanoscienze, National Research Council (CNR-NANO), Pisa 56127, Italy; Scuola Normale Superiore, Pisa 56127, Italy; INFN Sezione di Pisa, Pisa 56127, Italy;* orcid.org/0000-0002-7586-5039

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.5c06850

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors acknowledge the support of Next Generation-EU (Piano Nazionale di Ripresa e Resilienza (PNRR), Missione 4, Componente 2, Ecosistemi dell'Innovazione) through the project Tuscany Health Ecosystem (THE-Spoke 1, grant ECS_00000017); INFN CSN5 through the FRIDA Call and the MIRO project; and the SimBioSys CNR-internal project. We acknowledge the CINECA awards IsB27_SETE (no. HP10BSTFDD) and IsCb7_AG-ABETA (no. HP10C8JNX2) under the ISCRA initiative for the availability of high-performance computing resources and support.

■ ADDITIONAL NOTES

[1]We note that, however, our model has actually charge − 29e: in order to prevents structural biases, we build the models for all IFP variants of the same size, by trimming the sometime present unstructured tails, which in scGFP include a charged residue.
[2]CALVADOS-2 is chosen. See the Supporting Information for details on this choice.
[3]The dependence of the results on r_{cut} was checked to be weak.
[4]Except Trovato in the case of wtGFP.

■ REFERENCES

(1) Uversky, V. N. Intrinsically Disordered Proteins in Overcrowded Milieu: Membrane-less Organelles, Phase Separation, and Intrinsic Disorder. *Curr. Opin. Struct. Biol.* **2017**, *44*, 18−30.
(2) Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The Amyloid State and Its Association with Protein Misfolding Diseases. *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 384−396.
(3) Heo, L.; Sugita, Y.; Feig, M. Protein assembly and crowding simulations. *Curr. Opin. Struct. Biol.* **2022**, *73*, 102340.
(4) Brancolini, G.; Tozzini, V. Building Minimalist Models for Functionalized Metal Nanoparticles. *Front. Mol. Biosci.* **2019**, *6*, 00050.
(5) Brancolini, G.; Rotello, V. M.; Corni, S. Role of Ionic Strength in the Formation of Stable Supramolecular Nanoparticle-Protein Conjugates for Biosensing. *Int. J. Mol. Sci.* **2022**, *23*, 2368.
(6) Montepietra, D.; Germelli, L.; Marchetti, L.; Tozzini, V.; Angeloni, E.; Giacomelli, C.; Storti, B.; Bizzarri, R.; Barresi, E.; Taliani, S.; et al. An intramolecular FRET Biosensor for the Detection of SARS-CoV-2 in biological fluids. *Nanoscale* **2025**, *17* (14), 8803−8815.
(7) Bini, M.; Tozzini, V.; Brancolini, G. Deconstructing Electrostatics of Functionalized Metal Nanoparticles from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2023**, *127*, 8226−8241.
(8) Bini, M.; Brancolini, G.; Tozzini, V. Aggregation Behavior of Nanoparticles: Revisiting the Phase Diagram of Colloids. *Front. Mol. Biosci.* **2022**, *9*, 986223.
(9) Rumyantsev, A. M.; Gavrilov, A. A.; Johner, A. Complete Diagram of Conformational Regimes for Polyampholytic Disordered Proteins. *Macromolecules* **2024**, *57*, 5533−5544.

(10) Dautel, D. R.; Champion, J. A. Protein Vesicles Self-Assembled from Functional Globular Proteins with Different Charge and Size. *Macromolecules* **2021**, *22*, 116−125.
(11) Palermo, G.; Bonvin, A. M. J. J.; Dal Peraro, M.; Amaro, R. E.; Tozzini, V. Editorial: Multiscale Modeling from Macromolecules to Cell: Opportunities and Challenges of Biomolecular Simulations. *Front. Mol. Biosci.* **2020**, *7*, 194.
(12) Sterpone, F.; Melchionna, S.; Tuffery, P.; Pasquali, S.; Mousseau, N.; Cragnolini, T.; Chebaro, Y.; St-Pierre, J.; Kalimeri, M.; Barducci, A.; Laurin, Y.; Tek, A.; Baaden, M.; Nguyen, P.; Derreumaux, P. The OPEP protein model: from single molecules, amyloid formation, crowding and hydrodynamics to DNA/RNA systems. *Chem. Soc. Rev.* **2014**, *43*, 4871−4893.
(13) Sterpone, F.; Derreumaux, P.; Melchionna, S. Protein Simulations in Fluids: Coupling the OPEP Coarse-Grained Force Field with Hydrodynamics. *J. Chem. Theory Comput.* **2015**, *11*, 1843−1853.
(14) Bari, D. D.; Timr, S.; Guiral, M.; Giudici-Orticoni, M.-T.; Seydel, T.; Beck, C.; Petrillo, C.; Derreumaux, P.; Melchionna, S.; Sterpone, F.; et al. Diffusive Dynamics of Bacterial Proteome as a Proxy of Cell Death. *ACS Cent. Sci.* **2023**, *9*, 93−102.
(15) Binder, J.; Zalar, M.; Huelsmeyer, M.; Siedler, M.; Curtis, R.; Friess, W. Enhancing Martini 3 for protein self-interaction simulations. *Eur. J. Pharm. Sci.* **2025**, *209*, 107068.
(16) Martins, G.; Galamba, N. Wild-Type α-Synuclein Structure and Aggregation: A Comprehensive Coarse-Grained and All-Atom Molecular Dynamics Study. *J. Chem. Inf. Model.* **2024**, *64*, 6115−6131.
(17) Tozzini, V. Minimalist Models for Proteins: A Comparative Analysis. *Q. Rev. Biophys.* **2010**, *43*, 333−371.
(18) Brancolini, G.; Lopez, H.; Corni, S.; Tozzini, V. Low-resolution models for the interaction dynamics of coated gold nanoparticles with β2-microglobulin. *Int. J. Mol. Sci.* **2019**, *20*, 3866.
(19) Delfino, F.; Porozov, Y.; Stepanov, E.; Tamazian, G.; Tozzini, V. Evolutionary Switches Structural Transitions via Coarse-Grained Models. *J. Comput. Biol.* **2020**, *27*, 189−199.
(20) Trovato, F.; Nifosì, R.; Fenza, A. D.; Tozzini, V. A Minimalist Model of Protein Diffusion and Interactions: The Green Fluorescent Protein within the Cytoplasm. *Macromolecules* **2013**, *46* (20), 8311−8322.
(21) Dignon, G. L.; Zheng, W.; Kim, Y. C.; Best, R. B.; Mittal, J. Sequence Determinants of Protein Phase Behavior from a Coarse-Grained Model. *PLoS Comput. Biol.* **2018**, *14*, No. e1005941.
(22) Tesei, G.; Schulze, T. K.; Crehuet, R.; Lindorff-Larsen, K. Accurate Model of Liquid-Liquid Phase Behavior of Intrinsically Disordered Proteins from Optimization of Single-Chain Properties. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, *118* (44), No. e2111696118.
(23) Valdes-Garcia, G.; Heo, L.; Lapidus, L. J.; Feig, M. Modeling Concentration-dependent Phase Separation Processes Involving Peptides and RNA via Residue-Based Coarse-Graining. *J. Chem. Theory Comput.* **2023**, *19*, 669−678.
(24) Nifosí, R.; Amat, P.; Tozzini, V. Pietro Amat Variation of Spectral, Structural, and Vibrational Properties within the Intrinsically Fluorescent Proteins Family: A Density Functional Study. *J. Comput. Chem.* **2007**, *28*, 2366−2377.
(25) Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and Characterization of a Superfolder Green Fluorescent Protein. *Nat. Biotechnol.* **2006**, *24*, 79−88.
(26) Mout, R.; Tonga, G. Y.; Wang, L.-S.; Ray, M.; Roy, T.; Rotello, V. M. Programmed Self-Assembly of Hierarchical Nanostructures through Protein-Nanoparticle Coengineering. *ACS Nano* **2017**, *11* (4), 3456−3462.
(27) Ray, M.; Brancolini, G.; Luther, D. C.; Jiang, Z.; Cao-Milán, R.; Cuadros, A. M.; Burden, A.; Clark, V.; Rana, S.; Mout, R.; Landis, R. F.; Corni, S.; Rotello, V. M. High Affinity Protein Surface Binding through Co-Engineering of Nanoparticles and Proteins. *Nanoscale* **2022**, *14*, 2411−2418.
(28) Ma, C.; Malessa, A.; Boersma, A. J.; Liu, K.; Herrmann, A. Supercharged Proteins and Polypeptides. *Adv. Mater.* **2020**, *32* (20), 1905309.
(29) Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **1994**, *263*, 802−805.
(30) Wall, M. A.; Socolich, M.; Ranganathan, R. The Structural Basis for Red Fluorescence in the Tetrameric GFP Homolog DsRed. *Nat. Struct. Biol.* **2000**, *7*, 1133−1138.
(31) Laber, J. R.; Dear, B. J.; Martins, M. L.; Jackson, D. E.; DiVenere, A.; Gollihar, J. D.; Ellington, A. D.; Truskett, T. M.; Johnston, K. P.; Maynard, J. A. Charge Shielding Prevents Aggregation of Supercharged GFP Variants at High Protein Concentration. *Mol. Pharmaceutics* **2017**, *14*, 3269−3280.
(32) Senior, A. W.; et al. Improved protein structure prediction using potentials from deep learning. *Nature* **2020**, *577*, 706−710.
(33) Jumper, J.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(34) Tozzini, V.; Giulivi, C. Editorial: Revolutionizing Life Sciences: The Nobel Leap in Artificial Intelligence-Driven Biomodeling. *Front. Mol. Biosci.* **2025**, *11*, 1540823.
(35) Ghosh, D.; Biswas, A.; Radhakrishna, M. Advanced computational approaches to understand protein aggregation. *Biophys. Rev.* **2024**, *5*, 021302.
(36) Abramson, J.; et al. VAccurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493−500.
(37) Elsliger, M.-A.; Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J. Structural and Spectral Response of Green Fluorescent Protein Variants to Changes in pH. *Biochemistry* **1999**, *38* (17), 5296−5301.
(38) Scott, D. J.; Gunn, N. J.; Yong, K. J.; Wimmer, V. C.; Veldhuis, N. A.; Challis, L. M.; Haidar, M.; Petrou, S.; Bathgate, R. A. D.; Griffin, M. D. W. A Novel Ultra-Stable, Monomeric Green Fluorescent Protein for Direct Volumetric Imaging of Whole Organs Using CLARITY. *Sci. Rep.* **2018**, *8* (1), 667.
(39) Stepanenko, O. V.; Sulatsky, M. I.; Mikhailova, E. V.; Kuznetsova, I. M.; Turoverov, K. K.; Stepanenko, O. V.; Sulatskaya, A. I. New findings on GFP-like protein application as fluorescent tags: Fibrillogenesis, oligomerization, and amorphous aggregation. *Int. J. Biol. Macromol.* **2021**, *192*, 1304−1310.
(40) Tozzini, V.; Rocchia, W.; McCammon, J. Mapping all-atom models onto one-bead coarse-grained models: general properties and applications to a minimal polypeptide model. *J. Chem. Theory Comput.* **2006**, *2*, 667−673.
(41) Fenza, A. D.; Rocchia, W.; Tozzini, V. Complexes of HIV-1 integrase with HAT proteins: Multiscale models, dynamics, and hypotheses on allosteric sites of inhibition. *Proteins: Struct., Funct., Bioinf.* **2009**, *76*, 946−958.
(42) Delfino, F.; Porozov, Y.; Stepanov, E.; Tamazian, G.; Tozzini, V. Structural Transition States Explored With Minimalist Coarse Grained Models: Applications to Calmodulin. *Front. Mol. Biosci.* **2019**, *6*, 104.
(43) Tesei, G.; Lindorff-Larsen, K. Improved Predictions of Phase Behaviour of Intrinsically Disordered Proteins by Tuning the Interaction Range. *Open Res. Eur.* **2023**, *2*, 94.
(44) Vlachy, V.; Kalyuzhnyi, Y. V.; Hribar-Lee, B.; Dill, K. A. Protein Association in Solution: Statistical Mechanical Modeling. *J. Chem. Phys.* **2023**, *13*, 1703.
(45) Thompson, A. P.; Aktulga, H. M.; Berger, R.; Bolintineanu, D. S.; Brown, W. M.; Crozier, P. S.; In't Veld, P. J.; Kohlmeyer, A.; Moore, S. G.; Nguyen, T. D.; et al. LAMMPS - a Flexible Simulation Tool for Particle-Based Materials Modeling at the Atomic, Meso, and Continuum Scales. *Comput. Phys. Commun.* **2022**, *271*, 108171.
(46) Bini, M. *Script for generating and analyzing coarse-grained (CG) simulations of protein aggregation*. GitHub repository; Section: gfp-aggregation-coarse-grained, 2025, https://github.com/MargheritaBini/GFPs-aggregation-coarse-grained
(47) von Bülow, S.; Bullerjahn, J. T.; Hummer, G. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure. *J. Chem. Phys* **2020**, *153* (2), 021101.

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356

(48) Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO-the Open Visualization Tool. *Modell. Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.
(49) https://alphafoldserver.com/about.

https://doi.org/10.1021/acsomega.5c06850
ACS Omega 2025, 10, 54341−54356