



UNIVERSITAT DE
BARCELONA

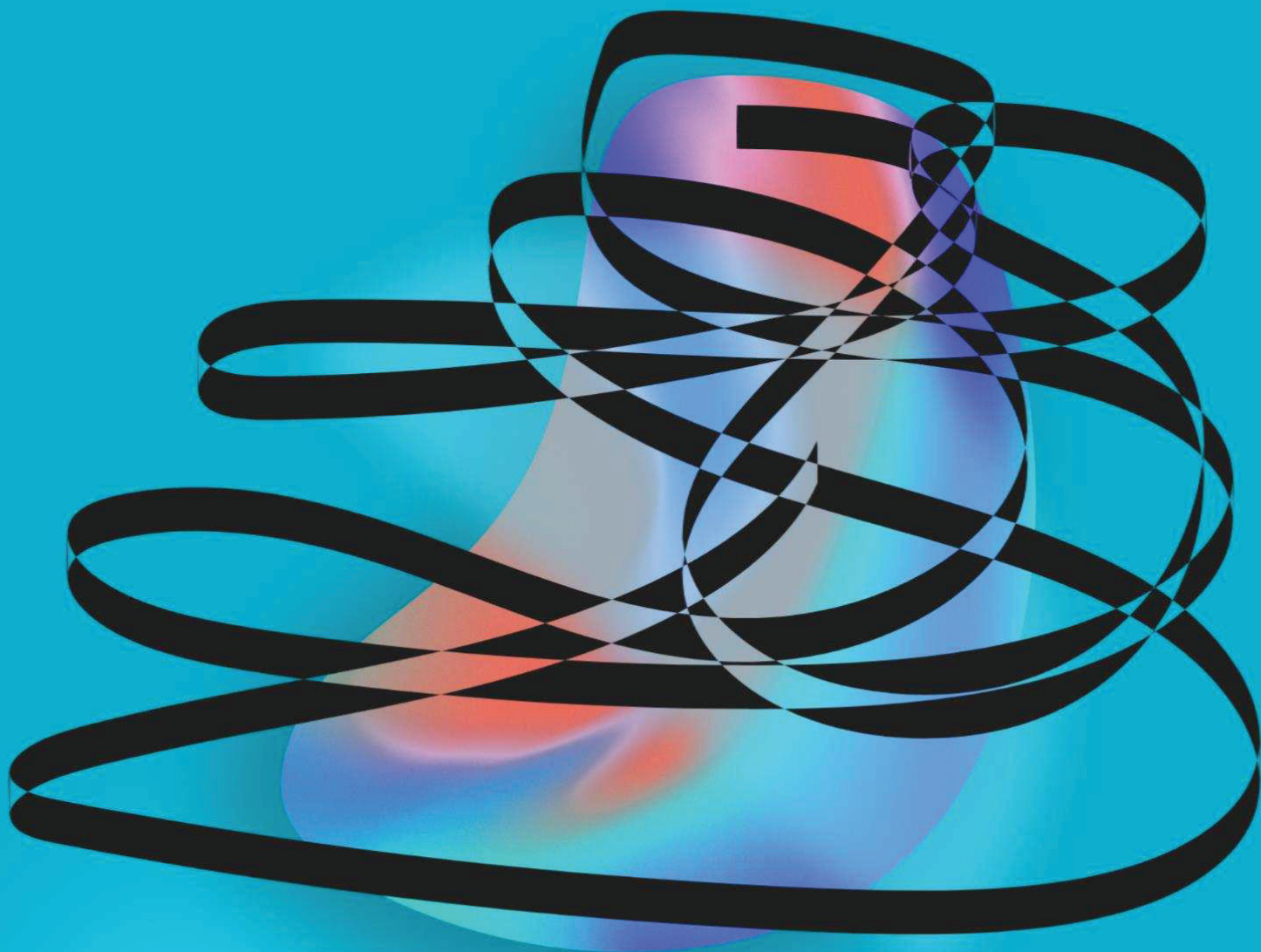
Characterization of protein conformational ensembles from Förster resonance energy transfer simulations

Daniel Gonzalo Palao

ADVERTIMENT. La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX (www.tdx.cat) i a través del Dipòsit Digital de la UB (diposit.ub.edu) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

ADVERTENCIA. La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR (www.tdx.cat) y a través del Repositorio Digital de la UB (diposit.ub.edu) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

WARNING. On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX (www.tdx.cat) service and by the UB Digital Repository (diposit.ub.edu) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.



Characterization Of Protein Conformational Ensembles From Förster Resonance Energy Transfer Simulations



UNIVERSITAT DE
BARCELONA

Daniel Gonzalo Palao



UNIVERSITAT DE
BARCELONA

UNIVERSITAT DE BARCELONA

FACULTAT DE QUÍMICA

INSTITUT DE QUÍMICA TEÒRICA I COMPUTACIONAL

DEPARTAMENT DE CIÈNCIA DE MATERIALS I QUÍMICA
FÍSICA

**CHARACTERIZATION OF PROTEIN
CONFORMATIONAL ENSEMBLES FROM FÖRSTER
RESONANCE ENERGY TRANSFER SIMULATIONS**

**CARACTERITZACIÓ DE CONJUNTS
CONFORMACIONALS DE PROTEÏNES A PARTIR DE
SIMULACIONS DE TRANSFERÈNCIA D'ENERGIA DE
RESSONÀNCIA DE FÖRSTER**

DANIEL GONZALO PALAO

BARCELONA, 2024



For the Degree of Doctor in
Theoretical Chemistry and Computational Modelling
at the University of Barcelona

**Characterization of protein conformational ensembles from Förster
resonance energy transfer simulations**

Memòria presentada per:

Daniel Gonzalo Palao

per optar al títol de Doctor per la Universitat de Barcelona

Dr. Carles Eduard Curutchet Barat

Director

**Departament de Farmàcia i Tecnologia
Farmacèutica i, Fisicoquímica**

Dra. Mercè Deumal Solé

Tutora

**Departament de Ciència de
Materials i Química Física**

Barcelona, desembre del 2024

*En memoria de mi abuela,
a mi familia y seres queridos,*

Abstract

This thesis focuses on the study of partially disordered proteins (IDPs), using calmodulin (CaM) as the system under study. Calmodulin is a crucial protein involved in numerous physiological processes, and its well-documented structure and function make it an ideal candidate for this research. The primary aim is to explore the conformational ensembles of calmodulin through molecular dynamics (MD) simulations, using two distinct force fields: Amber ff14SB, widely used for proteins, and a99SB-disp, specifically refined for disordered proteins like CaM. By generating and analyzing FRET observables, the study evaluates which force field better reflects the dynamic behavior of calmodulin.

A central component of this work is the application of Förster Resonance Energy Transfer (FRET) to gain insights into protein conformational dynamics. Advanced computational methods, including TrESP-MMPol and QM/MMPol, are employed to model the energy transfer process and assess how environmental factors affect FRET observables. A particular emphasis is placed on environmental screening, which describes how interactions between the chromophores and their surroundings such as solvent molecules or protein structures modify energy transfer. This screening effect addresses limitations in traditional models, offering a more accurate representation of real-world systems.

In addition to environmental screening, this thesis explores key factors influencing FRET, such as Coulombic dipole interactions and orientational anisotropy between the chromophores. A novel screening function is also introduced to study how environmental effects depend on distance, enhancing our understanding of energy transfer in complex biological systems. By connecting theoretical models with experimental data, this research deepens our understanding of calmodulin's conformational and functional behavior. Through the integration of computational simulations and FRET methodologies, this work advances our ability to study partially disordered proteins and contributes valuable insights to the broader field of protein dynamics and energy transfer.

Keywords: disordered proteins, molecular dynamics, FRET, excited states, multiscale QM/MM models

Resum

Aquesta tesi doctoral es centra en l'estudi de proteïnes parcialment desordenades (IDPs), utilitzant la calmodulina (CaM) com a sistema sota estudi. La calmodulina és una proteïna clau involucrada en nombrosos processos fisiològics, i la seva estructura i funció ben documentades la converteixen en un candidat ideal per a aquesta investigació. L'objectiu principal és explorar els conjunts conformacionals de la calmodulina mitjançant simulacions de dinàmica molecular (MD), utilitzant dos camps de força diferents: Amber ff14SB, àmpliament utilitzat per a proteïnes, i a99SB-disp, refinat específicament per a proteïnes desordenades com la CaM. Mitjançant la generació i l'anàlisi d'observables FRET, l'estudi avalua quin camp de forces reflecteix millor el comportament dinàmic de la calmodulina.

Un component central d'aquest treball és l'aplicació de la Transferència d'Energia per Resonància de Förster (FRET) per obtenir informació sobre les dinàmiques conformacionals de proteïnes. Es fan servir mètodes computacionals avançats, com TrESP-MMPol i QM/MMPol, per modelar el procés de transferència d'energia i avaluar com els factors ambientals afecten els observables FRET. Un èmfasi particular es posa en l'apantallament de l'ambient, que descriu com les interaccions entre els cromòfors i el seu entorn (com molècules de solvent o estructures proteïques) modifiquen la transferència d'energia. Aquest efecte d'apantallament aborda les limitacions dels models tradicionals, oferint una representació més precisa dels sistemes reals. A més de l'apantallament ambiental, aquesta tesi explora factors clau que influeixen en els observables FRET, com les interaccions dipolars coulòmbiques i l'anisotropia orientacional relacionada amb els cromòfors. També es presenta una nova funció d'apantallament per estudiar com els efectes ambientals depenen de la distància, millorant la comprensió de la transferència d'energia en sistemes biològics complexos. Aquesta investigació aprofundeix en la comprensió del comportament conformacional i funcional de la calmodulina. Amb la integració de simulacions computacionals i metodologies FRET, aquest treball avança la capacitat d'estudiar proteïnes parcialment desordenades i aporta coneixements valuosos al camp més ampli de les dinàmiques proteïques i la transferència d'energia.

Paraules clau: proteïnes desordenades, dinàmica molecular, FRET, estats excitats, models QM/MM multiescala

Resumen

Esta tesis se centra en el estudio de proteínas parcialmente desordenadas (IDPs), utilizando la calmodulina (CaM) como el sistema de estudio. La calmodulina es una proteína clave involucrada en numerosos procesos fisiológicos, y su estructura y función bien documentadas la convierten en un candidato ideal para esta investigación. El objetivo principal es explorar los conjuntos conformacionales de la calmodulina mediante simulaciones de dinámica molecular (MD), utilizando dos campos de fuerza distintos: Amber ff14SB, ampliamente utilizado para proteínas, y a99SB-disp, refinado específicamente para proteínas desordenadas como la CaM. A través de la generación y análisis de observables FRET, el estudio evalúa cuál de los campos de fuerza refleja mejor el comportamiento dinámico de la calmodulina.

Un componente central de este trabajo es la aplicación de la Transferencia de Energía por Resonancia de Förster (FRET) para obtener información sobre las dinámicas conformacionales de proteínas. Se emplean métodos computacionales avanzados, como TrESP-MMPol y QM/MMPol, para modelar el proceso de transferencia de energía y evaluar cómo los factores ambientales afectan los observables FRET. Se pone un énfasis particular en el apantallamiento ambiental, que describe cómo las interacciones entre los cromóforos y su entorno (como moléculas de solvente o matrices proteicas) modifican la transferencia de energía. Este efecto de apantallamiento aborda las limitaciones de los modelos tradicionales, ofreciendo una representación más precisa de los sistemas reales. Además del apantallamiento ambiental, esta tesis explora factores clave que influyen en los observables FRET, como las interacciones dipolares coulombicas y la anisotropía orientacional relacionada con los cromóforos. También se presenta una nueva función de apantallamiento para estudiar cómo los efectos ambientales dependen de la distancia, mejorando la comprensión de la transferencia de energía en sistemas biológicos complejos. Esta investigación profundiza en la comprensión del comportamiento conformacional y funcional de la calmodulina. Con la integración de simulaciones computacionales y metodologías FRET, este trabajo avanza en la capacidad de estudiar proteínas parcialmente desordenadas y aporta conocimientos valiosos al campo más amplio de las dinámicas proteicas y la transferencia de energía. **Palabras clave:** proteínas desordenadas, dinámica molecular, FRET, estados excitados, modelos QM/MM multiescala

Acknowledgements

Las personas tendemos a almacenar recuerdos de muchos modos distintos y no siempre tenemos a nuestro alcance la posibilidad de revivirlos con intensidad. A veces, oler el simple aroma de un perfume o escuchar la melodía de una canción parece devolvernos a aquellas épocas o instantes que formaron parte de nuestra vida, y solemos pensar: *“Ya no era consciente de lo mal o lo bien que lo pasé durante aquella temporada”*. En mi caso, quiero escribir estos agradecimientos de todo corazón para que sean un testimonio cuando necesite recordar las lecciones que he aprendido durante esta tesis doctoral. No tengo ningún reparo en plasmar en esta sección lo complicada que ha sido esta etapa de mi vida a nivel personal. Me gustaría que estas palabras fueran testigo de que, sin la inestimable ayuda de algunas personas, esto no hubiera sido posible.

Me gustaría empezar dedicando unas palabras (y no por puro convencionalismo) a mi director de tesis, el Dr. Carles Curuchet.

Carles, quanta paciència has necessitat amb mi... Sempre t'estaré eternament agraït per confiar en mi més del que jo mateix ho he fet. Mai has vacil·lat a l'hora d'animar-me, contagiar-me del teu optimisme i no dubtar mai de les meves capacitats, quan potser hi havia motius per fer-ho. Has estat un director impecable tant en l'àmbit professional com en el personal. Sé prou com per reconèixer el teu gran nivell com a científic, sempre tens una bona resposta per a tot i la humilitat que et caracteritza fa que vegis qualsevol ocasió com una bona oportunitat per aprendre coses noves. Ha estat un plaer treballar amb tú, moltíssimes gràcies de tot cor!

Day-to-day life at the university has always been improved by the people who have surrounded me during these years. I am grateful for the warm welcome Marina gave me when, at the beginning, it was just the two of us. The additions of Renato, Beste, Özge, Blanca and Jhonatan brought many good moments and lots of laughter that made a happier journey. I can't forget the X-men, who are now the J-men: Álvaro, Patri, Marina, and Varbina. Thank you so much for welcoming me into your lab when smoke was practically coming out of my ears, and I needed to rest. I think we've had a great time together, and I know deep down I also brought some laughs along with me. I would also like to mention Jordi and Salo. It has always been fun to share moments with you both. I will never forget Jordi joking about my miserable state during the chaotic days of writing.

Mención especial para Carlos y Pol, siempre reunidos para la hora de la comida, ha sido un placer poder compartir todo este tiempo con vosotros. Las risas a vuestro lado han sido eternas y estoy tremendamente agradecido de haberos conocido.

Pablo, Xavi, Juan, Raúl y todo el UAB team nunca me habéis dejado de lado, estar a vuestro lado durante algunas temporadas me sigue inspirando para siempre intentar hacerlo un poco mejor.

Ibano, Niko y todo el barro team, han sido compañeros de vida y me han hecho pasar momentos que me han hecho disfrutar como un niño. Brindo por ellos y por los que están por venir.

Varbina, thank you so much for becoming someone so special to me. You have always given me your unconditional support and tried to understand me. Your sweetness always reminds me how beautiful my life is by your side, even during the most stressful moments. I hope to slowly give back the happiness you bring to all my days now that I will be a little calmer.

Àlex siento haber dado tanto la lata con esta tesis, siempre te has prestado a conversar sobre cualquier cosa y has estado en todo momento cuando lo he necesitado. Nunca has cambiado tu actitud conmigo cuando he pasado largas temporadas sin ganas de nada y siempre has sido fiel a nuestra eterna amistad. No tengo palabras para agradecerte lo mucho que me has ayudado.

Ikersito, que hayamos realizado el doctorado a la vez ha hecho que entendamos mejor que le ocurría al otro. No quiero repetir de nuevo todo lo que ya sabes, solo quiero que sepas que todos te echamos mucho de menos en Barcelona. Siempre nos vas a tener cuando lo necesites.

Alainsito... que decir de ti. Has llegado a querer con más fuerza que yo mismo que este día llegue solo porque sabías lo mucho que yo lo quería. Has sido mi compañero de batalla incondicionalmente y me has ayudado en todo lo que ha estado en tu mano. "No importa lo que se venga, lo haremos juntos".

Mi familia por supuesto ha sido una pieza clave desde que me embarque en el intrépido camino de la ciencia, pues sin saber mucho a lo que yo me enfrentaba siempre me han arropado y cuidado como los que más. Vosotros habéis hecho que sea como soy, para siempre agradecido. Papa, mama, hermana, os quiero muchísimo.

TABLE OF CONTENTS

CHAPTER 1: Introduction.....	27
1.1 Förster theory of electronic resonance energy transfer.....	32
1.1.1 Applications of Förster Resonance Energy Transfer (FRET)	32
1.2 Golden rule formulation for EET rate expression	37
1.3 The FRET tool in dynamic structural biology.....	45
1.4 Techniques to study the conformational preferences of disordered proteins	50
1.4.1 Experimental techniques for protein structure characterization	51
1.4.2 FRET and computational simulations: Exploring protein dynamics	56
1.5 Calmodulin (CaM).....	59
1.5.1 Dye-attached calmodulin FRET simulations.....	61
1.5.2 Roles of calmodulin in human body.....	62
1.5.3 Previous studies on Calmodulin	64
CHAPTER 2: Methodology	67
2.1 Techniques to simulate protein dynamics	69
2.1.1 Classical force fields	69
2.1.2 Classical molecular dynamics	73
2.1.3 Challenges in simulating disordered systems	74
2.2 Excited states	79
2.2.1 Excited states calculations in biological environments.....	79
2.2.2 QM/MMPol method	81
2.2.2.1 QM/MMPol model for excited states.....	84
2.2.2.2 Formulation based on transition densities: QM/MMPol.....	85
2.2.3 Formulation based on transition charges: TrESP-MMPol.....	87
2.2.4 Comparison of TrESP-MMPol and QM/MMPol Methods.....	89
2.2.5 Beyond Förster approximation: Dipole approximation and screenings effects.....	90

2.2.6 Generating FRET observables from MD trajectories.....	92
2.2.6.1 Coupling FRET dynamics and dye orientational dynamics (dynamic, static and intermediate averaging regimes).....	92
2.2.6.2 Derivation of J from R_0 and calculation of instantaneous $k(t)$ rates and efficiency and lifetimes along MD trajectory	96
2.2.6.3 Combining PDA and TrESP-MMPol approach	99
2.3 Computational protocol	101
CHAPTER 3: Objectives	103
CHAPTER 4: On the breakdown of Förster energy transfer theory due to solvent effects: Atomistic simulations unveil distance-dependent dielectric screening in calmodulin	109
4.1 Introduction.....	111
4.2 Methods: Multiscale MD/TrESP-MMPol approach.....	115
4.3 Computational details	116
4.3.1 MD simulations.....	116
4.3.2 Electronic coupling calculations	117
4.4 Conformational ensembles of CaM.....	120
4.5 Beyond Förster coupling: dielectric screening effects	125
4.6 Beyond Förster coupling: Deviations from the ideal isotropic distribution	130
4.7 Fluorescence Lifetime and Efficiency Distributions	134
CHAPTER 5: Impact of calcium binding and dye labelling in the structural ensemble of calmodulin studied using atomistic energy transfer simulations	143
5.1 Introduction.....	145
5.2 Methods: Multiscale MD/TrESP-MMPol approach	148
5.3 Computational details	149
5.3.1 MD simulations.....	149
5.3.2 Electronic coupling calculations	150
5.4 Conformational ensembles of CaM.....	152

5.5 Beyond Förster coupling: dipole approximation, dielectric screening and isotropic approximation	160
5.6 Fluorescence lifetime distributions.....	171
CHAPTER 6: Conclusions.....	177
APPENDIX 1	183
APPENDIX 2	191
BIBLIOGRAPHY.....	209

TABLE OF FIGURES

Figure 1.1: Different energy pathways after donor excitation	37
Figure 1.2: Basic FRET principle.....	37
Figure 1.3: Spectral overlap between donor emission and acceptor absorption..	40
Figure 1.4: Orientational factor illustration	41
Figure 1.5: Overview on FRET rates measure	42
Figure 1.6: Eq (1.12) plot distance vs efficiency dependence representation	43
Figure 1.7: Number of publications per year containing the keywords “FRET”	48
Figure 1.8: Representation of FRET applications in biomolecular studies	48
Figure 1.9: Results from NMR corresponding to the secondary structure of a protein	51
Figure 1.10: Conventional followed procedure in X-ray crystallography	53
Figure 1.11: CaM structure	59
Figure 1.12: The figure shows CaM in 2 different conformational states	60
Figure 1.13: Dyes molecular structure	61
Figure 1.14: Calmodulin dye labeled system.....	62
Figure 1.15: sMFRET CaM distributions	65
Figure 2.1: All atom versus coarse-grained representation in the CG model MARTINI of a protein section and water	77
Figure 2.2: Example of a QM/MM calculation of a biomolecular system	83
Figure 2.3: Representative scheme on QMMMPol transition densities	87
Figure 2.4: Representative scheme on TrESP-MMPol transition atomic charges..	88
Figure 2.5: MC scheme on FRET process	94
Figure 2.6: Computational protocol procedure workflow	102
Figure 4.1: Structure of <i>holo</i> Calmodulin tagged with the Texas Red C2 maleimide and Alexa Fluor 488 C5 maleimide dyes	114
Figure 4.2: Distribution of radius of gyration computed along the MD trajectories of <i>holo</i> CaM-AF488-TRC2 and CaM-TRC2-AF488 compared to the experimental value	122
Figure 4.3: Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings.....	127
Figure 4.4: Graphical representation of water and amino acid (residue) contributions to the MMPol environment-mediated term	129

Figure 4.5: Probability distribution of orientation factors derived from MD simulations for holo CaM-AF488-TRC2 and CaM-TRC2-AF488	130
Figure 4.6: Density distribution of orientation factors as a function of D/A separation	131
Figure 4.7: Distribution of donor-acceptor distances directly measured along MD trajectories	132
Figure 4.8: Ratio between PDA and TrESP coulombic coupling contributions ...	133
Figure 4.9: Distribution of fluorescent lifetimes	136
Figure 4.8: Fluorescence decay from MD/TrESP-MMPol	139
Figure 4.11: FRET rates averaged over MD/TrESP-MMPol data for the two CaM systems	140
Figure 5.1. A Structure of holo Calmodulin tagged with the Alexa Fluor 488 C5 maleimide and the Texas Red C2 maleimide dyes..	147
Figure 5.2. Conformational ensembles derived from MD simulations for holo CaM.	153
Figure 5.3. Conformational ensembles derived from MD simulations for apo CaM..	154
Figure 5.4. Distribution of radius of gyration computed along the MD trajectories of CaM compared to the experimental value	156
Figure 5.5. Ratio between PDA and TrESP coulombic coupling contributions.....	161
Figure 5.6. Density distribution of orientation factors as a function of D/A separation derived from MD simulations for CaM-AF488-TRC2/TRC2-AF488 systems	161
Figure 5.7. Density distribution of orientation factors as a function of D/A separation derived from MD simulations for CaM-AF488-AF594/AF594-AF488 systems	163
Figure 5.8. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for CaM-AF488-TRC2/TRC2-AF488	166
Figure 5.9. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for CaM-AF488-AF594/AF594-AF488 systems	168
Figure 5.10. Distribution of fluorescence lifetimes computed for holo and apo CaM-AF488-TRC2/TRC2-AF488 systems	174
Figure 5.11. Distribution of fluorescence lifetimes computed for holo and apo CaM-AF488-AF594/AF594-AF488 systems	175

Table 1-A1. Structural and FRET properties averaged over MD trajectories for holo CaM-AF488-TRC2 and CaM-TRC2-AF488 systems.	185
Figure 1-A1. Mean absolute percentage error (MAPE) in screening factors computed using QM/MMPol for 9 selected structures of CaM-AF488-TRC2 using different values of MMPol cutoff radius.	185
Figure 2-A1. Comparison between QM/MMPol screening factors, obtained using a fixed 20 Å MMPol cutoff radius, with TrESP/MMPol values obtained for different values of the cutoff.	186
Figure 3-A1. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for holo CaM-AF488-TRC2 and CaM-TRC2-AF488 systems along MD trajectories	187
Figure 4-A1. Contribution of the AF488 and TRC2 dyes to the total solvent-accessible surface area (SASA) of the holo CaM-AF488-TRC2 and CaM-TRC2-AF488 systems computed along MD trajectories.	189
Figure 5-A1. Electronic couplings and screening factors computed along MD trajectories for <i>holo</i> CaM-AF488-TRC2 and CaM-TRC2-AF488 systems	190
Figure 1-A2. Structural characterization of the ensemble of <i>holo</i> CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on ff14SB	193
Figure 2-A2. Structural characterization of the ensemble of <i>holo</i> CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on a99SB-disp.	194
Figure 3-A2. Structural characterization of the ensemble of <i>holo</i> CaM-AF488-AF594/AF594-AF488 from MD simulations based on ff14SB.	195
Figure 4-A2. Structural characterization of the ensemble of <i>holo</i> CaM-AF488-AF594/AF594-AF488 from MD simulations based on a99SB-disp.	196
Figure 5-A2. Structural characterization of the ensemble of <i>apo</i> CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on ff14SB.	197
Figure 6-A2. Structural characterization of the ensemble of <i>apo</i> CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on a99SB-disp.	198
Figure 7-A2. Structural characterization of the ensemble of <i>apo</i> CaM-AF488-AF594/AF594-AF488 from MD simulations based on ff14SB.	199
Figure 8-A2. Structural characterization of the ensemble of <i>apo</i> CaM-AF488-AF594/AF594-AF488 from MD simulations based on a99SB-disp.	200
Figure 9-A2. Probability distribution of orientation factors derived from MD simulations for CaM-AF488-TRC2/TRC2-AF488 systems	201
Figure 10-A2. Probability distribution of orientation factors derived from MD simulations for CaM-AF488-AF594/AF594-AF488 systems.	202

Figure 11-A2. Contribution of the AF488 and TRC2 dyes to the total solvent-accessible surface area (SASA) of the CaM-AF488-TRC2 and CaM-TRC2-AF488 systems	203
Figure 12-A2. Contribution of the AF488 and AF594 dyes to the total solvent-accessible surface area (SASA) of the CaM-AF488-AF594 and CaM-AF594-AF488 systems	204
Figure 13-A2. Distribution of fluorescence lifetimes computed for <i>holo</i> CaM-AF488-TRC2 and CaM-TRC2-AF488 systems using different electronic coupling models	205
Figure 14-A2. Distribution of fluorescence lifetimes computed for <i>apo</i> CaM-AF488-TRC2 and CaM-TRC2-AF488 systems using different electronic coupling models	206
Figure 15-A2. Distribution of fluorescence lifetimes computed for <i>holo</i> CaM-AF488-AF594 and CaM-AF594-AF488 systems	207
Figure 16-A2. Distribution of fluorescence lifetimes computed for <i>apo</i> CaM-AF488-AF594 and CaM-AF594-AF488 systems	208

ABBREVIATIONS

A	Acceptor
AMBER	Assisted Model Building with Energy Refinement
ATPase	Adenine Triphosphatase
AWSEM IDP	Coarse-Grained Force Field for Intrinsically Disordered Proteins.
BEMD	Bias-Exchange Metadynamics
Ca ²⁺	Calcium ion
CaM	Calmodulin
CAMB3LYP	Counterpoise Corrected for Adsorption Model Becke, 3-parameter, Lee–Yang–Parr
CASPT2	Complete Active Space-Perturbation Theory
CDZ	Calmidazolium
CFP	Cian Fluorescent Proteins
CHARMM	Chemistry at HARvard Molecular Mechanics
Cryo-EM	Cryo Electron Microscopy
D	Donor
DFT	Density functional theory
DNA	deoxyribonucleic acid
EET	Electron energy transfer
EF-hand	E- α -helix F α -helix hand
FCWD	Franck-Condon Weight Density of states
FLIM	Förster Lifetime Imaging Microscopy
FMO	Fenna-Matthews-Olson
FRET	Förster Resonance Energy Transfer
GaMD	Gaussian Accelerated Molecular Dynamics
GFP	Green Fluorescent Proteins
HF	Hartree-Fock
IDPs	Intrinsically Disordered Proteins
LH2	Light harvesting complex
LR	Linear response
MD	Molecular dynamics
MOPH-IDP	Maximum Entropy Optimized force field
NEVPT2	n-electron valence state perturbation theory
NMR	Nuclear magnetic resonance
OLEDs	Organic Light-Emitting Diodes
OPLS	Optimized Potentials for Liquid Simulations
PCM	polarizable continuum model
PDA	Point dipole approximation

PDPs	Partially Disordered Proteins
QM/MM	Quantum Mechanic Molecular Mechanics
QM/MMPol	Quantum Mechanic Molecular Mechanics Polarizable
REMD	Replica Exchange Molecular Dynamics
REST2	Replica Exchange with Solute Tempering
RNA	ribonucleic acid
SAXS	Small Angle X ray scattering
SIRAH	Structurally Unbiased Coarse-Grained Force Field
smFRET	Single Molecule FRET
TBP	TATA-binding protein
TDDFT	Time dependent density functional theory
TrESP-MMPol	Transition Electrostatic Potencial Molecular Mechanics Polarizable
WT-IMTD	Well-Tempered Metadynamics
XC	Exchange-correlation
YFP	Yellow Fluorescent Proteins
ZINDO	Zerner's Intermediate Neglect of Differential Overlap

CHAPTER 1: Introduction

The study of molecular systems has been a pillar of scientific research for decades, with experimental and computational methods evolving to provide deeper insights into their complex behavior. Among these systems, proteins are particularly significant due to their essential roles in all biological processes. Understanding proteins, including Intrinsically Disordered Proteins (IDPs) and Partially Disordered Proteins (PDPs), is fundamental to understanding how and why processes in our bodies work. Proteins function through their structure, and studying their shape and dynamics is key to unravel their mechanisms.¹

One protein of particular interest is calmodulin (CaM), a highly conserved calcium-binding protein that plays a critical role in regulating numerous physiological processes in the human body.² CaM is essential for calcium signaling and influences activities such as muscle contraction, cell cycle progression, and neurotransmitter release. Dysregulation of this protein activity has been linked to severe conditions, including cardiac arrhythmias and certain neurodegenerative disorders,³ underscoring the importance of understanding its structural and functional dynamics.

Computational techniques, especially molecular dynamics (MD) simulations, have gained significant importance in protein research. MD simulations generate detailed, time-evolving ensembles of protein conformations, revealing how proteins fluctuate structures over time.⁴ This ability to model three-dimensional structures and their dynamic changes has made MD simulations indispensable for studying proteins like calmodulin, as their function is intimately linked to their shape and conformational transitions.⁵

By integrating MD simulations with Förster resonance energy transfer (FRET) experiments, researchers can go beyond static structural information to study the dynamic conformational landscapes of proteins. MD simulations allow the calculation of FRET theoretical observables, such as distance distributions, energy transfer efficiencies, and fluorophore dynamics,

connecting structural models and experimental FRET data.⁶ This combined approach enables a detailed interpretation of how distance-dependent FRET efficiencies correlate with specific structural states or transitions.

FRET experiments also complement MD simulations by providing critical benchmarks to validate the accuracy of computational models, including the force fields and parameters used in the simulations. Moreover, FRET can serve as a powerful technique to refine force fields, ensuring that they better capture the real behavior of proteins in dynamic and heterogeneous environments.⁷ This feedback loop between FRET and MD simulations enhances both methods, allowing for more reliable and comprehensive insights into protein behavior.⁸

However, classical Förster theory for electronic energy transfer (EET) relies on approximations that can limit its accuracy. Förster's original formulation describes energy transfer as a dipole-dipole interaction, simplified by an isotropic averaging assumption for the orientation factor and a dielectric screening factor ($1/n^2$) to account for the environment.⁹ While effective in many cases, this framework cannot fully capture the complexities of real biological systems, where the environment plays a critical role in modulating couplings through polarization effects and atomistic interactions.

To address these limitations, advanced computational methods such as Quantum Mechanics/Molecular Polarizable Mechanics (QM/MMPol) and TrESP-MMPol have been developed. These approaches incorporate an atomistic description of the environment and explicitly account for electronic polarization, providing a more accurate representation of environmental effects on FRET couplings.¹⁰ The QM/MMPol method rigorously calculates electronic transition densities and their interactions with a polarizable environment, enabling detailed insight into screening effects and their impact on energy transfer rates. The TrESP-MMPol method, while computationally more efficient, approximates the transition densities using

transition electrostatic charges derived from quantum mechanical calculations.¹¹ This approach retains much of the accuracy of QM/MMPol while being better suited for large-scale systems or long MD trajectories.

The ability to move beyond Förster approximations is critical for ensuring that computationally generated FRET results accurately match experimental data.¹² Including the environmental atomistic description and polarizable effects not only improves the reliability of coupling calculations but also provides deeper insights into the interplay between donor-acceptor interactions and their surroundings. By combining these advanced methods with MD simulations and FRET experiments, researchers can achieve a more comprehensive and precise understanding of protein dynamics and the mechanisms driving energy transfer.

1.1 Förster theory of electronic resonance energy transfer

EET is a key process in materials science and biology, involving the nonradiative transfer of electronic excitation energy from a donor molecule to a nearby acceptor molecule over distances from 10 nm to 100 nm approximately. This concept was significantly advanced about 70 years ago by Theodor Förster,^{9,13} whose pioneering work established the foundation for understanding energy transfer based on spectroscopic principles.

Förster's work has had a profound impact across several disciplines. In the fields of chemistry and biology, the term FRET is commonly used to describe both the underlying process and the fluorescence-based technique that employs this energy transfer to measure distances between molecules, due to the sensitivity of EET efficiency to the distance and orientation between donor and acceptor molecules. This relationship allows FRET to serve as a molecular ruler,^{14,15} enabling precise measurements at the nanoscale and providing valuable insights about molecular interactions and structural dynamics.

FRET's sensitivity to molecular interactions and its broad adaptability have made it an essential tool in a variety of scientific and technological fields.¹⁶ The following section will address several applications on advanced materials science that enable breakthroughs in biological research.

1.1.1 Applications of Förster Resonance Energy Transfer (FRET)

The wide applicability of FRET has enabled advancements in various scientific and technological fields. Its ability to measure intermolecular distances (able to reveal interactions) with high sensitivity and specificity has been instrumental in both fundamental research and innovative applications

Organic Light-Emitting Diodes (OLEDs)

FRET has contributed significantly to the development of OLED technology, which is widely used in display and lighting applications.

- **Color Tuning:** By facilitating energy transfer between organic molecules with distinct emission wavelengths, precise control over the color output is achieved.¹⁷ This improves the vibrancy, purity, and consistency of emitted light, which is critical for high-quality displays.
- **Enhanced Efficiency:** Energy management within OLED layers is optimized, minimizing energy loss and maximizing brightness.^{18,19} This not only extends the lifespan of OLED devices but also reduces energy consumption, making them more sustainable.

Sensors

FRET has become indispensable in sensing technologies, particularly in biosensors and chemical sensors, due to its ability to detect molecular interactions with high sensitivity and specificity.

- **Biosensors:** Donor and acceptor fluorophores are used in FRET-based biosensors to monitor molecular interactions, such as protein-protein binding or enzymatic activity. When a target molecule binds, the distance or orientation between the donor and acceptor changes, leading to detectable shifts in fluorescence. These biosensors are extensively used in medical diagnostics, drug discovery, and environmental monitoring.²⁰
- **Chemical Sensors:** Enhancing chemical sensing, energy transfer detects environmental changes such as pH, ion concentrations, or solvent polarity.

These sensors are valuable in real-time monitoring of chemical reactions, detecting toxic substances, and studying cellular microenvironments.²¹

Photovoltaics

In photovoltaic technology, FRET offers innovative strategies to enhance solar energy harvesting.

- **Energy Conversion:** Solar cells benefit from facilitated energy transfer between different materials or layers within the cell.²² This ensures that absorbed light energy is efficiently channeled to the active regions where it is converted into electrical power.
- **Light Management:** By optimizing the distribution and absorption of light within solar cells, reflection losses are reduced, and the overall efficiency of energy capture is enhanced.¹⁸

Single-molecule and cellular imaging

In biology, FRET is a critical technique for studying molecular interactions and dynamics at the single-molecule level.

- **Protein Interactions:** Molecular folding, conformational changes, and complex formation are monitored by tagging specific sites on a protein with donor and acceptor fluorophores, allowing researchers to observe how a protein changes shape or interacts with other molecules.²³
- **Live-Cell Imaging:** FRET-based imaging enables real-time visualization of molecular interactions within living cells.²⁴ For example, changes in distances between signaling proteins during cellular responses can be tracked, offering

insights into processes like gene expression, signal transduction, and membrane dynamics.

Drug discovery and molecular diagnostics

Techniques based on energy transfer have significantly advanced drug discovery and diagnostics by providing precise tools to study molecular mechanisms.

- **High-Throughput Screening:** In drug discovery, large libraries of compounds are screened using FRET assays, which detect molecular interactions or changes in enzymatic activity.²⁵
- **Disease Diagnostics:** In diagnostic tests, energy transfer is employed to detect disease biomarkers.²⁶ For instance, DNA or RNA sequences associated with specific genetic conditions or infections are identified, offering sensitive and accurate detection methods.

Nanotechnology and advanced materials

Applications in nanotechnology and advanced materials have been significantly enriched by FRET.

- **Nanoparticle-Based Sensors:** Nanosensors designed using energy transfer principles can detect specific chemical or biological targets with high sensitivity. Often, functionalized nanoparticles act as donors or acceptors, enabling precise detection at the nanoscale.²³
- **Energy Transfer in Nanomaterials:** Processes like energy transfer are employed to study the optical and electronic properties of nanostructured

materials, such as quantum dots or carbon nanotubes.²⁷ This provides valuable insights for the development of next-generation materials.

In summary, Förster Resonance Energy Transfer is not only a fundamental concept in physics and chemistry but also a powerful tool with a wide range of applications. Its ability to provide precise nanoscale measurements and its adaptability across diverse disciplines highlight its importance in advancing both scientific understanding and technological innovation.

1.2 Golden rule formulation for EET rate expression

Electronic energy transfer describes the non-radiative transfer of excitation energy from a Donor (D) on its excited state $|D^*A\rangle$ coupled to the excitation of an Acceptor (A) molecule promoting it to an excited state $|DA^*\rangle$ ²⁸

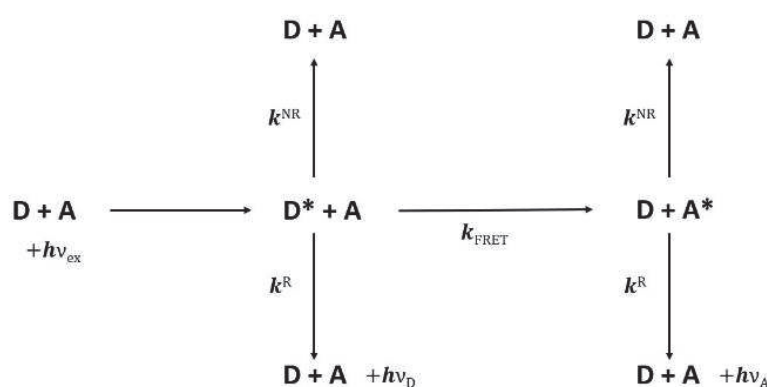


Figure 1.1: Different energy pathways after donor: excitation ($h\nu_{ex}$) possibly leading to a non-radiative decay, luminescence emission of D ($h\nu_D$) or A ($h\nu_A$) for FRET phenomena.

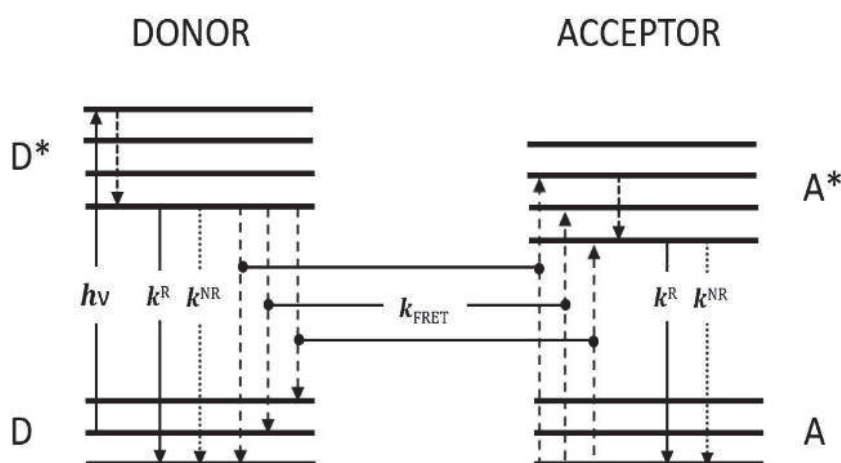


Figure 1.2: Basic FRET principle: Simplified energy level scheme representing the excitation of the donor from an electronic ground state (D) to an excited state (D^*), followed by inner relaxation to the ground vibrational state of the excited state, followed by radiative decay (k^R), nonradiative decay (k^{NR}), or FRET (k_{FRET}) process from D^* to A^* . The difference between

the respective energy levels need to be equal (resonance condition). After FRET, the acceptor is in an excited state (A^*), followed by radiative or nonradiative decay to its ground state (A).

Theodor Förster theory provides a framework for understanding the rate of energy transfer between a pair of weakly-coupled donor and acceptor molecules, grounded in the principles of the Fermi Golden Rule.²⁹ The rate of Förster Resonance Energy Transfer (FRET) is given by:

$$k_{FRET} = \frac{2\pi}{\hbar} V^2 FCWD \quad (1.2)$$

Where V represents the electronic coupling between the donor and acceptor, and FCWD stands for the Franck-Condon weighted density of states.³⁰

To derive the Fermi Golden Rule in the context of a molecular system, we consider two sets of vibronic states, v and v' , corresponding to the initial and final states (denoted as i and f):

$$\begin{aligned} k_{FRET} &= \frac{2\pi}{\hbar} |V_{if}|^2 \sum_{v,v'} p(E_v) |\langle v|v' \rangle|^2 \delta(E_{iv} - E_{fv'}) \\ &= \frac{2\pi}{\hbar} |V_{if}|^2 FCWD \end{aligned} \quad (1.3)$$

Here, the summation encompasses the vibrational states of i and f , incorporating the Boltzmann population $p(E_v)$ of the initial state v and the Franck-Condon factors $|\langle v|v' \rangle|^2$. The Condon approximation is employed, allowing us take out of the summation the electronic coupling V_{if} .

In the context of electronic energy transfer, if the initial and final states of the donor and acceptor are localized with independent vibrations, the states

v can be represented as direct products of the vibrational states of the excited donor v'_D and the ground state acceptor v_A , and vice versa. Consequently, the Franck-Condon weighted density of states can be simplified using the donor emission and acceptor absorption functions:²⁸

$$f_D(E) = \sum_{v_D, v'_D} p(E_{v_D}) |\langle v_D | v'_D \rangle|^2 \delta(E - \Delta E_{0v_D, 1v'_D}) \quad (1.4)$$

$$f_A(E) = \sum_{v_A, v'_A} p(E_{v_A}) |\langle v_A | v'_A \rangle|^2 \delta(E - \Delta E_{0v_A, 1v'_A}) \quad (1.5)$$

where $\Delta E_{0v_D, 1v'_D}$ represents the excitation energy of D from its ground state in the v_D vibrational state to its excited state in the v'_D vibrational state and similarly for the acceptor.

Finally, the FCWD can be expressed as the spectral overlap of these donor and acceptor functions:

$$FCWD = \int_{-\infty}^{\infty} dE f_D(E) f_A(E) E q(X) \quad (1.6)$$

This integral quantifies how the vibrational states of the donor and acceptor molecules overlap, contributing to the overall rate of energy transfer in the system.

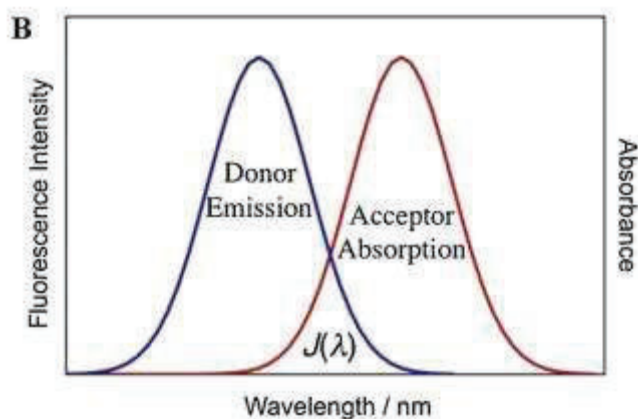


Figure 1.3: Spectral overlap between donor emission and acceptor absorption. Adapted figure from Förster resonance energy transfer – A spectroscopic nanoruler: Principle and applications.³¹

The functions $f_D(E)$ and $f_A(E)$ are related to the Franck-Condon factors for the acceptor's absorption spectrum $A_A(\omega) \propto |\mu_A^{tr}|^2 \omega f_A(\hbar\omega)$ and the donor's emission spectrum $F_D(\omega) \propto |\mu_D^{tr}|^2 \omega^3 f_D(\hbar\omega)$ through specific prefactors. Where μ_A^{tr} and μ_D^{tr} denotes the transition dipole moment for the acceptor and donor molecules and ω is the angular frequency related to the transition energy. Spectral overlap can be computed via Eq 1.6³² from experimental spectra and dividing each spectrum by its frequency prefactor and using the $f_D(E)$ and $f_A(E)$ normalized functions.

After expressing the FCWD as the spectral overlap, we can focus on the Förster theory formulation in the weak coupling limit between donor and acceptor molecules:²⁹

$$k_{FRET} = \frac{2\pi}{\hbar} V^2 J \quad (1.7)$$

Where V stands as the electronic coupling between D/A molecules and J the spectral overlap factor obtained from the area-normalized donor emission and acceptor absorption spectra.

In the Förster theory coupling is approximated assuming a point dipole approximation (PDA)²⁸ and a simple screening factor $s: \frac{1}{n^2}$

$$V \approx sV_{PDA} = \frac{1}{n^2}V_{PDA} = \frac{1}{n^2} \frac{\kappa \mu_D \mu_A}{R^3} \quad (1.8)$$

$$\kappa = \hat{\mu}_D \cdot \hat{\mu}_A - 3(\hat{\mu}_D \cdot \hat{R})(\hat{\mu}_A \cdot \hat{R}) \quad (1.9)$$

Where μ_D and μ_A are the D/A transition dipole moments, κ the dipole orientation factor,³³ R the D/A center-to-center distance, and n the refractive index of the medium. Central to the formulation is the assumption that the transition densities of the donor and the acceptor interact as point transition dipoles, where the screening of this Coulomb interaction by the surrounding medium is described with a factor $\frac{1}{n^2}$, with n^2 being the square of the refractive index, which represents the optical component of the dielectric constant of the medium (ϵ_{opt}).³⁴ In addition, the chromophores are assumed to dispose of free rotational motion allowing to assume a value 2/3 for the orientational factor.

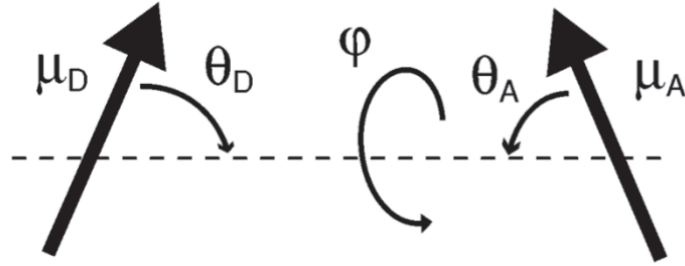


Figure 1.4: Orientational factor illustration.

Alternatively, Förster rate can be formulated from purely spectroscopic data³⁵

$$k_{FRET} = k_D \left(\frac{R_o}{R} \right)^6 = \frac{1}{\tau_D} \left(\frac{R_o}{R} \right)^6 \quad (1.10)$$

Where $k_D = \frac{1}{\tau_D}$ describes the decay rate of the excited donor molecule in the absence of the acceptor molecule, R is the center-to-center distance between chromophores and R_o the Förster radius. The Förster radius denotes the distance at which $E = 0.5$, that is, 50% of the donor excitation is transferred to the acceptor and 50% decays to the ground state of the donor.

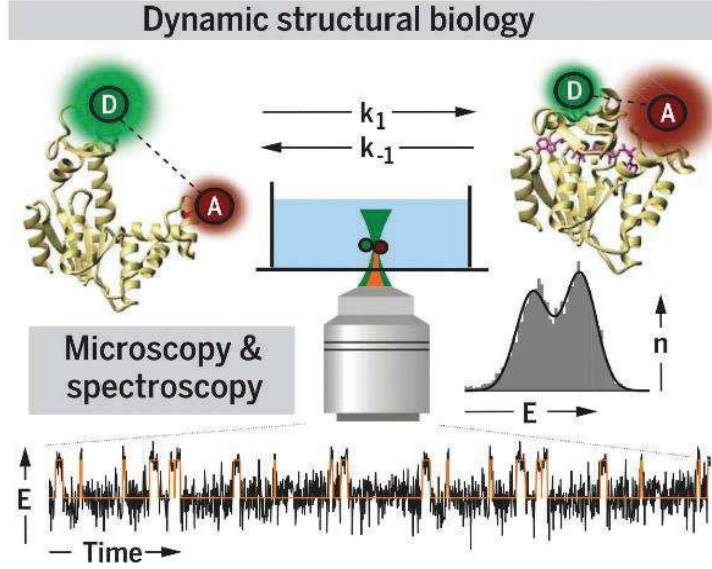


Figure 1.5: Overview on FRET rates mesure and posterior efficiency density distribution analysis. Figure adapted from: Toward dynamic structural biology: Two decades of single-molecule förster resonance energy transfer.¹⁶

Förster radius can also be expressed with the following expression^{28,36}

$$R_o^6 = \frac{9(\ln 10)(\kappa^2 \varphi_D)}{128\pi^5 N_A n^4} I \quad (1.11)$$

Where φ_D is the fluorecence quantum yield, I the spectral overlap and N_A the Avogadro number. From these derivations and definitions, transfer efficiency can be expressed as³⁷

$$E_{FRET} = \frac{k_{FRET}}{k_D + k_{FRET}} = \frac{1}{1 + (R/R_0)^6} \quad (1.12)$$

Where E_{FRET} is the efficiency of the energy transfer phenomena and it is calculated by using the decay rate of the donor molecule. Below is shown the derivation by adding the transfer energy rate to the donor molecule decay rate in absence of acceptor molecules.

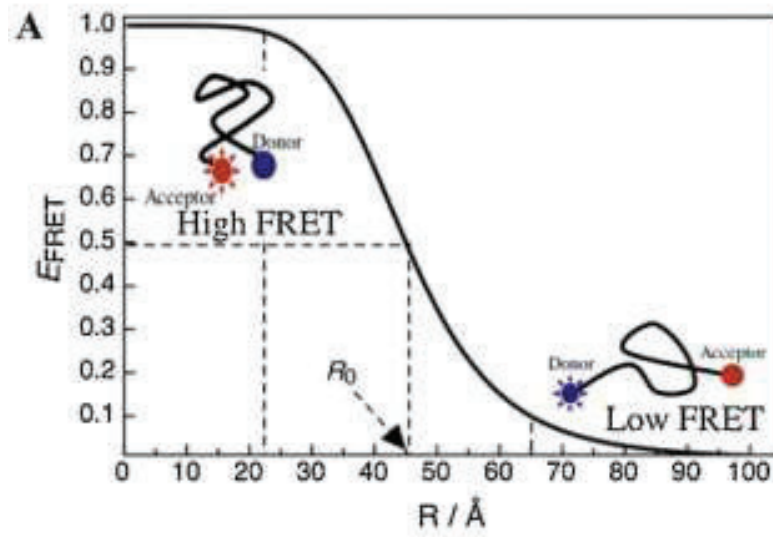


Figure 1.6: Plot of distance vs efficiency dependence, represented by Eq 1.12. Adapted figure from: Förster resonance energy transfer – A spectroscopic nanoruler: Principle and applications.³¹

Without taking in to account electron energy transfer, the rate constant can be expressed following the summation of the donor rates

$$k_D = k_{Radiative} + k_{Non-Radiative} \quad (1.13)$$

Where the radiative and non-radiative decay from the excited state are included. If we incorporate the FRET process, then we must add the energy transfer constant rate. Knowing as well that donor lifetime is the inverse of the donor rate constant we can derive

$$k_{Total} = k_{Radiative} + k_{Non-Radiative} + k_{FRET} \quad (1.14)$$

$$\tau_{R+n-R} = \frac{1}{k_{Radiative} + k_{Non-Radiative}} \quad (1.15)$$

Alternatively, the Efficiency can be expressed as

$$E_{FRET} = \frac{k_{FRET}}{k_{Radiative} + k_{Non-Radiative} + k_{FRET}} \quad (1.16)$$

And simplify the expressions as

$$E_{EET} = \frac{k_{EET}}{\frac{1}{\tau_D} + k_{EET}} \quad ; \quad E_{EET} = \frac{1}{1 + \frac{1}{\tau_D k_{EET}}} \quad (1.17)$$

1.3 The FRET tool in dynamic structural biology

FRET has become an essential technique for investigating molecular interactions and conformations at nanometer scales. Its conceptual origins date back to the 1920s when Jean Perrin first described excitation energy transfer between molecules in proximity, a phenomenon he termed "transfert d'activation".³⁸ Perrin's work established the role of oscillating dipole moments in this process but lacked a detailed theoretical framework. This gap was addressed in 1948 by Theodor Förster,^{13,39} who formalized the mechanism by assuming identical dipole moments and small interaction energies compared to the spectral transitions involved. Förster's theory provided a mathematical foundation for understanding the distance dependence of energy transfer, which later proved critical for biological applications.

The experimental validation of FRET principles began in earnest during the 1970s and 1980s. Stryer and Haugland's seminal studies demonstrated that FRET could act as a molecular ruler,³⁶ precisely measuring distances between donor and acceptor fluorophores with nanometer accuracy. Their work confirmed the inverse sixth-power dependence of FRET efficiency on distance, as predicted by Förster, and highlighted its potential for studying biomolecular conformations and interactions.

Further advancements during this period refined FRET theory and its experimental implementation. Dale, Eisinger, and Blumberg in 1979 introduced the concept of the orientation factor (κ^2)⁴⁰ clarifying the role of dipole orientation in FRET efficiency. Their work underscored how the relative orientation of donor and acceptor dipoles could significantly influence energy transfer, providing greater accuracy in distance measurements and reinforcing the theoretical foundations of FRET.

The integration of time-resolved fluorescence techniques in the 1980s, pioneered by researchers like Selvin, was another milestone.⁴¹ By measuring fluorescence lifetimes rather than intensities, these methods reduced artifacts and enhanced precision in quantifying FRET efficiency. This innovation opened the door to studying dynamic processes, such as protein folding and signaling protein conformational changes, in real-time. For example, FRET studies on calmodulin revealed its calcium-induced structural transitions, revealing its regulatory mechanisms.⁴²

The 1990s witnessed the introduction of single-molecule FRET (smFRET),⁴³ a transformative advancement pioneered by researchers like Ha, Weiss, and Deniz.⁴⁴ SmFRET enabled the observation of individual molecular events rather than ensemble averages, providing insights into molecular heterogeneity and dynamic conformational changes. Early applications, such as Deniz et al.'s work on protein folding, revealed intermediate states in folding pathways that were previously inaccessible, marking a significant advance in our understanding of molecular dynamics.

In parallel, the discovery and engineering of fluorescent proteins, such as Green Fluorescent Proteins (GFP) and its derivatives by Chalfie, Shimomura, and Tsien, revolutionized the field.⁴⁵ These genetically encoded fluorophores facilitated live-cell FRET studies, enabling the visualization of protein-protein interactions and intracellular signaling pathways.⁴⁶ Cyan Fluorescent Proteins and Yellow Fluorescent Proteins (CFP/YFP) pairs became widely used for creating biosensors capable of monitoring real-time molecular events with spatial resolution.⁴⁷

FRET's adaptability extended to nucleic acid research, where it was employed to study Ribonucleic Acid (RNA) folding and Deoxyribonucleic Acid (DNA) dynamics, offering unprecedented insights into these essential biomolecular processes. Notably, FRET was used to monitor the folding of ribozymes, as demonstrated by Zhuang et al. in 2000,⁴⁸ who revealed key intermediate

states and folding pathways for these RNA enzymes at the single-molecule level. This work underscored the technique's ability to dissect complex conformational transitions in RNA that are crucial for its biological function.

In the study of DNA, FRET has been instrumental in understanding bending and looping dynamics, which are essential for processes like transcription regulation and DNA repair. For example, Ha et al. in 1999 used single-molecule FRET to investigate DNA looping mediated by the lac repressor,⁴⁹ directly observing loop formation and dissociation in real time. These studies provided a deeper understanding of how regulatory proteins interact with DNA to modulate gene expression. Similarly, Wang et al. in 2002 applied FRET to study the conformational flexibility of nucleosomes,⁵⁰ revealing how histone-DNA interactions influence chromatin structure and accessibility.

FRET has also been used to probe the interactions between DNA and transcription factors, such as TATA-Binding Protein (TBP),⁵¹ elucidating how these factors recognize and bind to specific sequences to initiate transcription. In addition, the technique has been applied to study the dynamics of DNA unwinding and strand separation during replication and repair processes. For example, Myong et al. in 2007⁵² used FRET to investigate the helicase-mediated unwinding of DNA, providing insights into the stepwise nature of helicase activity.

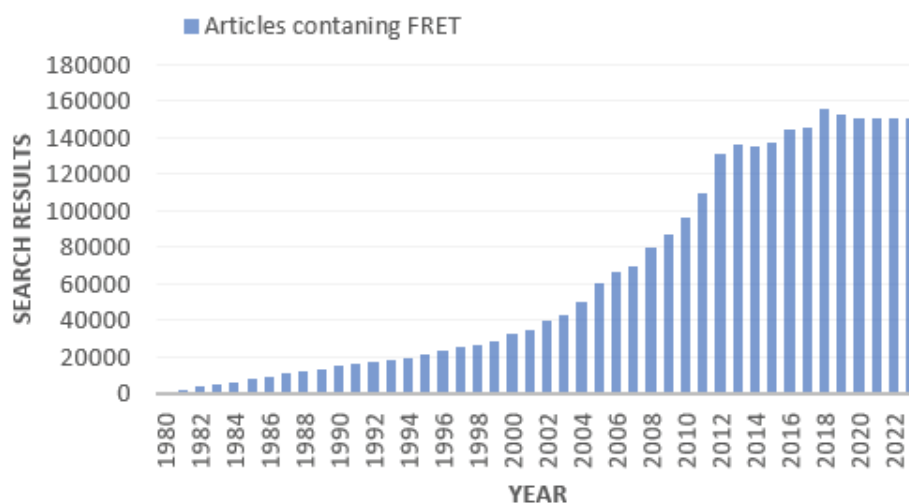


Figure 1.7: Number of publications per year containing the keywords “FRET” or “single molecule FRET” in the title or abstract (Source: Google Scholar).

These applications highlight FRET’s unique ability to capture dynamic transitions and conformational changes that are inaccessible to static or ensemble averaged structural methods like X-ray crystallography or nuclear magnetic resonance.⁵³ By offering real-time insights into nucleic acid behavior and interactions, FRET has significantly advanced our understanding of RNA and DNA dynamics, reinforcing its role as a powerful tool in structural and molecular biology.

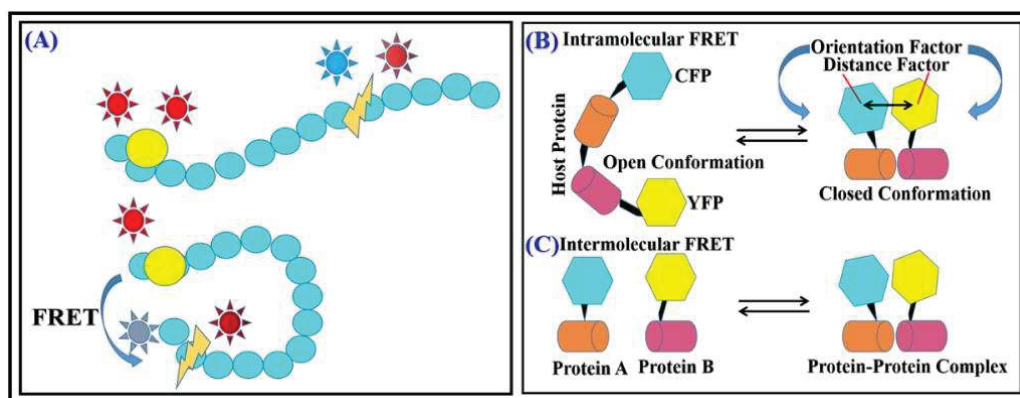


Figure 1.8: Schematic representation of FRET applications in biomolecular studies. (A) FRET in a protein folding process, where donor and acceptor fluorophores report conformational changes. (B) Intramolecular FRET used to study open and closed conformations of a host protein. (C) Intermolecular FRET applied to monitor protein-protein complex formation. Figure adapted from: The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo.⁵⁴

The integration of FRET with advanced imaging techniques, such as confocal microscopy and Förster Lifetime Imaging Microscopy (FLIM), further expanded its applications.⁵⁵ These combinations enabled researchers to study protein-protein interactions and complex molecular assemblies in living cells with high spatial and temporal resolution. For example, early imaging studies revealed receptor clustering and molecular organization in cell membranes.

These foundational contributions, from Perrin's early observations to the advancements by Stryer, Dale, Selvin, Ha, and others, have cemented FRET as a standard technique of structural biology. By connecting theoretical principles with advanced applicability, FRET has become a versatile tool for studying the dynamic behaviors of proteins, nucleic acids, and molecular complexes, complementing traditional structural methods by providing insights into the time-evolving nature of biomolecular interactions.

1.4 Techniques to study the conformational preferences of disordered proteins

Proteins are typically associated with well-defined three-dimensional structures that determine their functions. However, intrinsically disordered proteins and intrinsically disordered protein regions challenge this conventional paradigm. Unlike ordered proteins, IDPs lack a stable tertiary structure under physiological conditions. Instead, they exist as dynamic ensembles of conformations, allowing them to adapt and interact with multiple partners. This structural flexibility distinguishes IDPs from well-ordered proteins, which rely on rigid structures for their biological roles.

The functional repertoire of IDPs includes roles in signaling, regulation, and molecular recognition, often involving transient interactions. Their lack of stable structure enables rapid response to environmental changes, making them crucial in processes like cell cycle regulation and stress response. However, this disorder can also predispose IDPs to aggregation, linking them to diseases such as Alzheimer's and Parkinson's. Understanding the unique properties of IDPs and how they contrast with ordered proteins is essential for unraveling their complex roles in cellular biology and disease.

Determining protein structures is fundamental to understanding their functions and interactions within biological systems. Over the years, a variety of experimental techniques have been developed to investigate proteins at various levels of detail. These approaches provide valuable insights into protein folding, flexibility, and functional domains, enhancing our understanding of their behavior and interactions.⁵⁶ This section explores the most widely used experimental techniques for studying protein structure such as: Nuclear Magnetic Resonance (NMR) Spectroscopy, Small-Angle X-ray Scattering (SAXS), Förster Resonance Energy Transfer (FRET), X-ray crystallography, and Cryo-Electron Microscopy (Cryo-EM). Each of these methods offers unique strengths and limitations, making them suited to specific applications in protein research.

1.4.1 Experimental techniques for protein structure characterization

Nuclear Magnetic Resonance (NMR) spectroscopy

NMR spectroscopy uses the magnetic properties of atomic nuclei to provide detailed structural and dynamic information about proteins in solution.⁵⁷ Unlike techniques that require protein crystallization, NMR allows the study of proteins in their natural, liquid environment, preserving their native dynamics. This is particularly advantageous for proteins that undergo conformational changes as part of their function.⁵³ NMR experiments involve placing a protein sample in a strong magnetic field and measuring how specific nuclei, such as hydrogen, carbon, or nitrogen, respond to radiofrequency pulses. These responses reveal interatomic distances and chemical environments.

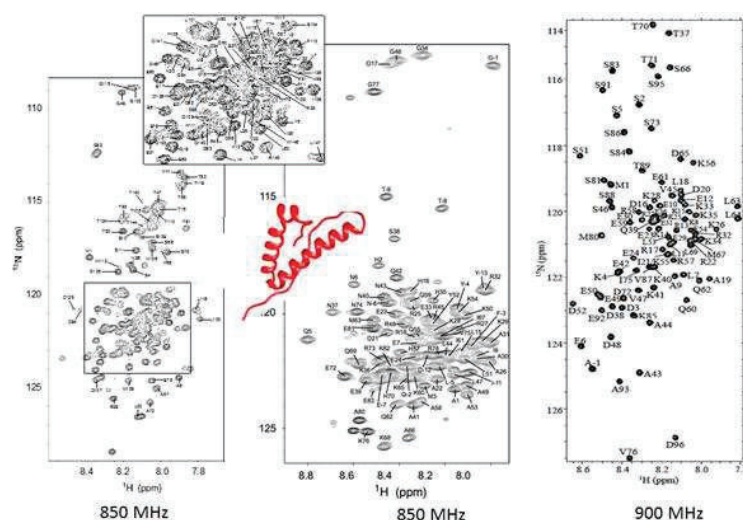


Figure 1.9: Results from NMR corresponding to the secondary structure of a protein, highlighting the limitations of the technique of the complex interpretation due to the larger size of biological systems.

NMR's ability to simultaneously provide structural and dynamic information makes it invaluable for studying how proteins move and interact. However, its application is typically limited to smaller proteins (up to ~30 kDa) due to

signal complexity in larger systems. For IDPs, NMR can reveal important details about their flexibility and transient structures.⁵⁸ However, interpreting results for disordered systems is particularly challenging because the spectra represent an ensemble of rapidly interconverting conformations, making it difficult to distinguish specific structural states. Additionally, for highly dynamic proteins, the averaging of NMR signals may obscure subtle conformational differences.

To address these limitations, methods such as isotope labeling and multidimensional NMR are used to enhance resolution and simplify data analysis. Despite these challenges, this technique remains a critical tool for characterizing proteins in solution and understanding their functional dynamics.

Small-Angle X-ray Scattering (SAXS)

SAXS is a technique that provides information about the overall shape and size of proteins or protein complexes in solution, offering low-resolution structural data.⁵⁹ Unlike X-ray crystallography, which requires proteins to be crystallized, SAXS allows researchers to study proteins in their native, often more flexible, state. This is particularly useful for proteins that are difficult or impossible to crystallize, such as those with flexible or disordered regions. In a SAXS experiment, an X-ray beam is directed at a solution of proteins, and the scattered X-rays are collected at small angles. The resulting scattering pattern provides information about the general shape and size of the protein.⁶⁰

While SAXS does not provide the detailed atomic resolution of techniques like X-ray crystallography or NMR, it can reveal important structural features, such as the protein's overall dimensions, the presence of flexible or disordered regions, and the relative arrangement of domains in multi-domain proteins.^{61,62} SAXS is also a valuable tool for studying large protein complexes or proteins in varying conformations, as it can capture structural

changes in real time. By combining SAXS data with high-resolution models or other techniques, researchers can obtain a more complete understanding of protein structure and dynamics, especially for complex or flexible systems.

Förster Resonance Energy Transfer (FRET)

FRET is a spectroscopic technique used to study protein interactions and conformational changes by measuring energy transfer between two fluorophores.⁶³ These fluorophores are strategically attached to specific protein sites, and energy transfer occurs when the donor fluorophore, upon excitation, transfers energy to a nearby acceptor fluorophore. The efficiency of this transfer is highly distance-dependent, making FRET a molecular ruler capable of measuring nanoscale distances (1–10 nm) within or between proteins.

FRET is particularly valuable for investigating the dynamical conformational properties of proteins, providing insights into their ensemble behaviors and capturing transitions between extended and compact states such as protein folding, conformational changes, or the binding interactions between proteins.⁶⁴ This makes it uniquely suited to studying proteins that adopt multiple conformations, including intrinsically disordered proteins (IDPs) or multi-domain proteins with flexible linkers. By analyzing FRET efficiencies across a range of conditions or time points, researchers can probe how proteins move and interact in real time.

A consideration when using FRET is that attaching fluorophores to the protein can potentially alter its natural behavior.⁶⁵ Modifications introduced during dye labeling, such as changes in charge or steric effects, might slightly perturb the protein's conformational ensemble. However, numerous studies have shown that with careful design, these effects are minimal, and the resulting data remain highly representative of the protein's native behavior. This balance between perturbation and utility has solidified FRET's role as a

versatile tool for studying the structural dynamics of proteins in both in vitro and in vivo contexts.⁶⁶

X-ray crystallography and Cryo-Electron Microscopy (Cryo-EM)

X-ray crystallography is a gold-standard technique for determining protein structures at atomic resolution. It involves crystallizing the protein and exposing it to an X-ray beam.⁵⁹ The diffraction patterns generated by the crystalline structure are analyzed to reconstruct a detailed 3D model of the protein.

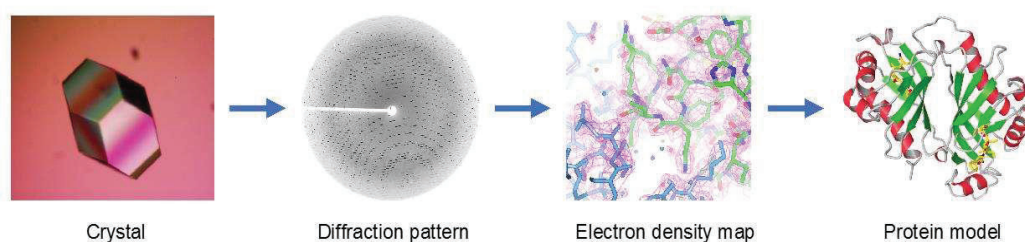


Figure 1.10: Conventional followed procedure in X-ray crystallography to obtain a protein model.

The primary strength of X-ray crystallography is its ability to achieve extremely high-resolution structural information. However, the technique requires high-quality crystals, which can be challenging to obtain for proteins that are large, flexible, or membrane bound.⁶⁷ Crystallization may restrict natural protein conformations, potentially hiding their dynamic behavior or ensemble of states in solution. Despite this limitation, X-ray crystallography remains a powerful technique in structural biology, particularly for well-ordered and stable proteins.⁶⁸

Cryo-EM has revolutionized the field of structural biology, particularly in the study of large macromolecular complexes and membrane proteins.⁶⁹ Unlike X-ray crystallography, which requires crystallized proteins, Cryo-EM allows researchers to study proteins in their native, non-crystallized state. Proteins

are rapidly frozen in a thin layer of ice, preserving them in a near-native state without the need for staining or other chemical treatments.⁷⁰ This makes it possible to visualize proteins and complexes that are difficult to study with other methods, such as those that are large, flexible, or heterogeneous. While Cryo-EM is well-suited for capturing multiple conformational states, sample heterogeneity can complicate image reconstruction and requires advanced computational methods to resolve.⁷¹

Finally, because Cryo-EM studies proteins in a vitrified, frozen state, it does not capture their real-time dynamics in solution. Despite these challenges, the technique has become an indispensable technique in modern structural biology, offering unique possibilities for studying complex biological systems in their native environments.

Although the previous mentioned techniques are gold-standard for determining protein structure, they are not usually applied to the study of disordered proteins, due to the applied conditions on the samples. These restrictions make them not suitable when the purpose is going into conformational ensembles insights.

1.4.2 FRET and computational simulations: Exploring protein dynamics

MD simulations have emerged as a core computational technique for studying the behavior of proteins at atomic resolution over time. By simulating the motion of atoms and molecules using physics-based force fields, MD provides detailed insights into protein dynamics, folding pathways, and conformational changes. Unlike experimental techniques that often capture static snapshots, MD reveals the continuous, time-evolving nature of protein structures, making it invaluable for understanding how proteins behave in different environments and conditions.

The integration of MD simulations with FRET experiments has proven particularly powerful for studying protein dynamics.⁷² FRET experiments provide experimental measurements of nanoscale distances between fluorophore pairs, offering valuable data on protein conformations and interactions. However, interpreting FRET data can be challenging, especially for proteins with flexible or disordered regions. Here, MD simulations play a critical role by generating ensembles of protein conformations that can be used to calculate FRET observables, such as distance distributions, energy transfer efficiencies, and fluorophore dynamics.

Combining MD and FRET integrates experimental and computational insights between experimental observations and computational models. MD-generated conformational ensembles allow for a more detailed interpretation of FRET data, enabling researchers to connect distance-dependent FRET efficiencies with specific structural states or transitions.⁷³ The computational models allow us to see how these distances change in the context of the protein's 3D structure, a powerful advantage for researchers.

At the same time, FRET experiments provide experimental benchmarks that can validate the accuracy of MD simulations, ensuring that the chosen force fields and simulation parameters capture the protein's real behavior. This

synergy is particularly useful for studying dynamic processes, such as protein folding, domain movements, or interactions within complexes.⁷⁴

By integrating MD and FRET, researchers can go beyond static structural information to study the dynamic conformational landscapes of proteins. This approach not only enhances our understanding of protein dynamics but also provides a framework for validating computational models against experimental data, ensuring a more comprehensive and accurate understanding of protein behavior.

Limitations of FRET simulations and advances in Förster theory

While the integration of MD simulations with FRET provides a powerful framework for studying protein dynamics, the accuracy of FRET simulations is dependent on the approximations made within Förster theory.¹² Förster theory assumptions on dipole-dipole, isotropic orientation and average continuum dielectric may fail under certain conditions, such as very short donor-acceptor distances, complex environmental effects, or deviations from the ideal dipole approximation.

A critical factor that significantly influences FRET simulations is the accurate modeling of the environment surrounding the fluorophores. The description of the environment using a polarizable force field, such as MM polarizable models, plays an essential role in determining FRET observables.⁷⁵ Environmental effects, including variations in the local refractive index and the dynamic interaction between fluorophores and their surroundings, can have a greater impact on FRET results than deviations in dipole alignment or inaccuracies in Coulomb coupling. Incorporating MM polarizable force fields ensures that these environmental influences are properly accounted for, allowing for more realistic and reliable simulations.⁷⁶

To further address the limitations of Förster theory, advanced computational methods such as QM/MM polarizable models and transition charge schemes with MM polarizable models are employed in this work.⁷⁷ These methods

provide detailed representations of the transition electronic density and allow the calculation of EET coupling and energy transfer rates with higher accuracy throughout MD trajectories. By integrating precise environmental modeling with refined electronic property calculations, this thesis aims to overcome limitations of Förster theory, offering a robust framework for interpreting FRET data and studying the dynamic conformational behavior of proteins.

1.5 Calmodulin (CaM)

Calmodulin (CaM) is a highly conserved calcium-binding protein that plays a central role in regulating numerous cellular processes in eukaryotic organisms.⁷⁸ It is a partially disordered protein composed of approximately 148 amino acids, CaM exists in two distinct forms: the *apo* form, which lacks bound calcium ions, and the *holo* form, in which can bind up to 4 calcium ions.⁷⁹ Calcium binding induces significant conformational changes in CaM, enabling it to interact with a diverse range of target proteins. Acting as a key intracellular messenger, CaM translates changes in intracellular calcium concentrations into vital cellular responses, including muscle contraction, cell division, and signal transduction.^{3,80}

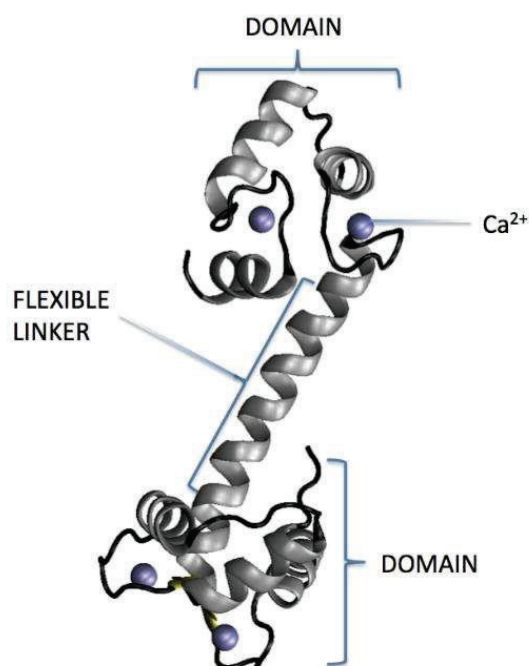


Figure 1.11: CaM structure, highlighted flexible linker (most disordered part), 2 domains and Ca^{2+} binding sites.

Structurally, CaM is characterized by two globular lobes (N-terminal and C-terminal), each containing two EF-hand motifs—helix-loop-helix domains specialized for calcium coordination—connected by a highly flexible central

linker.⁷⁸ The linker is the most disordered region of the protein, imparting significant flexibility to the overall structure. This flexibility allows CaM to adopt a wide range of conformations, enabling it to bind targets of various sizes and orientations.

Upon calcium binding, the E- α -helix F- α -helix hand (EF-hand) motifs undergo a conformational shift, exposing hydrophobic surfaces that facilitate interactions with target proteins in a calcium-dependent manner. Notably, the C-lobe is more flexible than the N-lobe, a property that provides additional selectivity in target recognition.^{81,82}

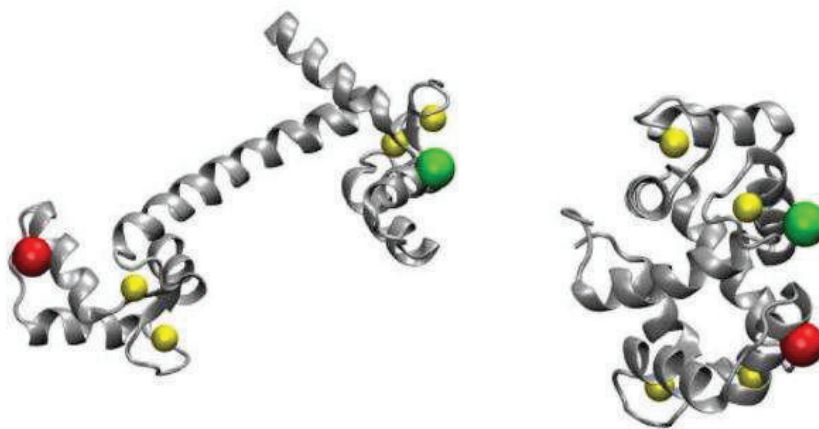


Figure 1.12: The figure shows CaM in 2 different conformational states, in the left the protein is adopting an extended conformational state while in the right a compact one. This picture evidences the different shape between the 2 different states that would lead to 2 different functions in any biological process. In red and green it is highlighted the position of the 34 and 110 methionine, where normally dyes for FRET experiments are attached.

The transition between the *apo* and *holo* forms is essential for CaM's ability to act as a molecular switch. In the *apo* form, the protein is less compact, and its hydrophobic surfaces are largely buried, while in the *holo* form, CaM adopts a more stable and compact conformation, optimized for target binding.⁸³ The flexible linker plays a pivotal role in these transitions, allowing

the lobes to move independently and facilitating the binding of structurally diverse targets. This dynamic behavior has been extensively studied through experimental techniques such as FRET and NMR, as well as computational approaches like MD simulations.⁸⁴ These studies consistently highlight the importance of the central linker in mediating the protein's conformational flexibility.

1.5.1 Dye-attached calmodulin FRET simulations

In this subsection, we introduce the chromophores, or dyes, used to simulate FRET phenomena in CaM. These dyes serve as the donor and acceptor molecules described in the EET theory introduced in Section 1.2. The selected FRET dye pairs are Alexa Fluor 488/Alexa Fluor 594 (AF488)/(AF594) and Alexa Fluor 488/Texas Red C2 (TRC2).⁸⁵ These pairs were chosen because they fulfill all the key conditions for efficient energy transfer, including:

- **Strong electronic transitions** in the UV, visible, or IR regions for both donor and acceptor chromophores.
- **Spectral overlap** between the donor's emission spectrum and the acceptor's absorption spectrum.

Where Alexa Fluor 488 always acts as the donor chromophore and Alexa Fluor 594 or Texas Red C2 always act as the acceptor chromophores.⁸⁶

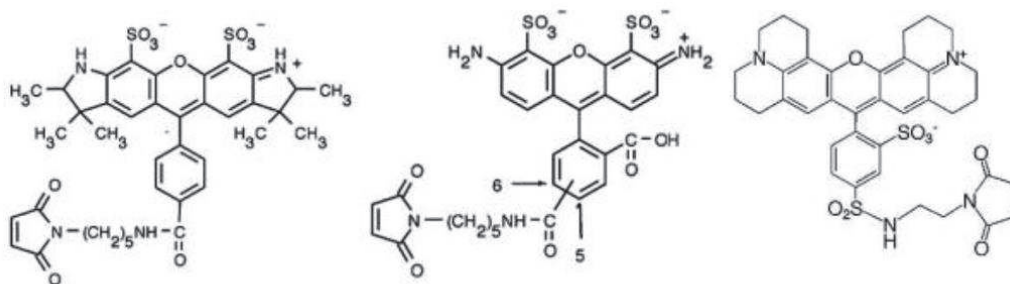


Figure 1.13: Dyes molecular structure. From left to right: Alexa Fluor 594 C5 maleimide, Alexa Fluor 488 C5 maleimide and Texas Red C2 maleimide.

The selection process involved an exhaustive review of dyes commonly used in experimental FRET studies involving CaM, which ensures that the simulated FRET observables can be directly compared with a broader range of experimental data, enhancing the validation of our theoretical models and ensures that Förster radius of the 2 dye-pairs is well known,⁸⁶ a fact that will contribute to simplify PDA equations (Section 2.2.6) to compute FRET observables.

The use of two distinct FRET dye pairs also allows us to investigate the impact of different dyes on calmodulin's theoretical conformational ensembles. By analyzing how each dye pair interacts with the protein, we aim to deepen our understanding of how dye choice influences both FRET observables and the conformational dynamics of calmodulin in theoretical simulations.

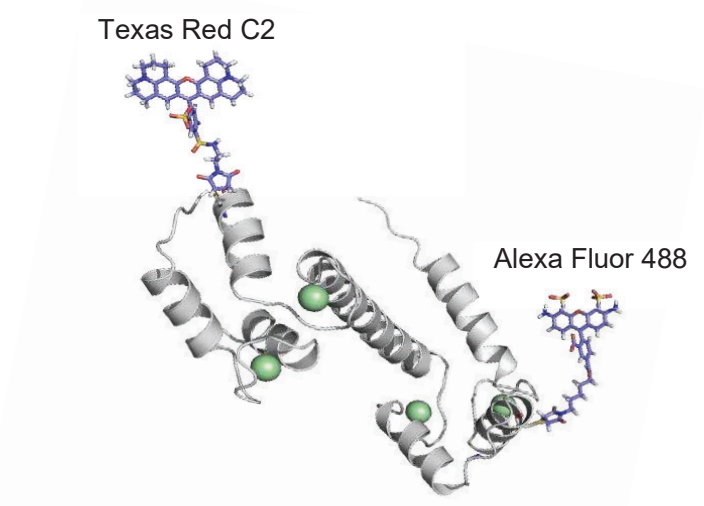


Figure 1.14: Calmodulin dye labeled system Alexa Fluor 488/Texas Red C2 extracted from *holo* CaM simulation.

1.5.2 Roles of calmodulin in human body

Calmodulin plays a fundamental role in several physiological processes within the human body, including muscle contraction, neurotransmitter release, and immune responses. In muscle cells, CaM is integral to the regulation of contraction by interacting with myosin light-chain kinase, which

phosphorylates myosin and facilitates contraction.³ In neurons, CaM is involved in synaptic signaling, where it regulates the activity of various ion channels and enzymes, influencing neurotransmitter release and neuronal excitability.⁸⁷ Moreover, it is crucial for immune responses, as it modulates the activity of calcineurin, a phosphatase that activates T-cell signaling pathways.⁸⁸

The importance of CaM in health and disease is underscored by its involvement in various pathologies. Dysregulation of CaM signaling has been linked to several conditions, including hypertension, heart disease, and neurodegenerative disorders. Understanding the molecular mechanisms of this PDP function is vital for developing therapeutic strategies targeting these diseases.

Recent reviews have discussed the significance of CaM in cellular signaling and its broader implications. For example, the article by Chin, D. et al⁸⁹ provides an overview of CaM's roles in various signaling pathways and its importance in cellular homeostasis. Another review by Krebs, J. et al.⁹⁰ details the structural aspects of CaM and its interactions with target proteins, emphasizing its versatility in cellular functions.

In the context of my research, understanding the structural and dynamic properties of CaM would be essential for investigating its function in calcium signaling. By employing computational methods, I aim to elucidate the conformational ensembles of this protein. Combining FRET simulations with MD will provide insights into the time-evolving behavior of CaM and the insights into the conformational ensembles might be helpful to understand its role in various cellular processes.

1.5.3 Previous studies on Calmodulin

This thesis aims to explore several contributions to understanding CaM's structural ensembles, emphasizing how experimental and computational approaches have converged to reveal the dynamic nature of this essential regulatory protein.

As it was introduced in the last section, CaM is a highly conserved calcium-binding messenger protein that plays a vital role in calcium signaling pathways. Its ability to bind calcium ions induces significant conformational changes, enabling CaM to interact with a diverse array of target proteins.⁹¹ These structural transitions, particularly between the *apo* (calcium-free) and *holo* (calcium-bound) states, have been the subject of extensive research to understand CaM's functional versatility.⁹²

The structural differences between *apo* and *holo* CaM are well-characterized. In its *apo* form, CaM typically adopts an extended dumbbell-shaped structure, with the N- and C-terminal lobes connected by a flexible central linker. Upon calcium binding, significant conformational changes occur, shifting the equilibrium toward more compact conformations that facilitate target binding by exposing hydrophobic patches on the protein surface.⁹³ This conformational adaptability is critical for CaM's ability to interact with a wide variety of targets, including enzymes such as myosin light-chain kinase and Ca²⁺-ATPase.

Molecular dynamics simulations have provided additional insights into CaM's conformational ensembles, revealing how local flexibility and interdomain rearrangements makes the protein rapidly interconvert into their conformational states. These simulations have also highlighted the differential behavior of the N- and C-terminal lobes between the *apo* and *holo* states, underscoring the dynamic nature of CaM in response to calcium binding.⁹⁴

Single-molecule Förster resonance energy transfer has been fundamental in exploring CaM's conformational dynamics at the nanosecond and millisecond time scale. Johnson et al. in 2016⁸⁶ used smFRET to demonstrate distinct conformational substates of *apo* and *holo* CaM, revealing populations of both extended and compact structures. In this publication Calmodulin FRET experiments revealed lifetime and efficiencies distributions very appropriate to compare with our FRET computational generated data.

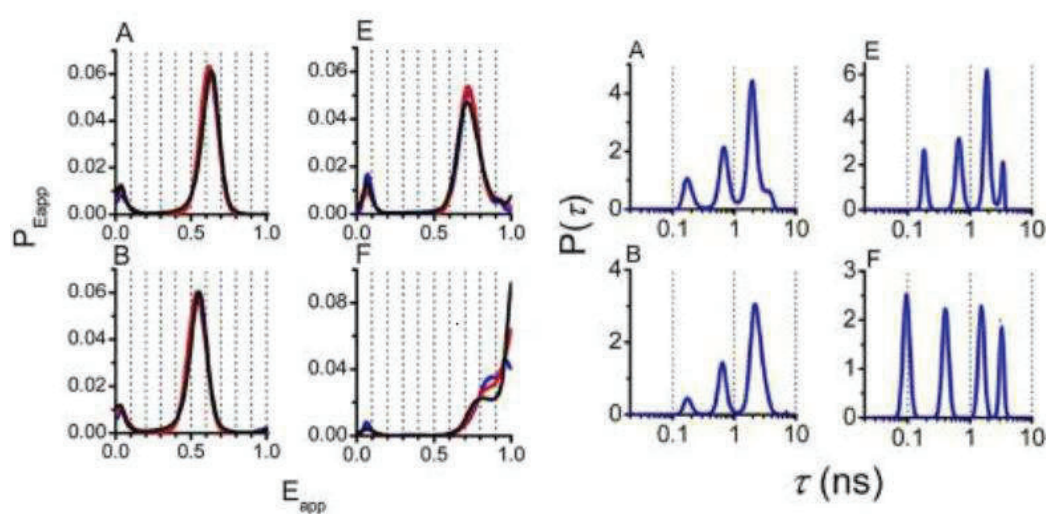


Figure 1.15: FRET CaM distributions. Left image belongs to smFRET efficiency distributions for CaM-AF488-AF594 (A and B) and CaM-AF488-TRC2 (E and F) measured in the microsecond timescale. The top row shows results at low Ca^{2+} and the bottom row at high Ca^{2+} . The black lines show analysis with 90 μs bins, the red lines with 300 μs bins, and the blue lines with 1000 μs bins. Right image belongs to fluorescence lifetime distributions for CaM-AF488AF594 (A and B) and CaM-AF488TRC2 (E and F), which correspond with FRET processes measured in the nanosecond timescale. The top row shows results at low Ca^{2+} and the bottom row at high Ca^{2+} (adapted by Johnson, C. K. et al.).⁸⁶

Their findings showed that calcium binding shifts the equilibrium toward compact conformations, critical for target protein recognition. These studies also captured fluctuations between these substates on millisecond timescales, emphasizing the dynamic nature of CaM in solution.

Additional smFRET studies reconstructed FRET states of CaM to investigate how dye interactions and the binding of peptides like CaMKII affect its conformational transitions. These analyses revealed the necessity of considering dye-protein interactions to accurately interpret FRET data and understand CaM's structural flexibility in different environments.⁹⁵

The binding of ligands, such as calmidazolium (CDZ), has been shown to stabilize CaM's compact conformation, reducing its structural dynamics over broad timescales. Ligand-binding studies have demonstrated how this stabilization facilitates effective interactions with target proteins, as highlighted by approaches combining molecular simulations and experimental techniques.⁹⁴

Structural studies using X-ray crystallography and NMR spectroscopy have complemented these findings by providing atomic-level details of CaM's conformational ensembles. Early crystal structures revealed the dumbbell-like shape of *holo* CaM, while solution NMR studies identified compact conformations that better represent the protein's behavior in cellular environments.⁶⁷ Environmental factors, such as macromolecular crowding and ionic conditions, have been shown to significantly influence CaM's helicity, EF-hand orientation, and overall structure, further underscoring the importance of studying CaM under native-like conditions.

Together, these studies have established a comprehensive understanding of CaM's conformational heterogeneity and its functional implications. Calcium binding drives transitions between *apo* and *holo* states, enabling the exposure of key hydrophobic surfaces required for interaction with a diverse range of targets. These conformational changes are critical for CaM's role as a universal calcium sensor, modulating cellular signaling pathways in response to fluctuating calcium levels

CHAPTER 2: Methodology

2.1 Techniques to simulate protein dynamics

Understanding protein dynamics is essential for unraveling protein functional mechanisms, particularly in the case of IDP or PDP proteins. Computational approaches such as MD simulations provide valuable insights into the time-dependent behavior of proteins, complementing experimental techniques like FRET. This section outlines the key aspects of simulating protein dynamics, including the use of classical force fields, the implementation of MD simulations, and challenges associated with modeling disordered systems.

2.1.1 *Classical force fields*

Classical force fields form the backbone of molecular dynamics simulations by providing a mathematical framework to describe the interactions between atoms in a molecular system.⁹⁶ These force fields typically include bonded interactions, such as bond stretching, angle bending, and torsional rotations, as well as non-bonded interactions, including van der Waals forces and electrostatic interactions. Popular force fields like AMBER,⁹⁷ CHARMM,⁹⁸ and OPLS⁹⁹ have been widely used for simulating biomolecules due to their balance between computational efficiency and accuracy. However, the performance of a force field depends on its parameterization and its ability to capture specific aspects of protein behavior, such as conformational flexibility or long-range electrostatics, which are particularly relevant for disordered systems.

Force field general expression

In this work, two different force fields were used to perform MD simulations, the ff14SB¹⁰⁰ force field from AMBER and the a99SB-disp.¹⁰¹ Both are based on the same potential energy expression, the primary differences between them lie in their force constants and the parameterization of electrostatic and van der Waals interactions.

The choice of force field is critical for accurately modeling protein dynamics. ff14SB is widely used for its balanced treatment of secondary protein structure elements, while a99SB-disp is a specially refined force field by Shaw et al. for disordered proteins that includes dispersion corrections improving the description of disordered regions. This makes it particularly suitable for studying intrinsically disordered proteins and flexible domains, such as those present in Calmodulin.

Force fields are parameterized based on experimental data and quantum mechanical calculations to balance computational efficiency with accuracy. However, they have limitations, such as an incomplete representation of electronic effects. By performing MD simulations of CaM using these two force fields, we can generate FRET observables and compare them with experimental FRET results. This comparison will allow us to identify which force field captures more accurately the dynamic behavior of CaM, providing a robust basis for validating our MD simulations and establishing a solid protocol to compare force fields performance on a partially disordered system.

The potential energy of the system is described by a series of mathematical terms representing bonded and non-bonded interactions. To describe the bond strength between two atoms, the harmonic oscillator expression is used:

$$E_{bond} = \sum_{bonds} K_b (r - r_0)^2 \quad (2.1)$$

Here, K_b is the bond force constant, r is the bond length, and r_0 is the equilibrium bond length.

For the angles formed by three consecutively bonded atoms, the potential energy is described as follows:

$$E_{angle} = \sum_{angles} K_{\theta} (\theta - \theta_0)^2 \quad (2.2)$$

In this equation, K_{θ} is the angle force constant, θ is the bond angle, and θ_0 is the equilibrium bond angle.

The energy associated with dihedral (torsion) angles between four atoms (e.g., A-B-C-D) is periodic and can be modeled with a trigonometric expansion:

$$E_{dihedral} = \sum_{dihedrals} V_n [1 + \cos(n\phi - \gamma)] \quad (2.3)$$

Here, V_n is the force constant for the dihedral, ϕ is the torsion angle, n is the periodicity, and γ is the phase angle.

The van der Waals and electrostatic interactions are the two remaining components of the potential energy. Both depend on the distances between atom pairs and contribute significantly to the total potential energy.

The van der Waals interactions are modeled using the Lennard-Jones potential:

$$E_{VdW} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] \quad (2.4)$$

In this expression, A_{ij} and B_{ij} are parameters derived from ϵ (the depth of the potential well) and σ (the distance at which the potential is zero), while R_{ij} is the distance between atoms i and j . The first term accounts for repulsion, while the second term describes dispersion forces.

The electrostatic interactions are calculated using Coulomb's law:

$$E_{ele} = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{q_i q_j}{\epsilon_{ij} R_{ij}} \quad (2.5)$$

Here, q_i and q_j are the atomic charges, R_{ij} is the distance between them, and ϵ_{ij} is the dielectric constant of the medium.

In addition to these terms, improper dihedral angles contribute to the potential energy. The expression for improper terms is:

$$E_{improper} = \sum_{improper} \frac{1}{2} K_{improper} (\theta - \theta_0)^2 \quad (2.6)$$

Where $K_{improper}$ determines the resistance to deviations from the equilibrium improper angle, θ the angle defined by four atoms, typically used to maintain planarity or stereochemical configurations, θ_0 the target value for the improper dihedral angle.

The total potential energy of the system is the sum of all these contributions:

$$V_e = E_{total} = E_{angle} + E_{tor} + E_{bond} + E_{vdW} + E_{ele} + E_{improper} \quad (2.7)$$

This formulation allows the force fields to effectively model the structural and dynamic properties of molecular systems, making them suitable for the MD simulations performed in this study.

2.1.2 Classical molecular dynamics

Molecular mechanics employs classical mechanics to model chemical systems, treating atoms as particles and bonds as harmonic oscillators. Classical MD, a methodology within this framework, utilizes computer simulations to analyze the physical movements of atoms over time, providing insights into the dynamic behavior of molecular systems.¹⁰² It is based on Newton's equations of motion, which describe how particles respond to forces in a system. The fundamental equation governing particle motion in MD is:

$$F_i(t) = M_i a_i(t) \quad (2.8)$$

In this equation, $F_i(t)$ denotes the force acting on particle i at a specific time t , and M_i is the mass of the particle. The force experienced by the particle is directly related to its acceleration $a_i(t)$.

The force $F_i(r)$ is calculated based on the potential energy function $V(r)$, which describes the energy of the system as a function of the positions of the particles. The potential energy function is critical because it determines the interactions between particles. The force on a particle is derived from the gradient of this potential energy function:

$$F_i(r) = -\nabla_i V(r) \quad (2.9)$$

Here, ∇ represents the gradient with respect to the position of particle i , and $V(r)$ is the potential energy function of the system. The negative sign indicates that the force acts in the direction of decreasing potential energy, aligning with the principle that particles move towards lower energy states.

As mentioned before, one of the primary challenges in classical MD is accurately describing these interactions through potential energy functions. That makes possible to generate trajectories that capture the dynamic behavior of particles in the system, allowing for the exploration of molecular properties such as conformational changes, binding events, and flexibility. These simulations are a key tool for understanding the physical basis of molecular interactions, complementing experimental methods like FRET. Different force fields provide various parameterizations and functional forms to model the potential energy accurately.

2.1.3 Challenges in simulating disordered systems

Simulating disordered systems remains a significant challenge in computational biology due to their lack of stable tertiary structures and highly dynamic conformational ensembles. These proteins often populate heterogeneous ensembles with multiple low-energy conformational states, making it difficult to accurately characterize their behavior using traditional MD techniques. Addressing these challenges requires advancements in force fields, sampling techniques, and multi-scale approaches to bridge computational and experimental gaps.

Force fields for disordered proteins

Traditional force fields, such as CHARMM, Amber, and OPLS, were designed primarily for folded proteins and often fail to accurately represent the conformational ensembles of IDPs. Common issues include over-stabilization of compact structures and biases toward unrealistic secondary structure content. Recent refinements, such as Amber ff99SB-disp, CHARMM36m, and OPLS-AA/M, have rebalanced protein-protein, protein-water, and water-water interactions to improve the accuracy of IDP simulations. These refined

force fields better capture the conformational heterogeneity of disordered states, bringing simulations closer to experimental observations.

Newer force fields, such as DES-Amber,¹⁰³ have been developed specifically for systems with highly polar interfaces. DES-Amber provides improved accuracy for both ordered and disordered protein complexes by addressing limitations in protein-protein interaction modeling. However, further validation is needed to fully assess its applicability across diverse IDP systems.

Polarizable force fields, such as AMOEBA¹⁰⁴ and Drude-2013¹⁰⁵, provide a more nuanced representation of electrostatic interactions by explicitly modeling electronic polarization. While these force fields are promising, they are computationally intensive and may destabilize native-like protein structures in certain cases. Continued refinement is required to optimize their use for IDP simulations, especially in dynamic cellular environments.

Enhanced sampling techniques

The large conformational space and flat energy landscapes of IDPs pose significant challenges for standard MD simulations, which often struggle to achieve sufficient sampling within practical timescales. Enhanced sampling methods are essential for overcoming these limitations. Techniques such as Replica Exchange Molecular Dynamics (REMD),¹⁰⁶ Well-Tempered Metadynamics (WT-MTD),¹⁰⁷ and Bias-Exchange Metadynamics (BEMD)¹⁰⁸ accelerate sampling by facilitating barrier crossing and providing broader coverage of high-energy states.

Gaussian Accelerated Molecular Dynamics (GaMD)¹⁰⁹ has emerged as a promising technique for IDP simulations, offering enhanced barrier crossing by applying smooth energy boosts to the potential surface. GaMD has been shown to improve the convergence of conformational ensembles while

maintaining detailed atomistic accuracy, making it well-suited for complex systems.

Replica Exchange with Solute Tempering (REST2)¹¹⁰ has proven particularly effective for IDPs by modulating solute interactions, ensuring efficient sampling of disordered ensembles. Advanced variants of REMD, such as integrated tempering sampling, combine multiple biasing strategies to achieve even higher sampling efficiency.

Coarse-grained models and multiscale approaches

Atomistic simulations of IDPs are computationally expensive due to the high dimensionality of their conformational space. Coarse-grained models address this limitation by simplifying protein representations while preserving essential structural and dynamic properties. Popular models, such as MARTINI¹¹¹ and SIRAH,¹¹² map groups of atoms to single beads, enabling simulations of large systems like biomolecular condensates and protein-protein assemblies.

Implicit solvent models, such as AWSEM-IDP,¹¹³ further enhance computational efficiency by approximating solvent effects without explicitly modeling solvent molecules. These models have been successfully applied to study aggregation processes, liquid-liquid phase separation, and protein folding pathways in IDPs.

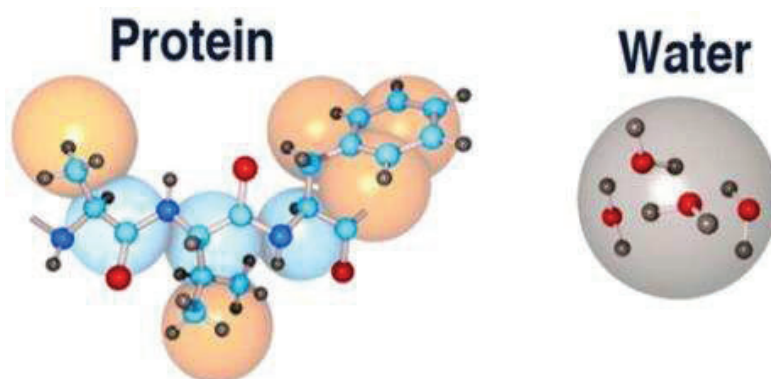


Figure 2.1: All atom versus coarse-grained representation in the CG model MARTINI of a protein section and water. All atom representation is shown in balls and sticks, while the CG representation is depicted in large spheres.

Recent developments, such as MOFF-IDP,¹¹⁴ incorporate experimental data like SAXS profiles into coarse-grained simulations, improving their ability to reproduce experimental observations. This hybrid approach connects between simulation and experimental validation, enhancing confidence in the results.

Multiscale approaches combine coarse-grained and atomistic models, allowing researchers to improve the efficiency of coarse-graining for large-scale phenomena while retaining atomistic detail for critical regions of interest. These methods are particularly valuable for studying phase transitions, such as liquid-liquid phase separation in IDPs, where both large-scale and fine-grained interactions are essential for accurate representation.

Emerging directions and remaining challenges

Despite significant progress, challenges remain in simulating IDPs with high accuracy. Force fields continue to require refinement to balance interactions across diverse environments and states, particularly for highly charged systems. Enhanced sampling methods, while effective, often require

extensive computational resources and careful tuning of parameters to achieve optimal results. Coarse-grained models must strike a delicate balance between simplicity and accuracy, especially as they are extended to more complex systems.

Machine learning approaches are increasingly being integrated into IDP simulations to optimize collective variables for enhanced sampling or predict conformational ensembles directly. For example, neural networks have been employed to identify low-dimensional representations of conformational landscapes, accelerating sampling and improving free energy calculations. These emerging techniques hold promise for addressing the high-dimensional complexity of IDP systems.

Towards comprehensive simulations

Despite advancements, challenges remain in fully capturing the structural and dynamic complexity of PDPs and IDPs. The integration of refined force fields, enhanced sampling techniques, and multiscale models continues to push the boundaries of what is possible. These developments are essential for achieving reliable simulations of disordered proteins, offering deeper insights into their roles in cellular processes and diseases.

2.2 Excited states

Accurate modeling of excited states in biological environments is essential for understanding key processes such as energy transfer, light absorption, and fluorescence, which are critical for techniques like FRET. In these systems, excited-state properties such as transition dipole moments, oscillator strengths, and spectral overlaps play an essential role in determining the efficiency and dynamics of energy transfer. Capturing these properties accurately requires advanced computational methods that account for the intrinsic electronic characteristics of chromophores and the complex environments they inhabit.

2.2.1 Excited states calculations in biological environments

Biological environments introduce additional challenges for excited-state calculations due to their dynamic and polarizable nature. Proteins, nucleic acids, and solvent molecules create local electric fields that strongly influence electronic transitions, modifying transition densities, dipole strengths, and excitation energies. Warshel and Levitt in 1976¹¹⁵ pioneered the hybrid QM/MM approach to address these complexities, providing a framework for coupling the electronic structure of chromophores to their biological surroundings. QM/MM approaches have since evolved to model excited states more accurately, particularly through polarizable embeddings that account for mutual polarization between chromophores and their environments, as described by Curutchet et al. in 2009.⁷⁵

Several computational methods are employed to study excited states in such systems, each with its strengths and limitations. Time-Dependent Density Functional Theory (TDDFT)¹¹⁶ is widely used for its computational efficiency and reasonable accuracy, particularly for localized valence excitations. However, time dependent density functional theory TDDFT faces challenges

with charge-transfer states, Rydberg states, and double excitations.¹¹⁷ Benchmarking studies on biochromophores, such as those by Jacquemin D. et al.,^{118,119} highlight the performance of hybrid and range-separated functionals like CAM-B3LYP and ω B97X-D, which improve TDDFT's accuracy for these problematic excitations. In addition, TrESP (Transition Charges from Electrostatic Potentials) methods, introduced by Renger^{120,121} and complemented by Mennucci et al. with TrESP-MMPol,¹¹ offer computationally efficient alternatives for calculating electronic couplings by approximating transition densities as fitted atomic charges, achieving near-Coulombic accuracy.

For high-accuracy calculations, Coupled Cluster methods,¹²² particularly CC3, remain the gold standard for single-excitation states, offering highly reliable results for excitation energies and oscillator strengths, as demonstrated in benchmarking studies from Bartlett & Musiał.¹²³ However, their computational expense restricts their application to small systems. For systems with significant multireference character, methods like CASPT2¹²⁴ and NEVPT2¹²⁵ provide a more practical alternative, achieving energy errors within 0.1–0.2 eV by Dreuw & Head-Gordon.¹²⁶

In dynamic biological systems, the coupling between electronic states and the environment often exceeds the capabilities of vacuum or continuum solvation models. Polarizable QM/MM models, such as those applied to the Fenna-Matthews-Olson (FMO) complex and light-harvesting (LH) systems,¹²⁷ allow for accurate representation of environmental effects, including electrostatic fields, hydrogen bonding, and chromophore interactions. These approaches, highlighted in studies on LH2 complexes, enable the prediction of spectral shifts and energy transfer mechanisms in complex environments.¹²⁸

For FRET studies, methods like TrESP-MMPol have proven invaluable. By integrating TrESP with polarizable molecular mechanics, Benedetta Mennucci et al.,¹¹ demonstrated improved accuracy in coupling calculations across MD

trajectories, enabling efficient simulations of dynamic systems like light-harvesting complexes. Similarly, benchmarking studies of TDDFT with CAM-B3LYP and ω hPBE0 functionals have provided critical insights into achieving accurate oscillator strengths and transition dipole moments, both essential for energy transfer efficiency in FRET.¹¹⁹

This chapter explores the computational approaches for modeling excited states in biological environments, emphasizing their application to FRET. It introduces relevant contributions in benchmarking QM methodologies for excited states and presents advanced methods like QM/MMPol and TrESP-MMPol, which are central to this thesis.

2.2.2 QM/MMPol method

The QM/MMPol approach, developed by C. Curutchet,^{75,77} combines quantum mechanics and molecular mechanics (QM/MM) to study electronic energy transfer (EET), addressing the limitations of continuum models. In this method, chromophores are treated using quantum mechanics, while the surrounding environment, including solvent molecules, is represented using a classical force field. This approach effectively captures solvent effects on molecular properties, such as solvatochromic shifts in optical spectra.

The method extends the polarizable continuum model (PCM) framework by incorporating a linear response (LR) scheme to describe chromophore excitations and EET couplings. A key feature of QM/MMPol is the explicit inclusion of electronic polarization in the environment, which is essential for accurately representing screening effects on EET couplings. This is achieved using a polarizable molecular mechanics force field, allowing the method to provide detailed insights into EET processes.

Applying the Born-Oppenheimer approximation, the electronic wave function of the solute will satisfy a corresponding equation, reflecting the interaction between the solute and the surrounding environment.

$$\hat{H}_{eff}|\Psi\rangle = (\hat{H}_o + \hat{H}_{env})|\Psi\rangle = E|\Psi\rangle \quad (2.10)$$

Where $\hat{H}_o$ is the Hamiltonian of the QM solute in gas phase and the $\hat{H}_{env}$ form is what differentiates MMPol from polarizable continuum model (PCM), where in this method introduces the coupling between the solute and the solvent of the system.

In this context, $\hat{H}_o$ represents the Hamiltonian for the QM solute system in the gas phase, while the operator $\hat{H}_{env}$ accounts for the coupling between the solute and the solvent.

In the QM/MMpol approach, the MM system is modeled using a classical polarizable force field based on the induced dipole model. Specifically, electrostatic interactions are represented by atomic partial charges, and polarization is treated explicitly by incorporating isotropic polarizabilities at chosen points within the solvent molecules.

The $\hat{H}_{env}$ stands as:

$$\hat{H}_{env} = \hat{H}_{QM/MM} + \hat{H}_{MM} \quad (2.11)$$

the solute-solvent interaction $\hat{H}_{QM/MM}$ follow the next expressions:

$$\begin{aligned} \hat{H}_{QM/MM} &= \hat{H}_{QM/MM}^{el} + \hat{H}_{QM/MM}^{pol} \\ &= \sum_m q_m \hat{V}(r_m) - \frac{1}{2} \sum_a \mu_a^{ind} \hat{E}_a^{solute}(r_a) \end{aligned} \quad (2.12)$$

$\hat{H}_{QM/MM}^{el}$ and $\hat{H}_{QM/MM}^{pol}$ describe the interaction between the quantum treated system and the MM charges and induced dipoles respectively.

The MM energy terms expression stands as:

$$\begin{aligned}\hat{H}_{MM} &= \hat{H}_{MM}^{el} + \hat{H}_{MM}^{pol} \\ &= \sum_m \sum_{n>m} \frac{q_m q_n}{r_{mn}} \\ &\quad - \frac{1}{2} \sum_a \mu_a^{ind} \sum_{n>m} \frac{q_m (r_a - r_m)}{|r_a - r_m|^3}\end{aligned}\quad (2.13)$$

Consequently, in Eq 2.13 $\hat{H}_{MM}^{el}$ describes the electrostatic self-energy of the MM charges and $\hat{H}_{MM}^{pol}$ does for the polarization interaction between MM charges and the induced dipoles.

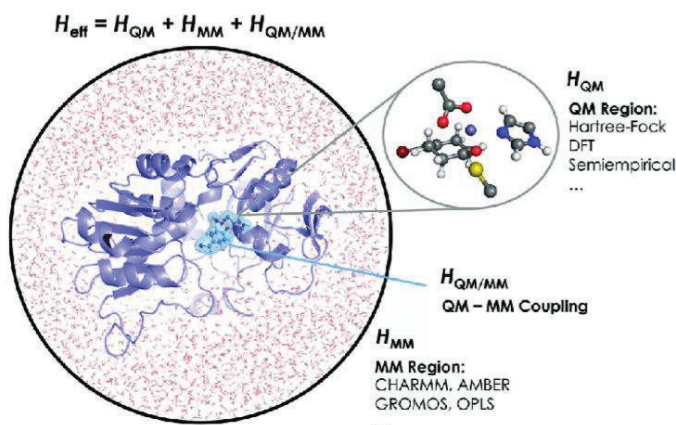


Figure 2.2: Example of a QM/MM calculation of a biomolecular system. The solvent and the backbone of the enzyme is simulated using an MM description, and the catalytic site is treated with QM.

Looking at both equations one can interpret that $\hat{V}(r_m)$ and $\hat{E}_a^{solute}(r_a)$ represent the electrostatic potential and electric field operators due to the electrons and nuclei of the QM (solute) system at the MM sites. The indices m, n and a refer to the MM charges represented as q_m and induced dipoles μ_a^{ind} on its respective coordinates r_m and r_a .

The induced dipoles on the MM description are given by

$$\mu_a^{ind} = \alpha_a (E_a^{solute} + E_a^{solvent}\{q; \mu^{ind}\}) \quad (2.14)$$

Where α_a is the isotropic polarizability for every MM point, $E_a^{solvent}$ is the total solvent electric field calculated at the given 'a' MM point and contains a sum of contributions from the point charges and the MM induced dipole moments of the system.

2.2.2.1 QM/MMPol model for excited states

This section develops the working expressions to incorporate the effects of a polarizable molecular mechanics environment within a TD-DFT linear response framework. Extending this approach to other electronic structure methods, such as Hartree-Fock (HF) or semiempirical models (e.g., TD-HF, CIS, or ZINDO), is straightforward. Within this framework, the excitation energies of a molecular system are determined by solving the following equations.

$$\begin{pmatrix} A & B \\ B^* & A^* \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} = \omega_n \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} X_n \\ Y_n \end{pmatrix} \quad (2.15)$$

The matrices AA and BB represent the Hessian of the electronic energy, while the transition vectors X_n and Y_n correspond to collective eigenmodes of the density matrix with eigenfrequencies ω_n . The Coulomb and exchange-correlation (XC) kernels contribute both diagonal and off-diagonal elements to AA and BB, refining the transitions between occupied and unoccupied levels of the ground-state potential into the true transitions of the system. The effect of the polarizable environment on the A and B matrices is incorporated by considering the molecular mechanics dipoles induced by the density matrix associated with the transition vectors X_n and Y_n .

The electronic polarization response of the environment is represented by a set of point dipoles induced by the relevant density matrix. Using the standard notation for molecular orbitals—where $i, j, \dots, i, j$, represent occupied orbitals and $a, b, \dots, a, b$, represent virtual orbitals—the matrices A and B can be expressed as follows:

$$A_{ai,bj} = \delta_{ab}\delta_{ij}(\varepsilon_a - \varepsilon_i) + K_{ai,bj} + C_{ai,bj}^{Pol} \quad (2.16)$$

$$B_{ai,bj} = K_{ai,bj} + C_{ai,bj}^{Pol} \quad (2.17)$$

Where ε_r are the orbital energies, $K_{ai,bj}$ is the coupling matrix and $C_{ai,bj}^{Pol}$ is the polarizable MM matrix, corresponding to the following expressions

$$K_{ai,bj} = \int dr \int dr' \phi_i(r') \phi_a^*(r') \left(\frac{1}{|r' - r|} + g_{xc}(r', r) \right) \phi_j(r) \phi_b^*(r) \quad (2.18)$$

$$C_{ai,bj}^{Pol} = - \sum_k \left(\int dr \phi_i(r) \phi_a^*(r) \frac{(r_k - r)}{|r_k - r|^3} \right) \mu_k^{ind}(\phi_j \phi_b^*) \quad (2.19)$$

Where the index k represents the total number of polarizable atoms in the MM region.

2.2.2.2 Formulation based on transition densities: QM/MMPol

The electronic coupling for singlet-singlet energy transfer involving bright states is dominated by coulombic and environment-mediated contributions, by some approximations it is possible to split the total electronic coupling into the sum of coulombic and environment contributions.

$$V = V_{Coul} + V_{env} \quad (2.20)$$

These contributions can be obtained using linear response time-dependent density functional theory (TD-DFT) in a polarizable quantum/molecular mechanical framework, where the chromophores and the environment are described by transition electronic densities and isotropic polarizabilities, respectively:

$$V_{Coul} = \int d\mathbf{r} \int d\mathbf{r}' \rho_D^{T*}(\mathbf{r}') \frac{1}{|\mathbf{r} - \mathbf{r}'|} \rho_A^T(\mathbf{r}) \quad (2.21)$$

where short-range contributions are neglected and the transition densities $\rho_{A/D}^T$ for the D and A molecules describe the diagonal part of the one-particle density matrix constructed from the ground and excited-state wave functions.

V_{env} describes the interaction between D and A mediated by the polarizable environment.

$$V_{env,TrESP} = - \sum_{i,l} d\mathbf{r}' \rho_D^{T*}(\mathbf{r}') \frac{(\mathbf{r}_i - \mathbf{r}_l)}{|\mathbf{r}_i - \mathbf{r}_l|^3} \boldsymbol{\mu}_l^{MMPol}(\rho_A^T(\mathbf{r})) \quad (2.22)$$

Where $\boldsymbol{\mu}_l^{MMPol}$ represents the magnitude introduced in Eq 2.14 from this QMMMPol description.

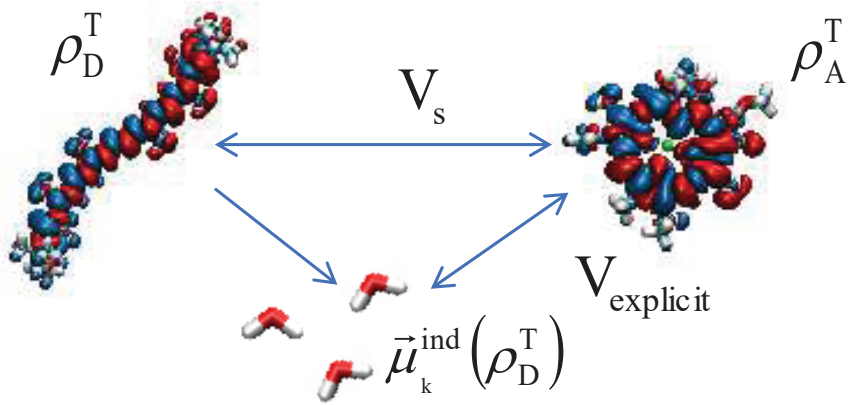


Figure 2.3: Representative scheme of QMMMPol transition densities, the coulombic coupling and mediated environment coupling.

One can then define an effective dielectric constant for the environment as

$$s = \frac{1}{\epsilon_{eff}} = \frac{V_{Coul} + V_{env}}{V_{Coul}} \quad (2.23)$$

Using this expression, the total coupling can be expressed as $V = sV_{Coul}$, providing a direct link to the coulombic term and the screening factor in Förster coupling expression

$$s V \approx sV_{PDA} = 1/\epsilon_{opt} \cdot V_{PDA} \quad (2.24)$$

2.2.3 Formulation based on transition charges: TrESP-MMPol

A more practical simplification, however, relies on approximating the transition densities as a distributed transition monopole distribution, i.e. a set of atomic transition charges like those widely used in classical force fields. Here, we adopt a TrESP-MMPol model,¹¹ in which charges are fitted to

reproduce the electrostatic potential generated by QM-derived transition densities, which leads to the following expressions:

$$V_{Coul,TrESP} = \sum_{i,j} \frac{q_{D,i}^T q_{A,j}^T}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2.25)$$

This expression describes the coulombic coupling, in analogy with Eq 2.21 from QM/MMPol method, where instead of electronic transition density we place electronic transition charges. And for the environment coupling then the analogous expression is

$$V_{env,TrESP} = - \sum_{i,l} \frac{q_{D,i}^T (\mathbf{r}_i - \mathbf{r}_l) \cdot \boldsymbol{\mu}_l^{MMPol}(\{q_A^T\})}{|\mathbf{r}_i - \mathbf{r}_l|^3} \quad (2.26)$$

Where we find the same substitution as in the previous equation on the transition atomic charges.

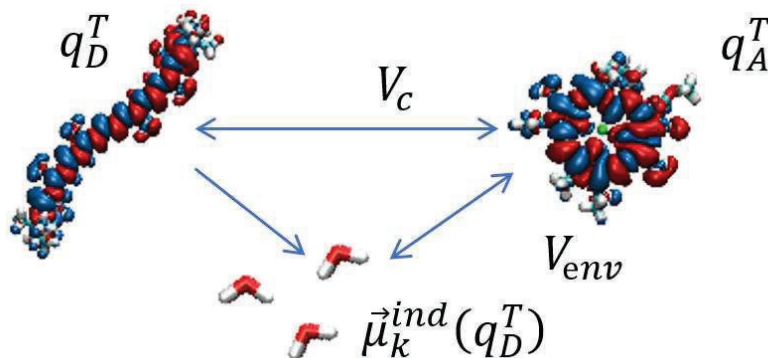


Figure 2.4: Representative scheme on TrESP-MMPol transition atomic charges, the coulombic coupling and mediated environment coupling.

The $\boldsymbol{\mu}_l^{MMPol}$, described as the induced environment mediated dipole, would be calculated as

$$\mu_k^{MMpol} = \alpha \cdot E(r_k) \quad (2.27)$$

Where α is the polarizabilities of the atoms and $E(r_k)$ the electric field created by environment atoms disposition.

2.2.4 Comparison of TrESP-MMPol and QM/MMPol Methods

The QM/MMpol and TrESP-MMPol methods both combine quantum mechanics and molecular mechanics to study electronic energy transfer, with a focus on capturing environmental effects. However, they differ in their computational requirements and ability to account for structural and electronic variations in the system.

The QM/MMpol method recalculates the electronic transition density for each frame of the molecular dynamics trajectory using time-dependent density functional theory. This approach ensures that the coupling reflects the exact excited-state properties for the specific configuration, including any structural variations in the fluorophores or their environment. By capturing the dynamic electronic polarization and screening effects, QM/MMpol provides highly accurate coupling values, making it the most rigorous approach. However, this comes at a substantial computational cost, making QM/MMpol impractical for long MD simulations or large systems.

The TrESP/MMpol method, on the other hand, simplifies the process by deriving electrostatic transition charges from a single TDDFT calculation. These charges are then applied throughout the MD trajectory, allowing for efficient coupling calculations without recalculating the electronic transition density for every frame. This makes TrESP-MMPol particularly suitable for studies involving long trajectories, such as those in this research. However, the method's main limitation lies in its static representation of the transition

charges, which cannot capture variations in the excited states caused by significant structural changes in the fluorophores or their interactions with the environment. This limitation becomes especially relevant when the fluorophores adopt unusual positions or conformations, as these changes can alter the excited states and couplings, leading to slight inaccuracies in the calculated values.

Errors in the TrESP method are most pronounced in systems where the fluorophores undergo dynamic structural changes or interact strongly with their surroundings. These effects, which are captured more accurately by the QM/MMpol approach, highlight the trade-off between computational efficiency and accuracy.

In summary, QM/MMpol offers unparalleled accuracy by recalculating the electronic transition density at every frame, but it is computationally demanding and unsuitable for extended MD simulations. TrESP-MMpol provides a practical compromise by enabling efficient coupling calculations over long trajectories, despite potential inaccuracies in systems with significant structural or environmental variations. TrESP-MMpol accounts of the atomistic heterogeneous environment of the dyes for the extensive sets of frames explored in MD simulations.

2.2.5 Beyond Förster approximation: Dipole approximation and screenings effects

Förster resonance energy transfer (FRET) describes the non-radiative transfer of electronic excitation energy between a donor and an acceptor molecule mediated by dipole-dipole Coulombic interactions. Theodor Förster developed the practical expressions for calculating energy transfer rates based on spectroscopic observables and the distance separating the donor

and acceptor. These equations provide insights into protein-protein interactions or conformational changes in nucleic acids. However, the classical Förster framework relies on key approximations that can limit its accuracy. Central to the theory is the assumption that the donor-acceptor coupling is governed by a dipole-dipole interaction attenuated by a simple dielectric screening factor $1/n^2$, where n is the refractive index of the medium. While this factor accounts for the optical dielectric properties of the environment, it treats the surroundings as a homogeneous continuum and does not capture the heterogeneous polarizability of real biological systems. This simplification can lead to significant deviations in calculated energy transfer rates, particularly in systems with complex or dynamic environments. Additionally, the assumption of isotropic averaging for the dipole orientation ($\langle \kappa^2 \rangle = 2/3$) can be inadequate when donor and acceptor molecules are not free to rotate or adopt fixed orientations due to steric constraints.

These limitations are particularly pronounced when the donor-acceptor distance approaches the size of the fluorophores themselves or when structural fluctuations significantly alter their excited-state properties. At such scales, the dipole approximation breaks down, and the actual electronic couplings depend on the transition densities of the donor and acceptor. Furthermore, the impact of environmental screening is more complex than a simple factor, as the local polarizability and spatial arrangement of the medium can either enhance or suppress coupling by factors far exceeding the classical predictions.

To overcome these challenges, we aim to use the previous methods explained before to check if the Förster approximations can actually change the FRET analysis on the study of CaM. By explicitly considering the electronic polarization of the medium and the full spatial distribution of donor-acceptor interactions, these methods provide a more accurate depiction of energy transfer processes. They can account for deviations from the conventional screening factor, particularly in heterogeneous or highly polarizable

environments, and adjust the coupling values based on the actual structural and electronic configurations of the donor and acceptor. Such approaches reveal that energy transfer rates can deviate significantly—sometimes by a factor of four or more—from those predicted by Förster theory.

These advancements have shown that screening effects, donor-acceptor orientations, and local environmental heterogeneity play a much larger role in FRET than previously assumed under classical Förster theory. By addressing the inherent approximations in the dipole-dipole coupling and dielectric screening, these methods allow for a more detailed and accurate understanding of energy transfer processes in biological systems, establishing for improved interpretations of FRET experiments.

Our research shows that screening impact on the FRET distribution interpretation plays a major role than dipole approximation and it will be further discussed on the results section.

2.2.6 Generating FRET observables from MD trajectories

This section will explain how we have adapted specific expressions from Förster theory and applied various approaches to derive the EET observables, which represent a substantial portion of the results in this thesis.

2.2.6.1 Coupling FRET dynamics and dye orientational dynamics (dynamic, static and intermediate averaging regimes)

The Coulombic and short-range contributions to electronic energy transfer are highly sensitive to the mutual distance and orientation between donor (D) and acceptor (A) molecules. As a result, electronic couplings often display significant fluctuations, which are crucial for describing electronic energy

transfer dynamics accurately. These fluctuations can be understood in two limiting regimes: the dynamic averaging regime and the static averaging regime.

In the dynamic averaging regime, coupling fluctuations occur on timescales much faster than the donor's excited-state lifetime. Here, the FRET rate can be expressed using the average squared coupling value, $\langle V^2 \rangle$. In contrast, the static averaging regime applies when structural changes are slower than the donor excited-state decay. Under these conditions, a distribution of transfer rates arises, reflecting the static conformations of the system. This distinction is particularly relevant in FRET studies of biosystems, where the primary source of uncertainty is often related to the orientation factor, κ . The common assumption of an isotropically averaged orientation factor, $\langle \kappa^2 \rangle = 2/3$, may not always hold, especially in cases where D/A molecules are constrained by the system's structure.

Molecular dynamics (MD) simulations have become a valuable tool for investigating the impact of electronic coupling fluctuations on FRET experiments and for characterizing the structural ensembles underlying static rate distributions. For instance, MD simulations have highlighted discrepancies with the isotropic approximation often used in FRET analyses. They have also revealed correlations between D/A distances and orientations, which are neglected under the assumption that $\langle \kappa^2 R^{-6} \rangle = \langle \kappa^2 \rangle \langle R^{-6} \rangle$.

Beyond FRET, MD simulations have been instrumental in studying EET processes in systems where D/A molecules are restricted to specific conformations, such as photosynthetic complexes, DNA, and protein–ligand assemblies. These studies demonstrate the importance of accounting for the

full range of coupling fluctuations, as they often deviate from the assumptions of Förster theory, leading to a more accurate characterization of EET dynamics in complex biological environments.

There are many proposed strategies to capture this FRET dynamics fluctuations. The most accurate would be to apply Monte Carlo (MC) algorithm into energy transfer process. By calculating the energy transfer rates, a random number is generated, and the dynamics of the system can be described using three distinct probabilities. The first is the probability of the donor undergoing de-excitation through photon emission by the donor's fluorescence lifetime. The second is the time-dependent that describes the transfer of excitation energy from the donor to the acceptor. Finally, the remaining probability, corresponds to the donor remaining in its excited state during the time interval. These probabilities illustrate the interplay of radiative and non-radiative processes in the donor-acceptor system.

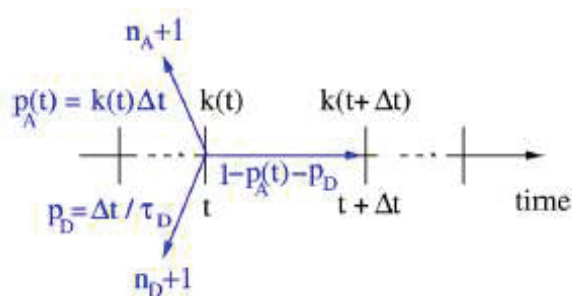


Figure 2.5: MC scheme on FRET process. Adapted from: Theory of FRET “Spectroscopic Ruler” for Short Distances: Application to Polyproline.³⁵

In our case, we proceeded to adopt a double average regime approximation,^{28,35,129} validated by MC simulations in many cases, to incorporate static and dynamic disorder by separating slow and fast fluctuations in instantaneous transfer rates. In FRET systems where coupling

fluctuations are fast compared to the donor lifetime for the excited state of interest (dynamic average regime), EET efficiency can be computed by the averaged rate constant:

$$E_{EET} = \left(1 + \frac{1}{\tau_D \langle k_{EET} \rangle_{fast}} \right)^{-1} \quad (2.28)$$

In static average regime, the ensemble efficiency will be able to be computed by averaging the efficiency over static conformations, as the EET rates will remain will be a narrow distribution of values:

$$E_{EET} = \left\langle \left(1 + \frac{1}{\tau_D k_{EET}} \right)^{-1} \right\rangle_{slow} \quad (2.29)$$

When both scenarios are relevant to the system, as in CaM, intermediate averaging regime is used, using a double average over static and dynamic disorder:

$$E_{EET} = \left\langle \left(1 + \frac{1}{\tau_D \langle k_{EET} \rangle_{fast}} \right)^{-1} \right\rangle_{slow} \quad (2.30)$$

Where $\langle \dots \rangle_{fast}$ is an average over fluctuations that are fast compared to the D excited state lifetime. The dynamic or fast average is made on the exact value of the lifetime value of the donor whereas the slow or static average is made to compute a unique value for that EET observable.

In this thesis particular case, we have assumed as the fast regime the average on the calculation of the rates by averaging the rates over 4 ns (which is the lifetime donor dye). After that, the static regime is included by averaging the EET efficiency over the set of averaged rates (like is shown in Eq 2.30). This

averaging regime already constitutes the intermediate averaging regime on our results.

2.2.6.2 Derivation of J from R_0 and calculation of instantaneous $k(t)$ rates and efficiency and lifetimes along MD trajectory

From the generated MD simulations we aim to calculate FRET properties such as EET rates, EET efficiencies or lifetime distributions. A mathematical derivation has been followed in order to be able to compare the generated observables from point dipole approximation and TrESP-MMPol. From the golden rule expression mentioned in the introduction

$$k_{FRET} = \frac{2\pi}{\hbar} V^2 J \quad (2.31)$$

It is possible to assume that

$$k_{FRET} = \frac{2\pi}{\hbar} \left(\frac{1}{n^2} \frac{\kappa \mu_D \mu_A}{R^3} \right)^2 ; J = \frac{1}{\tau_D} \left(\frac{R_0}{R} \right)^6 \quad (2.32)$$

By assuming Förster's approximation and by combining Eq 2.31 and Eq 2.32 when $k_2=2/3$, one can derive that the spectral overlap factor J from the experimental R_0 value, tabulated for specific dye pairs is equivalent to the following expression

$$\frac{2\pi}{\hbar} J = \frac{3n^4 R_0^6}{2\tau_D \mu_D^2 \mu_A^2} \quad (2.33)$$

Then, it is possible to compute the instantaneous transfer rate using different approximations for the coupling V computed at time t of an MD trajectory using the following expression, which combines Eq. 2.31 and 2.33:

$$k_{theo}(t) = V(t)^2 \frac{3n^4 R_0^6}{2\tau_D \mu_D^2 \mu_A^2} \quad (2.34)$$

If we adopt PDA approximation a simpler expression follows

$$k_{PDA}(t) = \frac{3R_0^6}{2\tau_D} \frac{\kappa(t)^2}{R(t)^6} \quad (2.35)$$

And from Eq 2.31 and Eq 2.34 we find that

$$k_{theo}(t) = V(t)^2 \frac{3n^4 R_0^6}{2\tau_D \mu_D^2 \mu_A^2} = V(t)^2 \cdot C \approx V(t)^2 \cdot J \quad (2.36)$$

That leads to complement TrESP-MMPol rate expression as

$$k_{TrESP-MMPol}(t) = V_{TrESP-MMPol}(t)^2 \frac{3n^4 R_0^6}{2\tau_D \mu_D^2 \mu_A^2} \quad (2.37)$$

That would make comparable generated observables from both methods, as any deviation on the electronic transition dipole moment of the donor or acceptor chromophore would be corrected by this approximation.

Following the previous expressions from section 1.3 Eq 1.17 and 2.30 and applying the double average regimes for dynamic and static disorder

$$E_{PDA} = \left\langle \left(1 + \frac{1}{\tau_D \langle k_{PDA} \rangle_{fast}} \right)^{-1} \right\rangle_{slow} \quad (2.38)$$

$$E_{TrESP-MMPol} = \left\langle \left(1 + \frac{1}{\tau_D \langle k_{TrESP-MMPol} \rangle_{fast}} \right)^{-1} \right\rangle_{slow} \quad (2.39)$$

And for PDA expression:

$$E_{\text{EET}}^{\text{PDA}(R,k)} = \left\langle \left(1 + \frac{2}{3R_0^6} \frac{1}{\langle \kappa^2 \rangle_f} \right)^{-1} \right\rangle_s \quad (2.40)$$

In addition, fluorescence lifetime distributions and decays were derived from the transfer rates using the following expression.

$$\tau_f = \frac{1}{k_D + \langle k_{\text{theo}}(t) \rangle_{\text{fast}}} \quad (2.41)$$

$$I = \langle e^{-(k_D + \langle k_{\text{theo}}(t) \rangle_{\text{fast}})t} \rangle_{\text{slow}} \quad (2.42)$$

Where K_D is the donor molecule decay rate introduced in Eq 1.15.

Another relevant analysis for this thesis, is to compute the theoretical results on how an experimental efficiency would be interpreted using the PDA approach instead of TrESP-MMPol approach.

In this context, from the distances obtained from the MD simulations we compute the efficiency using TrESP-MMPol and intermediate average regime as it was explained before. If we assume that this data generation is the most accurate one, from this efficiency, we can re-calculate the distance by adapting the Förster theory. In this way, we are able to compare the distance distribution generated by PDA and TrESP-MMPol on the same data, making them directly comparable and giving the opportunity to appreciate the differences using the two different approaches.

The analysis is based on adapting the expression of the PDA efficiency with the Förster radius

$$E_{FRET} = \frac{1}{1 + (R/R_0)^6} \rightarrow R(E) = R_0(E^{-1} - 1)^{1/6} \quad (2.43)$$

2.2.6.3 Combining PDA and TrESP-MMPol approach

Throughout this manuscript, it has been demonstrated that the electronic energy transfer (EET) coupling can be decomposed into a Coulombic component and an environment-dependent component. We have discussed how advanced techniques allow us to go beyond the traditional Förster approximations, introducing refined formulations for these couplings. In this section, we aim to explain how these PDA and TrESP-MMPol couplings are combined to differentiate the significant impacts that the various approaches have on the resulting observables, providing a deeper understanding of their contributions.

As previously shown, the definitions for the environment coupling are

$$V_{PDA\ environment} = \frac{1}{n^2} \quad (2.44)$$

And

$$\begin{aligned} V_{TrESP-MMPol\ environment} &= S = \frac{1}{\epsilon_{eff}} \\ &= \frac{V_{TrESP-MMPol\ Coul} + V_{TrESP-MMPol\ env}}{V_{TrESP-MMPol\ Coul}} \end{aligned} \quad (2.45)$$

For the coulombic coupling

$$V_{PDA\ coulombic} = \frac{\kappa\mu_D\mu_A}{R^3} \quad (2.46)$$

And

$$V_{TrESP\ coulombic} = \sum_{i,j} \frac{q_{D,i}^T q_{A,j}^T}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2.47)$$

It is possible to define a combination such as

$$V_{PDA-MMPol} = \frac{V_{TrESP-MMPol\ Coul} + V_{TrESP-MMPol\ env}}{V_{TrESP-MMPol\ Coul}} \cdot \frac{\kappa\mu_D\mu_A}{R^3} \quad (2.48)$$

And

$$V_{TrESP} = \frac{1}{n^2} \cdot \sum_{i,j} \frac{q_{D,i}^T q_{A,j}^T}{|\mathbf{r}_i - \mathbf{r}_j|} \quad (2.49)$$

In that way now, lifetimes distribution values will be compared from computing PDA, TrESP-MMPol, TrESP and PDA-MMPol, from combining the decomposed coupling contributions. This procedure allows to differentiate weather the impact is coming from in a very straightforward manner.

2.3 Computational protocol

This section aims to clarify the multiscale computational protocol adopted in this thesis to simulate FRET properties from MD simulations on the CaM protein. We thus give a quick overview on the computational pipeline followed. First, atomistic classical MD simulations of several CaM variants were performed using two force fields: ff14SB and a99SB-disp. FRET properties from these MD trajectories were then estimated by computing electronic couplings along the frames of the trajectories using the TrESP/MMPol method, described in Section 2.2.3, or using the PDA approximation using transition dipoles calculated for the donor and acceptor dyes. The TrESP/MMPol method allows a very fast estimate of couplings along extensive trajectories using TrESP charges precomputed for the dyes from a previous QM calculation. Before applying this method, however, TrESP/MMPol calculations were benchmarked against rigorous QM/MMPol calculations (Section 2.2.2) performed on hundreds of structures, in which couplings were computed from transition densities obtained from TD-DFT calculations. In this way, we verified the accuracy of TrESP/MMPol and determined the optimal MMPol cutoff radius for performing the TrESP/MMPol calculations on the MD trajectories, balancing computational precision and calculation time. Then, the equations described in Sections 2.2.6.2 and 2.2.6.3 were applied to generate all FRET observables. In Section 2.2.6.3, we specifically described how the PDA and TrESP/MMPol methods were combined to derive four levels of approximation for the coupling estimates, to help in assessing the validity of either the point dipole or screening approximations in Förster theory: PDA, TrESP-MMPol, TrESP and PDA-MMPol.

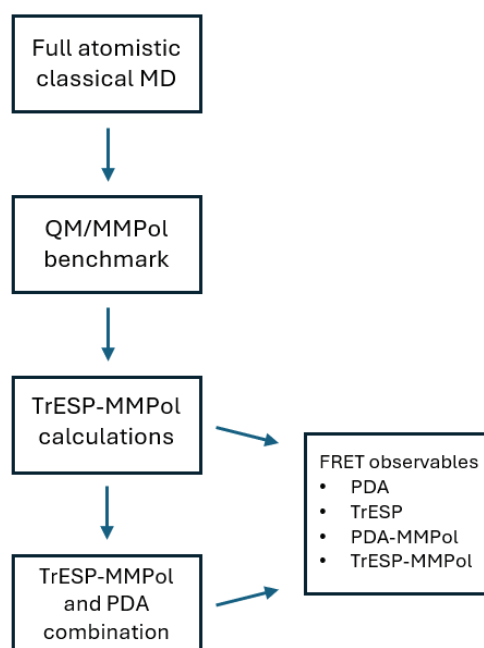


Figure 2.6: Computational protocol procedure workflow.

CHAPTER 3: Objectives

OBJECTIVES

This thesis focuses on the study of partially disordered proteins (IDPs), with calmodulin (CaM) chosen as the model system due to its central role in numerous physiological processes and the availability of extensive structural and experimental data for validation. A primary objective is the characterization of the conformational ensembles of calmodulin using molecular dynamics (MD) simulations with two distinct force fields: Amber ff14SB, a standard and widely recommended force field for proteins, and a99SB-disp, a refined force field specifically optimized for partially disordered proteins like CaM. By generating reliable FRET observables, this work seeks to validate which force field better captures the conformational behavior of calmodulin.

By exploring the applicability of Förster Resonance Energy Transfer (FRET) in providing insights of conformational ensembles, we use advanced computational approaches such as TrESP-MMPol and QM/MMPol. We seek to evaluate how different descriptions of Coulombic coupling and environmental polarization influence energy transfer calculations. A particular focus is placed on environmental screening, which arises from the interaction between the chromophores transition densities and their surrounding environment, such as solvent molecules or protein matrices. This screening provides a more realistic representation of the dynamic interactions between chromophores and their environment, addressing limitations of classical methods like Förster's Point Dipole Approximation (PDA), which typically neglect these effects. By integrating environmental screening into the calculation of FRET observables, this study aims to refine the theoretical understanding of energy transfer processes and improve the connection between simulations and experimental data.

Beyond evaluating environmental screening, this thesis investigates additional contributions to FRET observables, including Coulombic dipole interactions, and orientational anisotropy (κ^2). A novel screening function is also developed to link FRET observables with distance, offering deeper

insights into how environmental effects modulates energy transfer in complex biological systems.

Ultimately, this research provides a comprehensive understanding of calmodulin's conformational and functional dynamics, by connecting theoretical and experimental results. By combining computational simulations and energy transfer methodologies, this work advances the field of FRET-based studies and deepens our knowledge of MD simulations on partially disordered proteins like calmodulin.

Specific objectives:

- Characterize the conformational ensembles of calmodulin: Perform a detailed analysis of the conformational behavior of calmodulin using MD simulations with two force fields: Amber ff14SB, a standard and widely recommended force field for proteins and a99SB-disp, a force field specifically refined for intrinsically disordered proteins.
- Generate comparable FRET observables: Compare and validate the MD simulations by generating and analyzing FRET observables, determining which force field better describes the conformational ensembles of calmodulin.
- Validate and compare computational models: Evaluate Förster's point dipole approximation (PDA) against advanced methods such as TrESP and QM/MMPol to identify the strengths and limitations of each approach in describing energy transfer in dye-labeled systems.
- Analyze the impact of environmental screening: Investigate how environmental polarization, modeled with the QM/MMPol and TrESP-MMPol method, affects EET coupling and FRET observables.
- Quantify contributions to FRET observables: Assess the contributions of Coulombic dipole interactions, environmental screening, and the orientational factor κ^2 to FRET observables.

- Develop a novel screening function: Derive and validate a new screening function to describe FRET observables as a function of distance, adjusting parameters based on MD simulations and FRET data.
- Bridge theory and experiment: Use the validated FRET observables to correlate theoretical predictions with experimental findings, improving the understanding of the conformational dynamics of dye-labeled calmodulin.

OBJECTIVES

CHAPTER 4: On the breakdown of Förster energy transfer theory due to solvent effects: Atomistic simulations unveil distance-dependent dielectric screening in calmodulin

4.1 Introduction

The content of this chapter presents the results associated with the first manuscript derived from the content of this thesis, which has been submitted for publication. It begins with an introduction to the context of the study, followed by a detailed analysis and discussion of the various aspects and findings of the project.

As it was previously mentioned in section 1.2, Förster theory screening effects are described by adding a $1/n^2$ prefactor in the dipole coupling rate expression:

$$V \approx sV_{PDA} = \frac{1}{n^2} V_{PDA} = \frac{1}{n^2} \frac{\kappa \mu_D \mu_A}{R^3} \quad (4.1)$$

where V is the electronic coupling between D/A molecules, J the spectral overlap factor obtained from the area-normalized donor emission and acceptor absorption spectra, μ_D and μ_A the D/A transition dipole moments, κ the dipole orientation factor, R the D/A center-to-center distance, and n the refractive index of the medium.

The $1/n^2$ screening prefactor leads to a dramatic fourfold attenuation of the FRET rate in typical biological environments, where n^2 is often assumed a value of 2, representing the optical component of the dielectric constant of the medium (ϵ_{opt}).¹³⁰ Förster derived this simple screening factor by assuming a continuum dielectric embedding the D/A point dipoles. Some refinements can be introduced if one assumes that the molecules are placed in spherical cavities inside the dielectric. This leads to screening factors ranging from $3/(2\epsilon_{opt} + 1)$ for dipolar transitions to $2/(\epsilon_{opt} + 1)$ for high-order multipoles.¹³¹ More realistic models have later been developed coupling a QM description of the chromophores to modern continuum solvation models like the integral equation formalism of the polarizable

continuum model (IEFPCM), which allows to account for the molecular-shaped cavities enclosing the molecules in the dielectric, or by adopting an atomistic polarizable molecular mechanics description of the environment in a quantum/molecular mechanics framework.^{132–134}

The application of multiscale QM/classical models has allowed unprecedented details on the impact of screening effects in FRET processes occurring in a variety of biological systems.¹³⁵ Application of the QM/IEFPCM model to pigment pairs from photosynthetic complexes of cyanobacteria, higher plants and cryptophyte algae, showed that screening is exponentially attenuated at D/A separations below 20 Å, thus modifying the well-known R^{-6} distance dependence of Förster FRET rate, a behavior occurring when the pigments start to form a common cavity inside the dielectric.^{136,137} In turn, application of the QM/MMPol model to photosynthetic complexes, nucleic acids and protein-ligand complexes,¹³⁵ indicated that the heterogeneous nature of the environment polarizability often leads to important deviations from the $1/\epsilon_{opt}$ Förster prefactor, profoundly tuning by a factor up to ~ 4 energy migration rates compared to the average continuum dielectric view that has been historically assumed.¹³⁸

Recently, Eder and Renger applied extensive Poisson-TrEsp calculations to study photosystem I trimers, concluding that exponential distance dependence only contributes to the dielectric screening in chlorophyll pairs with in-line dipole arrangements.¹³⁹ In disordered proteins, the exchange between folded and more solvent-exposed extended structures means both effects – distance-dependent screening and dielectric heterogeneity – could play an important role in energy transfer processes underlying FRET structural studies.

Thus, the objective of this work is to assess how environment effects deviate from Förster $1/\epsilon_{opt}$ assumption in a disordered system, an approximation that biases the characterization of the underlying conformational ensemble.

The cost associated to multiscale QM/MMPol models has so far limited its application to the massive number of conformations needed to describe, for example, a disordered protein. Here we address this challenge using a rigorous theoretical framework that overcomes Förster limits adopting a TrESP-MMPol approach, which describes the fluorophores through electrostatic potential-fitted transition charges coupled to a polarizable molecular mechanics description of the environment.¹⁴⁰ While this approach allows an efficient processing of thousands of structures extracted from extensive MD simulations, it allows to retain the accurate description of screening effects provided by more costly QM/MMPol calculations. We apply this strategy to investigate the impact of Förster approximations on the Ca^{2+} -dependent conformational preferences of calmodulin (CaM), a protein that plays a major role in the transmission of calcium signals in eukaryotes.^{141,142}

In particular, we simulate the FRET properties of *holo* CaM tagged with Alexa Fluor 488 and Texas Red dyes, shown in Fig. 4.1, which was investigated by the Johnson group using time-resolved and single-molecule fluorescence spectroscopy.¹⁴³ Our results indicate a profound impact of dielectric screening on CaM FRET distributions. Strong distance-dependent screening effects induce an important shift on fluorescence lifetime distributions, leading to corrections much larger than those due to breakdown of the dipole approximation, especially at close D/A separations. Interestingly, our results warn that the common practice of ignoring distance-dependent screening effects masks the common tendency of classical force fields to overstabilize compact folded structures, with important consequences in the refinement and validation of classical potentials for disordered proteins based on FRET data.

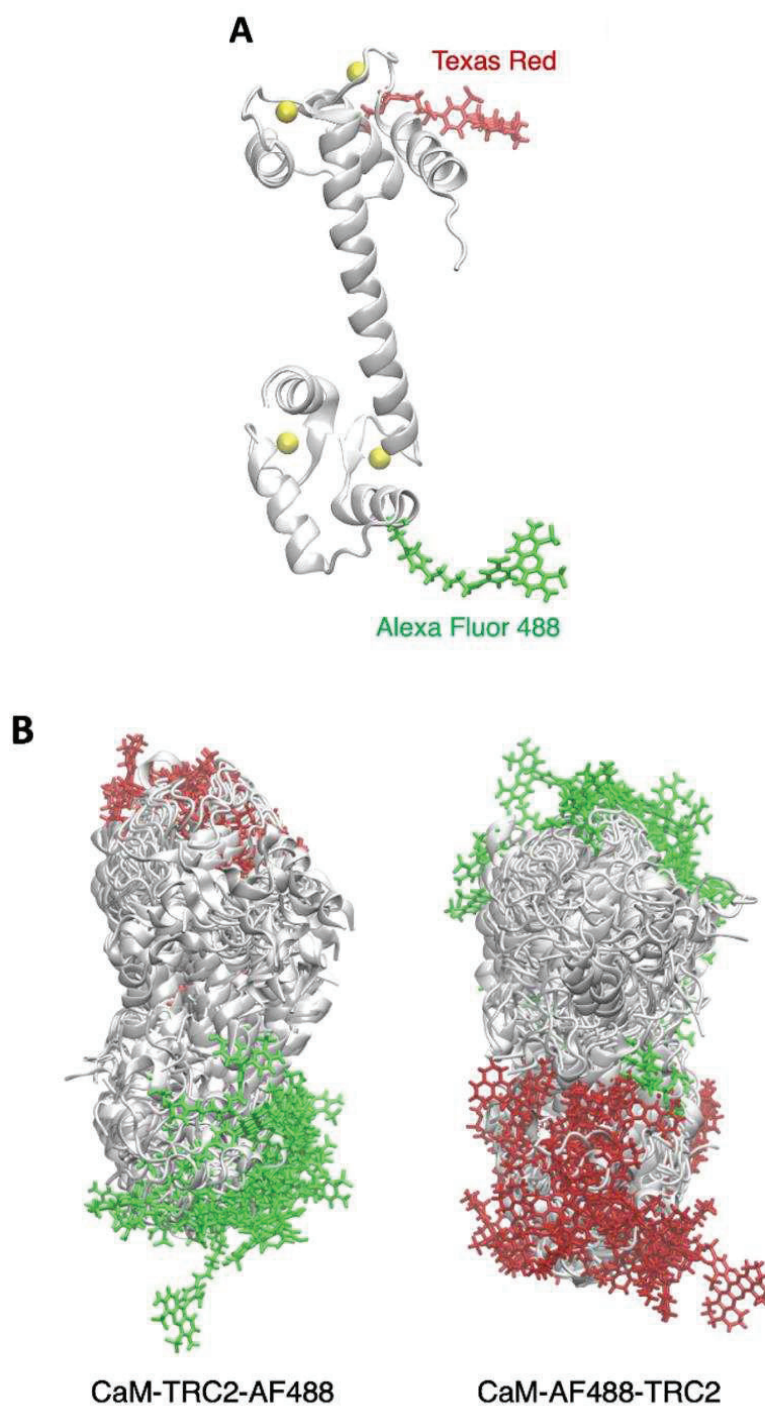


Figure 4.1. A Structure of *holo* Calmodulin tagged with the Texas Red C2 maleimide and Alexa Fluor 488 C5 maleimide dyes. B Conformational ensembles derived from MD simulations for *holo* CaM-TRC2-AF488 and CaM-AF488-TRC2 systems.

4.2 Methods: Multiscale MD/TrESP-MMPol approach

We adopt a multiscale pipeline based on performing classical MD simulations and postprocessing the trajectories with extensive calculations of FRET electronic couplings based on the TrESP-MMPol method.¹⁴⁰ Compared to using other multiscale approaches like QM/IEFPCM, based on a continuum solvation model, this strategy allows to account for heterogeneities in the polarizability of the environment in the calculation of FRET couplings by using an atomistic polarizable description of the protein-water environment. On the other hand, it allows to estimate couplings with similar quality compared to the QM/MMPol approach at a much cheaper cost by using TrESP charges parametrized from QM calculations, which avoids the need to perform QM/MM calculations on many structures. In the following we describe the expressions adopted to estimate FRET properties and couplings using this protocol, as well as the details of the MD and TrESP-MMPol calculations performed.

The FRET observables generation for the results follow the expression and equations derivations explained in the section 2.2.6.

4.3 Computational details

4.3.1 MD simulations

Initial systems were prepared starting from the human *holo* structure of CaM solved at 1.7 Å resolution (PDB ID 1CLL).¹⁴⁴ We considered CaM with mutations T34C and T110C labelled at these positions with donor/acceptor dyes Alexa Fluor 488 C5 maleimide and Texas Red C2 maleimide. We prepared two systems, *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems, with donor/acceptor dyes placed on 34/110 or 110/34 positions, in order to reproduce the experimental mixture.

Missing residues and the C- and N-terminal ends were added manually, and all amino acids were considered in their standard protonation state as predicted by PROPKA3 calculations.¹⁵¹ Simulations were based on the Amber ff14SB protein force field,¹⁴⁶ the TIP3P water model,¹⁴⁷ the Joung-Cheatham parameters for monovalent ions¹⁴⁸ and the Li-Merz parameters for Ca²⁺.¹⁴⁹ AF488 and TRC2 dyes were described using the Amber GAFF2 force field, with RESP charges derived at the HF/6-31G(d) level of theory on geometries obtained using CAM-B3LYP/6-31G(d) including Grimme D3 dispersion correction.¹⁵² All systems were solvated in a 0.1 M KCl aqueous solution octahedral box, minimized and then thermalized from 0 K to 300 K during 250 ps NVT simulations followed by 250 ps equilibration in the NPT ensemble. Then, NVT production runs were extended for 5 μ s. This procedure was performed for a total of $3 \times 5 \mu$ s replicas for each system CaM-AF488-TRC2 and CaM-TRC2-AF488.

All simulations were performed using the Amber20 software¹⁵⁰ using periodic boundary conditions, the SHAKE algorithm to restrain bonds involving hydrogen, the particle-mesh Ewald approach to account for long-range electrostatics, a nonbonded cutoff equal to 10 Å, and the Hydrogen Mass Repartitioning scheme,¹⁵³ which allowed an integration time step of 4

fs. RMSD and Rg analysis were performed on trajectories with frames saved every 20 ps.

4.3.2 Electronic coupling calculations

We performed extensive calculations of TrESP-MMPol electronic couplings every 20 ps of the MD trajectories using the Trespcoup software,¹⁵⁴ leading to a total of 750,000 calculations for each system CaM-AF488-TRC2 and CaM-TRC2-AF488. In these calculations, the MMPol environment was described using the Amber pol12 AL polarizable force field,^{155,156} whereas the dyes were described using atom-centered TrESP charges obtained from a fit of the electrostatic potential obtained from TD-CAM-B3LYP/6-31G(d)/IEFPCM transition densities using the TraDA tool.¹⁵⁷

Alternatively, one can also train a regression approach from multiple QM/MMPol calculations able to derive the charges directly from the dye MD geometries.¹⁴⁰ Because here FRET rates are estimated through Eq. 2.35, in which couplings are effectively normalized by the transition dipoles of D and A, results are nevertheless not expected to be much sensitive to fluctuations in TrESP charges. In addition, TrESP-MMPol couplings were benchmarked with TD-DFT QM/MMPol calculations performed using a locally modified development version of the Gaussian package.¹⁵⁸

As mentioned before, the MMPol environment was described using the Amber pol12 AL polarizable force field and the dyes were described either using TD-DFT (QM/MMPol) or adopting a distributed model of atom-centered transition charges (TrESP-MMPol). We explored the performance of different DFT functionals to describe the couplings between the relevant π - π^* excited states of the dyes, including M06, B3LYP, CAM-B3LYP, PBE, PBE0, M06-2X. We selected CAM-B3LYP, which led to a simpler identification of the state of interest among different structures, although our results indicated that electronic couplings are rather robust regarding the choice of QM method and basis set, in line with previous studies.^{135,159}

After test calculations, we also defined the boundaries between MMPol and the QM regions in the bond between the linker and the phenyl ring bound to the chromophoric system of the dyes using the link atom scheme.¹⁶⁰ Atomic charges for water were taken from previous work,¹⁶¹ whereas polarization-consistent ESP charges for AF488 and TRC2 dyes were derived using the Polchat tool at the B3LYP/6-31G(d) level of theory.¹⁶² For K^+ and Ca^{2+} cations not defined in the pol12 AL polarizable force field,^{155,156} atomic polarizabilities were computed at the CCSD(T)/6-311++G(d,p) level of theory. On the other hand, TrESP charges corresponding to the transition between the ground and excited state of the dyes were derived from TD-CAM-B3LYP/6-31G(d) calculations in IEFPCM water solvent on the optimized geometries described in the previous section.

The electrostatic potential (ESP) from the transition densities was computed using Gaussian at points selected according to the Merz-Singh-Kollman scheme using 10 van der Waals layers and a density of 10 points per unit area, and were fitted to a set of atomic charges using the TraDA tool.¹⁵⁷ QM/MMPol calculations were performed using a locally modified development version of the Gaussian package,¹⁵⁸ while TrESP-MMPol calculations were run using the Trespcoup software.¹⁵⁴

The TrESP-MMPol model was initially benchmarked with full QM/MMPol calculations.¹³² Because screening effects markedly depend on the choice of MMPol cutoff radius, which defines the spatial extent of the polarizable MM region,¹³⁸ we first explored this dependence for 9 structures extracted from the trajectory of CaM-AF488-TRC2 spanning a range of interdye distances between 15-70 Å. In these calculations, explicit polarization was limited to residues within the MMPol cutoff radius from the QM region, whereas remaining residues up to a 35 Å were described through the additive Amber force field. As shown in Figures 1-A1, acceptable errors were obtained using a value ~15-20 Å for the MMPol cutoff radius, thus we fixed this value to 20 Å.

We then benchmarked TrESP-MMPol couplings with QM/MMPol TD-CAM-B3LYP/6-31G(d) couplings for an extended set of 100 structures extracted from the MD trajectories. TrESP-MMPol couplings were computed adopting a MMPol cutoff radius equal to 10, 15 or 20 Å. The results, shown in Figure 2-A1, indicate that a value of 15 Å provides accurate results while keeping computational cost low, so this value was adopted for all further TrESP-MMPol calculations performed every 20 ps on the 3×5 ms MD trajectories obtained for each system: CaM-AF488-TRC2 and CaM-TRC2-AF488.

FRET rates defined in Eq. 2.36 and 2.37 were obtained using a refractive index $n^2 = 2$,¹⁶³ a critical Förster radius $R_0 = 54$ Å¹⁴³ and a fluorescence lifetime for AF488 $\tau_D = 4$ ns.¹⁶⁴ Transition dipole moments for the AF488 donor and TRC2 acceptor equal to $\mu_D = 7.56 D$ and $\mu_A = 10.19 D$ were extracted from the TD-DFT calculations used to parametrize the TrESP charges. The center-to-center distance between dyes $R(t)$ and the $\kappa(t)$ orientation factors were computed from the structures sampled along the MD trajectories. Finally, to dissect the impact of Förster approximations, we defined four coupling models (PDA, PDA-MMPol, TrESP, TrESP-MMPol) using the general expression $V = s \cdot V_{Coul}$, which combine a Coulomb term computed using the TrESP or the PDA approximation with a screening factor given by Förster $1/n^2$ value or derived using TrESP/MMPol calculations using the expressions reported on section 2.2.6.3.

4.4 Conformational ensembles of CaM

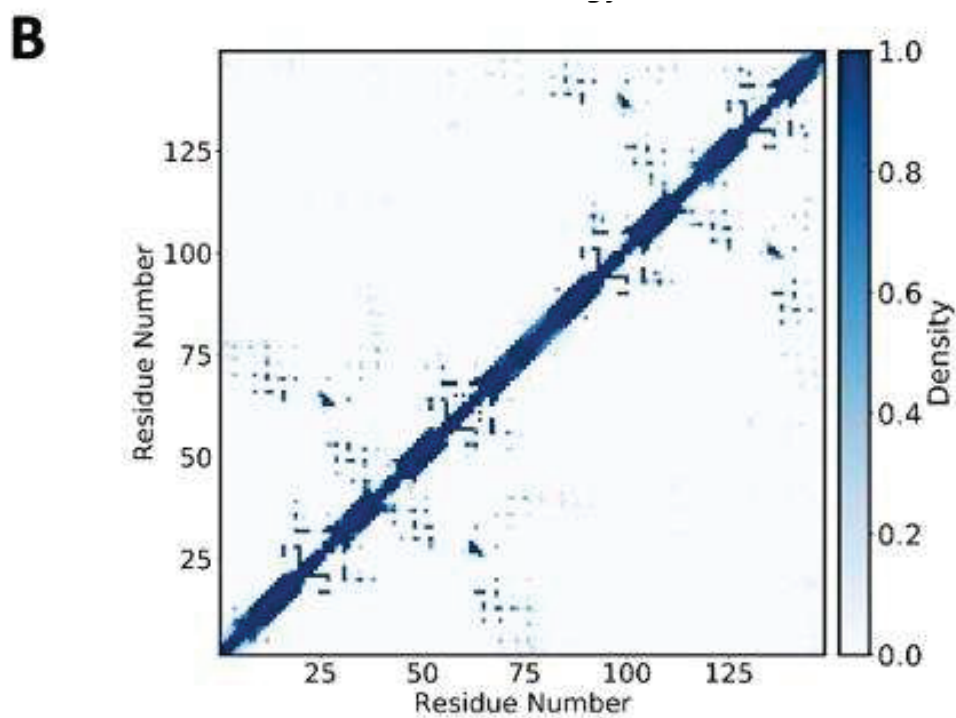
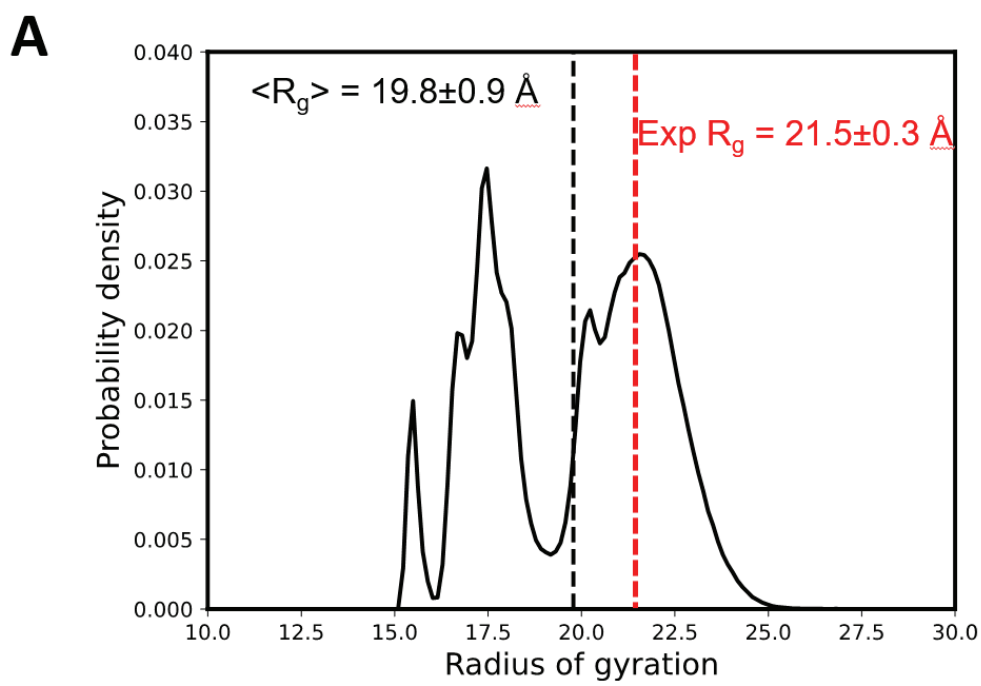
The structure of CaM is characterized by two lobes connected by a linker, each domain containing two E-F hands. Ca^{2+} binding to the EF hands then leads to the exposure of the lobe's hydrophobic interfaces, which allows CaM to bind a variety of protein targets in a Ca^{2+} dependent manner. The flexible linker allows binding targets of different size anchored to the N and C-terminal domains, but the lobes are also flexible, and both structural data and MD simulations suggest that the C-lobe is more flexible and provides more selectivity.¹⁶⁵ In Fig. 4.1, we show an illustration of the MD ensembles we obtained for *holo* CaM-TRC2-AF488 and CaM-AF488-TRC2, which correspond to a total of 15 μs per system. For clarity, we only display the superposition of a subset of the structures sampled, but this suffices to illustrate the conformational heterogeneity of CaM, which leads the dyes to explore a rich variety of relative arrangements, in line with previous MD simulations^{165–170} and FRET studies that show distinct configurations of the N- and C-lobes about the central linker, which is crucial for target recognition and binding.^{143,164,171}

The structural heterogeneity underlying the MD ensembles lead to a distribution of radius of gyration with values from 15 to 25 Å, as shown in Fig. 4.2(A), leading to an average value $\langle R_g \rangle$ equal to 19.8 Å, which slightly underestimates the experimental R_g 21.5 Å.¹⁷² This could arise from an overpopulation of compact structures, an ongoing problem in classical force fields, but is also affected by the limited sampling in our simulations.

To get deeper insights into the conformational ensemble, we performed a clustering analysis to determine the four more populated structures, with the aim of relating them to the four main FRET states identified in experimental lifetime distributions by DeVore and co-workers.¹⁴³ In Fig. 4.2(C) we display representative structures for the clusters, along with their R_g values, which allows to characterize the structures underlying the R_g distribution in Fig.

4.2(A). The most populated Cluster 1 (51%) has an $\langle R_g \rangle$ value of 21.2 Å, thus it accounts for the peak centered around the experimental value in Fig. 4.2(A), together with a minor contribution from Cluster 4 (12%). In these clusters, the linker keeps the two CaM domains spatially separated, resembling the main conformation of *holo* CaM displayed in crystal structures, although in Cluster 4 the linker helix is bent. In contrast, Clusters 2 and 3, which amount to a total 37% population, display $\langle R_g \rangle$ values of 16.9 Å and 17.7 Å. Thus, these subpopulations account for the band centered at ~17.5 Å and the minor peak at ~15.5 Å in Fig. 4.2(A) and display a structural collapse of the extended crystal arrangement leading to compact shapes like those observed in complexes of CaM with peptides.¹⁶⁸

Because in this study we aim at characterizing CaM states observed through FRET, it is interesting to examine the D/A separations in the centroids of the clusters. We estimate this separation using the C34-C110 C α -C α distance (the Cys residues linking the AF488 and TRC2 dyes), as the specific position of the dyes among members of the cluster can vary considerably. The resulting distances in the extended Clusters 1 and 4 are similar to the more compact Cluster 3, with values ~49 Å, suggesting similar FRET data can be obtained for them despite remarkable differences in terms of overall CaM structure. In contrast, Cluster 2 adopts a different mutual arrangement of the N and C-lobes, leading to a closer distance of 16.8 Å between Cys34 and Cys110 which allows the dyes to get in contact. We expect these conformations to explain the states characterized by the largest FRET efficiencies and shortest fluorescence lifetimes, as will be discussed in the next sections.



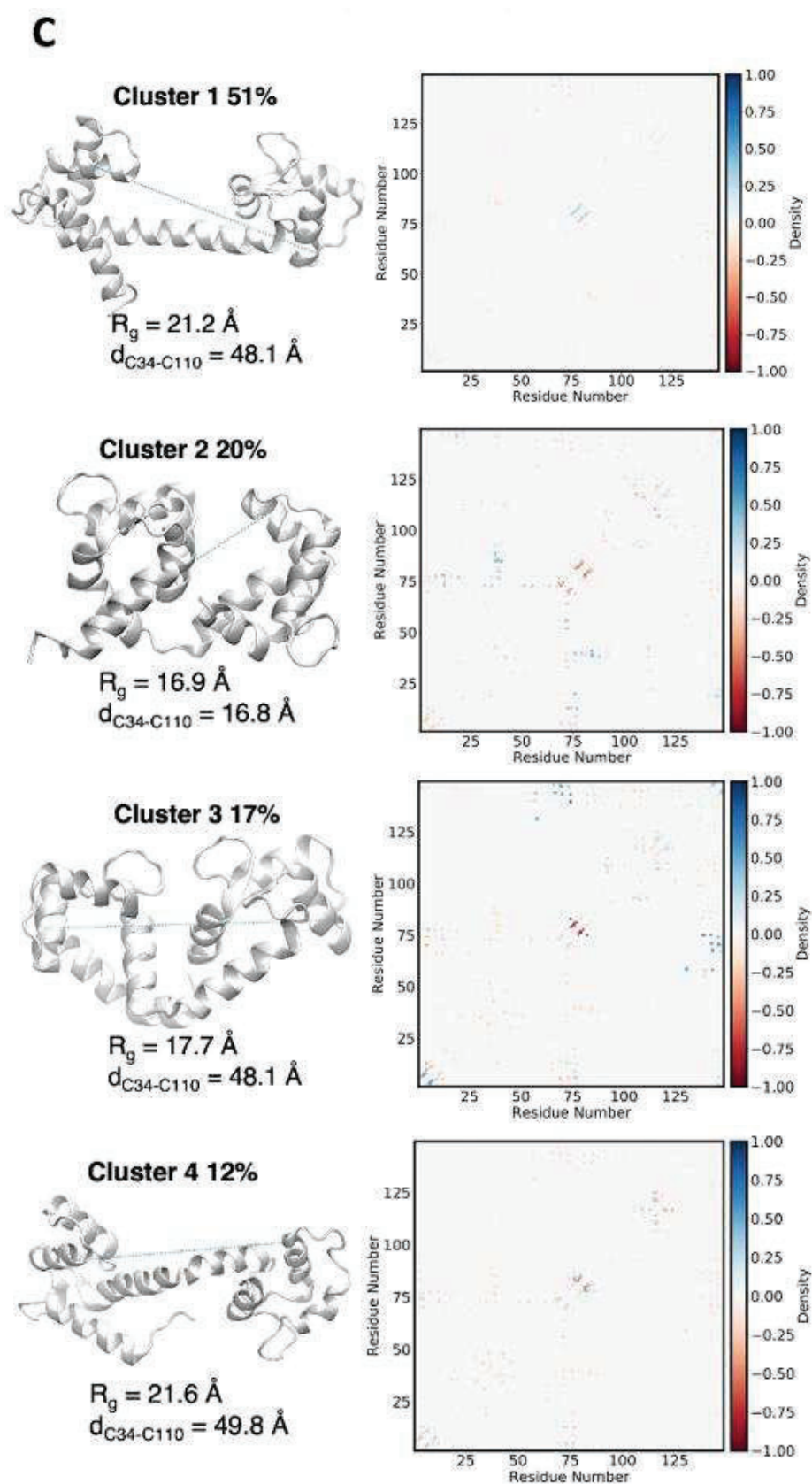


Figure 4.2. A Distribution of radius of gyration computed along the MD trajectories of both *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 compared to the experimental value.¹⁷² Note that the experiments contain a mixture of both D-A and A-D systems due to unselective dye labeling, so the overall simulated distribution is shown. Standard error of the mean (SEM) for theoretical values obtained from the mean values of the six 5 μ s replicas. **B Intramolecular contact map** averaged over the MD showing the folding patterns of the two domains in CaM. **C Major clusters for the structural ensemble of CaM.** For each cluster we draw the most representative structure (centroid) and report its population, the MD-averaged radius of gyration and the intramolecular contact map relative to the ensemble-averaged contact map shown in panel B. Blue and red contacts indicate those populated more and less frequently than the average, respectively. The C34-C110 $\text{C}\alpha$ - $\text{C}\alpha$ distances between Cys residues linking the AF488 and TRC2 dyes in the centroid structures are also reported and represented by dashed lines.

4.5 Beyond Förster coupling: dielectric screening effects

Beyond the dipole approximation, the R^{-6} dependence of FRET rates also relies on the validity of the $1/n^2$ screening prefactor in Eq. 1.10, which describes how the interaction between D/A dipoles is screened by the polarizable environment. As discussed in the introduction, the accurate yet efficient formulation of the TrESP-MMPol model allows us to investigate the impact of dielectric screening over long timescale MD trajectories in atomic detail. In Fig. 4.3, we plot the distribution of environment screening factors computed according to Eq. 2.45 from these atomistic calculations.

Our results display a strong exponential attenuation of screening effects at D/A separations below 20 Å, reaching a value around ~ 0.6 , which increases to ~ 0.7 beyond 30 Å. The dependence of dielectric screening on D/A separation has been previously examined in studies of photosynthetic pigment-protein complexes^{136,137,139,173} and to analyze single-molecule experiments on a polyproline helix.¹⁷⁴ Although those studies were based on continuum solvent formulations of the D/A environment in the calculation of the coupling, quite strikingly the initial exponential decay until ~ 10 Å observed in our atomistic simulations strongly resembles the behavior in those studies, suggesting a universal trend. This is somewhat unexpected, as here we deal with different dye molecules, the dyes are significantly more solvent-exposed than the pigments in the photosynthetic complexes, and we adopt an atomistic description of the environment based on a polarizable force field instead of a continuum model, which is able to account for heterogeneities in the protein + solvent dyes environment. Our findings thus reinforce the notion that distance-dependent dielectric screening alters the R^{-6} distance dependence of FRET (Eq. 1.8 and 1.10).

We then fitted our results to an exponentially decaying empirical screening function that other researchers can use to account for dielectric screening,

albeit in an approximate way. In analogy with previous studies, we first fitted our results to the following expression:^{136,137}

$$s = 30.0e^{-0.61R} + 0.70 \quad (4.2)$$

Where R is the D/A separation in Å. This expression indicates negligible screening effects ($s = 1$) at ~ 7.5 Å, as there is little or no intervening solvent between the dyes, a value similar to previous estimates.¹³⁷

Recently, however, Eder and Renger applied extensive Poisson-TrEsp calculations to study photosystem I, concluding that exponential decay only contributes to the screening in chlorophyll pairs with in-line dipole arrangements. Thus, they proposed a modified version of Eq. 4.2 in which the $Ae^{-\beta R}$ term only survives for geometries with a κ dipole orientation factor above a certain threshold.¹³⁹ We then examined the dependence of screening factors on κ in our simulations on CaM, and compared exponential decaying functions fitted for specific subsets of data characterized by different ranges of orientation factors (Group 1 $0.0 \leq \kappa \leq 0.6$); Group 2 $0.6 \leq \kappa \leq 1.2$; Group 3 $1.2 \leq \kappa \leq 2.0$). Although the results (see Appendix 1 Fig. 3-A1) suggest a somewhat steeper exponential decay for in-line geometries (Group 3), all subsets clearly show the exponential decay, so we choose not to include any explicit dependence of the empirical screening function on κ .

Beyond the exponential decay, however, our data indicates that screening is maximized at separations ~ 15 Å, where s values adopt a minimum ~ 0.6 , which afterwards increases to ~ 0.7 at distances beyond ~ 30 Å. We ascribe this trend to the decreased polarizability of water compared to a protein environment, given that at distances >30 Å the dyes are more exposed to the water solvent, as exemplified by the structures shown in Fig. 4.4. In contrast, in more compact conformations in the 15-25 Å range, the interaction of the dyes with the protein becomes more important.

In the figure 4-A1 we report the contribution of the dyes to the total solvent-accessible surface area (SASA) of the system, which confirms this increased solvent exposure beyond 30 Å. In Fig. 4.3 we then indicate as dashed lines the Onsager screening factor estimated for protein and water environments (ϵ_{opt} 2.0 and 1.776, respectively), which illustrate the fact that a smaller screening is expected in water. It is also worth noting that, although our data tends to the Onsager factor at long separations, there is a strong deviation from Förster's $1/\epsilon_{opt}$ assumption at all distance ranges.

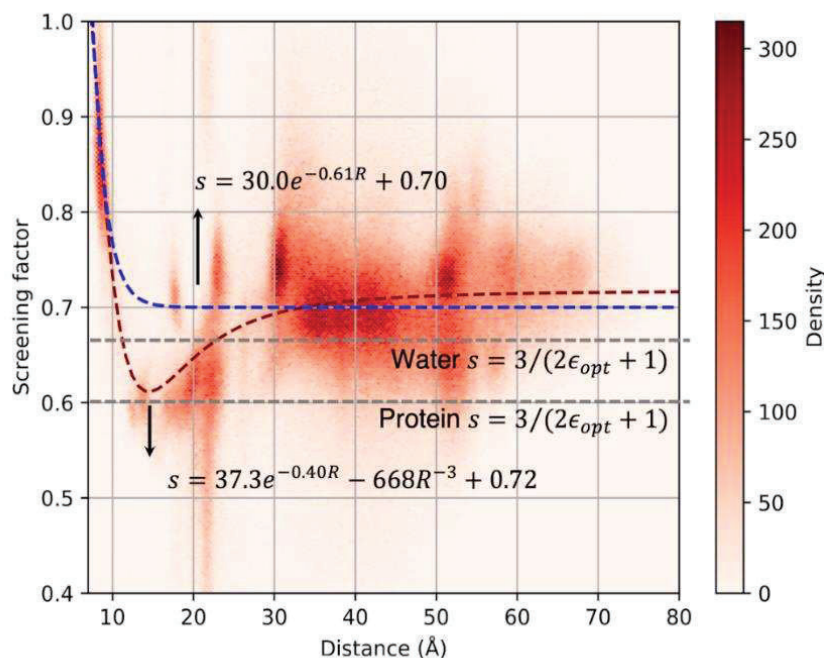


Figure 4.3. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems along MD trajectories. Empirically fitted functions are compared to estimates from Onsager model $s = 3/(2\epsilon_{opt} + 1)$ using ϵ_{opt} values of 2.0 and 1.776 for protein and water environments, respectively.

To account for the minimum at ~ 15 Å, in which the fluorophores are partially excluded from the solvent and thus screening is dictated by the more polarizable protein environment, we decided to add an R^{-3} term, leading to this modified fitted empirical function:

$$s = 37.3e^{-0.40R} - 668R^{-3} + 0.72 \quad (4.2)$$

We recall nevertheless the approximate nature of this empirical expression, as screening effects also depend on the local environment and D/A orientation, as visible in the distribution of values in Fig. 4.3 that deviate from the fitted curve.

To illustrate the origin of the screening distance dependence, in Fig. 4.4 we graphically represent the environment-mediated contributions estimated using the atomistic TrESP-MMPol model for two limiting cases at short and long separations ($R = 8.4$ Å and $s = 0.92$; $R = 64$ Å and $s = 0.71$). In the top picture with $R = 8.4$ Å, screening is largely attenuated, as blue regions that contribute to screen the interaction, located surrounding the dyes, are counterbalanced by other surrounding red residues that significantly enhance the interaction. For example, Asn111, which contributes with -13 cm $^{-1}$ to the MMPol environment-mediated coupling term V_{env} . In contrast, in the $R = 64$ Å case, blue residues contributing to the screening of the interaction are located in between the AF488 and TRC2 dyes, including residues Arg106, Lys115 and Leu116 near TRC2 and a water molecule next to AF488. In contrast, red residues enhancing the interaction like Asn111, Leu112 and Gly113 show in this case smaller contributions, thus leading to an overall significant screening effect.

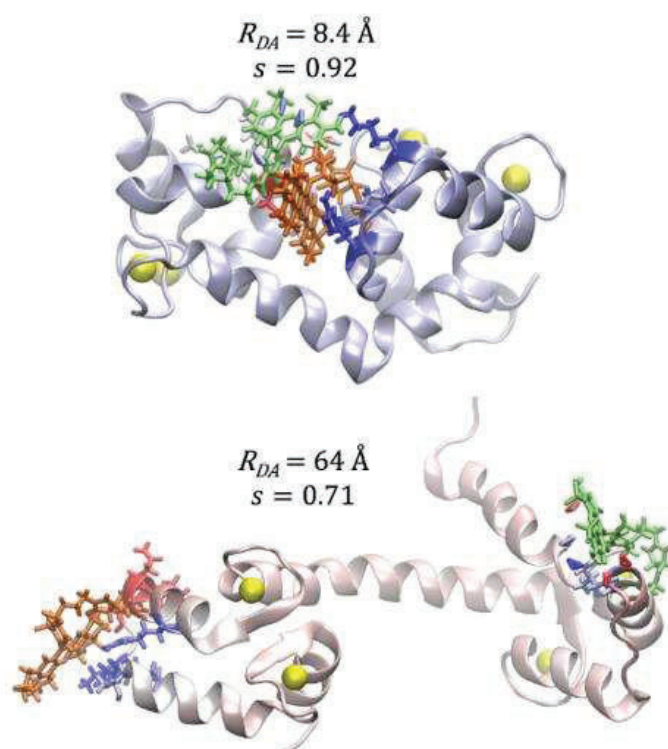


Figure 4.4: Graphical representation of water and amino acid (residue) contributions to the MMPol environment-mediated term (Eq. 2.26) in the electronic coupling between AF488 (lime) and TRC2 (orange) for two limiting cases at short and long D/A separations. Blue colored residues contribute to screen the interactions, whereas red ones reinforce the Coulomb interaction between D and A.

4.6 Beyond Förster coupling: Deviations from the ideal isotropic distribution

In cases where the specific relative orientation between D and A is unknown, it is common to assume an isotropic distribution, in which the dyes randomly adopt all possible mutual orientations, leading to $\langle \kappa^2 \rangle = 2/3$. This allows to simplify Eq. 1.8 and derive Eqs. 1.9 and 1.10, which allows to extract D/A distances directly from FRET efficiencies, an approximation that typically performs well for flexible linkers. The atomistic detail of MD simulations, however, allows direct calculation of the orientation factors and also tackling eventual correlations between κ and R dictated by dye and protein dynamics.¹⁷⁵ Thus, to examine this further assumption in Förster description of the coupling, in Fig. 4.5 we report the distribution of orientation factors sampled along the MD trajectories, which turns to be very close to the random isotropic distribution, with an average $\langle \kappa^2 \rangle = 0.62$ close to the isotropic limit. As shown in Fig. 4.6, however, at D/A separations below 20 Å we find significant deviations from the isotropic average value. This is probably due to the more compact structures adopted by CaM in this separation range, which restrict the orientational freedom of the dyes.

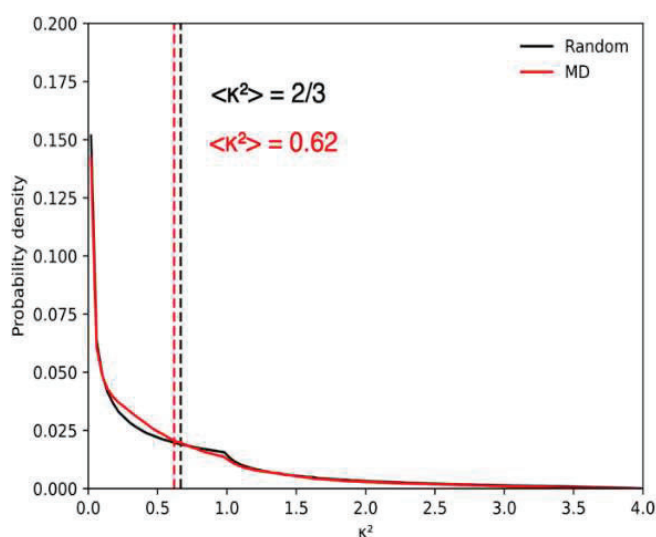


Figure 4.5. Probability distribution of orientation factors derived from MD simulations for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems compared to the random isotropic distribution.

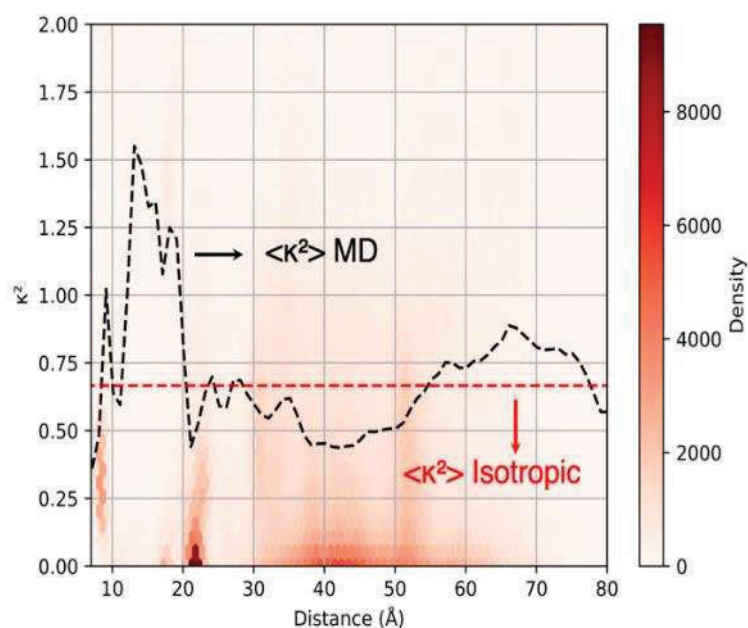


Figure 4.6: Density distribution of orientation factors as a function of D/A separation (average values over distance bins indicated by black dashed curve).

To examine the joint impact that the deviations (screening effect, dipole approximation, isotropic assumption) introduce, we show a comparison of the distribution of distances directly measured from the MD simulation and those derived using Eq. 2.43 from the distribution of TrESP-MMPol FRET rates in Fig 4.7,¹⁷⁶ the latter rates thus emulating eventual experimental FRET efficiencies. Compared to the actual MD data, the distribution estimated from FRET efficiencies assuming Förster approximations displays a global shift to lower D/A separations. This probably arises from the fact that, for a given estimate of FRET efficiency, Förster theory tends to overestimate screening effects, thus leading to a concomitant underestimation of the true D/A distance. Indeed, the average D/A separation derived from FRET efficiencies is 38.6 Å, which is 2.7 Å lower than the value 41.3 Å directly measured along the MD. A more complex picture however emerges at separations below 30 Å, and here the distances derived from FRET are in contrast slightly overestimated. This is because in this region all deviations from Förster theory come into play, including the attenuation of screening at

close separations, incomplete isotropic averaging and significant breakdown of the dipole approximation.

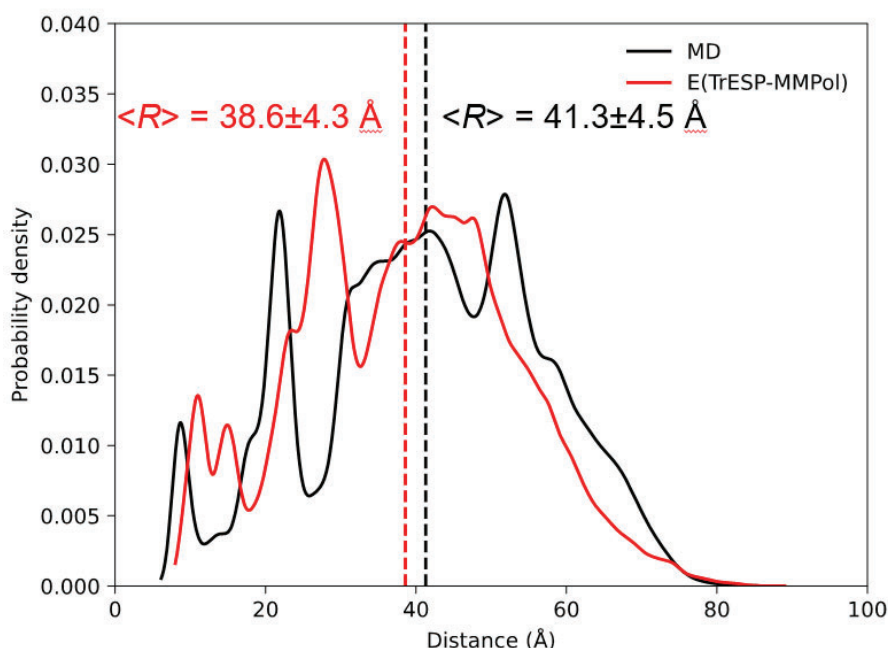


Figure 4.7: Distribution of donor-acceptor distances directly measured along MD trajectories estimated using Eq. 2.45 from TrESP-MMPol FRET efficiencies. Standard error of the mean (SEM) for theoretical values obtained from the mean values of the six 5 μ s replicas.

Regarding the dipole approximation, a rule of thumb is to consider that it will breakdown at separations close to the dye's dimensions. In Fig. 4.8, we compare the Coulomb coupling contribution obtained using point dipoles (PDA) or adopting atomic transition charges (TrESP). Our results confirm that more significant deviations are found at close separations, below 30 Å. According to average values of the ratios indicated by the dashed line, overall, the PDA tends to slightly overestimate the Coulomb term for all range of D/A separations. But this deviation is largely increased at distances below 30 Å, where the PDA is expected to introduce more deviations in FRET-derived structural data.

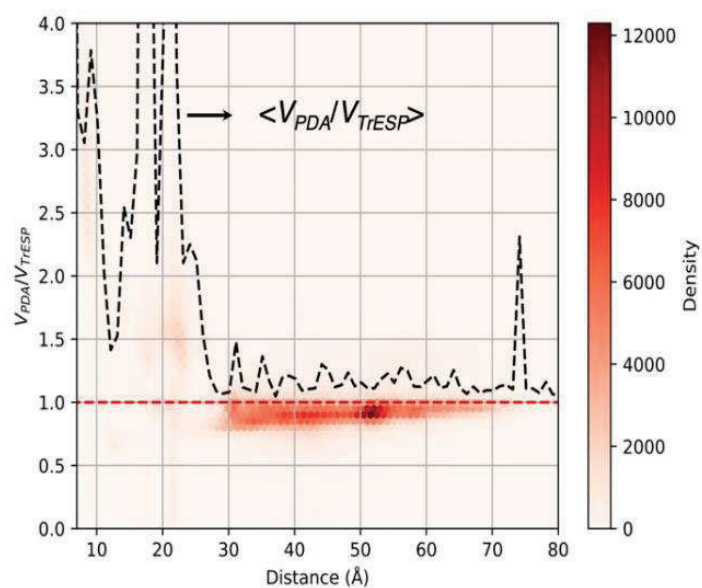


Figure 4.8: Ratio between PDA and TrESP coulombic coupling contributions (average values over distance bins indicated by black dashed curve).

4.7 Fluorescence Lifetime and Efficiency Distributions

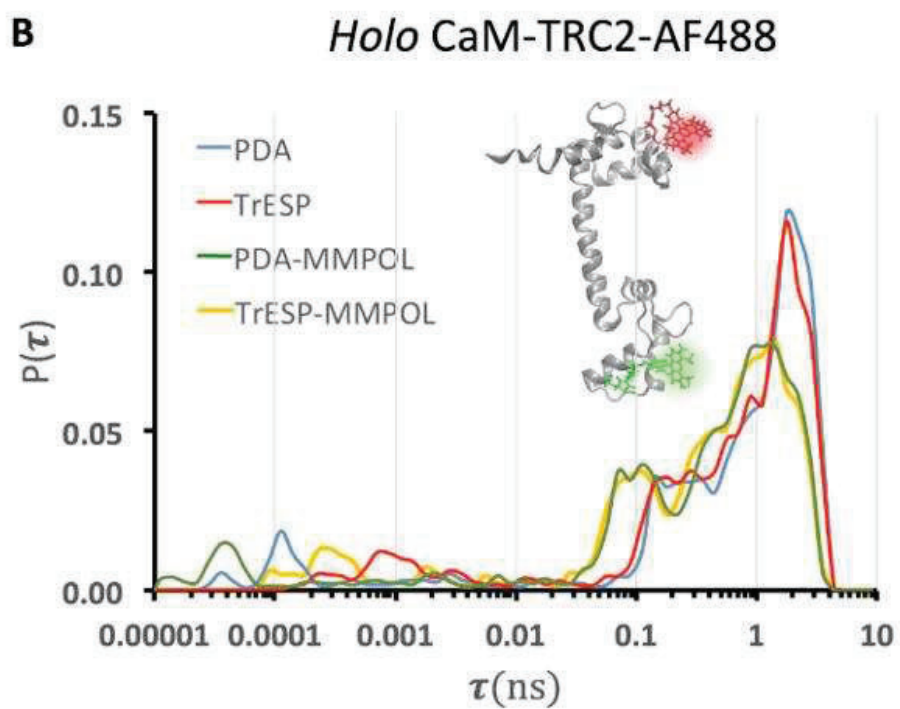
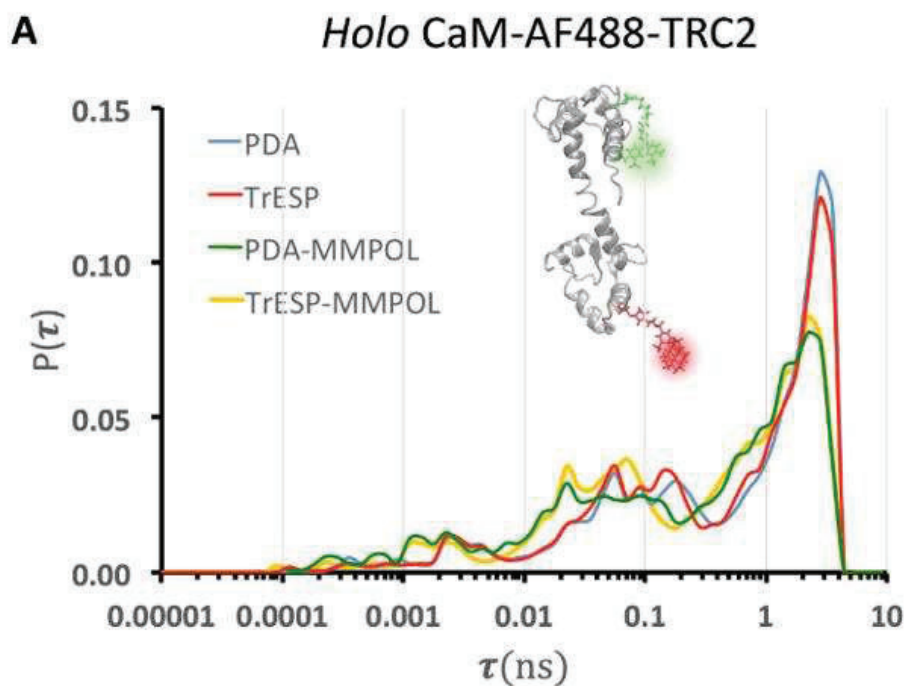
In the previous sections, we have shown how screening effects impact FRET couplings at all distance ranges, whereas the dipole approximation and the isotropic assumption lead to significant deviations only at short D/A separations. Here, we examine how all these effects impact the analysis of FRET data. We first analyzed a comparison of the distribution of distances directly measured from the MD simulation and those estimated using Eq. 2.42 from the distribution of TrESP-MMPol FRET efficiencies. The estimated distances display a global shift to lower D/A separations compared to the actual MD data from the overestimation of screening effects in Förster theory, leading to an underestimation of the D/A distances.

On the other hand, it is also interesting to investigate how Förster approximations impact the simulation of FRET distributions. We focus on FRET efficiency histograms, but also on fluorescence lifetime distributions, as they also reflect the underlying FRET distribution because energy transfer decreases the donor fluorescence lifetime. We then compare both distributions to the experimental data measured from donor fluorescence decays and smFRET.¹⁴³

The results in Fig. 4.9(A) and 4.9(B) allow to gauge the performance of different electronic coupling models in describing the lifetime distributions, which were generated using the expression in Eq. 2.41. To dissect the impact of errors due to dielectric screening or the PDA approximation, we compare several models of increasing accuracy (PDA, TrESP, PDA-MMPol and TrESP-MMPol), (Eq. 2.45-Eq-2.49) where PDA/TrESP indicates the method used to compute the Coulomb term, whereas MMPol indicates that atomistic screening factors obtained from TrESP-MMPol calculations are used. Thus, how breakdown of the dipole approximation impacts results can be assessed by comparing the PDA and TrESP curves, or alternatively the PDA-MMPol and TrESP-MMPol curves. The results in Fig. 4.9(A) and 4.9(B) indicate that

adoption of the PDA leads to small changes in the theoretical lifetime distributions, although some significant deviations are observed in the sub-ps lifetime region of CaM-TRC2-AF488, where close donor/acceptor separations are expected to worsen the performance of the PDA. Thus, even though significant deviations were observed in the ratios V_{PDA}/V_{TrESP} shown in Fig. 4.8, the actual outcome of these deviations in terms of lifetime distributions is rather small.

In stark contrast, the adoption of Förster screening factor $s = 1/\epsilon_{opt}$ significantly shifts the complete lifetime distributions compared to results that incorporate atomistic MMPol screening effects estimated from Eq. 2.26. Overall, Förster model overestimates screening effects, leading to underestimated FRET rates and efficiencies, and longer lifetimes. We can observe this trend when comparing PDA and TrESP-MMPol lifetimes in Fig. 4.9(C), or by comparing the efficiency distributions plotted in Fig. 4.9(D). In both cases, the incorporation of atomistic screening effects improves the agreement with experimental data. The average efficiency obtained with TrESP-MMPol ($E = 0.79$) is closer to the experimental distribution, which displays a maximum near ~ 0.8 .^{143,176}



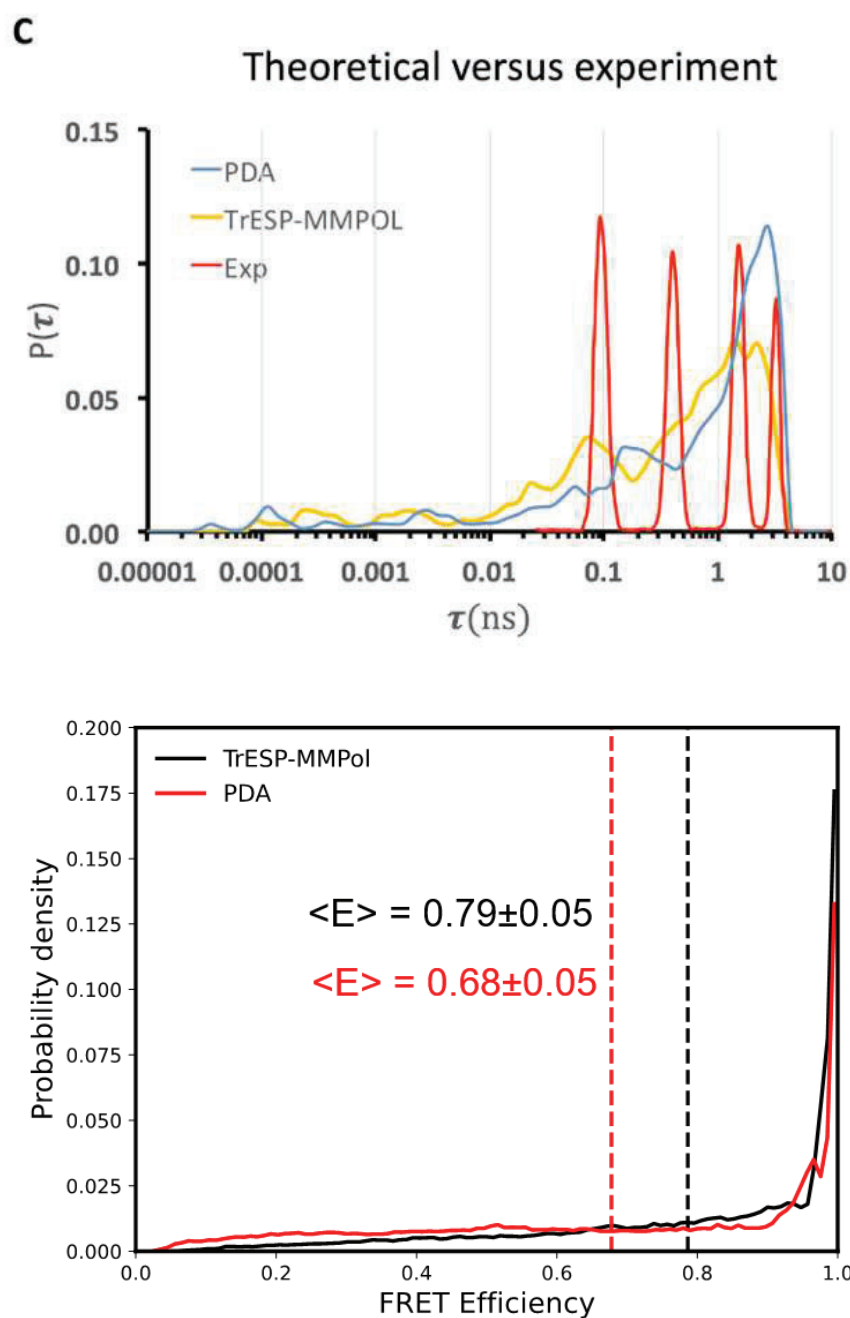


Figure 4.9: Distribution of fluorescence lifetimes computed for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems using different electronic coupling models (PDA: Point dipole approximation + Förster screening factor. PDA-MMPol: Point dipole approximation + MMPol screening factor. TrESP: TrESP coulombic term + Förster screening factor. TrESP-MMPol: TrESP coulombic term + MMPol screening factor): **A** *holo* CaM-AF488-TRC2, **B** *holo* CaM-TRC2-AF488 and **C** *Holo* CaM-AF488-TRC2 + CaM-TRC2-AF488 versus experiment.¹⁴³ Note that experiments contain a mixture of both D-A and A-D systems due to unselective dye labelling. **D** Distribution of FRET efficiencies computed for the two CaM systems using the PDA and the TrESP-MMPol model. Standard error of the mean from the six 5 μ s replicas.

In addition, in Fig. 4.9(C), the relative intensity of the main FRET states suggested by the experimental lifetime distributions improves when MMPol effects are included, as FRET states with lifetimes near 0.1 ns and 0.4 ns increase their probability. This comparison, however, must be done with care, given that the experimental distribution is derived from a fitting of fluorescence decays using classic maximum entropy (cMEM) analysis. Because only 4 FRET populations were recovered in these fits, the experimental distribution only displays these peaks, but this probably does not reflect the complete distribution and misses intermediate states. Beyond the overall agreement of simulations in terms of average efficiency and shape of the lifetime distribution in the 0.1-4 ns range, Fig. 4.9(C) gives the impression that our simulations describe a set of rather compact conformations not observed experimentally. Significant populations are observed of efficient FRET states with lifetimes <10 ps, whereas experimentally the fastest component of the decay was fitted to a lifetime of 100 ps. This disagreement however could arise from the fact that fluorescence decays were measured with an instrument response function (IRF) with a full width at half-maximum <50 ps, so it is unlikely to observe those fast components in this experimental set-up. Comparison of simulated and experimental decays (Fig. 4.10) nevertheless clearly indicates that the overall MD ensemble of CaM is too compact, leading to a decay slightly faster than in experiments.

However, this problem also arises, for example, from the underestimation of populations with lifetimes ~3-4 ns, as shown in Fig. 4.9(C), and not only from the states with lifetimes <10 ps.

Current additive force fields widely used for the simulation of proteins tend to introduce a bias that favors folded states, thus breaking the balance between compact and more extended or disordered states, and this could still be happening here for CaM. Because of this, important efforts are being carried out to refine protein force fields for disordered systems.^{177,178} Our

findings are an illustrative example of the potential of FRET and MD to refine current force fields, as simulations provide rather accurate ensemble properties, like FRET efficiencies, but the sub-ensemble lifetime distributions indicate the presence of a set of compact states that could be unrealistic, although more accurate experiments would be needed to confirm this hypothesis, as discussed above. Interestingly, such compact states are however in agreement with the hypothesis put forward by DeVore and co-workers, suggesting that the TRC2 dye, in contrast to other acceptors, may stabilize a compact conformation of CaM by conformational selection.¹⁴³ Nevertheless, our results indicate that it becomes particularly important to account for atomistic screening effects if FRET is used to examine the balance between extended and folded conformations, given the strong attenuation of screening effects at close separations. Neglecting this distance-dependence contribution in the PDA model artificially reduces the coupling found for compact structures, in this way counteracting their overestimated interaction and masking eventual disagreements between simulations and experiment.

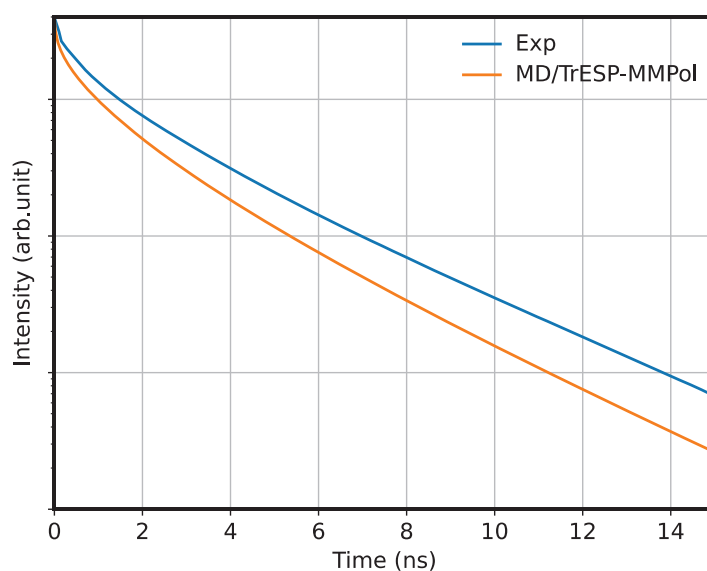


Figure 4.10: Fluorescence decay from MD/TrESP-MMPol data calculated for the two CaM systems compared to experiment

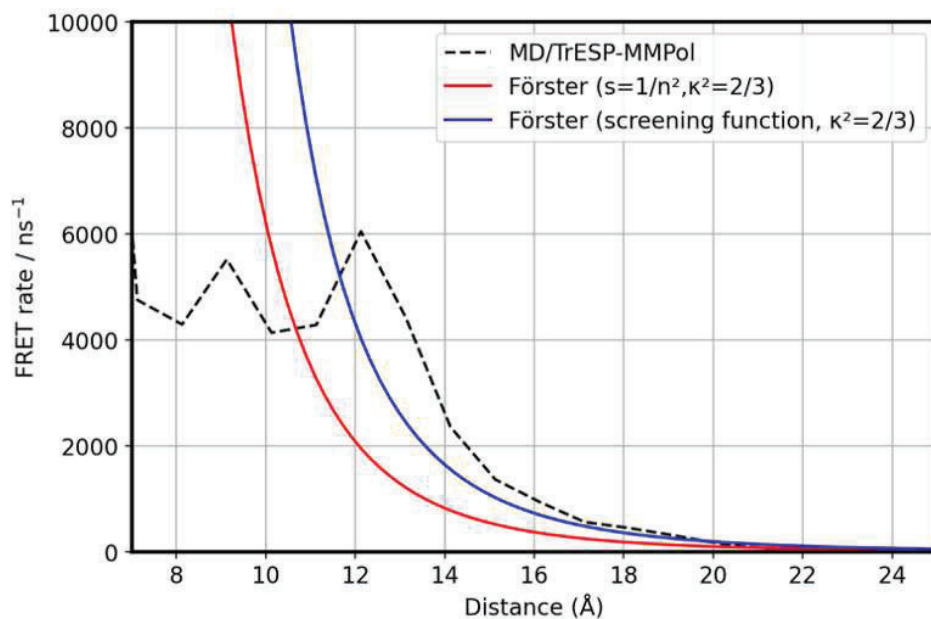


Figure 4.11: FRET rates averaged over MD/TrESP-MMPol data for the two CaM systems as a function of dye separation compared to rates computed using Förster theory either using the screening factor or adopting the empirical screening formula $s = 30.0e^{-0.61R} + 0.70$ derived from MD/TrESP-MMPol calculations.

Fig. 4.11 also helps understanding why Förster expression tends to underestimate D/A distances for a given FRET rate, except at short separations, where other effects beyond screening come into play. The figure illustrates how dielectric screening modifies the R^{-6} distance dependence of FRET by comparing rates obtained using Förster theory using the screening factor or the empirical screening function fitted here (Eq.4.2), in both cases assuming an isotropic orientation factor $\kappa^2 = 2/3$. Not only the distance dependence is modified, but FRET rates are significantly underestimated at all distances ranges, given the strong deviation of screening factors from Förster's assumption at all distance ranges, as shown in Fig. 4.3. The adoption of the empirical screening function together with the isotropic κ^2 , albeit approximate, allows to predict rates from D/A distances that are in much better agreement with the average values estimated from atomistic

MD/TrESP-MMPol calculations. As expected, however, this agreement starts to degrade at distances below ~ 16 Å, due to breakdown of the dipole approximation and increasing deviations from the $\kappa^2 = 2/3$ limit observed in Fig. 4.6.

CHAPTER 5: Impact of calcium binding and dye labelling in the structural ensemble of calmodulin studied using atomistic energy transfer simulations

5.1 Introduction

The content of the following chapter presents the results associated with the second manuscript derived from this thesis, which is currently in preparation and will be submitted for publication soon. Calmodulin plays a major role in the transmission of calcium signals in eukaryotes. This is achieved by a subtle control of the conformation of its two lobes, each containing two E-F hands, connected by a linker. Ca^{2+} binding leads to exposure of the lobe's hydrophobic interfaces, and this allows Ca^{2+} dependent binding to a variety of targets, whereas the flexible linker allows to anchor targets of different size to the N and C-terminal domains. Molecular dynamics simulations and FRET studies have shown that distinct configurations of the N- and C-lobes about the central linker, which is crucial for target recognition and binding.

Here, as in the previous chapter 4, we address this challenge by using a rigorous theoretical framework based on electrostatic potential-fitted transition charges coupled to an atomistic polarizable classical environment. In previous chapter, we show the importance of distance-dependent dielectric screening effects to simulate FRET data from MD ensembles of *holo* calmodulin tagged with Alexa Fluor 488 and Texas Red dyes derived using the ff14SB protein force field. We extend that study by exploring the impact of Ca^{2+} binding, dye labelling and force field used, by studying *holo* and *apo* calmodulin tagged with both Alexa Fluor 488 — Texas Red C2 and Alexa Fluor 488 — Alexa Fluor 594 dye pairs. We compare results with the Amber ff14SB protein force field with a99SB-disp, a force field specifically optimized for disordered proteins. In addition, we study how deviations from Förster energy transfer theory impact the assessment of MD ensembles with FRET data. We use a rigorous theoretical framework that overcomes Förster limits adopting a TrESP-MMPol approach, which describes the fluorophores

through electrostatic potential-fitted transition charges coupled to a polarizable molecular mechanics description of the environment,¹⁴⁰ thereby allowing to tackle large amounts of structures extracted from long time scale MD simulations, which keeping the accuracy of more costly polarizable QM-MMPol calculations.

An ongoing question in FRET studies regards the impact of fluorophore labelling in the conformational preferences of the protein. DeVore and co-workers¹⁴³ investigated the role of conformational averaging on different measurement time scales (nanosecond or microsecond) and the nature of the dyes in the conformational states of CaM studied using single molecule FRET and time-resolved fluorescence spectroscopy. Interestingly, they found that smFRET distributions measured for *holo* CaM are significantly different upon incorporation of different FRET pairs, indicating that CaM states and the dynamics of conformational interchange are affected by the presence of the dye Texas Red compared to other acceptors like Alexa Fluor 594.

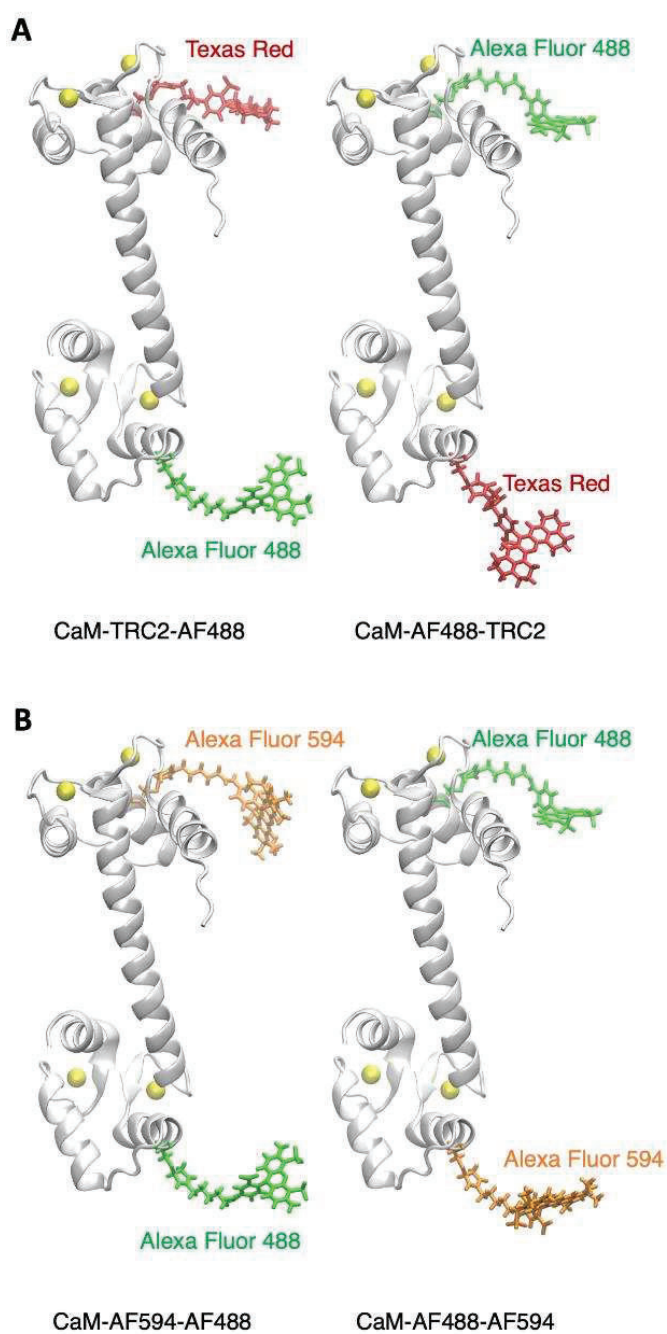


Figure 5.1. A Structure of *holo* Calmodulin tagged with the Alexa Fluor 488 C5 maleimide and the Texas Red C2 maleimide dyes. B Structure of *holo* Calmodulin tagged with the Alexa Fluor 488 C5 maleimide and the Alexa Fluor 594 C5 maleimide dyes. For both systems, we considered *apo* and *holo* CaM tagged with D/A or A/D dyes at Cys34 and Cys110 positions.

5.2 Methods: Multiscale MD/TrESP-MMPol approach

As in the previous chapter 4, we adopt a multiscale pipeline based on performing classical MD simulations and postprocessing the trajectories with extensive calculations of FRET electronic couplings based on the TrESP/MMPol method.¹⁴⁰ Compared to using other multiscale approaches like QM/IEFPCM, based on a continuum solvation model, this strategy allows to account for heterogeneities in the polarizability of the environment in the calculation of FRET couplings by using an atomistic polarizable description of the protein-water environment. On the other hand, it allows to estimate couplings with similar quality compared to the QM/MMPol approach at a much cheaper cost by using TrESP charges parametrized from QM calculations, which avoids the need to perform QM/MM calculations on many structures. In the following we describe the expressions adopted to estimate FRET properties and couplings using this protocol, as well as the details of the MD and TrESP-MMPol calculations performed.

The FRET observables generation for the results follow the expression and equations derivations explained in the section 2.2.6.

5.3 Computational details

5.3.1 MD simulations

Simulations were performed starting from the structure of human *holo* CaM solved at 1.7 Å resolution (PDB ID 1CLL).¹⁴⁴ We considered both wild type (WT) *apo* CaM and *holo* CaM (with 4 bound Ca^{2+} ions), as well as *apo* and *holo* proteins with mutations T34C and T110C labelled at these positions with the donor Alexa Fluor 488 C5 maleimide and two different acceptors, Alexa Fluor 594 C5 maleimide and Texas Red C2 maleimide, as shown in Fig. 5.1. For each dye-labelled system, we considered donor/acceptor dyes placed on 34/110 or 110/34 positions to reproduce the experimental mixture. Missing residues and the C- and N-terminal ends were added manually, and all amino acids were considered in their standard protonation state as predicted by PROPKA3 calculations.¹⁵¹

Simulations were based on the Amber ff14SB protein force field,¹⁴⁶ the TIP3P water model,¹⁴⁷ the Joung-Cheatham parameters for monovalent ions¹⁴⁸ and the Li-Merz parameters for Ca^{2+} .¹⁴⁹ Then, the same simulations were also performed using the a99SB-disp force field optimized for disordered proteins,¹⁰¹ using the a99SB-disp water model based on modified TIP4P-D parameters, the Joung-Cheatham parameters for monovalent ions and the Li-Merz parameters for Ca^{2+} . AF488, AF594 and TRC2 dyes were described using the Amber GAFF2 force field, with RESP charges derived at the HF/6-31G(d) level of theory on geometries obtained using CAM-B3LYP/6-31G(d) including Grimme D3 dispersion correction.¹⁵² All systems were solvated in a 0.1 M KCl aqueous solution octahedral box, minimized and then thermalized from 0 K to 300 K during 250 ps NVT simulations followed by 250 ps equilibration in the NPT ensemble. Then, NVT production runs were

extended for 5 μ s. This procedure was performed for a total of $3 \times 5 \mu$ s replicas for each *apo* and *holo* version of CaM-AF488-TRC2, CaM-TRC2-AF488, CaM-AF488-AF594 and CaM-AF594-AF488. For WT *apo* and *holo* CaM, we performed $12 \times 1 \mu$ s replicas plus $3 \times 5 \mu$ s replicas.

All simulations were performed using the Amber20 software¹⁵⁰ using periodic boundary conditions, the SHAKE algorithm to restrain bonds involving hydrogen, the particle-mesh Ewald approach to account for long-range electrostatics, a nonbonded cutoff equal to 10 Å, and the Hydrogen Mass Repartitioning scheme,¹⁵³ which allowed an integration time step of 4 fs. Radius of gyration analysis were performed on trajectories with frames saved every 20 ps. The simulations for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems have been taken from previous chapter 4.

5.3.2 Electronic coupling calculations

In TrESP/MMPol electronic coupling calculations, the MMPol environment was described using the Amber pol12 AL polarizable force field,^{155,156} whereas the dyes were described adopting TrESP charges. We defined the boundaries between MMPol and the TrESP regions in the bond between the linker and the phenyl ring bound to the chromophoric system of the dyes. Atomic charges for water, K^+ and Ca^{2+} cations, and the AF488 and TRC2 dyes were taken from the work performed in the previous chapter and another study,¹⁶¹ whereas polarization-consistent ESP charges for AF594 were derived using the Polchat tool at the B3LYP/6-31G(d) level of theory.¹⁶² TrESP charges for AF488 and TRC2 were also taken from that previous work, whereas for AF594 we used the same protocol, based on a fit of the electrostatic potential computed from TD-CAM-B3LYP/6-31G(d) transition densities using the TraDA tool.¹⁵⁷ TrESP-MMPol calculations were run using the Trespcoup software¹⁵⁴ using an MMPol cutoff radius equal to 15 Å and

performing calculations every 80 ps along the MD trajectories obtained for each system.

FRET rates defined in Eq. 2.35 were obtained using a refractive index $n^2 = 2$,¹⁶³ a critical Förster radius $R_0 = 54 \text{ Å}$ for AF488-TRC2 and $R_0 = 56 \text{ Å}$ for AF488-AF594,¹⁴³ and a fluorescence lifetime for AF488 $\tau_D = 4 \text{ ns}$.¹⁶⁴ Transition dipole moments for the AF488 donor and AF594 and TRC2 acceptors equal to $\mu_{AF488} = 7.56 D$, $\mu_{TRC2} = 10.19 D$ and $\mu_{AF594} = 10.31 D$ were extracted from the TD-DFT calculations used to parametrize the TrESP charges. The center-to-center distance between dyes $R(t)$ and the $\kappa(t)$ orientation factors were computed from the structures sampled along the MD trajectories. Finally, to dissect the impact of Förster approximations, we defined four coupling models (PDA, PDA-MMPol, TrESP, TrESP-MMPol) using the general expression $V = s \cdot V_{Coul}$, which combine a Coulomb term computed using the TrESP or the PDA approximation with a screening factor given by Förster $1/n^2$ value or derived using TrESP/MMpol calculations using the equations reported in the section 2.2.6.3.

5.4 Conformational ensembles of CaM

In Fig. 5.2 and 5.3 we illustrate the structural ensembles derived for WT CaM as well as for CaM- CaM-AF488-TRC2, TRC2-AF488, CaM-AF488-AF594 and CaM-AF594-AF488 systems. We note however that only a superposition of 300 frames sampled from equally spaced frames of the trajectories are shown. The simulations indicate that the linker explores a variety of conformations. This flexibility allows CaM to bind different targets anchored to the N and C-terminal domains. As has been observed in previous MD simulations^{165–170} and FRET studies, also the N- and C-lobes adopt different arrangements about the central linker, which is considered to be important for target recognition.^{143,164,171}

Comparison of the ensembles shown for WT CaM and CaM labelled with AF488 and TRC2 suggest that similar overall CaM structures are sampled, both in the absence and in the presence of Ca²⁺. In addition, simulations based on the ff14SB and a99SB-disp force fields also suggest similar ensembles. This situation however significantly changes when CaM is labelled with AF488 and AF594. In this case, the ensembles obtained using ff14SB resemble those found for WT CaM or CaM-AF488-TRC2.

However, simulations based on the a99SB-disp force field, which was optimized for disordered proteins, display a much larger flexibility, suggesting that more unfolded conformations are sampled with this force field. Thus, the presence of the AF594 donor is found to induce a larger disorder and unfolding compared to TRC2. This is somewhat unexpected, given that the linker for AF594 is longer, so one could expect a smaller impact on the intrinsic properties of CaM.

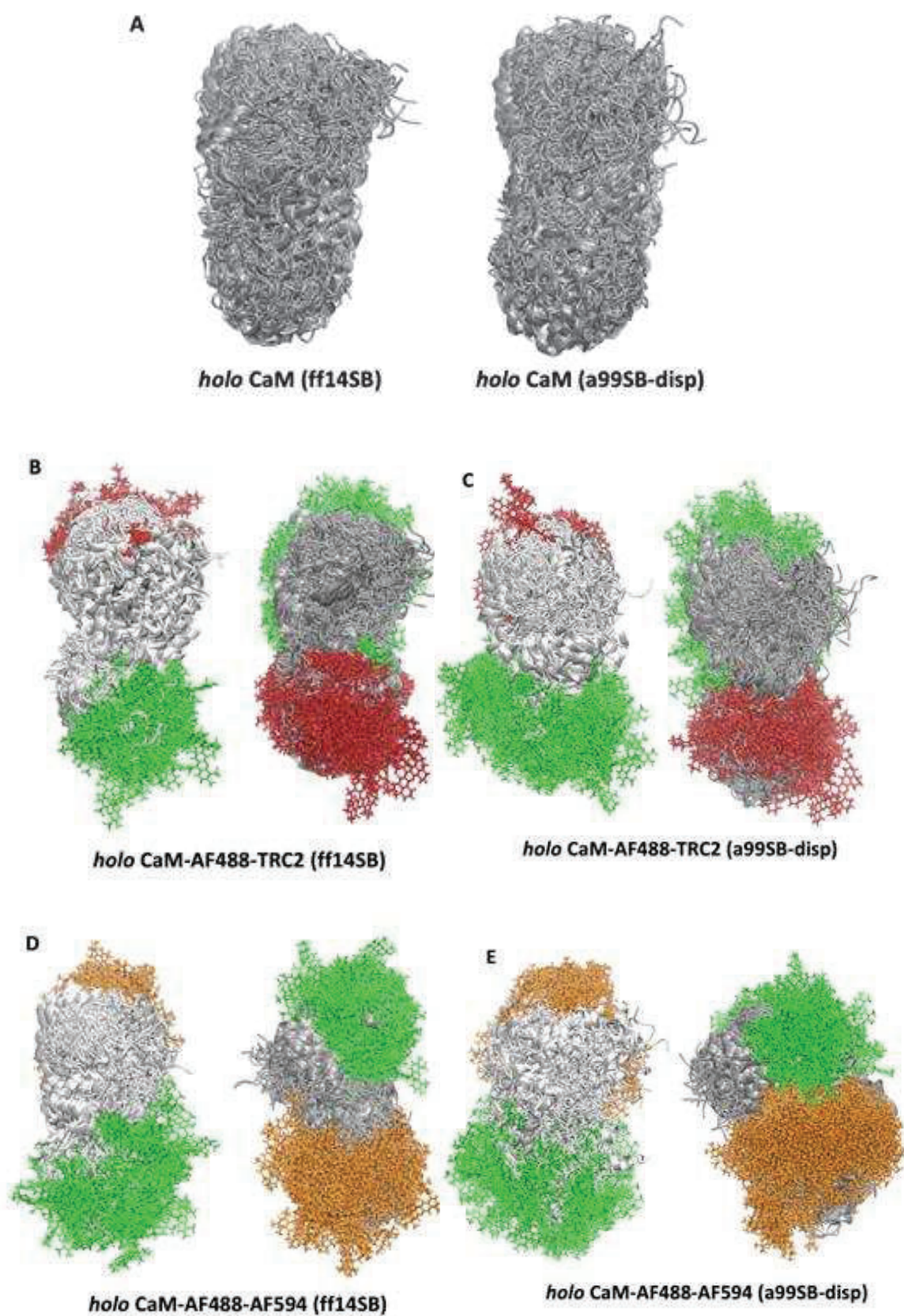
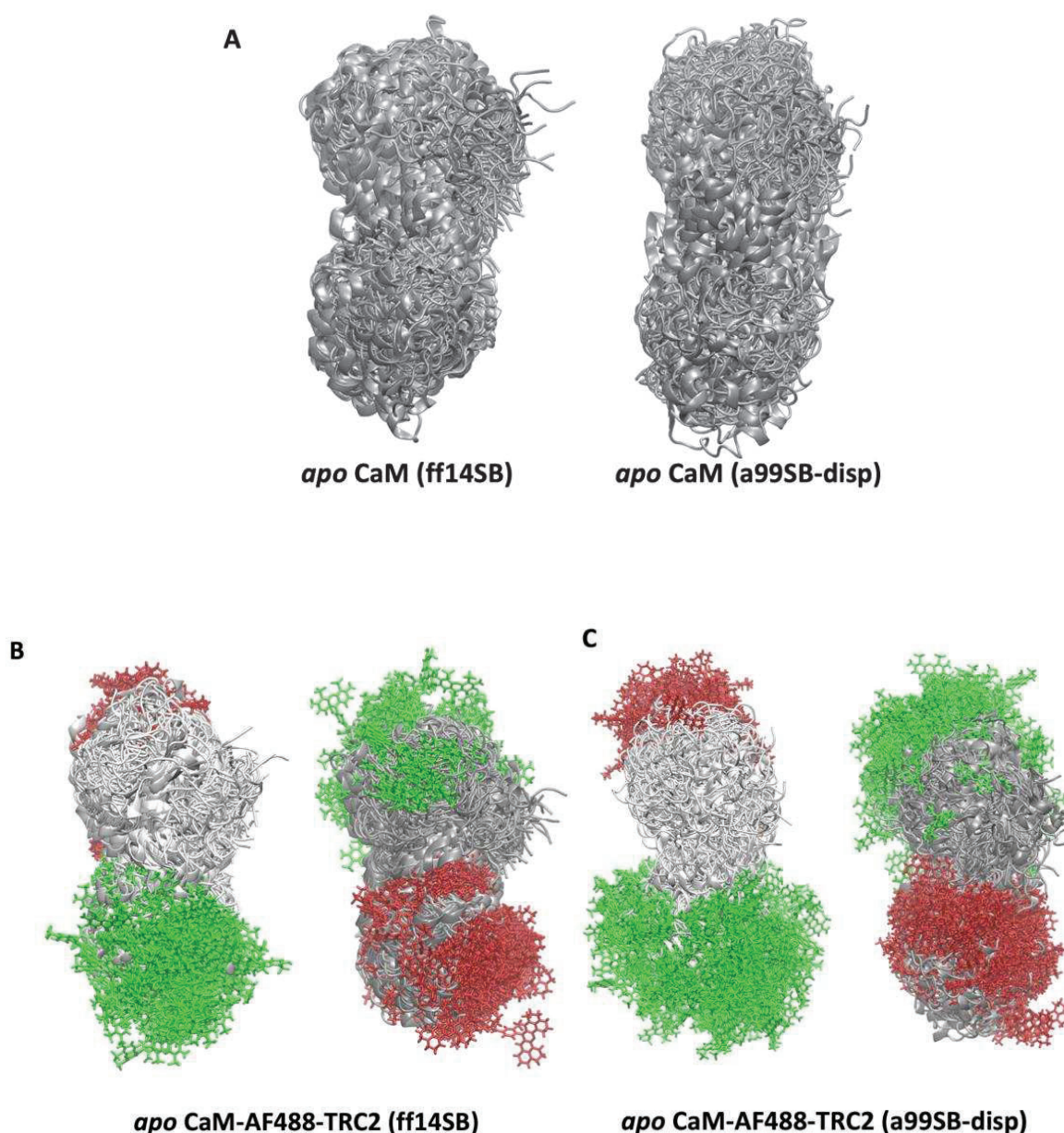


Figure 5.2. Conformational ensembles derived from MD simulations for *holo* CaM. A Wild type CaM ensembles derived from ff14SB and a99SB-disp force fields. B CaM-TRC2-AF488 and CaM-AF488-TRC2 ensembles derived from ff14SB. C CaM-TRC2-AF488 and CaM-AF488-TRC2 ensembles derived from a99SB-disp. D CaM-AF594-AF488 and CaM-AF488-AF594 ensembles derived from ff14SB. E CaM-AF594-AF488 and CaM-AF488-AF594 ensembles derived from a99SB-disp. AF488 in green, TRC2 in red and AF594 in orange.



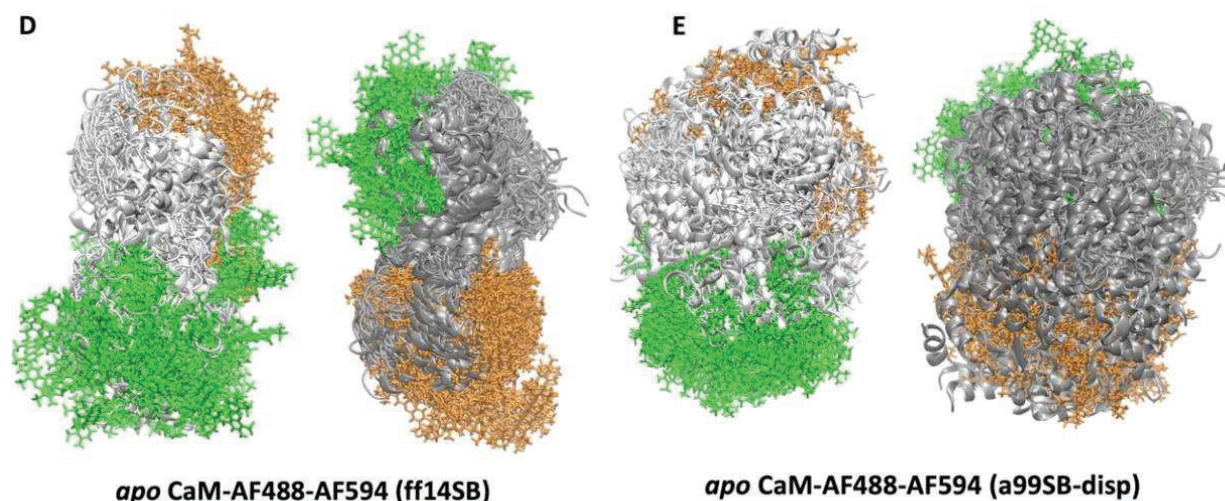
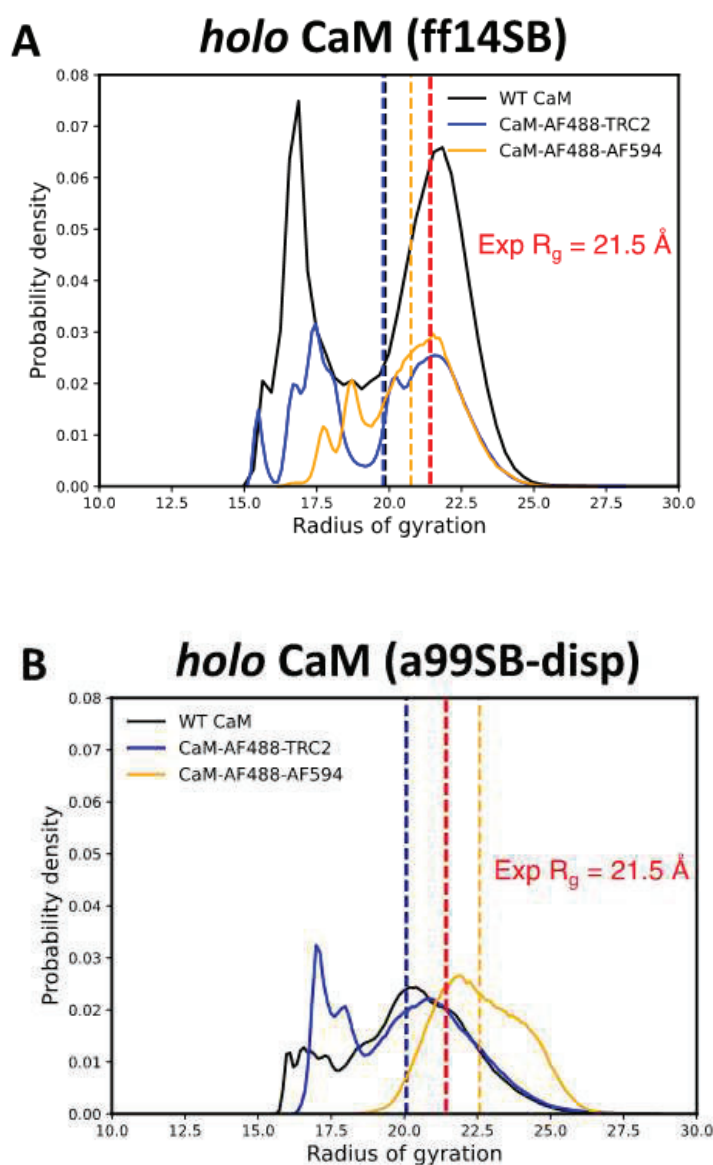


Figure 5.3. Conformational ensembles derived from MD simulations for *apo* CaM. **A** Wild type CaM ensembles derived from ff14SB and a99SB-disp force fields. **B** CaM-TRC2-AF488 and CaM-AF488-TRC2 ensembles derived from ff14SB. **C** CaM-TRC2-AF488 and CaM-AF488-TRC2 ensembles derived from a99SB-disp. **D** CaM-AF594-AF488 and CaM-AF488-AF594 ensembles derived from ff14SB. **E** CaM-AF594-AF488 and CaM-AF488-AF594 ensembles derived from a99SB-disp. AF488 in green, TRC2 in red and AF594 in orange.

To inspect the ensembles, in Fig. 5.4 and Table 5.1 we report the distribution of radius of gyration computed along the MD trajectories, which is compared to the experimental values of R_g 21.5 and 21.2 Å for *holo* and *apo* CaM, respectively.¹⁷² The simulations explore a range of radius of gyration values from 15 to 25 Å. As noted before, however, simulations based on a99SB-disp in the presence of the AF594 donor explore a larger range of R_g values up to ~27 Å. Overall, the Amber ff14SB provides a more consistent picture, with similar distributions found both for *apo* and *holo* forms for the WT protein or its dye-labelled counterparts. In all ff14SB simulations, however, the average R_g values are slightly underestimated compared to experiment. The performance of a99SB-disp is more complicated. For the WT and AF488-TRC2 labelled protein, the distributions are quite consistent, indicating as when using ff14SB that the presence of the dyes does not significantly alter the

conformational preferences. In this case, *apo* simulations describe an average R_g value in excellent agreement with experiment, whereas *holo* ones indicate a slight underestimation similar to that found with ff14SB. In contrast, as noted before, a99SB-disp simulations with the AF594 donor however significantly overestimate the radius of gyration, both for *apo* and *holo* CaM, confirming the observations in Fig. 5.2 and 5.3 that a more disordered and unfolded ensemble is found.



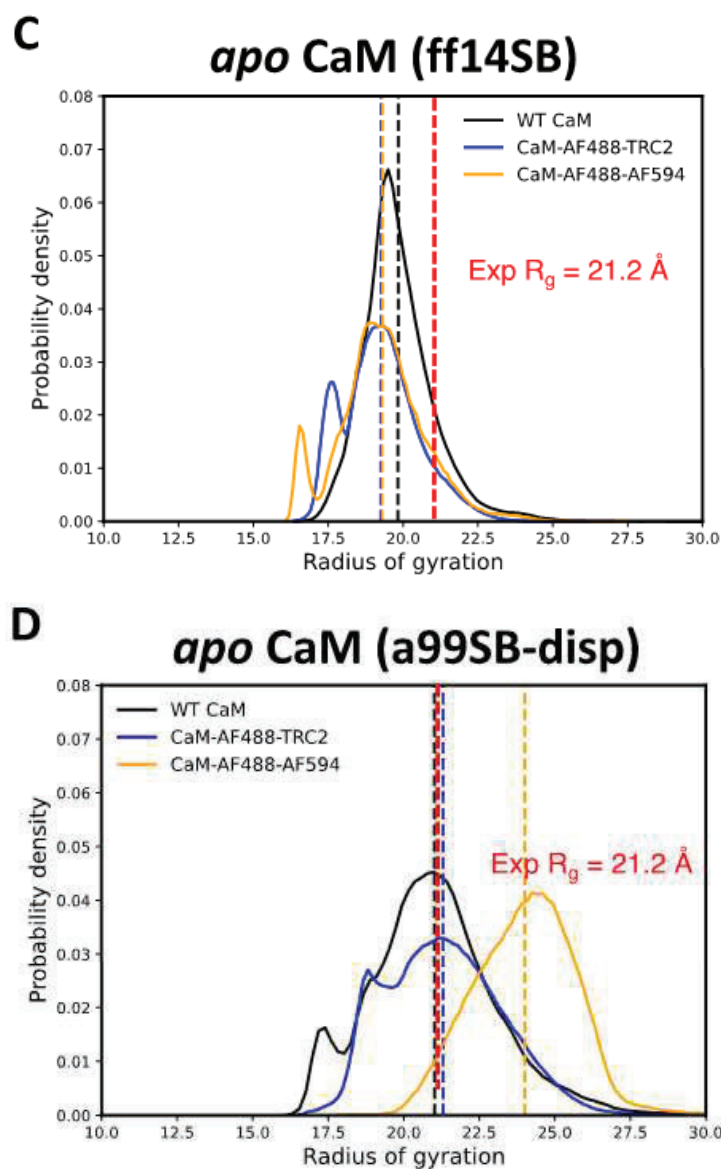


Figure 5.4. Distribution of radius of gyration computed along the MD trajectories of CaM compared to the experimental values $21.2 \pm 0.3 \text{ \AA}$ and $21.5 \pm 0.3 \text{ \AA}$ for *apo* and *holo* CaM, respectively.¹⁷² **A** *holo* WT CaM, CaM-TRC2-AF488/AF488-TRC2 and CaM-AF594-AF488/AF488-AF594 simulations based on ff14SB. **B** *holo* WT CaM, CaM-TRC2-AF488/AF488-TRC2 and CaM-AF594-AF488/AF488-AF594 simulations based on a99SB-disp. **C** *apo* WT CaM, CaM-TRC2-AF488/AF488-TRC2 and CaM-AF594-AF488/AF488-AF594 simulations based on ff14SB. **D** *apo* WT CaM, CaM-TRC2-AF488/AF488-TRC2 and CaM-AF594-AF488/AF488-AF594 simulations based on a99SB-disp.

Table 5.1. Average radius of gyration computed along the MD trajectories of CaM compared to experimental values.¹⁷² Note the experiments contain a mixture of both D-A and A-D systems due to unselective dye labelling, so the overall simulated distribution values are shown. Standard error of the mean (SEM) for theoretical values has been obtained from the mean values of the six 5 μ s replicas.^a

		<i>ff14SB</i>	<i>a99SB-disp</i>	<i>Exp.</i> ¹⁷²
<i>holo CaM</i>	<i>WT</i>	19.3±0.4	20.1±0.5	21.5±0.3
	<i>AF488-TRC2</i>	19.9±0.9	20.1±0.7	
	<i>AF488-AF594</i>	20.7±0.3	22.6±0.2	
	<i>WT</i>	19.8±0.3	20.7±0.3	21.2±0.3
<i>apo CaM</i>	<i>AF488-TRC2</i>	19.3±0.4	21.3±0.2	
	<i>AF488-AF594</i>	19.2±0.4	24.0±0.2	

^aAll provided values are in Å units.

In order to provide more insights into the ensembles, we have performed a clustering analysis for all systems, which is shown in Figs. 1-A2 to 8-A2 of the appendix 2. We classified the ensembles into four clusters, with the aim of relating them to the four main FRET states identified in experimental lifetime distributions by DeVore and co-workers.¹⁴³ The centroids of clusters, along with their R_g values reported in the figures, allows to characterize the main conformations that can contribute to the R_g distribution in Fig. 5.2 and 5.3. In some centroids, the linker keeps the C- and N-terminal domains separated, resembling the main conformation of *holo* CaM displayed in the crystal structure in Fig. 5.1. We also find conformations in which the N- and C-lobes directly interact, leading to compact shapes similar to those observed in complexes of CaM with peptides.¹⁶⁸

In Figs. 1-A2 to 8-A2 of the appendix 2, we also display intramolecular contact maps averaged over the MD showing the folding patterns of the two domains

in CaM. This contact maps confirm the fact that the presence of the AF594 donor in MDs based on a99SB-disp force field induce a significant unfolding of CaM, apparent in the smaller number of contacts in the C- and N-terminal domains.

5.5 Beyond Förster coupling: dipole approximation, dielectric screening and isotropic approximation

One of the insights of Förster theory relies on the ability to derive the D/A distance from the measured FRET efficiency and the R_0 value tabulated for a given dye pair, using the well-known R^{-6} distance dependence of the rate. This however involves approximating the coupling through the point dipole approximation, including a simple dielectric screening factor, and assuming an isotropic dipole orientation factor, $\langle \kappa^2 \rangle = 2/3$.

Regarding the dipole approximation, an unwritten rule is to consider that it will breakdown at separations close to the dye's dimensions.¹⁷⁹ In Fig. 5.5, we report the ration between the Coulomb coupling contribution obtained using PDA or adopting atomic transition charges TrESP as estimated from simulations of CaM systems labelled with TRC2 and AF488 dyes or AF594 and AF488 dyes. Our results confirm that more significant deviations are found at close separations, below 30 Å. According to average values of the ratios $\langle V_{PDA}/V_{TrESP} \rangle$ indicated by the dashed line, overall, the PDA provides a faithful representation of the coupling, with average ratios close to one until distances below 30 Å, where the PDA introduces significant deviations. We note that uncomplete orientational averaging is also increased at close separations, whereas complete averaging also contributes to improve the agreement of the PDA when results are averaged over multiple orientations.

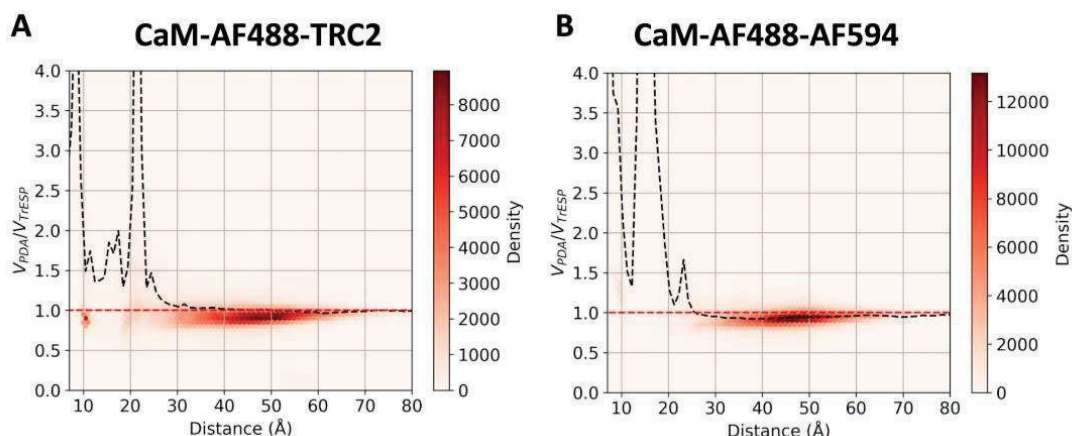
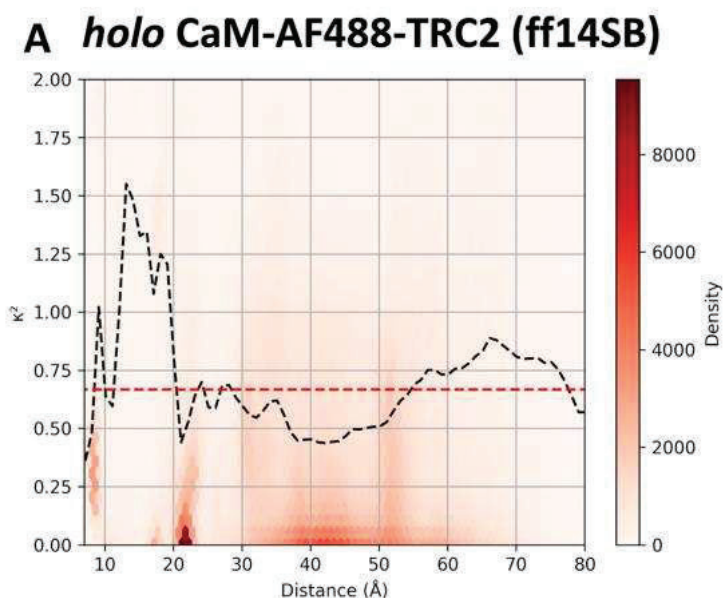


Figure 5.5. Ratio between PDA and TrESP coulombic coupling contributions (average values over distance bins indicated by black dashed curve). **A Data from CaM systems labelled with TRC2 and AF488 dyes. B Data from CaM systems labelled with AF594 and AF488 dyes.**

Then, additional deviations are expected if the dyes do not sample a complete isotropic distribution. In Fig. 9-A2 and 10-A2 of the appendix 2, we report the distribution of orientation factors sampled along the MD trajectories, which are very close to the random isotropic distribution in all cases, with average $\langle \kappa^2 \rangle$ values close to the isotropic limit $2/3$. As shown in Fig. 5.6 and 5.7, however, at D/A separations below $30 \text{ \AA}$ we find significant deviations from the isotropic average value. This is probably due to the more compact structures adopted by CaM in this separation range, which restrict the orientational freedom of the dyes.



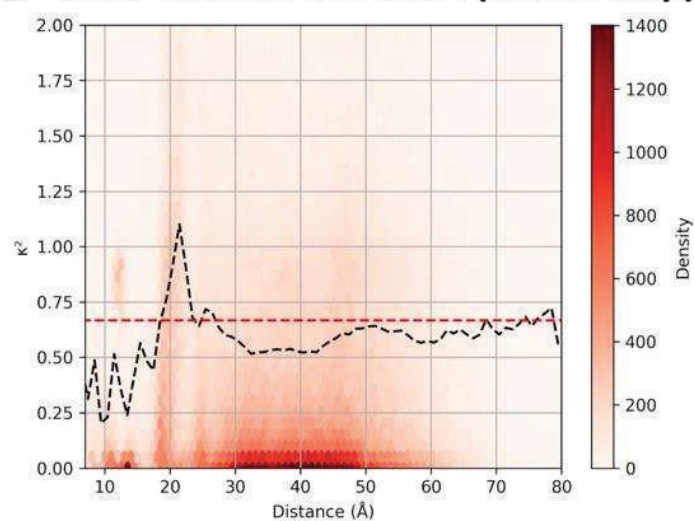
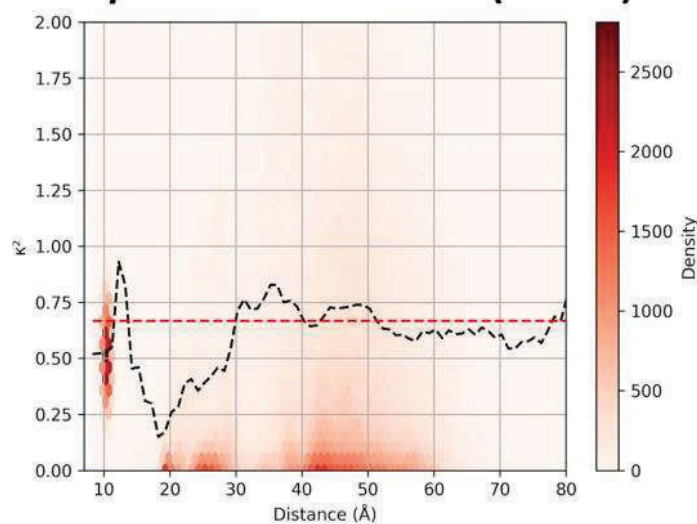
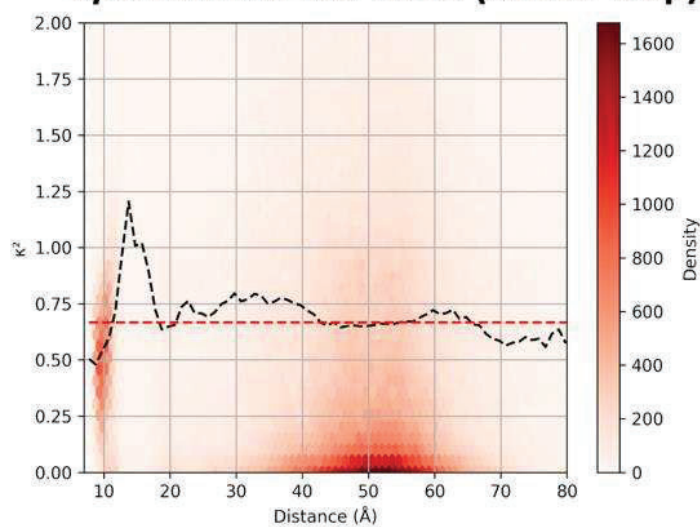
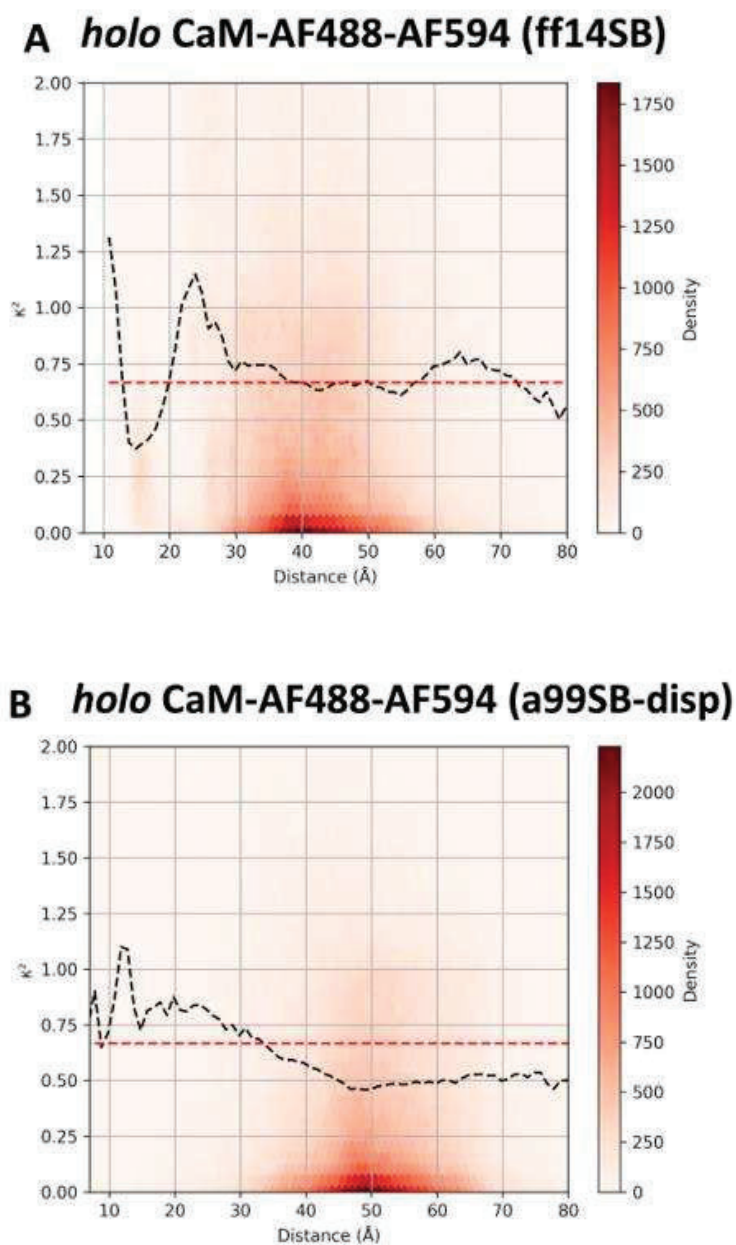
B *holo* CaM-AF488-TRC2 (a99SB-disp)**C** *apo* CaM-AF488-TRC2 (ff14SB)**D** *apo* CaM-AF488-TRC2 (a99SB-disp)

Figure 5.6. Density distribution of orientation factors as a function of D/A separation derived from MD simulations for CaM-AF488-TRC2/TRC2-AF488 systems (average values over distance bins indicated by black dashed curve, isotropic value indicated by red line). **A** *holo* CaM-AF488-TRC2 ff14SB. **B** *holo* CaM-AF488-TRC2 a99SB-disp. **C** *apo* CaM-AF488-TRC2 ff14SB. **D** *apo* CaM-AF488-TRC2 a99SB-disp.



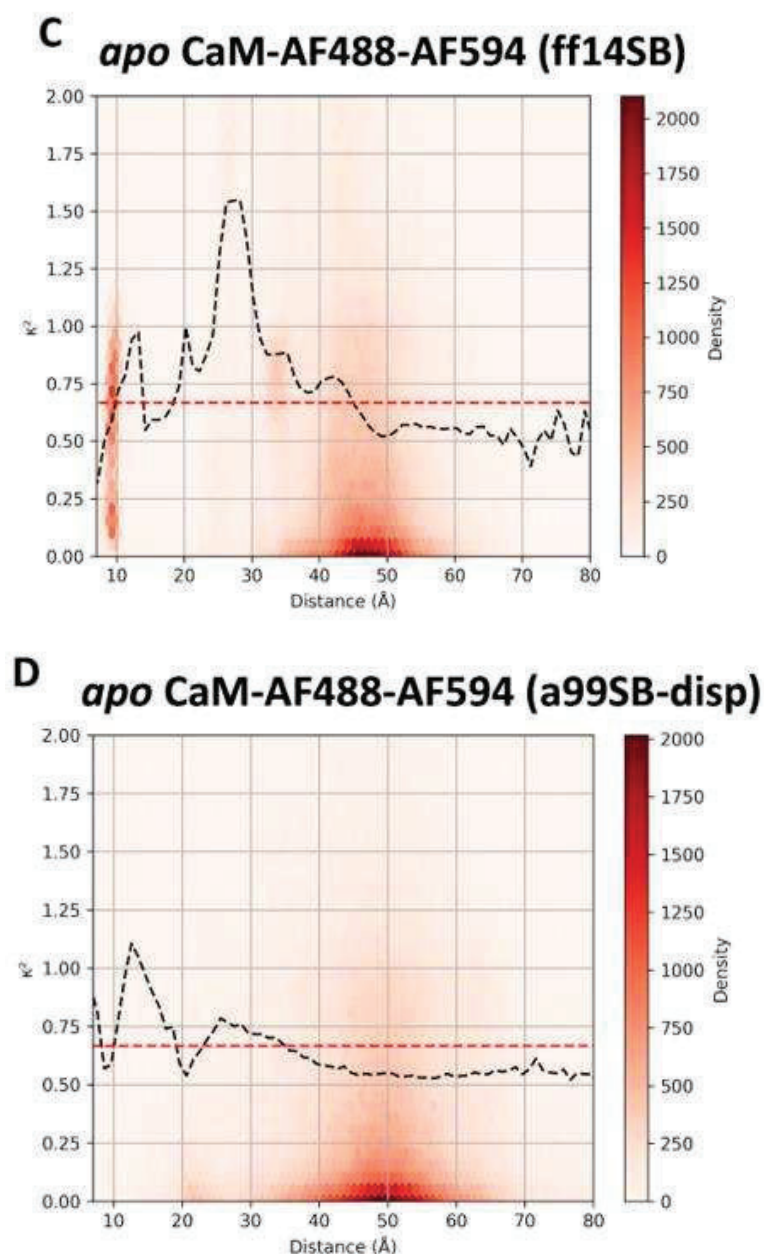


Figure 5.7. Density distribution of orientation factors as a function of D/A separation derived from MD simulations for CaM-AF488-AF594/AF594-AF488 systems (average values over distance bins indicated by black dashed curve, isotropic value indicated by red line). **A *holo* CaM-AF488-AF594 ff14SB. **B** *holo* CaM-AF488-AF594 a99SB-disp. **C** *apo* CaM-AF488-AF594 ff14SB. **D** *apo* CaM-AF488-AF594 a99SB-disp.**

Beyond the dipole approximation, the R^{-6} dependence of FRET rates also relies on the validity of the $1/n^2$ screening factor that multiplies the PDA

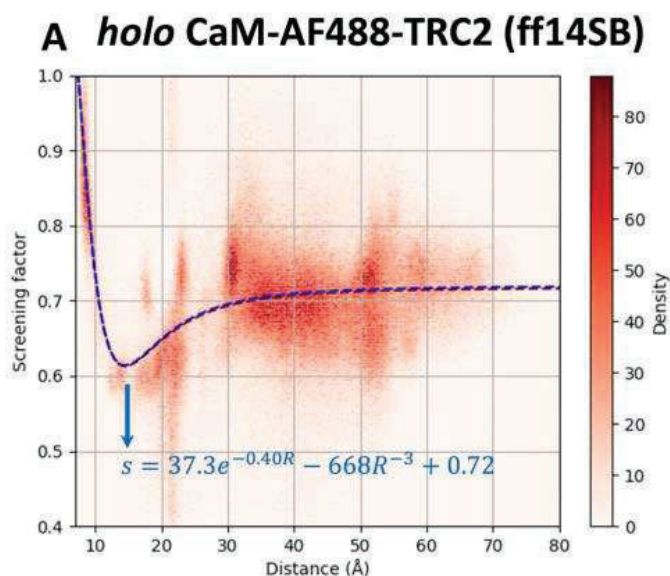
coupling in Förster theory. This accounts for dielectric screening effects exerted by the surrounding polarizable environment, which attenuate the interaction between D/A dipoles. In the previous chapter we studied in detail the impact of dielectric screening in the *holo* CaM-AF488-TRC2 system explored using the ff14SB force field.

Here, we investigate these effects for a wider variety of systems, including the *apo* form of CaM, systems labelled with the AF594 donor, and from simulations performed both from ff14SB and a99SB-disp force fields. We aim at investigating if the distance-dependent screening effects are also observed in these systems, and we examine the transferability of the empirical screening function we fitted from previous data. Beyond the distance dependence, we also found a minimum for the screening factor at D/A separations ~ 15 Å, in which the fluorophores are partially excluded from the solvent and thus screening is dictated by the more polarizable protein environment. To account for both effects, we included an R^{-3} term, leading to this empirical function:

$$s = 37.3e^{-0.40R} - 668R^{-3} + 0.72 \quad (5.1)$$

As discussed before, we expected the position of the minimum to effectively depend on the specific system, as this minimum arises from the difference between protein and water environment. In Fig. 5.8 and 5.9 we report the density distribution of screening factors as a function of D/A separation derived from TrESP-MMPol electronic couplings computed for *apo* and *holo* CaM-AF488-TRC2 and CaM-AF488-AF594 systems along the MD trajectories based on ff14SB and a99SB-disp force fields. We compare empirically fitted functions for each case, displayed in red, to the function in Eq. 5.1 fitted to results on *holo* CaM-AF488-TRC2 based simulated with ff14SB. For the simulations on CaM-AF488-TRC2, we found a convincing agreement with previous data. In all cases we fit a similar function, which mostly varies regarding the 0.70-0.80 asymptotic value adopted by the screening factor at

large separations. We note, however, that the most different case, with a value 0.80, is found for the simulations of *apo* CaM sampled with the a99SB-disp. This force field explores less compact conformations, and the higher value can be ascribed to a more relevant role of the water environment (less polarizable) compared to the protein. For the simulations on CaM-AF488-AF594, the change of the acceptor dye induces larger variations in the screening factors obtained, indicating that the screening effects also depend on the specific dyes investigated. Nevertheless, quite strikingly we found the same decaying parameter $0.40 \text{ \AA}^{-1}$ in these simulations, which suggests the distance dependence at close separations is quite general. The most different results are found for the protein labelled with AF488-AF594 dyes when sampled using the a99SB-disp force field. As was discussed in the previous section, these systems experience a rather important unfolding, which the N- and C-lobes partially unfolding. The result is that the ensembles do not explore close D/A separations, so these simulations do not allow to assess with confidence the exponential decay at close distances. In addition, the D/A dyes are much more exposed to the solvent, as shown in solvent-accessible surface areas (SASA) reported in Fig. 11-A2 and 12-A2 in appendix 2. Whereas in all other simulations the dyes reach SASAs up to 1200-1400 $\text{\AA}^2$, in CaM-AF488-AF594 trajectories with a99SB-disp the SASAs adopt values up to 1600-1800 $\text{\AA}^2$. These explains why the empirical fitted functions adopt asymptotic values 0.84-0.88.



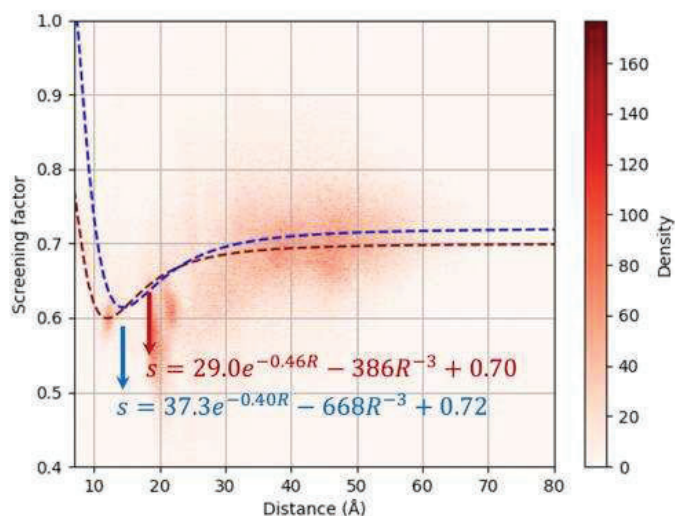
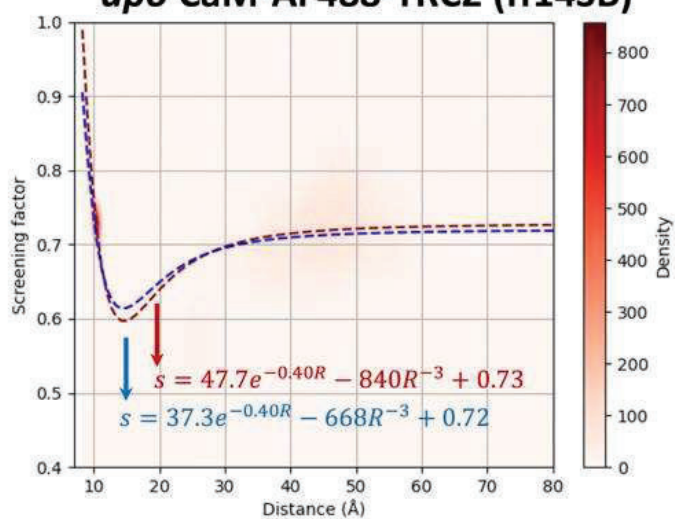
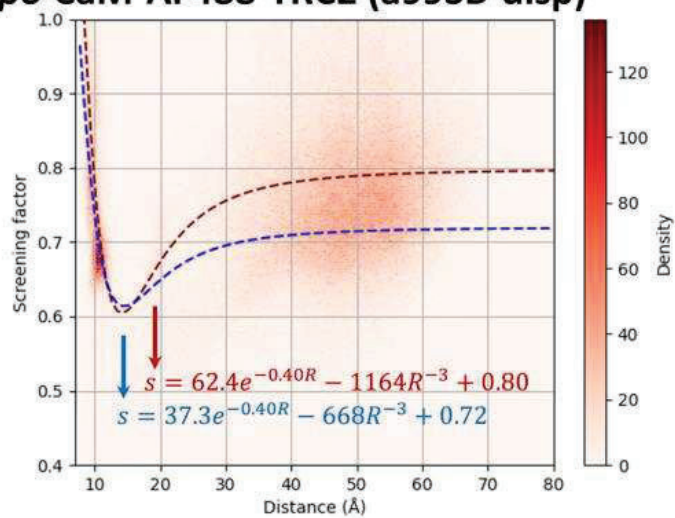
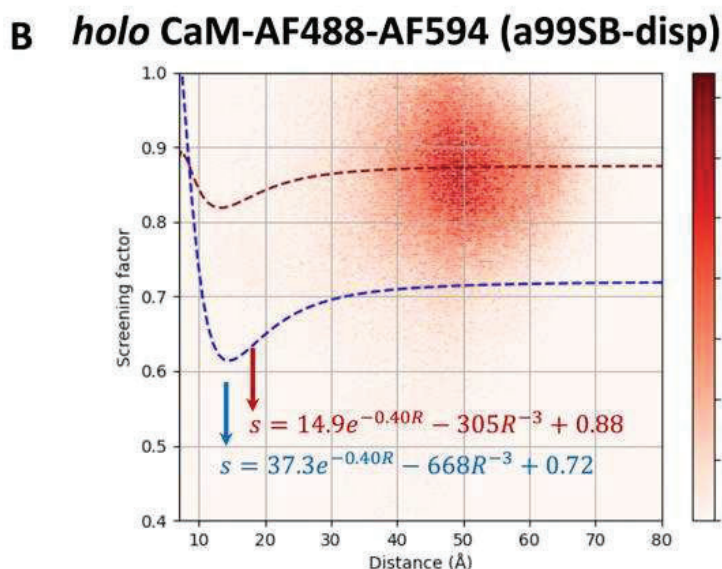
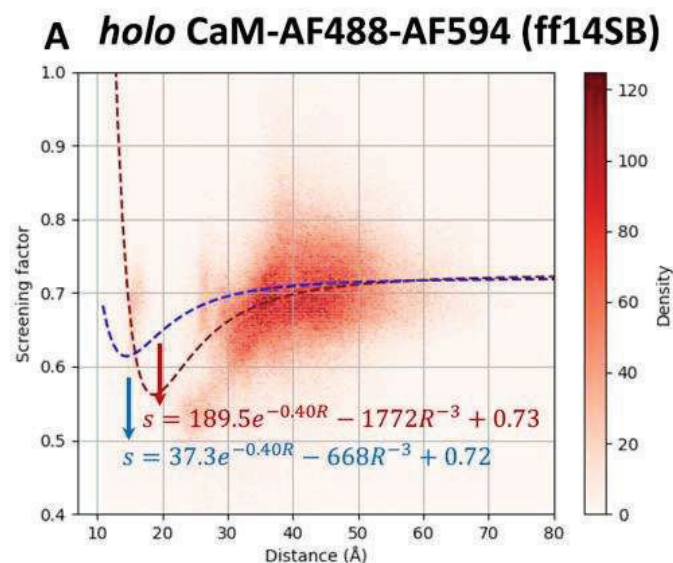
B *holo* CaM-AF488-TRC2 (a99SB-disp)**C** *apo* CaM-AF488-TRC2 (ff14SB)**D** *apo* CaM-AF488-TRC2 (a99SB-disp)

Figure 5.8. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for CaM-AF488-TRC2/TRC2-AF488 systems along MD trajectories based on ff14SB and a99SB-disp force fields. Empirically fitted functions for each case (in red) are compared to the function fitted from simulations on *holo* CaM-AF488-TRC2 based on ff14SB. **A** *holo* CaM-AF488-TRC2 ff14SB. **B** *holo* CaM-AF488-TRC2 a99SB-disp. **C** *apo* CaM-AF488-TRC2 ff14SB. **D** *apo* CaM-AF488-TRC2 a99SB-disp.



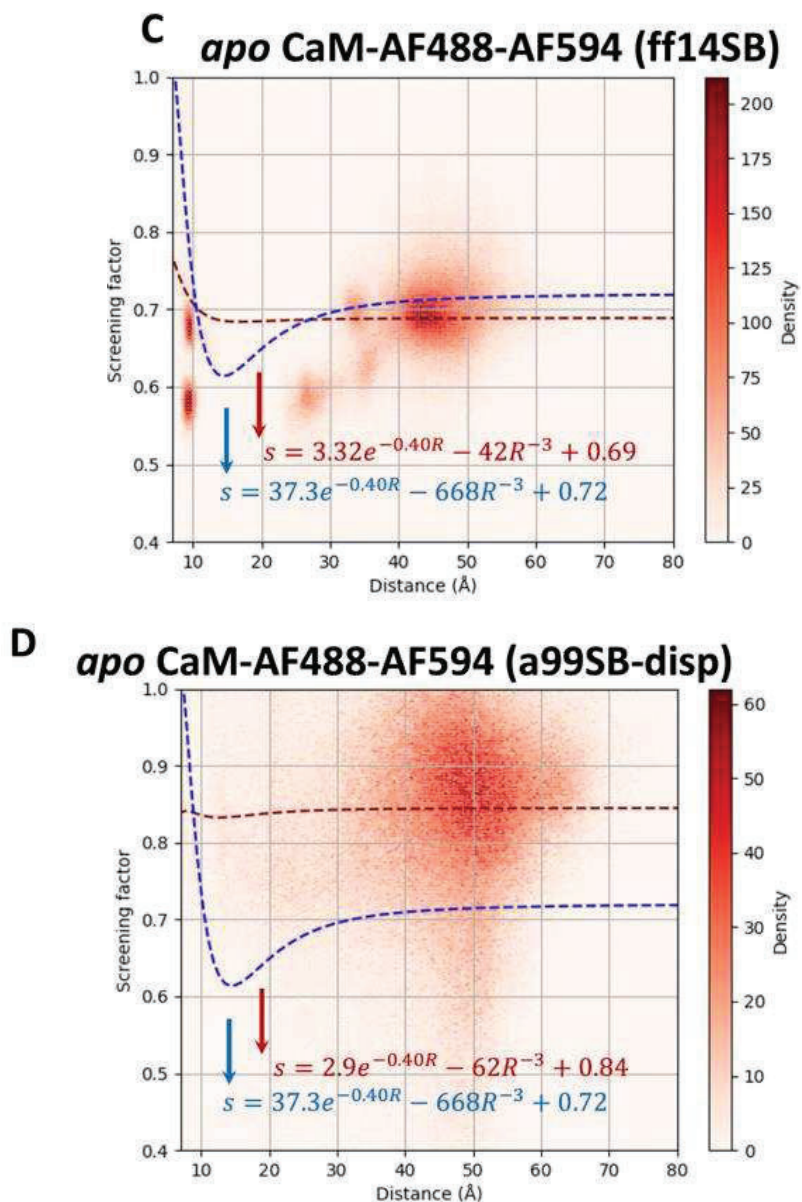


Figure 5.9. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for CaM-AF488-AF594/AF594-AF488 systems along MD trajectories based on ff14SB and a99SB-disp force fields. Empirically fitted functions for each case (in red) are compared to the function fitted from simulations on *holo* CaM-AF488-TRC2 based on ff14SB. **A** *holo* CaM-AF488-AF594 ff14SB. **B** *holo* CaM-AF488-AF594 a99SB-disp. **C** *apo* CaM-AF488-AF594 ff14SB. **D** *apo* CaM-AF488-AF594 a99SB-disp.

Overall, these data support the universality of the close exponential distance dependence of screening factors. At the same time, however, they indicate that the actual position of the minimum depends on the nature of the interacting dyes and the distance where screening factor goes up due to a more pronounced exposure to the water solvent. Moreover, they indicate that the actual screening factors depends on the specific conformations sampled by a particular force field, in this case a99SB-disp inducing a larger exposure of the dyes in the CaM-AF488-AF594 system that consequently leads to higher screening factors.

5.6 Fluorescence lifetime distributions

In the previous sections of this chapter, we have investigated how different approximations underlying Förster theory impact electronic couplings in the CaM systems. Here, we aim at examining the overall impact of Förster approximations in the simulation of FRET distributions arising from the MD ensembles. As previously discussed, we found that the dipole approximation, as well as the assumption of complete rotational freedom of the dyes through an isotropic orientation factor, mostly impact the estimation of couplings at short separations below 30 Å. However, we observed that dielectric screening effects are challenging to predict *a priori*, and although there is a clear exponential distance dependence that we corroborated in this chapter, the asymptotic behavior of screening effects at large separations, for example, depend on the specific system under study or even the force field adopted, as it depends on the degree of exposure to the water solvent of the dyes.

We focus on simulating fluorescence lifetime distributions, which reflect the underlying FRET distribution because energy transfer decreases the donor fluorescence lifetime. We then compare both distributions to the experimental data measured from donor fluorescence decays and smFRET.⁴²

In Fig. 5.10 and 5.11, we report the lifetime distributions obtained using electronic couplings computed using the PDA or the TrESP-MMPol model compared to experiment.¹⁴³ As was already observed in the previous chapter for holo CaM-AF488-TRC2, the PDA leads to an underestimation of the lifetimes at all ranges. This can be ascribed to the limitations of using a simple screening factor $1/n^2$ in Förster theory, which neglects its exponential distance dependence at short D/A separations, but also strongly exaggerates

screening even at long D/A distances. Thus, the complete lifetime distributions imply slower FRET and longer lifetimes based on the PDA compared to the more rigorous data based on TrESP-MMPol couplings, as visible in the trends in Fig. 5.10 and 5.11. The incorporation of atomistic screening effects improves the agreement with experimental data, which based on the PDA tends to display large peaks in the 1-4 ns lifetime ranges. In contrast, the TrESP-MMPol distributions display less steep tails toward shorter lifetimes, in better agreement with experimental distributions. Caution must be taken however in this comparison, as the experimental distribution is derived from a fitting of fluorescence decays using classic maximum entropy (cMEM) analysis where only four FRET states were recovered. Thus, the experimental distributions reported in the figures only show four peaks and does not reflect the complete distribution and misses intermediate states.

On the other hand, compared to experiment, theoretical distributions display some additional features in the sub-ps lifetime region for most CaM systems considered, except in CaM-AF488-AF594 when simulated using the a99SB-disp force field. We note however that these lifetimes have small populations, hardly contributing to the overall fluorescence decay, and in any case shorter than the typical time resolution of fluorescence experiments.⁶⁹ The results based on the ff14SB force field and a99SB-disp indicate that ff14SB displays stronger of such sub-ps lifetime features, linked to rather compact conformations not observed experimentally. Significant populations are observed of efficient FRET states with lifetimes <10 ps, whereas experimentally the fastest component of the decay was fitted to a lifetime of 100 ps. It is not clear if this disagreement is caused by inaccuracies in ff14SB, as the a99SB-disp force field optimized for disordered systems displays less of such features. However, this could also arise from the fact

that fluorescence decays were measured with an instrument response function (IRF) with a full width at half-maximum <50 ps, so it is unlikely to observe those fast components in this experimental set-up.

Additive force fields used for biosimulations, like ff14SB, tend to introduce a bias that favors folded states, thus breaking the balance between folded and extended states, and efforts are being done to improve the description of disordered systems.^{70,71} Our results however, independently of the force field used, indicate that the TRC2 donor induces compact conformations not observed for example for the AF488-AF594 labelled proteins, which agrees with the suggestion that the TRC2 dye, in contrast to other acceptors, may stabilize a compact conformation of CaM by conformational selection.⁴² Nevertheless, our results for the additional CaM systems tackled in this chapter confirm the importance of accounting for atomistic screening effects if FRET is used to examine the balance between extended and folded conformations, given the strong attenuation of screening effects at close separations. Indeed, neglecting this distance-dependence contribution in Förster theory artificially reduces the coupling found for compact structures, in this way counteracting their overestimated interaction and masking possible disagreements between simulations and experiment.

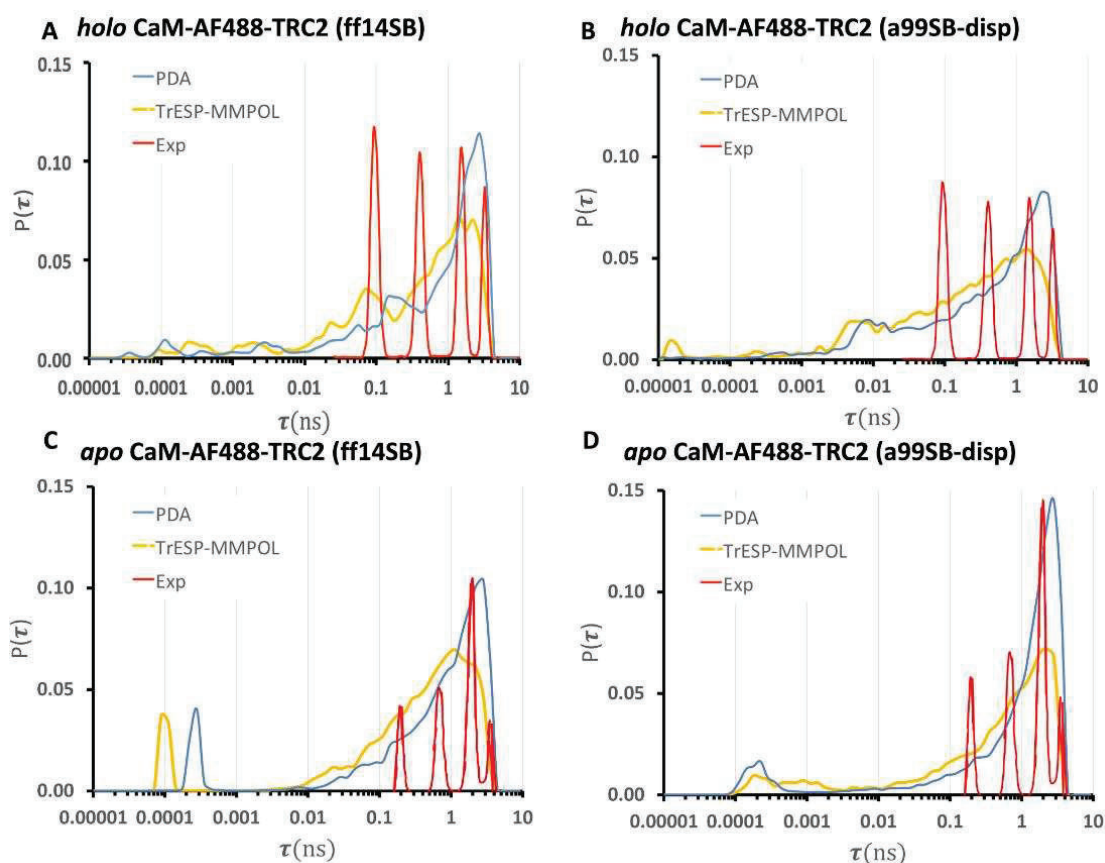


Figure 5.10. Distribution of fluorescence lifetimes computed for *holo* and *apo* CaM-AF488-TRC2/TRC2-AF488 systems using electronic couplings computed using the PDA or the TrESP-MMPol model compared to experiment.¹⁴³ **A** *holo* CaM-AF488-TRC2 (ff14SB). **B** *holo* CaM-AF488-TRC2 (a99SB-disp). **C** *apo* CaM-AF488-TRC2 (ff14SB). **D** *apo* CaM-AF488-TRC2 (a99SB-disp).

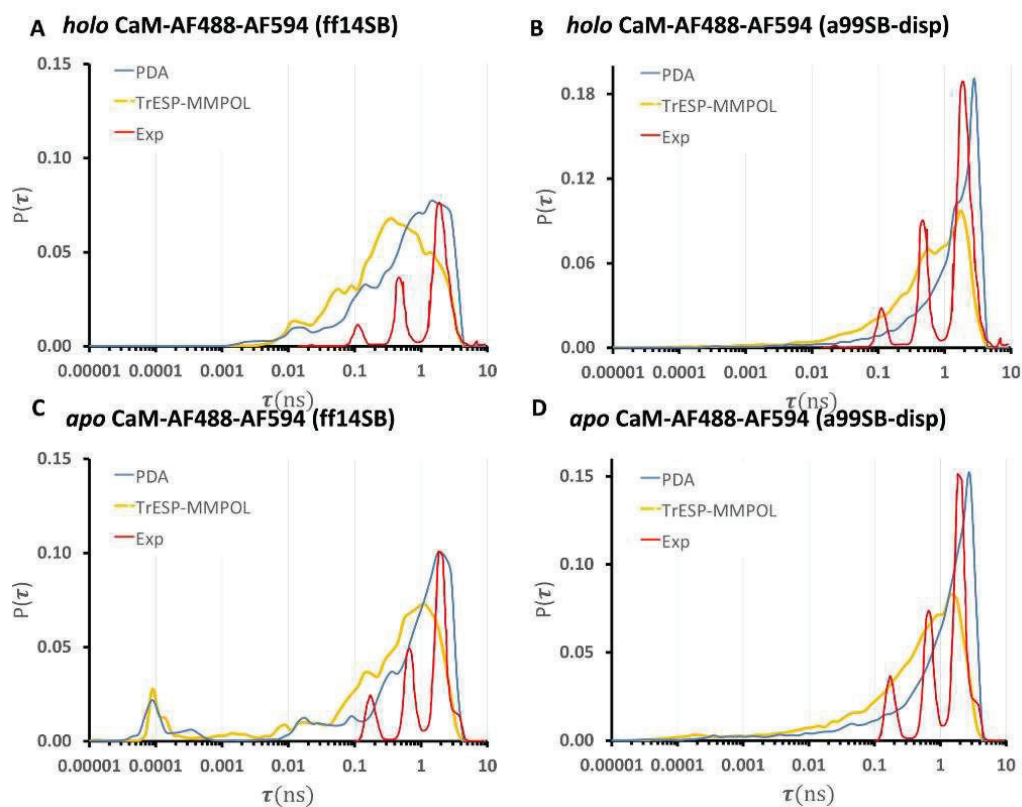


Figure 5.11. Distribution of fluorescence lifetimes computed for *holo* and *apo* CaM-AF488-AF594/AF594-AF488 systems using electronic couplings computed using the PDA or the TrESP-MMPol model compared to experiment.¹⁴³ **A** *holo* CaM-AF488-AF594 (ff14SB). **B** *holo* CaM-AF488-AF594 (a99SB-disp). **C** *apo* CaM-AF488-AF594 (ff14SB). **D** *apo* CaM-AF488-AF594 (a99SB-disp).

CHAPTER 6: Conclusions

CONCLUSIONS

In this thesis, we have examined for the first time how the heterogeneous environment impacts dielectric screening effects and thus the determination of the conformational properties of a partially disordered protein, CaM, studied combining MD and FRET simulations. We have also examined other limits of Förster theory relevant for the simulation of FRET properties from MD ensembles. This was made possible by the adoption of a rigorous yet efficient theoretical method that describes the fluorophores through distributed atomic transition charges coupled to a polarizable atomistic description of the protein and solvent environment. While this approach allows an efficient processing of thousands of structures extracted from extensive MD simulations, it retains the accurate description of screening effects provided by more costly QM/MMPol calculations. We can derive the following specific conclusions from the thesis:

- Our results indicate that the Förster dielectric screening approximation introduces systematic deviations much larger than those caused by the point dipole approximation, which mostly breaks down at short D/A separations where FRET lifetimes are anyway very short.
- We find a strong exponential attenuation of screening effects at D/A separations below 20 Å, which is in striking agreement with the behavior previously observed in pigment pairs of photosynthetic complexes from calculations based on continuum solvent models, despite the fact that we deal with different dye molecules located in a different protein-solvent environment, and more importantly, we describe the environment in atomic detail. Our findings thus indicate that dielectric screening alters the well-known R^{-6} distance

dependence of Förster energy transfer rate, which affects the practical derivation of structural information in FRET experiments.

- Despite the general exponential distance decay of screening, however, we observed that dielectric effects are challenging to predict a priori, and although there is a clear exponential distance dependence that we corroborated for multiple CaM systems, the asymptotic behaviour of screening effects at large separations, for example, depend on the specific system under study or even the force field adopted, as it depends on the degree of exposure to the water solvent of the dyes.
- The accurate description of screening effects seem to be of particular relevance to describe the balance between folded and unfolded conformations, characterized by either short or large inter dye separations. In this context, our results warn that the common practice of ignoring distance-dependent screening effects masks the tendency of classical force fields to overstabilize compact folded structures, with consequences in the refinement and validation of classical potentials for disordered proteins based on FRET data.
- Our simulations on apo and holo CaM labelled with different dyes suggest that labelling with AF594 donors, characterized by a long linker, seem to impact more strongly the conformational preferences of CaM compared to simulations based on the TRC2 donor, as observed for example in average radius of gyration values. We find, however, that a99SB-disp simulations in the presence of AF594 lead to partial unfolding of the N- and C-lobes of CaM, which could be an artifact caused by inaccuracies of the force field.
- Our results, however, independently of the force field used, indicate that the TRC2 donor induces compact conformations not observed

for example for the AF488-AF594 labelled proteins, which agrees with the suggestion that the TRC2 dye, in contrast to other acceptors, may stabilize a compact conformation of CaM by conformational selection.

CONCLUSIONS

APPENDIX 1

Appendix 1: for Chapter 4: On the breakdown of Förster energy transfer theory due to solvent effects: Atomistic simulations unveil distance-dependent dielectric screening in calmodulin

Table 1-A1. Structural and FRET properties averaged over MD trajectories for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems.^a

	$\langle R \rangle$	$\langle \kappa^2 \rangle$	$\langle s \rangle^b$	$\langle V_{PDA}^2 \rangle$	$\langle V^2 \rangle^b$	$\langle k_{theo} \rangle^b$	$\langle \tau_D \rangle^b$	$\langle E_{theo} \rangle^b$
TRC2-AF488	41.0	0.56	0.71	3145	1487	368	0.78	0.80
AF488- TRC2	41.6	0.67	0.69	270	514	127	0.92	0.76
Total	41.3	0.62	0.70	1708	1001	247	0.85	0.78

^a Distances R in Å, squared electronic couplings V_{PDA}^2 and V^2 in cm², rates k_{theo} in ns⁻¹ and excited state lifetimes τ_D in ns. ^bProperties derived from TrESP-MMPol electronic coupling calculations.

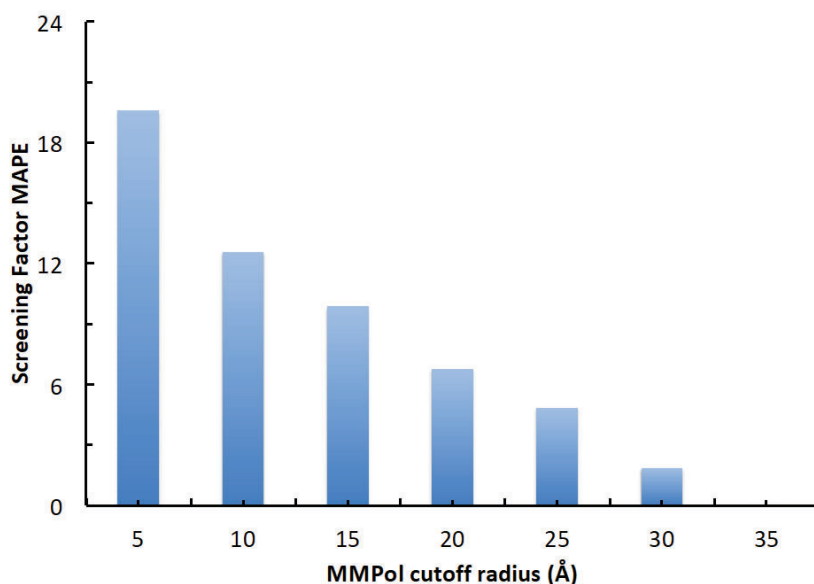


Figure 1-A1. Mean absolute percentage error (MAPE) in screening factors computed using QM/MMPol for 9 selected structures of CaM-AF488-TRC2 using different values of MMPol cutoff radius.

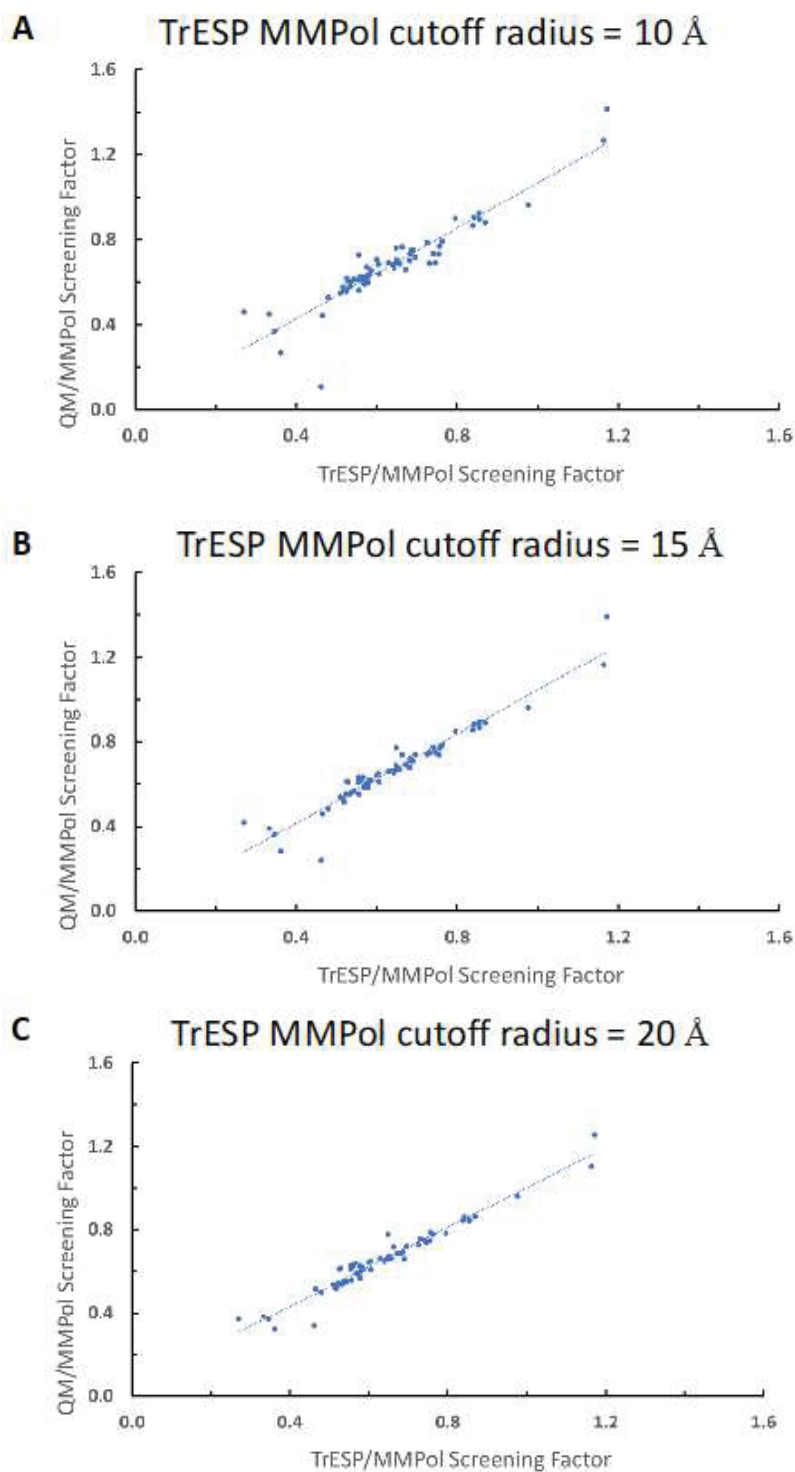
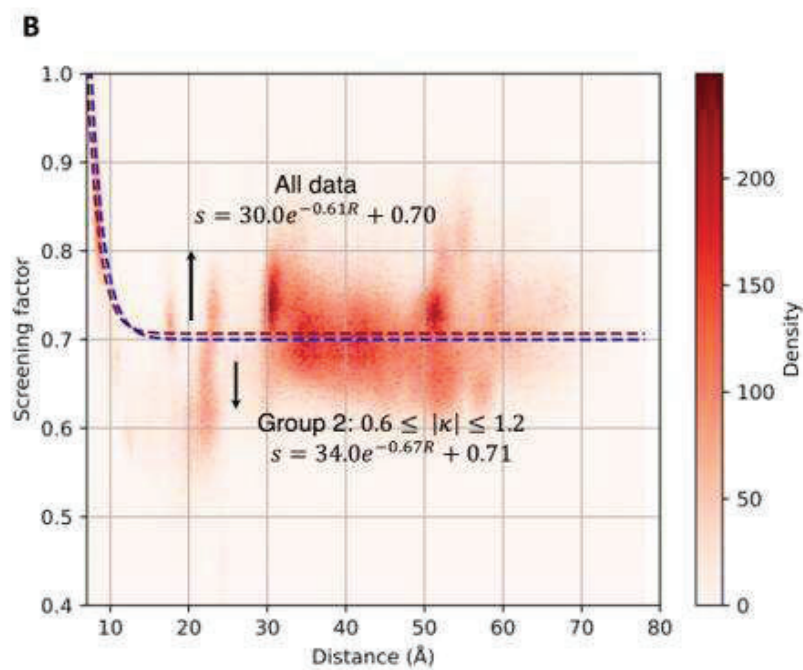
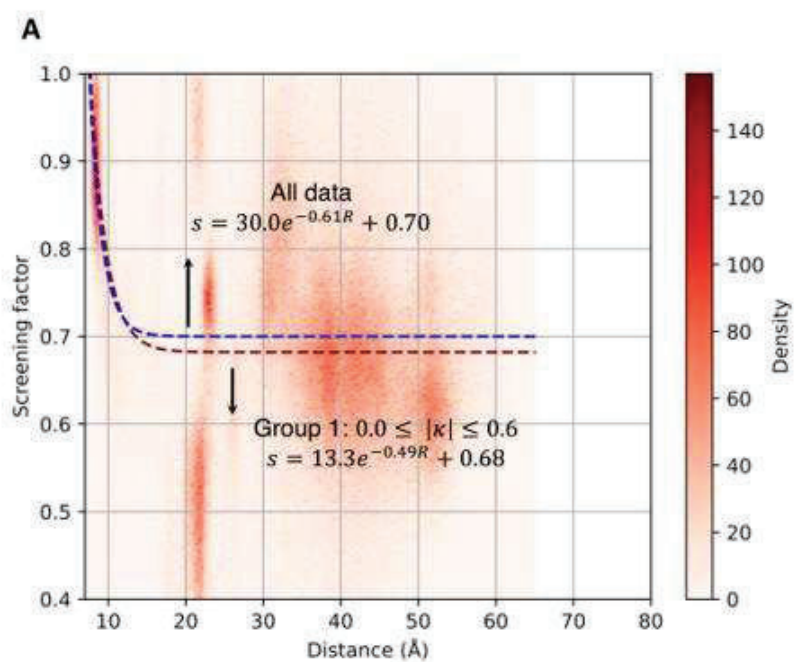


Figure 2-A1. Comparison between QM/MMPol screening factors, obtained using a fixed 20 Å MMPol cutoff radius, with TrESP/MMPol values obtained for different values of the cutoff.



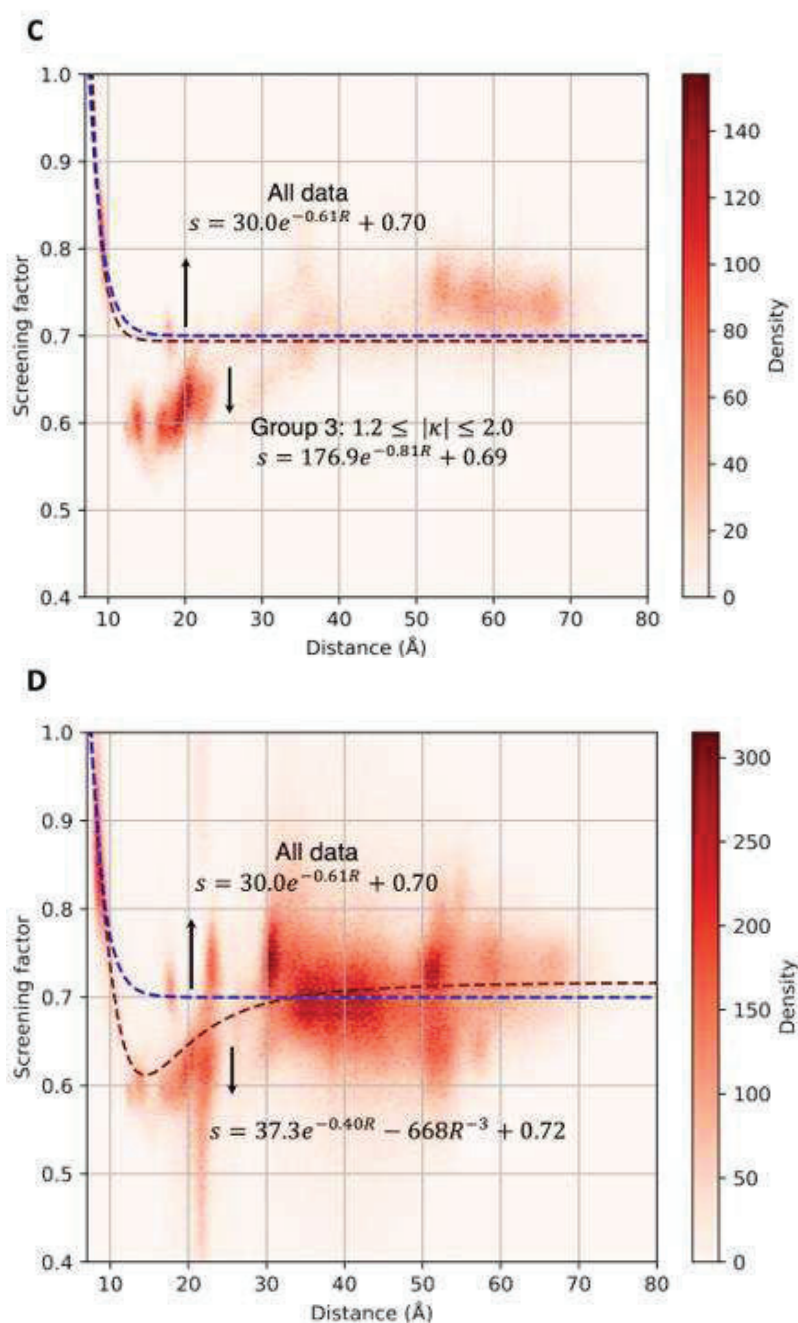


Figure 3 A-1. Density distribution of screening factors as a function of D/A separation derived from TrESP/MMPol electronic couplings computed for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems along MD trajectories. Overall results are fitted to the empirical formula $s = 30.0e^{-0.61R} + 0.70$ and compared to analogous expressions fitted to subsets of the data based on κ dipole orientation factor: A Group 1 ($0.0 \leq \kappa \leq 0.6$); B Group 2: $0.6 \leq \kappa \leq 1.2$ and C Group 3: $1.2 \leq \kappa \leq 2.0$. D Comparison of the exponential empirical function with the

analogous expression with an additional R^{-3} term to account for the minimum at separations ~ 15 Å.

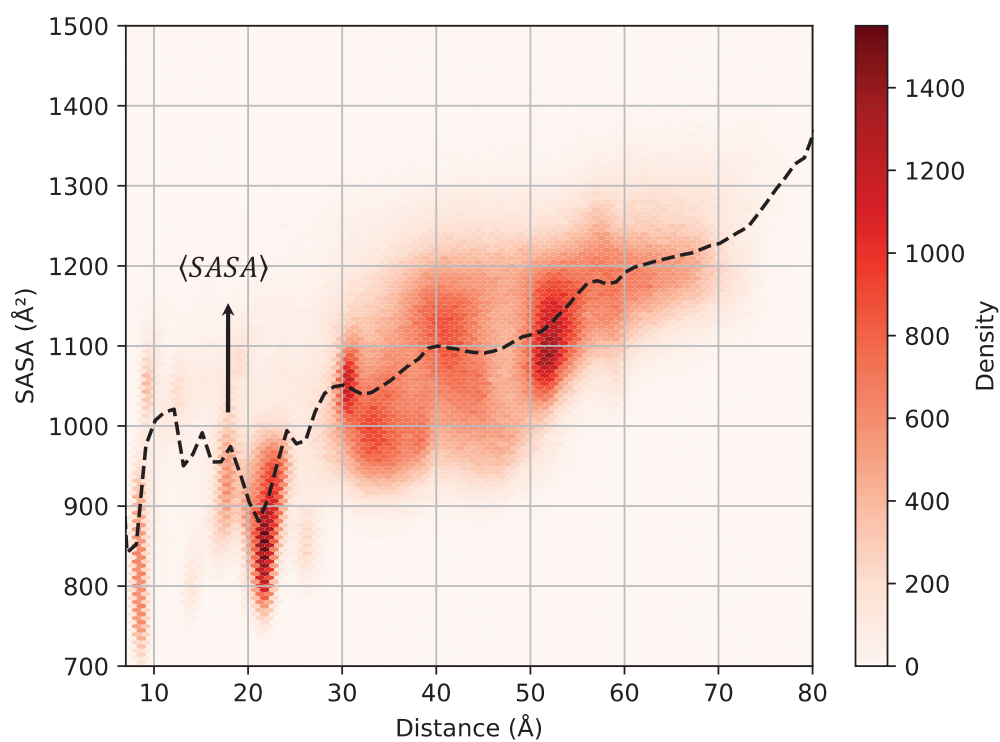


Figure 4-A1. Contribution of the AF488 and TRC2 dyes to the total solvent-accessible surface area (SASA) of the *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems computed along MD trajectories. Surface area computed using the LCPO algorithm of Weiser et al.¹⁸⁰

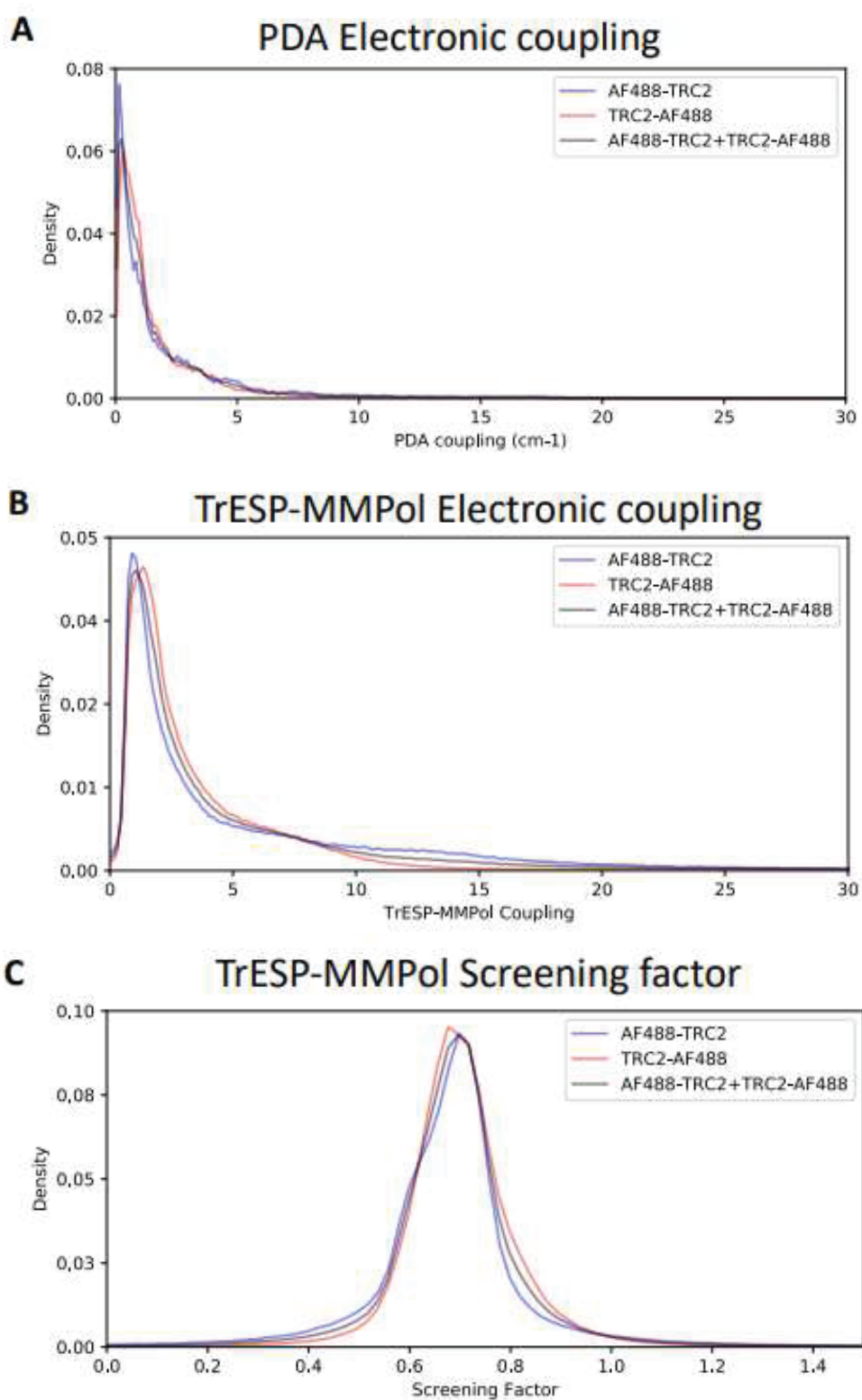


Figure 5-A1. Electronic couplings and screening factors computed along MD trajectories for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems.

APPENDIX 2

APPENDIX 2: for Chapter 5: Impact of calcium binding and dye labelling in the structural ensemble of calmodulin studied using atomistic energy transfer simulations

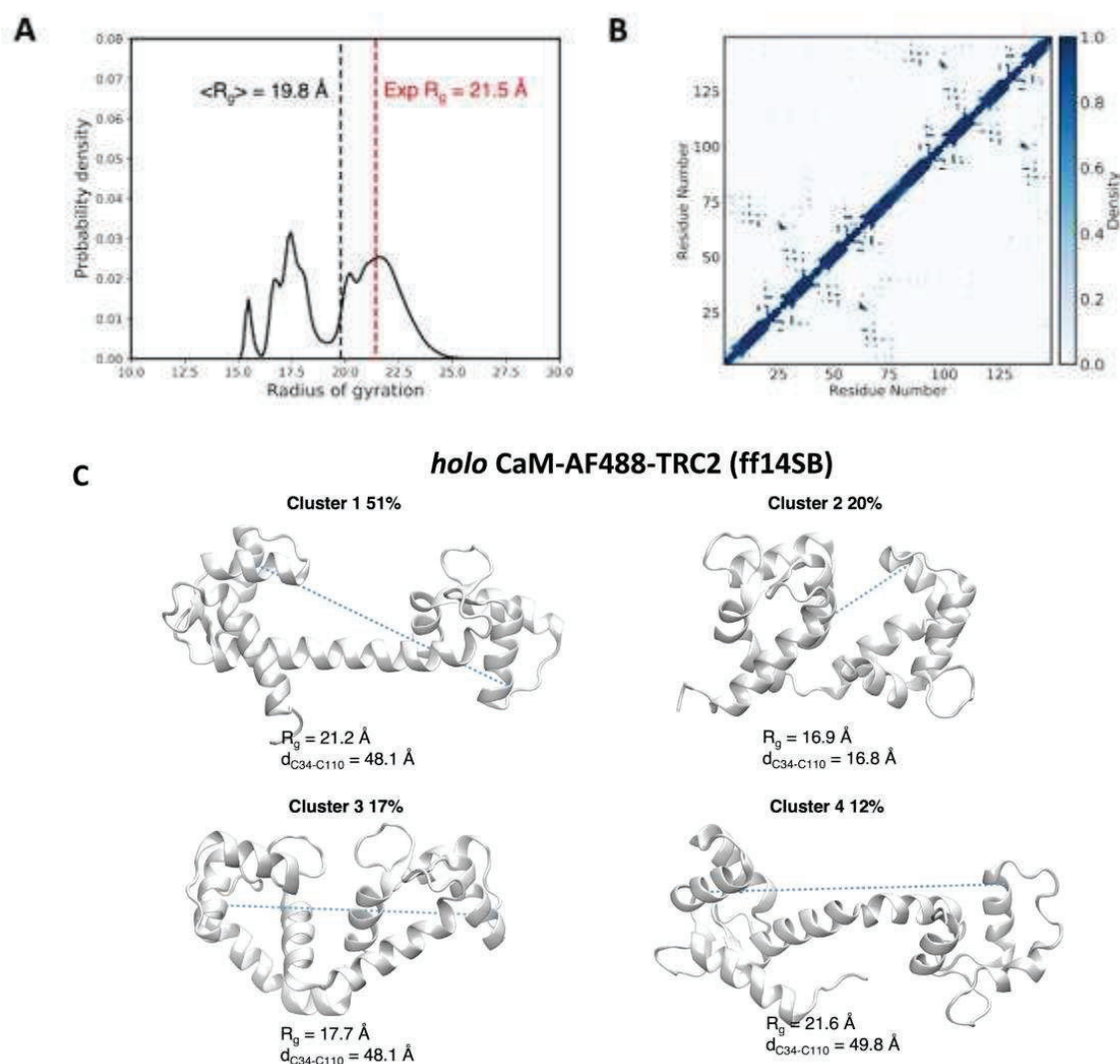


Figure 1-A2. Structural characterization of the ensemble of *holo* CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on ff14SB. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

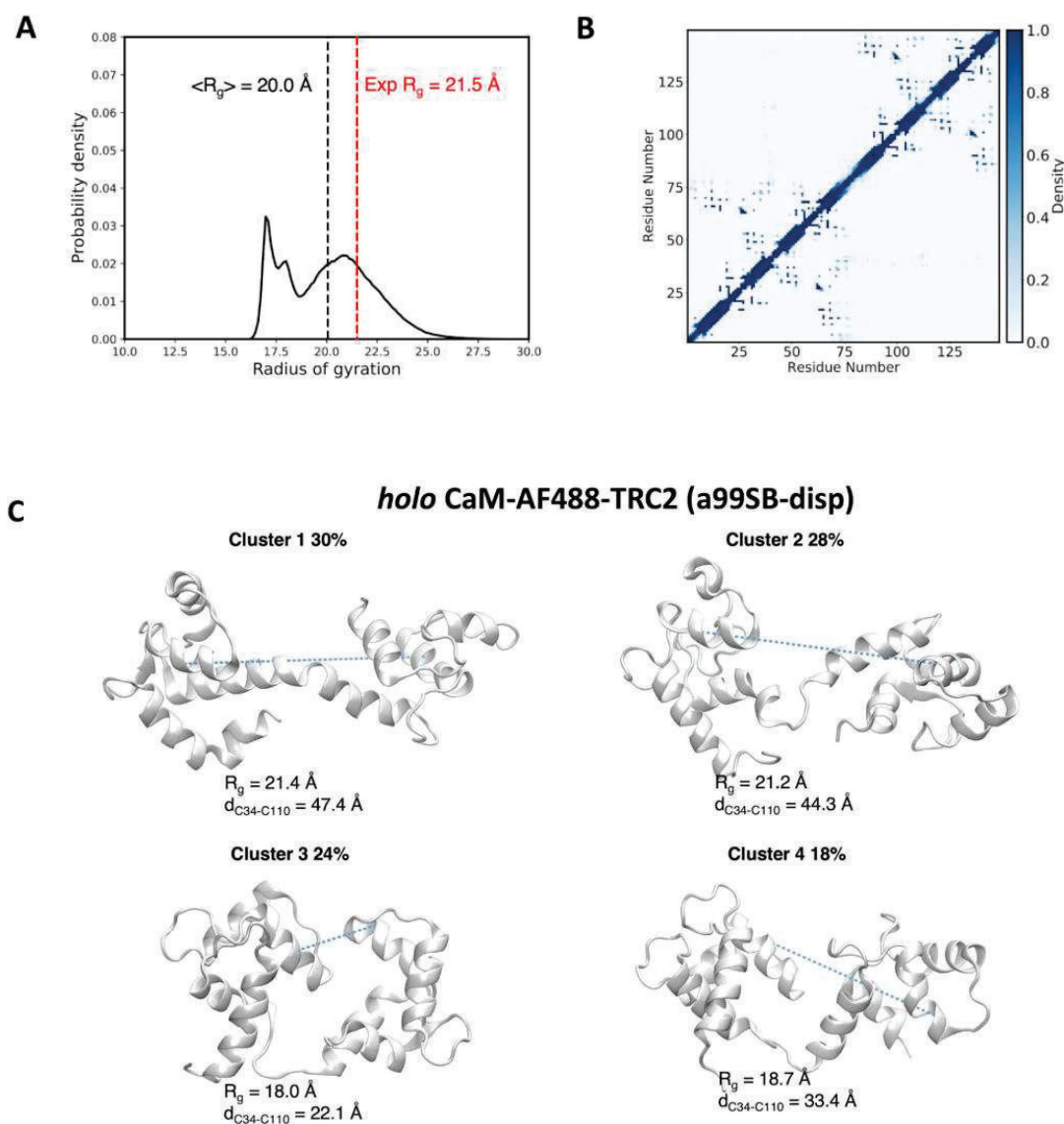


Figure 2-A2. Structural characterization of the ensemble of holo CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on a99SB-disp. A Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. B Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. C Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

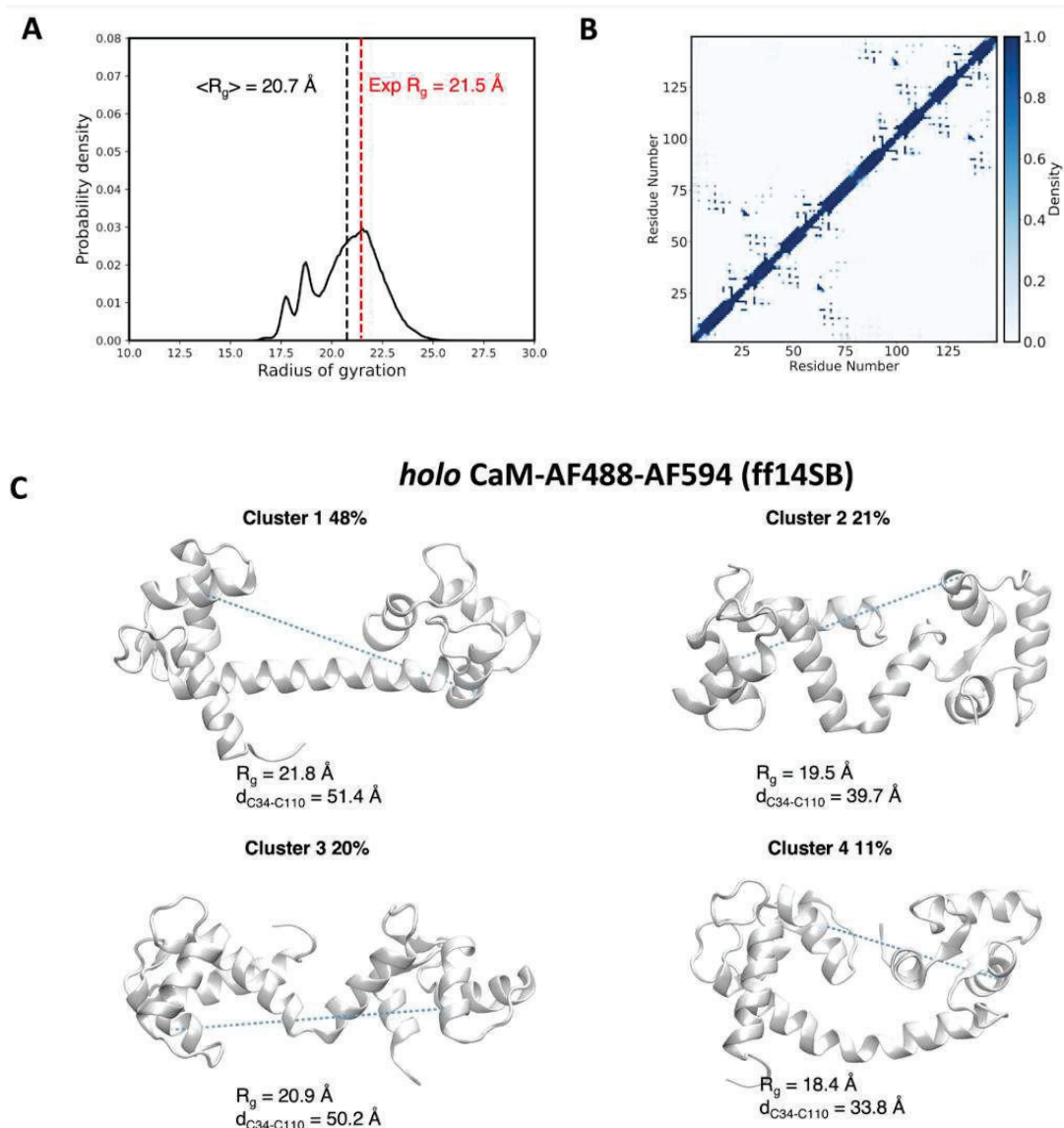


Figure 3-A2. Structural characterization of the ensemble of *holo* CaM-AF488-AF594/AF594-AF488 from MD simulations based on ff14SB. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

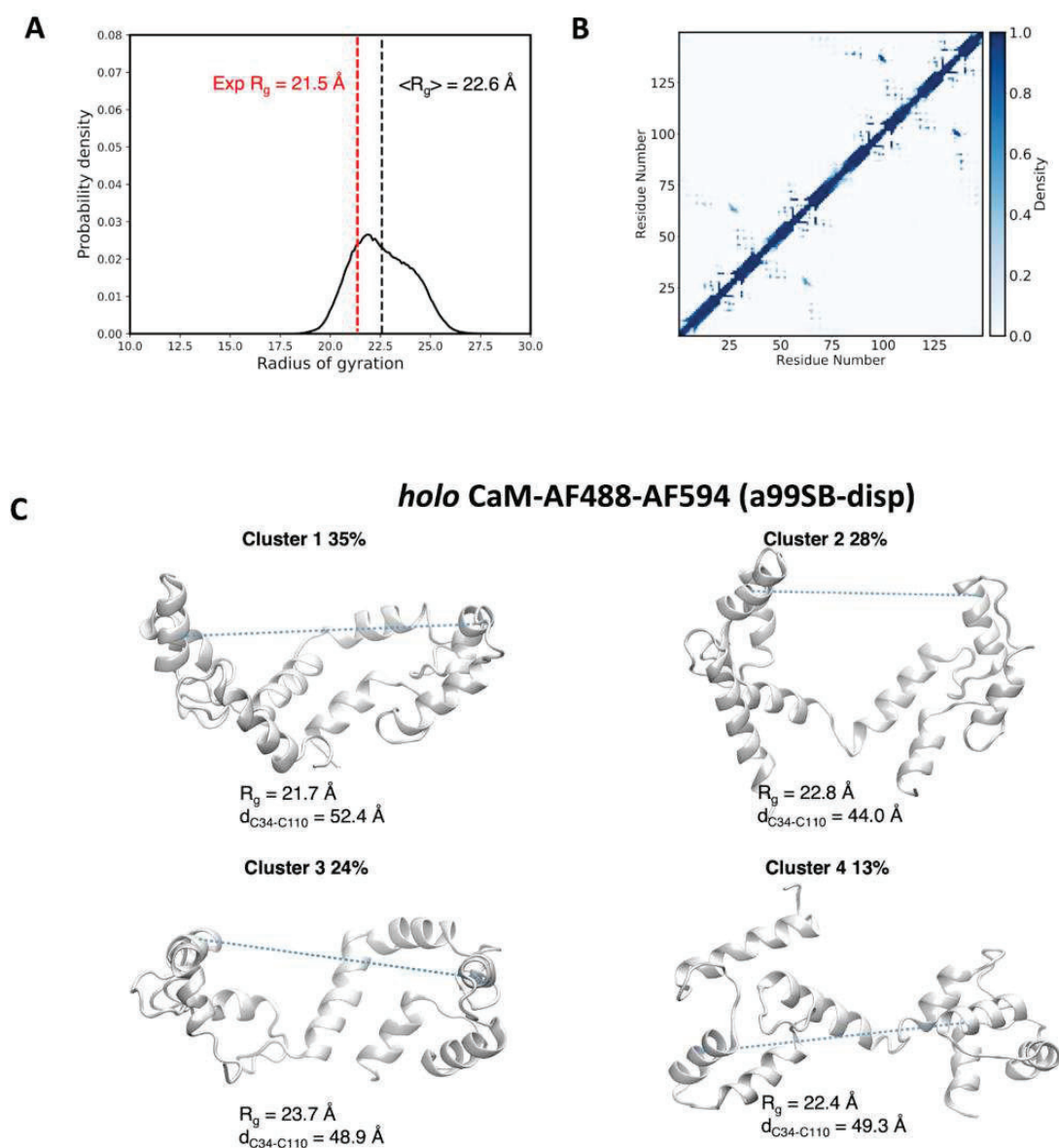


Figure 4-A2. Structural characterization of the ensemble of *holo* CaM-AF488-AF594/AF594-AF488 from MD simulations based on a99SB-disp. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

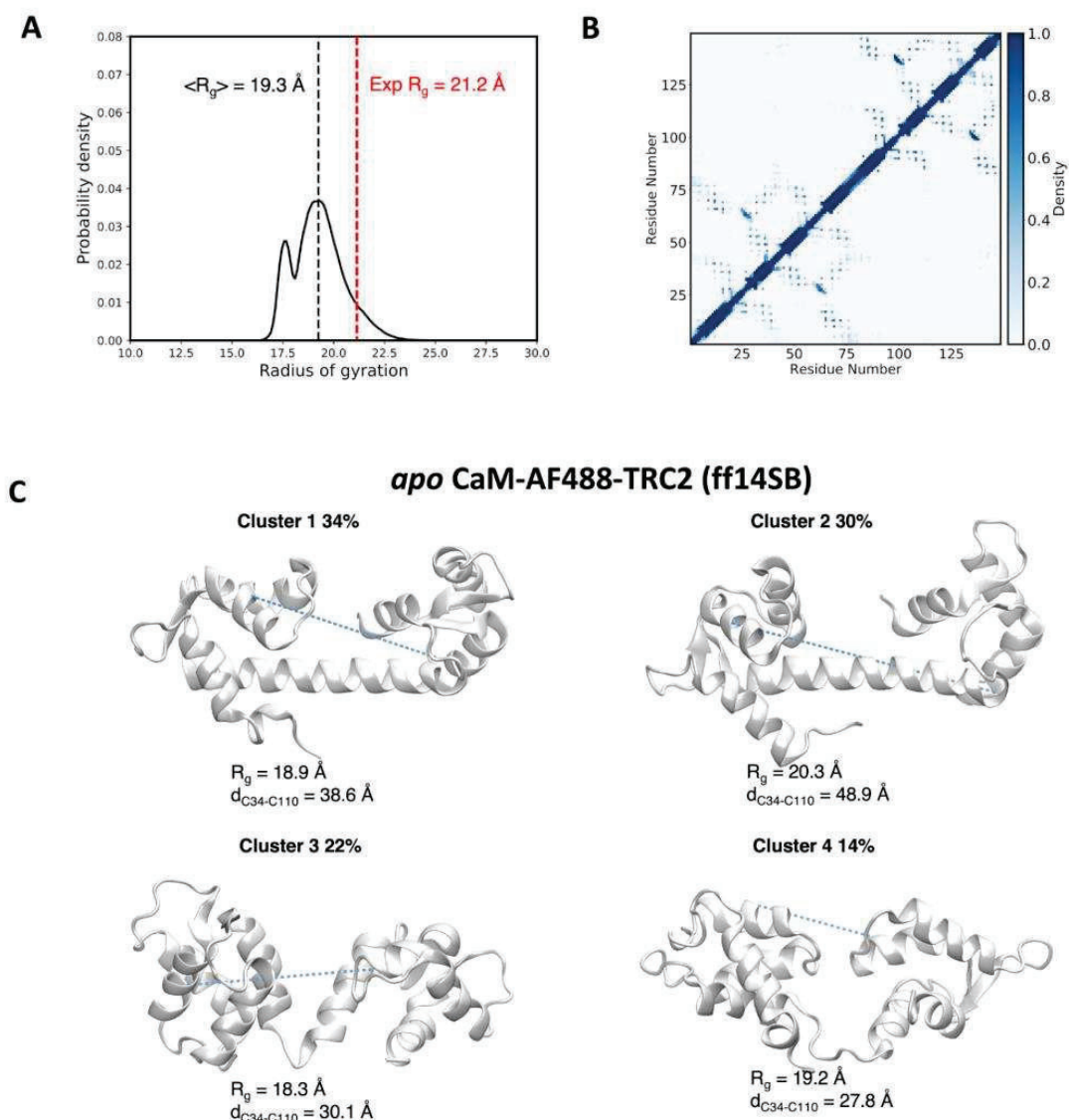


Figure 5-A2. Structural characterization of the ensemble of *apo* CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on ff14SB. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

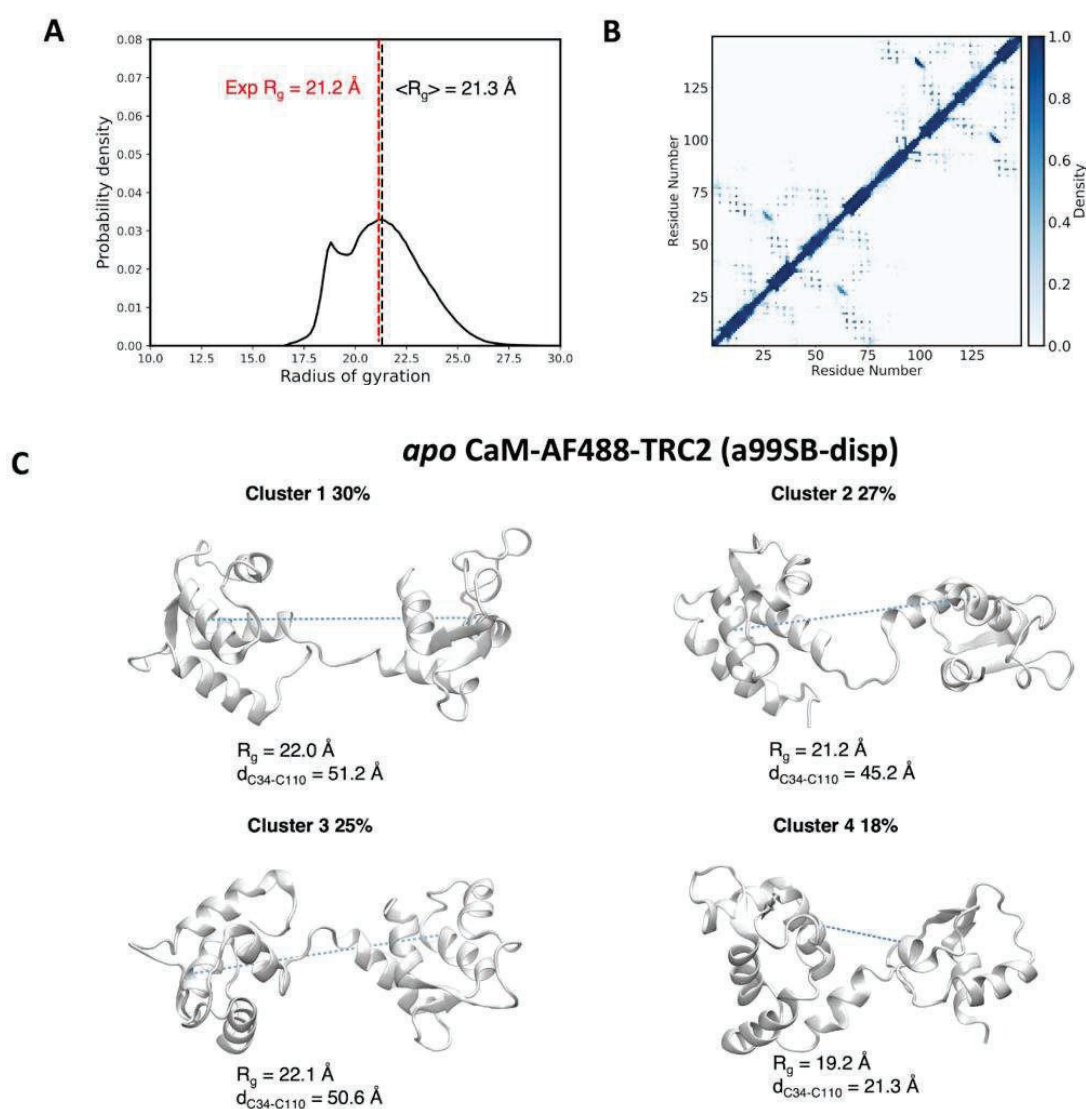


Figure 6-A2. Structural characterization of the ensemble of *apo* CaM-AF488-TRC2/TRC2-AF488 from MD simulations based on a99SB-disp. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 C α -C α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

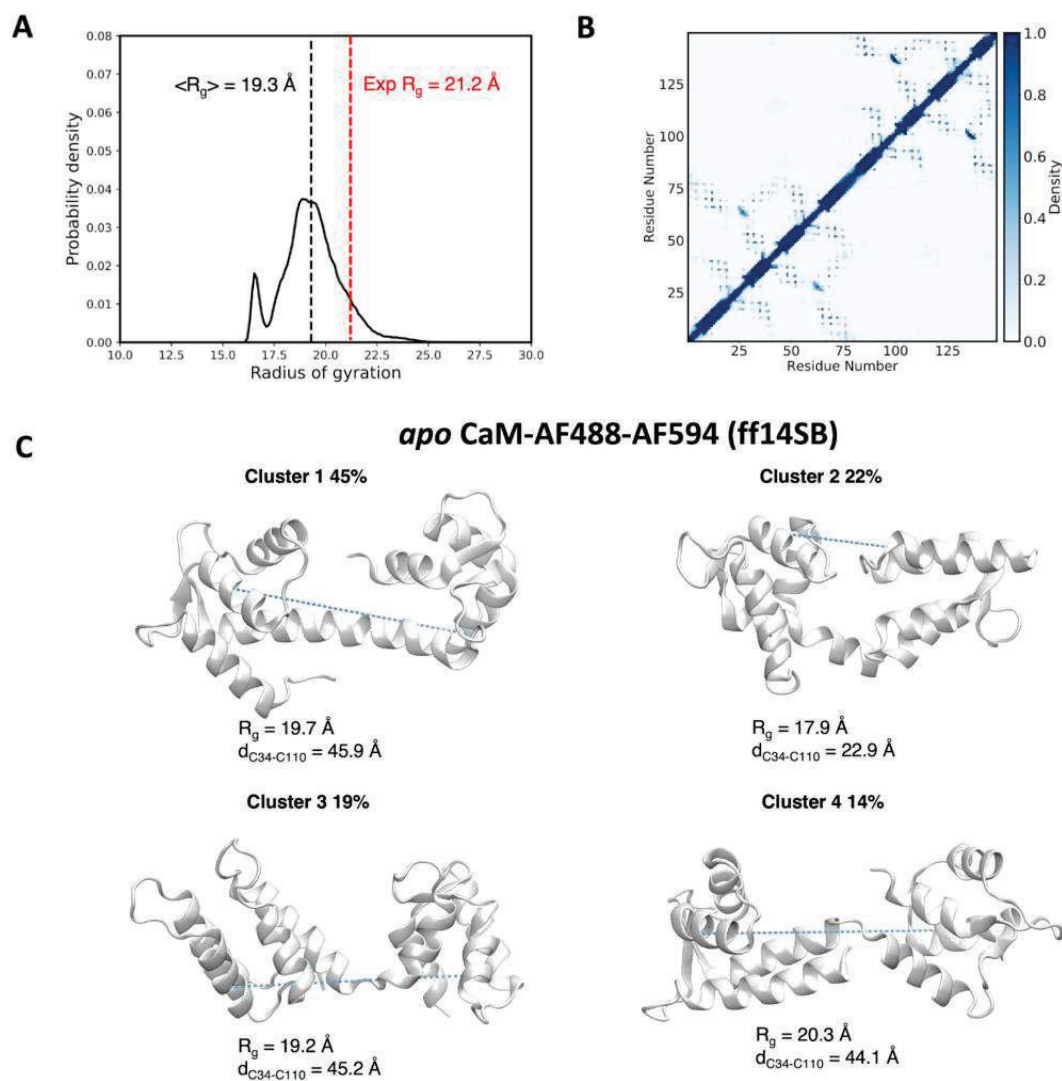


Figure 7-A2. Structural characterization of the ensemble of *apo* CaM-AF488-AF594/AF594-AF488 from MD simulations based on ff14SB. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

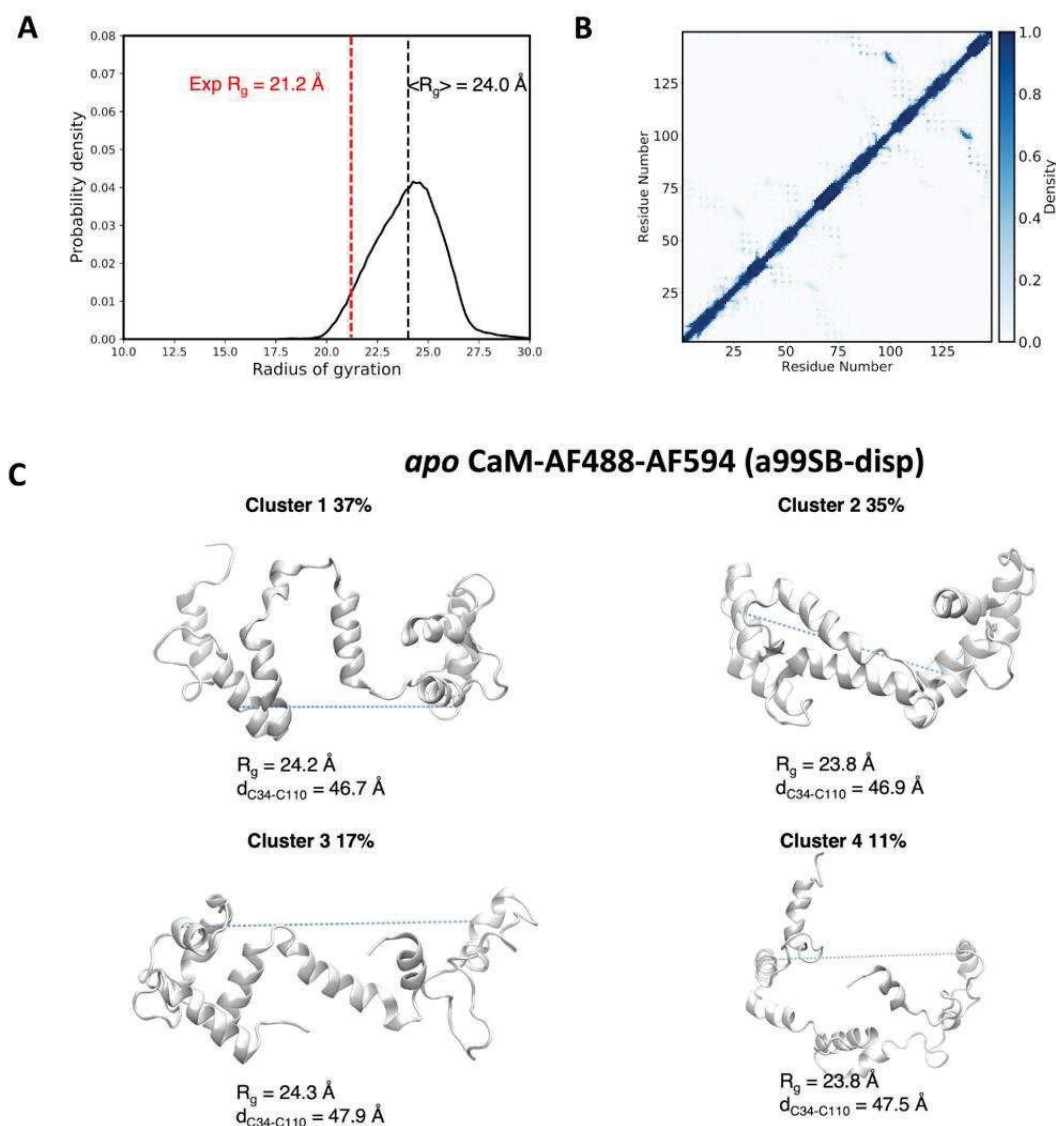


Figure 8-A2. Structural characterization of the ensemble of *apo* CaM-AF488-AF594/AF594-AF488 from MD simulations based on a99SB-disp. **A** Distribution of radius of gyration computed along the MD trajectories compared to the experimental value. **B** Intramolecular contact map averaged over the MD showing the folding patterns of the two domains in CaM. **C** Major clusters for the structural ensemble. For each cluster we draw the most representative structure (centroid) and report its population and the MD-averaged radius of gyration and C34-C110 α - α distances between Cys residues linking the dyes in the centroid structures (distances represented by dashed lines).

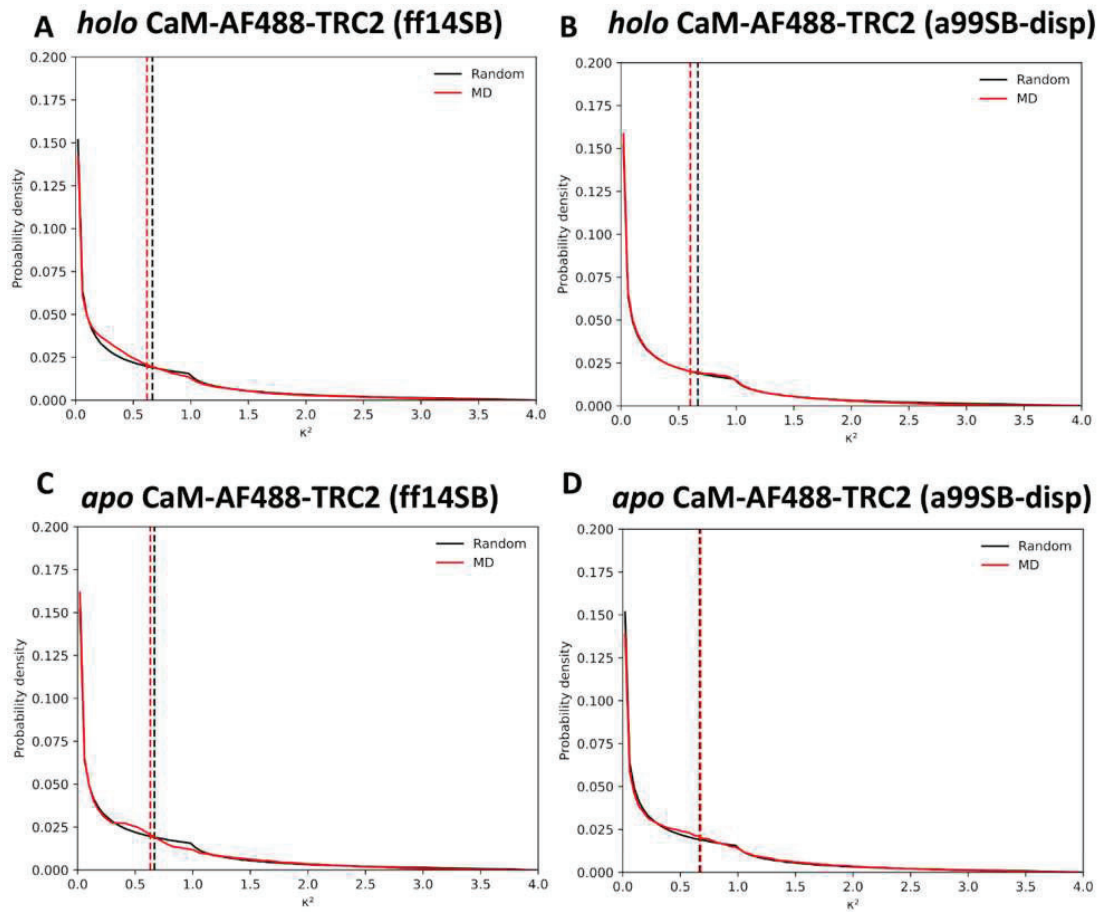


Figure 9-A2. Probability distribution of orientation factors derived from MD simulations for CaM-AF488-TRC2/AF488 systems compared to the random isotropic distribution (average values indicated by dashed vertical lines). **A** *holo* CaM-AF488-TRC2 ff14SB. **B** *holo* CaM-AF488-TRC2 a99SB-disp. **C** *apo* CaM-AF488-TRC2 ff14SB. **D** *apo* CaM-AF488-TRC2 a99SB-disp.

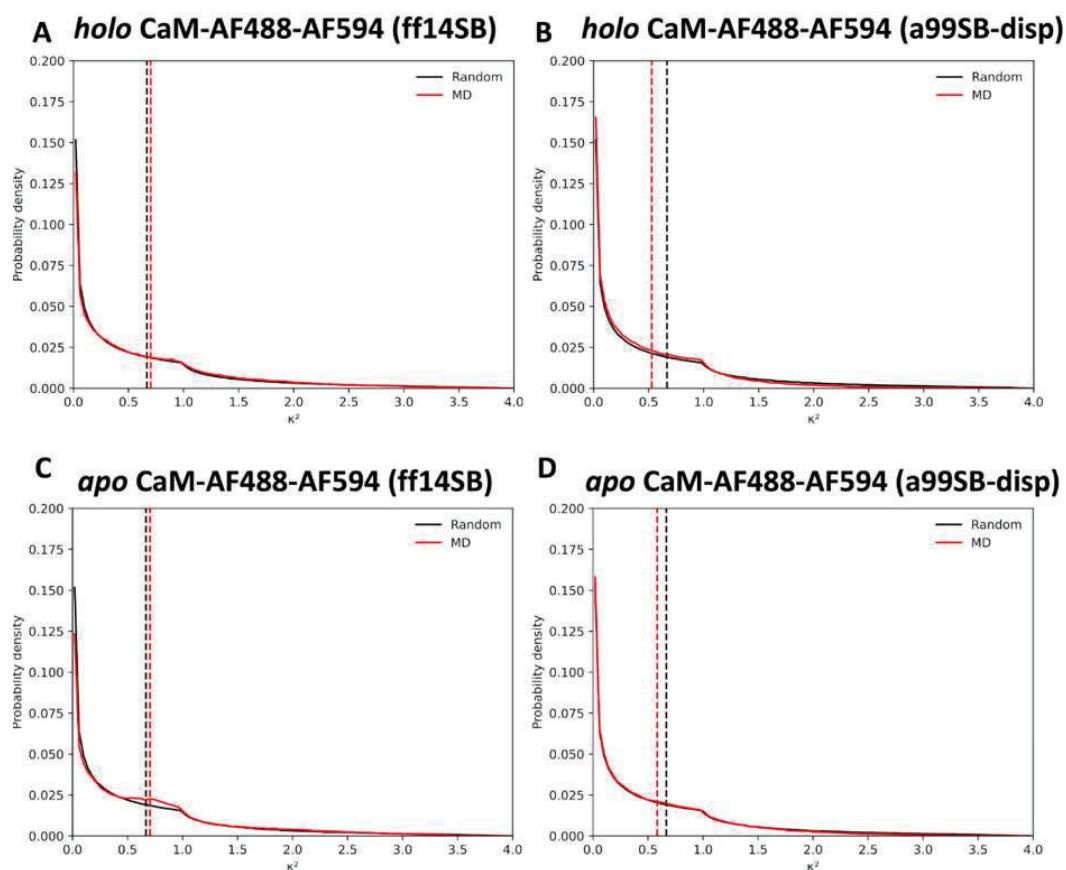


Figure 10-A2. Probability distribution of orientation factors derived from MD simulations for CaM-AF488-AF594/AF594-AF488 systems compared to the random isotropic distribution (average values indicated by dashed vertical lines). **A** *holo* CaM-AF488-AF594 ff14SB. **B** *holo* CaM-AF488-AF594 a99SB-disp. **C** *apo* CaM-AF488-AF594 ff14SB. **D** *apo* CaM-AF488-AF594 a99SB-disp.

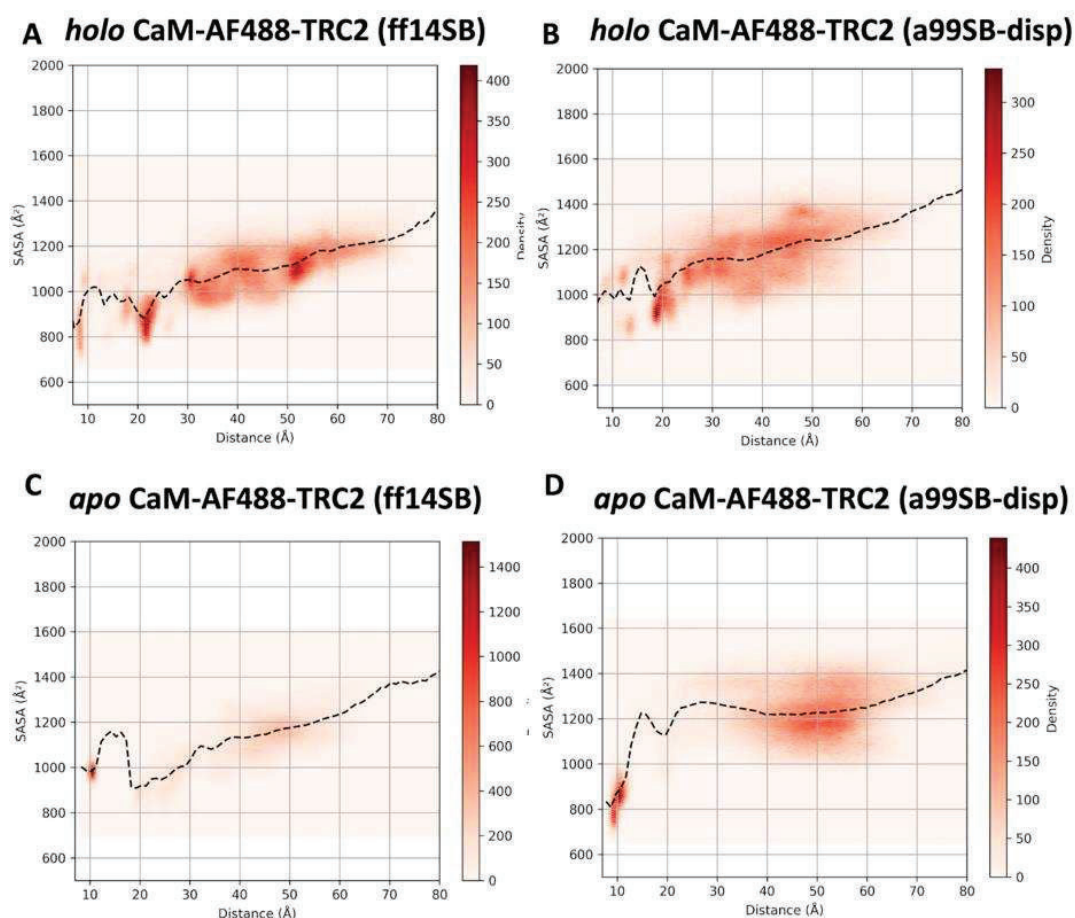


Figure 11-A2. Contribution of the AF488 and TRC2 dyes to the total solvent-accessible surface area (SASA) of the CaM-AF488-TRC2 and CaM-TRC2-AF488 systems computed along MD trajectories. Surface area computed using the LCPO algorithm of Weiser et al.¹⁸⁰ **A** *holo* CaM-AF488-TRC2 ff14SB. **B** *holo* CaM-AF488-TRC2 a99SB-disp. **C** *apo* CaM-AF488-TRC2 ff14SB. **D** *apo* CaM-AF488-TRC2 a99SB-disp.

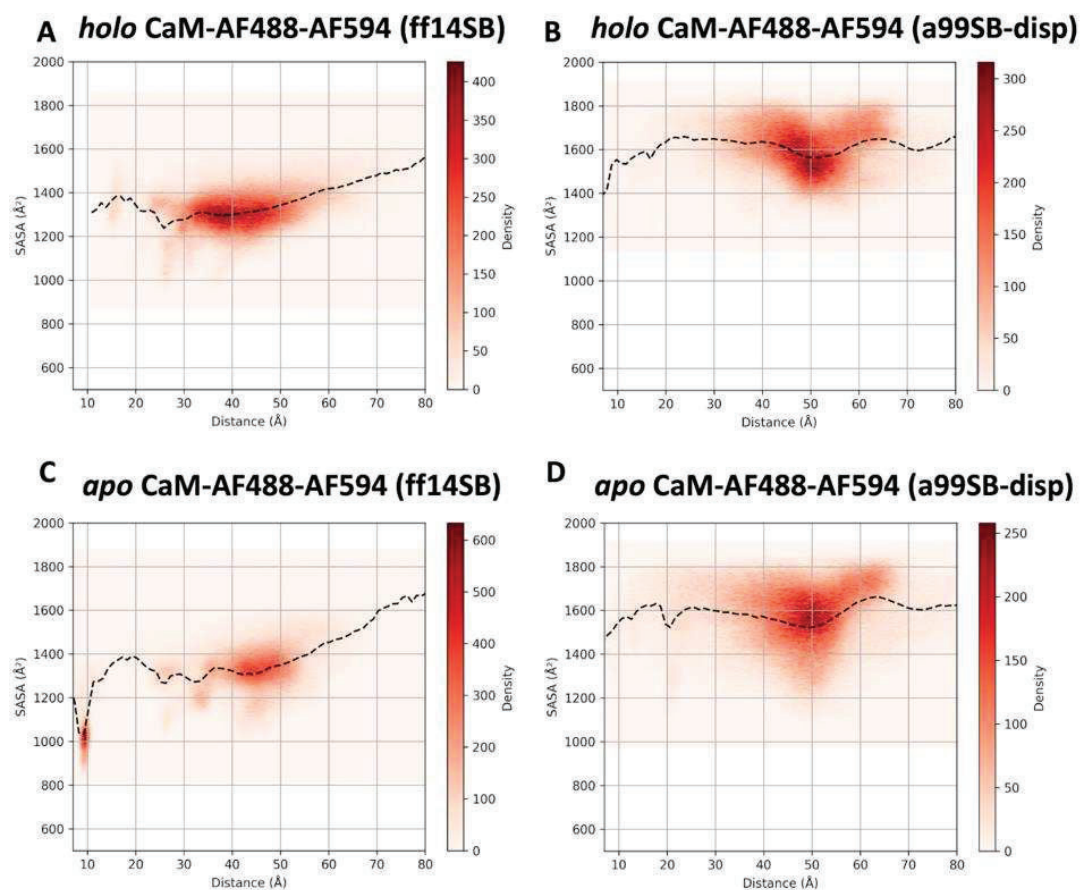


Figure 12-A2. Contribution of the AF488 and AF594 dyes to the total solvent-accessible surface area (SASA) of the CaM-AF488-AF594 and CaM-AF594-AF488 systems computed along MD trajectories. Surface area computed using the LCPO algorithm of Weiser et al.¹⁸⁰ **A** *holo* CaM-AF488-AF594 ff14SB. **B** *holo* CaM-AF488-AF594 a99SB-disp. **C** *apo* CaM-AF488-AF594 ff14SB. **D** *apo* CaM-AF488-AF594 a99SB-disp.

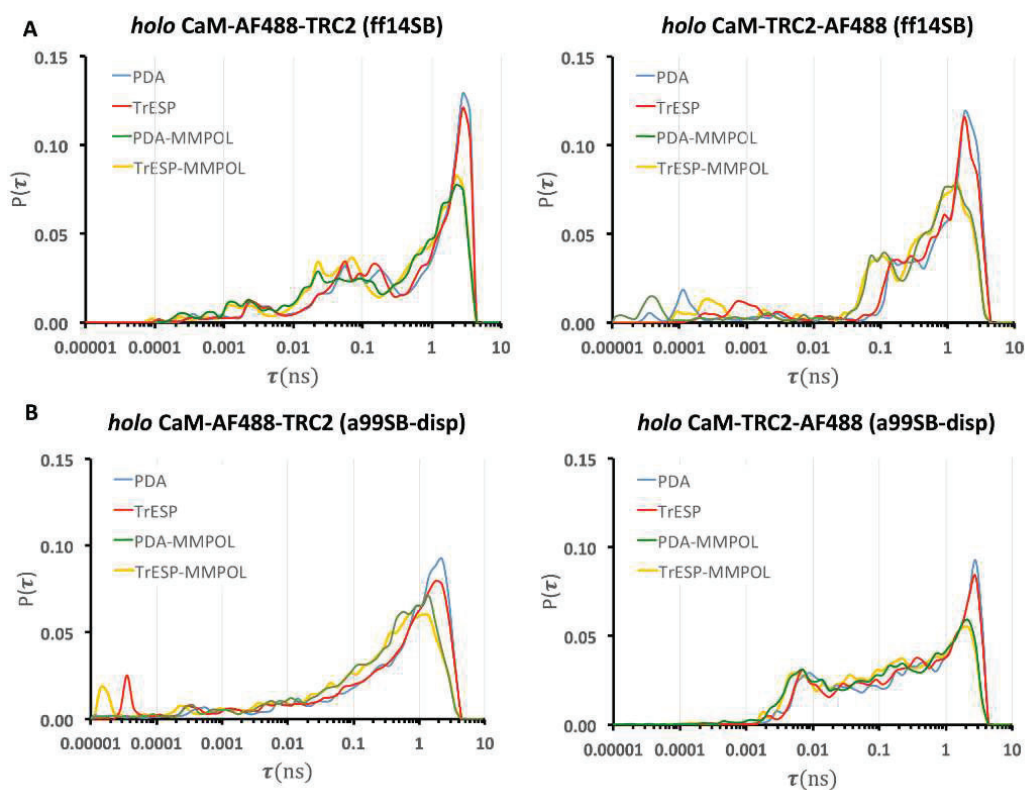


Figure 13-A2. Distribution of fluorescence lifetimes computed for *holo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems using different electronic coupling models (PDA: Point dipole approximation + Förster screening factor. PDA-MMPOL: Point dipole approximation + MMPOL screening factor. TrESP: TrESP coulombic term + Förster screening factor. TrESP-MMPOL: TrESP coulombic term + MMPOL screening factor): **A** *holo* CaM-AF488-TRC2/TRC2-AF488 (ff14SB). **B** *holo* CaM-AF488-TRC2/TRC2-AF488 (a99SB-disp).

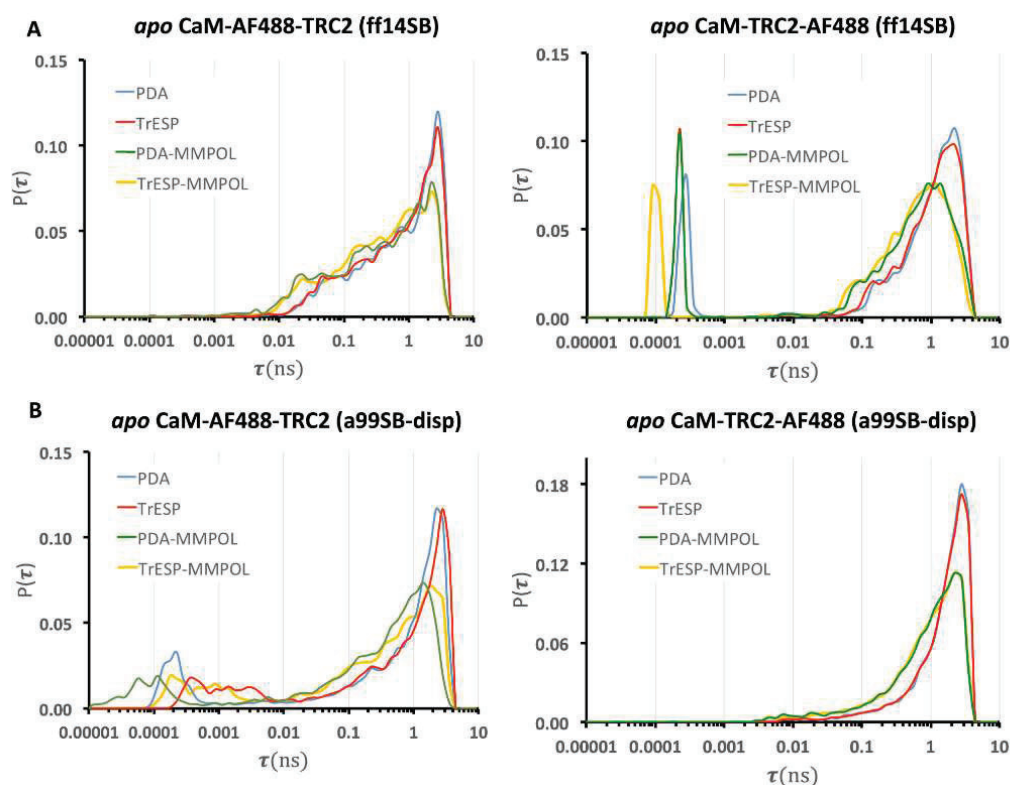


Figure 14-A2. Distribution of fluorescence lifetimes computed for *apo* CaM-AF488-TRC2 and CaM-TRC2-AF488 systems using different electronic coupling models (PDA: Point dipole approximation + Förster screening factor. PDA-MMPol: Point dipole approximation + MMPol screening factor. TrESP: TrESP coulombic term + Förster screening factor. TrESP-MMPol: TrESP coulombic term + MMPol screening factor): **A** *apo* CaM-AF488-TRC2/TRC2-AF488 (ff14SB). **B** *apo* CaM-AF488-TRC2/TRC2-AF488 (a99SB-disp).

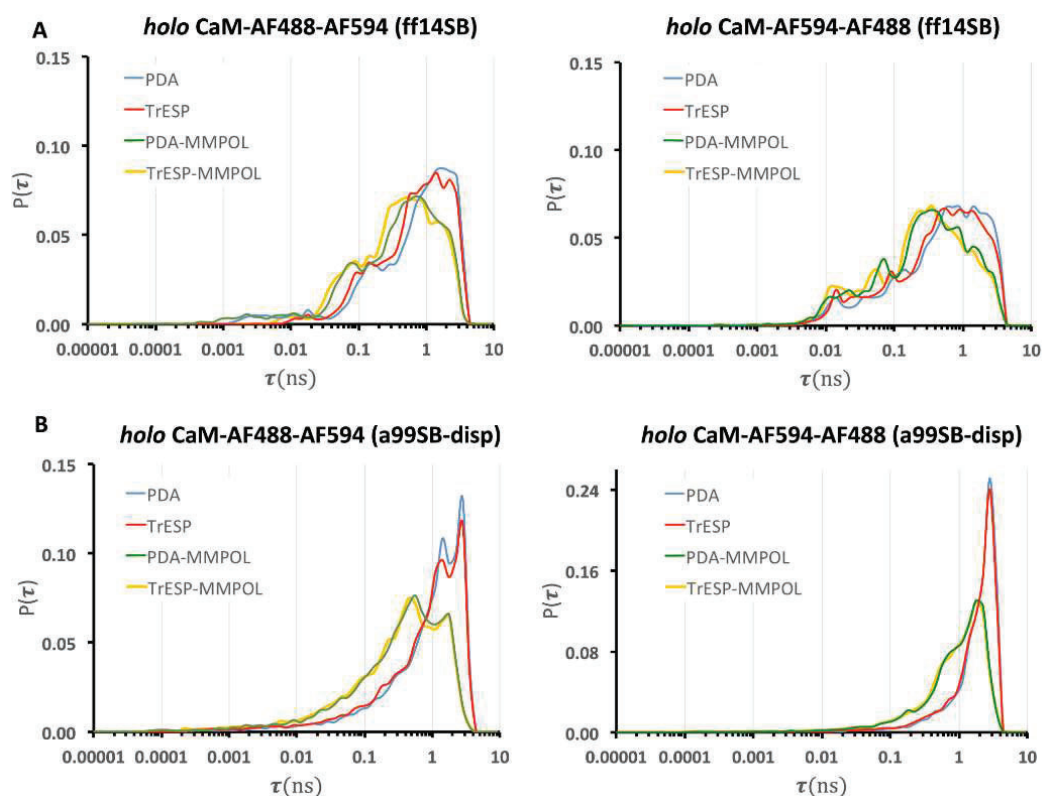


Figure 15-A2. Distribution of fluorescence lifetimes computed for *holo* CaM-AF488-AF594 and CaM-AF594-AF488 systems using different electronic coupling models (PDA: Point dipole approximation + Förster screening factor. PDA-MMPol: Point dipole approximation + MMPol screening factor. TrESP: TrESP coulombic term + Förster screening factor. TrESP-MMPol: TrESP coulombic term + MMPol screening factor): **A** *holo* CaM-AF488-AF594/ AF594-AF488 (ff14SB). **B** *holo* CaM-AF488-AF594/AF594-AF488 (a99SB-disp).

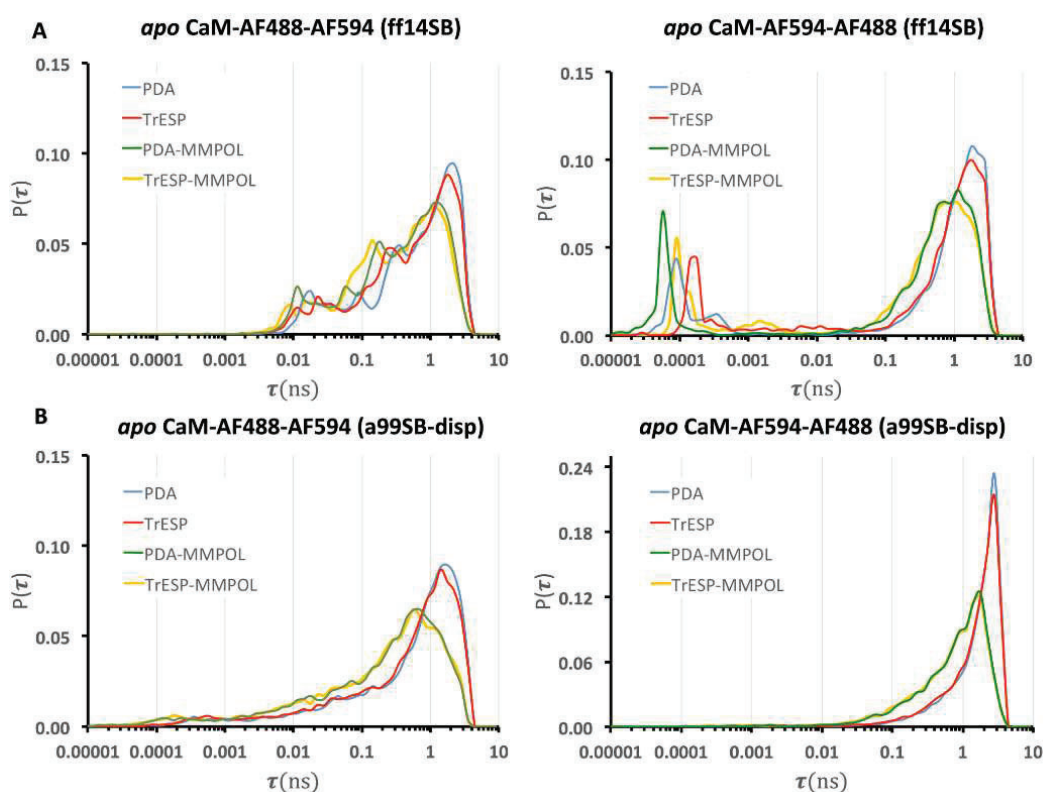


Figure 16-A2. Distribution of fluorescence lifetimes computed for *apo* CaM-AF488-AF594 and CaM-AF594-AF488 systems using different electronic coupling models (PDA: Point dipole approximation + Förster screening factor. PDA-MMPol: Point dipole approximation + MMPol screening factor. TrESP: TrESP coulombic term + Förster screening factor. TrESP-MMPol: TrESP coulombic term + MMPol screening factor): **A** *apo* CaM-AF488-AF594/ AF594-AF488 (ff14SB). **B** *apo* CaM-AF488-AF594/AF594-AF488 (a99SB-disp).

BIBLIOGRAPHY

- 1 Fisher, C.K. & Stultz, C.M. Constructing ensembles for intrinsically disordered proteins. *Curr Opin Struct Biol.* **21**: 426–431 (2011).
- 2 Clapham, D.E. Calcium signaling. *Cell.* **80**: 259–268 (1995).
- 3 Takemoto-Kimura, S., Suzuki, K., Horigane, S., Kamijo, S., Inoue, M., Sakamoto, M., *et al.* Calmodulin kinases: essential regulators in health and disease. *J Neurochem.* **141**: 808–818 (2017).
- 4 Patodia, S. Molecular Dynamics Simulation of Proteins: A Brief Overview. *J Phys Chem Biophys.* **4** (2014).
- 5 Pandey, K., Dhoke, R.R., Rathore, Y.S., Nath, S.K., Verma, N., Bawa, S., *et al.* Low pH overrides the need of calcium ions for the shape–function relationship of calmodulin: resolving prevailing debates. *J Phys Chem B.* **118**: 5059–5074 (2014).
- 6 Okamoto, K. & Sako, Y. Recent advances in FRET for the study of protein interactions and dynamics. *Curr Opin Struct Biol.* **46**: 16–23 (2017).
- 7 Best, R.B., Hofmann, H., Nettels, D. & Schuler, B. Quantitative Interpretation of FRET Experiments via Molecular Simulation: Force Field and Validation. *Biophys J.* **108**: 2721–2731 (2015).
- 8 Deplazes, E., Jayatilaka, D. & Corry, B. Testing the use of molecular dynamics to simulate fluorophore motions and FRET. *Physical Chemistry Chemical Physics.* **13**: 11045 (2011).
- 9 Kramer, H.E.A. & Fischer, P. The scientific work of Theodor Förster: A brief sketch of his life and personality. *ChemPhysChem.* **12** (2011).
- 10 Caprasecca, S., Jurinovich, S., Viani, L., Curutchet, C. & Mennucci, B. Geometry Optimization in Polarizable QM/MM Models: The Induced Dipole Formulation. *J Chem Theory Comput.* **10**: 1588–1598 (2014).
- 11 Cignoni, E., Cupellini, L. & Mennucci, B. A fast method for electronic couplings in embedded multichromophoric systems. *Journal of Physics Condensed Matter.* **34** (2022).

- 12 Beljonne, D., Curutchet, C., Scholes, G.D. & Silbey, R.J. Beyond förster resonance energy transfer in biological and nanoscale systems. *Journal of Physical Chemistry B*. **113**: 6583–6599 (2009).
- 13 Förster, Th. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann Phys*. **437**: 55–75 (1948).
- 14 Medintz, I.L. & Hildebrandt, N. *FRET-Förster resonance energy transfer: from theory to applications*. John Wiley & Sons, 2013.
- 15 Algar, W.R., Hildebrandt, N., Vogel, S.S. & Medintz, I.L. FRET as a biomolecular research tool — understanding its potential while avoiding pitfalls. *Nat Methods*. **16**: 815–829 (2019).
- 16 Lerner, E., Cordes, T., Ingargiola, A., Alhadid, Y., Chung, S.Y., Michalet, X., *et al.* Toward dynamic structural biology: Two decades of single-molecule förster resonance energy transfer. *Science* (1979). . **359** (2018).
- 17 Roy, P., Sury, A.S. & Pillai, P.P. Resonance energy transfer in electrostatically assembled donor-acceptor system based on blue-emitting InP quantum dots. *Chemical Physics Impact*. **7**: 100334 (2023).
- 18 Burrows, H.D., Fonseca, S.M., Dias, F.B., de Melo, J.S., Monkman, A.P., Scherf, U., *et al.* Singlet Excitation Energy Harvesting and Triplet Emission in the Self-Assembled System Poly{1,4-phenylene-[9,9-bis (4-phenoxy-butylsulfonate)]fluorene-2,7-diyl} copolymer/tris(bipyridyl)ruthenium(II) in Aqueous Solution. *Advanced Materials*. **21**: 1155–1159 (2009).
- 19 Bauri, J. & Choudhary, R.B. FRET mechanism to enhance the quantum yield of the PCz/gC₃N₄ nanocomposite, an emissive material for OLED applications. *Physical Chemistry Chemical Physics*. **25**: 22195–22210 (2023).
- 20 Roy, P., Sury, A.S. & Pillai, P.P. Resonance energy transfer in electrostatically assembled donor-acceptor system based on blue-emitting InP quantum dots. *Chemical Physics Impact*. **7**: 100334 (2023).
- 21 Gopal, A.R., Joy, F., Dutta, V., Devasia, J., Dateer, R. & Nizam, A. Carbon Dot-Based Fluorescence Resonance Energy Transfer (FRET)

- Systems for Biomedical, Sensing, and Imaging Applications. *Particle & Particle Systems Characterization*. **41** (2024).
- 22 Xiang, H., Valandro, S.R. & Hill, E.H. 2D Nanomaterial-Directed Molecular Aggregation and Energy Transfer between Edge-Bound Donor–Acceptor Pairs. *The Journal of Physical Chemistry C*. **127**: 15416–15422 (2023).
 - 23 Berdnikova, D. V. & Chernikova, E.Y. Fluorescence and Phosphorescence Energy Transfer in Cucurbituril-Based Supramolecular Systems. *ChemPhotoChem*. **8** (2024).
 - 24 Lezhennikova, K., Rustomji, K., Kuhlmei, B.T., Antonakakis, T., Jomin, P., Glybovski, S., *et al.* Experimental evidence of Förster energy transfer enhancement in the near field through engineered metamaterial surface waves. *Commun Phys*. **6**: 229 (2023).
 - 25 Kumar, V., Chunchagatta Lakshman, P.K., Prasad, T.K., Manjunath, K., Bairy, S., Vasu, A.S., *et al.* Target-based drug discovery: Applications of fluorescence techniques in high throughput and fragment-based screening. *Heliyon*. **10**: e23864 (2024).
 - 26 Zheng, C., Zhang, X., Gao, A., Ju, M., Hou, A. & Xie, K. Adjustable multimodality fluorescence emissions and photochromism of functional molecular device containing double chromophores with intramolecular FRET process. *Dyes and Pigments*. **228**: 112238 (2024).
 - 27 Chen, B., Su, Q., Kong, W., Wang, Y., Shi, P. & Wang, F. Energy transfer-based biodetection using optical nanomaterials. *J Mater Chem B*. **6**: 2924–2944 (2018).
 - 28 Cupellini, L., Corbella, M., Mennucci, B. & Curutchet, C. Electronic energy transfer in biomacromolecules. *WIREs Computational Molecular Science*. **9** (2019).
 - 29 Braslavsky, S.E., Fron, E., Rodríguez, H.B., Román, E.S., Scholes, G.D., Schweitzer, G., *et al.* Pitfalls and limitations in the practical use of Förster’s theory of resonance energy transfer. *Photochemical and Photobiological Sciences*. **7** (2008).
 - 30 De Filippis, G., Cataudella, V., Mishchenko, A.S., Perroni, C.A. & Devreese, J.T. Validity of the Franck-Condon Principle in the Optical

- Spectroscopy: Optical Conductivity of the Fröhlich Polaron. *Phys Rev Lett.* **96**: 136405 (2006).
- 31 Sahoo, H. Förster resonance energy transfer – A spectroscopic nanoruler: Principle and applications. *Journal of Photochemistry and Photobiology C: Photochemistry Reviews.* **12**: 20–30 (2011).
 - 32 Scholes, G.D. Long-Range Resonance Energy Transfer in Molecular Systems. *Annu Rev Phys Chem.* **54**: 57–87 (2003).
 - 33 der Meer, B.W. Kappa-squared: from nuisance to new sense. *Reviews in Molecular Biotechnology.* **82**: 181–196 (2002).
 - 34 Knox, R.S. & Van Amerongen, H. Refractive Index Dependence of the Förster Resonance Excitation Transfer Rate. . (2002).
 - 35 Sobakinskaya, E., Schmidt am Busch, M. & Renger, T. Theory of FRET “Spectroscopic Ruler” for Short Distances: Application to Polyproline. *J Phys Chem B.* **122**: 54–67 (2018).
 - 36 Stryer, L. & Haugland, R.P. Energy transfer: a spectroscopic ruler. *Proceedings of the National Academy of Sciences.* **58**: 719–726 (1967).
 - 37 Lindsey, J.S., Taniguchi, M., Bocian, D.F. & Holten, D. The fluorescence quantum yield parameter in Förster resonance energy transfer (FRET)—Meaning, misperception, and molecular design. *Chemical Physics Reviews.* **2** (2021).
 - 38 Perrin, F. La fluorescence des solutions. *Ann Phys (Paris).* **10** (1929). .
 - 39 Förster, Th. & Forster, Th. Transfer Mechanisms of Electronic Excitation Energy. *Radiat Res Suppl.* **2**: 326 (1960).
 - 40 Dale, R.E., Eisinger, J. & Blumberg, W.E. The orientational freedom of molecular probes. The orientation factor in intramolecular energy transfer. *Biophys J.* **26**: 161–193 (1979).
 - 41 Selvin, P.R. [13] Fluorescence resonance energy transfer. In: *Methods in enzymology.* Elsevier, 1995, pp 300–334.
 - 42 Johnson, C.K. Calmodulin, Conformational States, and Calcium Signaling. A Single-Molecule Perspective. *Biochemistry.* **45**: 14233–14246 (2006).

- 43 Nettels, D., Galvanetto, N., Ivanović, M.T., Nüesch, M., Yang, T. & Schuler, B. Single-molecule FRET for probing nanoscale biomolecular dynamics. *Nature Reviews Physics*. **6**: 587–605 (2024).
- 44 Deniz, A.A., Laurence, T.A., Beligere, G.S., Dahan, M., Martin, A.B., Chemla, D.S., *et al.* Single-molecule protein folding: Diffusion fluorescence resonance energy transfer studies of the denaturation of chymotrypsin inhibitor 2. *Proceedings of the National Academy of Sciences*. **97**: 5179–5184 (2000).
- 45 Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. & Prasher, D.C. Green Fluorescent Protein as a Marker for Gene Expression. *Science* (1979). **263**: 802–805 (1994).
- 46 Padilla-Parra, S. & Tramier, M. FRET microscopy in the living cell: Different approaches, strengths and weaknesses. *BioEssays*. **34**: 369–376 (2012).
- 47 Karpova, T.S., Baumann, C.T., He, L., Wu, X., Grammer, A., Lipsky, P., *et al.* Fluorescence resonance energy transfer from cyan to yellow fluorescent protein detected by acceptor photobleaching using confocal microscopy and a single laser. *J Microsc.* **209**: 56–70 (2003).
- 48 Zhuang, X., Bartley, L.E., Babcock, H.P., Russell, R., Ha, T., Herschlag, D., *et al.* A Single-Molecule Study of RNA Catalysis and Folding. *Science* (1979). **288**: 2048–2051 (2000).
- 49 Ha, T., Rasnik, I., Cheng, W., Babcock, H.P., Gauss, G.H., Lohman, T.M., *et al.* Initiation and re-initiation of DNA unwinding by the Escherichia coli Rep helicase. *Nature*. **419**: 638–641 (2002).
- 50 Wang, J.-P.Z. Improved alignment of nucleosome DNA sequences using a mixture model. *Nucleic Acids Res.* **33**: 6743–6755 (2005).
- 51 Wang, X. & Hayes, J.J. Physical methods used to study core histone tail structures and interactions in solutionThis paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. *Biochemistry and Cell Biology*. **84**: 578–588 (2006).

- 52 Myong, S., Bruno, M.M., Pyle, A.M. & Ha, T. Spring-Loaded Mechanism of DNA Unwinding by Hepatitis C Virus NS3 Helicase. *Science* (1979). **317**: 513–516 (2007).
- 53 Isvoran, A., Badel, A., Craescu, C.T., Miron, S. & Miteva, M.A. Exploring NMR ensembles of calcium binding proteins: Perspectives to design inhibitors of protein-protein interactions. *BMC Struct Biol.* **11**: 24 (2011).
- 54 Truong, K. & Ikura, M. The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes in vivo. *Curr Opin Struct Biol.* **11**: 573–578 (2001).
- 55 Wallrabe, H. & Periasamy, A. Imaging protein molecules using FRET and FLIM microscopy. *Curr Opin Biotechnol.* **16**: 19–27 (2005).
- 56 Carrillo-Carrion, C., Carril, M. & Parak, W.J. Techniques for the experimental investigation of the protein corona. *Curr Opin Biotechnol.* **46**: 106–113 (2017).
- 57 Wüthrich, K. The way to NMR structures of proteins. *Nat Struct Biol.* **8**: 923–925 (2001).
- 58 Gibbs, E.B., Cook, E.C. & Showalter, S.A. Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys.* **628**: 57–70 (2017).
- 59 Mertens, H.D.T. & Svergun, D.I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol.* **172**: 128–141 (2010).
- 60 Tants, J. & Schlundt, A. Advances, Applications, and Perspectives in Small-Angle X-ray Scattering of RNA. *ChemBioChem.* **24** (2023).
- 61 Kikhney, A.G. & Svergun, D.I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett.* **589**: 2570–2577 (2015).
- 62 Mohammed, A.S.A., Soloviov, D. & Jeffries, C.M. Perspectives on solution-based small angle X-ray scattering for protein and biological macromolecule structural biology. *Physical Chemistry Chemical Physics.* **26**: 25268–25286 (2024).

- 63 Mazal, H. & Haran, G. Single-molecule FRET methods to study the dynamics of proteins at work. *Curr Opin Biomed Eng.* **12**: 8–17 (2019).
- 64 Song, Y., Madahar, V. & Liao, J. Development of FRET Assay into Quantitative and High-throughput Screening Technology Platforms for Protein–Protein Interactions. *Ann Biomed Eng.* **39**: 1224–1234 (2011).
- 65 Dietrich, A., Buschmann, V., Müller, C. & Sauer, M. Fluorescence resonance energy transfer (FRET) and competing processes in donor–acceptor substituted DNA strands: a comparative study of ensemble and single-molecule data. *Reviews in Molecular Biotechnology.* **82**: 211–231 (2002).
- 66 Gravier, J., Sancey, L., Hirsjärvi, S., Rustique, E., Passirani, C., Benoît, J.-P., *et al.* FRET Imaging Approaches for *in Vitro* and *in Vivo* Characterization of Synthetic Lipid Nanoparticles. *Mol Pharm.* **11**: 3133–3144 (2014).
- 67 Wall, M.E., Clarage, J.B. & Phillips, G.N. Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering. *Structure.* **5**: 1599–1612 (1997).
- 68 Bertini, I., Kursula, P., Luchinat, C., Parigi, G., Vahokoski, J., Wilmanns, M., *et al.* Accurate Solution Structures of Proteins from X-ray Data and a Minimal Set of NMR Data: Calmodulin–Peptide Complexes As Examples. *J Am Chem Soc.* **131**: 5134–5144 (2009).
- 69 Callaway, E. Revolutionary cryo-EM is taking over structural biology. *Nature.* **578**: 201 (2020).
- 70 Yip, K.M., Fischer, N., Paknia, E., Chari, A. & Stark, H. Atomic-resolution protein structure determination by cryo-EM. *Nature.* **587**: 157–161 (2020).
- 71 Topf, M., Lasker, K., Webb, B., Wolfson, H., Chiu, W. & Sali, A. Protein Structure Fitting and Refinement Guided by Cryo-EM Density. *Structure.* **16**: 295–307 (2008).
- 72 Wang, W. Recent advances in atomic molecular dynamics simulation of intrinsically disordered proteins. *Physical Chemistry Chemical Physics.* **23**: 777–784 (2021).

- 73 Barth, A., Opanasyuk, O., Peulen, T.-O., Felekyan, S., Kalinin, S., Sanabria, H., *et al.* Unraveling multi-state molecular dynamics in single-molecule FRET experiments. I. Theory of FRET-lines. *J Chem Phys.* **156** (2022).
- 74 Torella, J.P., Holden, S.J., Santoso, Y., Hohlbein, J. & Kapanidis, A.N. Identifying Molecular Dynamics in Single-Molecule FRET Experiments with Burst Variance Analysis. *Biophys J.* **100**: 1568–1577 (2011).
- 75 Curutchet, C., Muñoz-Losa, A., Monti, S., Kongsted, J., Scholes, G.D. & Mennucci, B. Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J Chem Theory Comput.* **5**: 1838–1848 (2009).
- 76 Bondanza, M., Nottoli, M., Cupellini, L., Lipparini, F. & Mennucci, B. Polarizable embedding QM/MM: the future gold standard for complex (bio)systems? *Physical Chemistry Chemical Physics.* **22**: 14433–14448 (2020).
- 77 Caprasecca, S., Curutchet, C. & Mennucci, B. Toward a Unified Modeling of Environment and Bridge-Mediated Contributions to Electronic Energy Transfer: A Fully Polarizable QM/MM/PCM Approach. *J Chem Theory Comput.* **8**: 4462–4473 (2012).
- 78 Andrews, C., Xu, Y., Kirberger, M. & Yang, J.J. Structural Aspects and Prediction of Calmodulin-Binding Proteins. *Int J Mol Sci.* **22**: 308 (2020).
- 79 Moorthy, A.K. & Murthy, M.R.N. Conformation and Structural Transitions in the EF-Hands of Calmodulin. *J Biomol Struct Dyn.* **19**: 47–57 (2001).
- 80 Ghosh, A. & Giese, K.P. Calcium/calmodulin-dependent kinase II and Alzheimer's disease. *Mol Brain.* **8**: 78 (2015).
- 81 Westerlund, A.M. & Delemotte, L. Effect of Ca²⁺ on the promiscuous target-protein binding of calmodulin. *PLoS Comput Biol.* **14**: e1006072 (2018).
- 82 Tripathi, S. & Portman, J.J. Inherent flexibility determines the transition mechanisms of the EF-hands of calmodulin. *Proceedings of the National Academy of Sciences.* **106**: 2104–2109 (2009).

- 83 Wang, Q., Liang, K.-C., Czader, A., Waxham, M.N. & Cheung, M.S. The Effect of Macromolecular Crowding, Ionic Strength and Calcium Binding on Calmodulin Dynamics. *PLoS Comput Biol.* **7**: e1002114 (2011).
- 84 Sun, B. & Kekenyes-Huskey, P.M. Assessing the Role of Calmodulin's Linker Flexibility in Target Binding. *Int J Mol Sci.* **22**: 4990 (2021).
- 85 Reinartz, I., Sinner, C., Nettels, D., Stucki-Buchli, B., Stockmar, F., Panek, P.T., *et al.* Simulation of FRET dyes allows quantitative comparison against experimental data. *J Chem Phys.* **148** (2018).
- 86 DeVore, M.S., Braimah, A., Benson, D.R. & Johnson, C.K. Single-Molecule FRET States, Conformational Interchange, and Conformational Selection by Dye Labels in Calmodulin. *J Phys Chem B.* **120**: 4357–4364 (2016).
- 87 O'Day, D.H., Eshak, K. & Myre, M.A. Calmodulin Binding Proteins and Alzheimer's Disease. *Journal of Alzheimer's Disease.* **46**: 553–569 (2015).
- 88 Zhang, R., Khoo, M.S.C., Wu, Y., Yang, Y., Grueter, C.E., Ni, G., *et al.* Calmodulin kinase II inhibition protects against structural heart disease. *Nat Med.* **11**: 409–417 (2005).
- 89 Chin, D. & Means, A.R. Calmodulin: a prototypical calcium sensor. *Trends Cell Biol.* **10**: 322–328 (2000).
- 90 Krebs, J. The role of calcium in apoptosis. *Biometals.* **11**: 375–382 (1998).
- 91 Masino, L., Martin, S.R. & Bayley, P.M. Ligand binding and thermodynamic stability of a multidomain protein, calmodulin. *Protein Science.* **9**: 1519–1529 (2000).
- 92 Xu, G., Zhao, J., Cheng, K., Wu, Q., Liu, X., Liu, M., *et al.* The Effects of Macromolecular Crowding on Calmodulin Structure and Function. *Chemistry – A European Journal.* **23**: 6736–6740 (2017).
- 93 Vetter, S.W. & Leclerc, E. Novel aspects of calmodulin target recognition and activation. *Eur J Biochem.* **270**: 404–414 (2003).

- 94 Léger, C., Pitard, I., Sadi, M., Carvalho, N., Brier, S., Mechaly, A., *et al.* Dynamics and structural changes of calmodulin upon interaction with the antagonist calmidazolium. *BMC Biol.* **20**: 176 (2022).
- 95 DeVore, M.S., Gull, S.F. & Johnson, C.K. Reconstruction of calmodulin single-molecule FRET states, dye interactions, and CaMKII peptide binding by MultiNest and classic maximum entropy. *Chem Phys.* **422**: 238–245 (2013).
- 96 González, M.A. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique.* **12**: 169–200 (2011).
- 97 Cardelli, L. Amber. 1986, pp 21–47.
- 98 Brooks, B.R., Brooks, C.L., Mackerell, A.D., Nilsson, L., Petrella, R.J., Roux, B., *et al.* CHARMM: The biomolecular simulation program. *J Comput Chem.* **30**: 1545–1614 (2009).
- 99 Siu, S.W.I., Pluhackova, K. & Böckmann, R.A. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. *J Chem Theory Comput.* **8**: 1459–1470 (2012).
- 100 He, X., Liu, S., Lee, T.-S., Ji, B., Man, V.H., York, D.M., *et al.* Fast, Accurate, and Reliable Protocols for Routine Calculations of Protein–Ligand Binding Affinities in Drug Design Projects Using AMBER GPU-TI with ff14SB/GAFF. *ACS Omega.* **5**: 4611–4619 (2020).
- 101 Robustelli, P., Piana, S. & Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences.* **115** (2018).
- 102 Hansson, T., Oostenbrink, C. & van Gunsteren, W. Molecular dynamics simulations. *Curr Opin Struct Biol.* **12**: 190–196 (2002).
- 103 Gong, X., Zhang, Y. & Chen, J. Advanced Sampling Methods for Multiscale Simulation of Disordered Proteins and Dynamic Interactions. *Biomolecules.* **11**: 1416 (2021).
- 104 Ponder, J.W., Wu, C., Ren, P., Pande, V.S., Chodera, J.D., Schnieders, M.J., *et al.* Current Status of the AMOEBA Polarizable Force Field. *J Phys Chem B.* **114**: 2549–2564 (2010).

- 105 Lemkul, J.A., Huang, J., Roux, B. & MacKerell, A.D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem Rev.* **116**: 4983–5013 (2016).
- 106 Abraham, M.J. & Gready, J.E. Ensuring Mixing Efficiency of Replica-Exchange Molecular Dynamics Simulations. *J Chem Theory Comput.* **4**: 1119–1128 (2008).
- 107 Barducci, A., Bussi, G. & Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys Rev Lett.* **100**: 020603 (2008).
- 108 Zerze, G.H., Miller, C.M., Granata, D. & Mittal, J. Free Energy Surface of an Intrinsically Disordered Protein: Comparison between Temperature Replica Exchange Molecular Dynamics and Bias-Exchange Metadynamics. *J Chem Theory Comput.* **11**: 2776–2782 (2015).
- 109 Wang, J., Arantes, P.R., Bhattarai, A., Hsu, R. V., Pawnikar, S., Huang, Y.M., *et al.* Gaussian accelerated molecular dynamics: Principles and applications. *WIREs Computational Molecular Science.* **11** (2021).
- 110 Wang, L., Friesner, R.A. & Berne, B.J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J Phys Chem B.* **115**: 9431–9438 (2011).
- 111 Periole, X. & Marrink, S.-J. The Martini Coarse-Grained Force Field. 2013, pp 533–565.
- 112 Machado, M.R., Barrera, E.E., Klein, F., Sónora, M., Silva, S. & Pantano, S. The SIRAH 2.0 Force Field: Altius, Fortius, Citius. *J Chem Theory Comput.* **15**: 2719–2733 (2019).
- 113 Wu, H., Wolynes, P.G. & Papoian, G.A. AWSEM-IDP: A Coarse-Grained Force Field for Intrinsically Disordered Proteins. *J Phys Chem B.* **122**: 11115–11125 (2018).
- 114 Latham, A.P. & Zhang, B. Maximum Entropy Optimized Force Field for Intrinsically Disordered Proteins. *J Chem Theory Comput.* **16**: 773–781 (2020).

- 115 Warshel, A. & Levitt, M. Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol.* **103**: 227–249 (1976).
- 116 Marques, M.A.L. & Gross, E.K.U. TIME-DEPENDENT DENSITY FUNCTIONAL THEORY. *Annu Rev Phys Chem.* **55**: 427–455 (2004).
- 117 Morzan, U.N., Alonso De Armiño, D.J., Foglia, N.O., Ramírez, F., González Lebrero, M.C., Scherlis, D.A., *et al.* Spectroscopy in Complex Environments from QM-MM Simulations. *Chem Rev.* . **118**: 4071–4113 (2018).
- 118 Loos, P.F., Scemama, A. & Jacquemin, D. The Quest for Highly Accurate Excitation Energies: A Computational Perspective. *Journal of Physical Chemistry Letters.* . **11**: 2374–2383 (2020).
- 119 Sarkar, R., Boggio-Pasqua, M., Loos, P.F. & Jacquemin, D. Benchmarking TD-DFT and Wave Function Methods for Oscillator Strengths and Excited-State Dipole Moments. *J Chem Theory Comput.* **17**: 1117–1132 (2021).
- 120 Renger, T. & Müh, F. Theory of excitonic couplings in dielectric media. *Photosynth Res.* **111**: 47–52 (2012).
- 121 Renger, T. & Marcus, R.A. On the relation of protein dynamics and exciton relaxation in pigment–protein complexes: An estimation of the spectral density and a theory for the calculation of optical spectra. *J Chem Phys.* **116**: 9997–10019 (2002).
- 122 Koch, H., Christiansen, O., Jørgensen, P., Sanchez de Merás, A.M. & Helgaker, T. The CC3 model: An iterative coupled cluster approach including connected triples. *J Chem Phys.* **106**: 1808–1818 (1997).
- 123 Bartlett, R.J. & Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev Mod Phys.* **79**: 291–352 (2007).
- 124 Pulay, P. A perspective on the CASPT2 method. *Int J Quantum Chem.* **111**: 3273–3279 (2011).
- 125 Sarkar, R., Loos, P.-F., Boggio-Pasqua, M. & Jacquemin, D. Assessing the Performances of CASPT2 and NEVPT2 for Vertical Excitation Energies. *J Chem Theory Comput.* **18**: 2418–2436 (2022).

- 126 Dreuw, A. & Head-Gordon, M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem Rev.* **105**: 4009–4037 (2005).
- 127 Jurinovich, S., Curutchet, C. & Mennucci, B. The Fenna–Matthews–Olson Protein Revisited: A Fully Polarizable (TD)DFT/MM Description. *ChemPhysChem.* **15**: 3194–3204 (2014).
- 128 Renger, T., Klinger, A., Steinecker, F., Schmidt am Busch, M., Numata, J. & Müh, F. Normal Mode Analysis of the Spectral Density of the Fenna–Matthews–Olson Light-Harvesting Protein: How the Protein Dissipates the Excess Energy of Excitons. *J Phys Chem B.* **116**: 14565–14580 (2012).
- 129 Merchant, K.A., Best, R.B., Louis, J.M., Gopich, I. V. & Eaton, W.A. Characterizing the unfolded states of proteins using single-molecule FRET spectroscopy and molecular simulations. *Proceedings of the National Academy of Sciences.* **104**: 1528–1533 (2007).
- 130 Knox, R.S. & van Amerongen, H. Refractive index dependence of the Förster resonance excitation transfer rate. *J Phys Chem B.* **106**: 5289–5293 (2002).
- 131 Hsu, C.-P., Fleming, G.R., Head-Gordon, M. & Head-Gordon, T. Excitation energy transfer in condensed media. *J Chem Phys.* **114**: 3065 (2001).
- 132 Curutchet, C., Munoz-Losa, A., Monti, S., Kongsted, J., Scholes, G.D. & Mennucci, B. Electronic Energy Transfer in Condensed Phase Studied by a Polarizable QM/MM Model. *J Chem Theory Comput.* **5**: 1838–1848 (2009).
- 133 Iozzi, M.F., Mennucci, B., Tomasi, J. & Cammi, R. Excitation energy transfer (EET) between molecules in condensed matter: A novel application of the polarizable continuum model (PCM). *J Chem Phys.* **120**: 7029–7040 (2004).
- 134 Curutchet, C. & Mennucci, B. Quantum Chemical Studies of Light Harvesting. *Chem Rev.* **117**: 294–343 (2017).
- 135 Cupellini, L., Corbella, M., Mennucci, B. & Curutchet, C. Electronic energy transfer in biomacromolecules. *Wiley Interdiscip Rev Comput Mol Sci.* **9**: e1392 (2019).

- 136 Curutchet, C., Scholes, G.D., Mennucci, B. & Cammi, R. How solvent controls electronic energy transfer and light harvesting: Toward a quantum-mechanical description of reaction field and screening effects. *J Phys Chem B*. **111**: 13253–13265 (2007).
- 137 Scholes, G.D., Curutchet, C., Mennucci, B., Cammi, R. & Tomasi, J. How Solvent Controls Electronic Energy Transfer and Light Harvesting. *J Phys Chem B*. **111**: 6978–6982 (2007).
- 138 Curutchet, C., Kongsted, J., Muñoz-Losa, A., Hossein-Nejad, H., Scholes, G.D. & Mennucci, B. Photosynthetic Light-Harvesting Is Tuned by the Heterogeneous Polarizable Environment of the Protein. *J Am Chem Soc*. **133**: 3078–3084 (2011).
- 139 Eder, M. & Renger, T. A Simple Expression for the Screening of Excitonic Couplings between Chlorophylls as Inferred for Photosystem I Trimers. *Int J Mol Sci*. **25** (2024).
- 140 Cignoni, E., Cupellini, L. & Mennucci, B. A fast method for electronic couplings in embedded multichromophoric systems. *Journal of Physics: Condensed Matter*. **34**: 304004 (2022).
- 141 Kursula, P. The many structural faces of calmodulin: A multitasking molecular jackknife. *Amino Acids*. **46**: 2295–2304 (2014).
- 142 Johnson, C.K. Calmodulin, Conformational States, and Calcium Signaling. A Single-Molecule Perspective †. *Biochemistry*. **45**: 14233–14246 (2006).
- 143 DeVore, M.S., Braimah, A., Benson, D.R. & Johnson, C.K. Single-Molecule FRET States, Conformational Interchange, and Conformational Selection by Dye Labels in Calmodulin. *J Phys Chem B*. **120**: 4357–4364 (2016).
- 144 Chattopadhyaya, R., Meador, W.E., Means, A.R. & Quiocho, F.A. Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol*. **228**: 1177–1192 (1992).
- 145 Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. & Case, D.A. Development and testing of a general amber force field. *J Comput Chem*. **25**: 1157–1174 (2004).
- 146 Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E. & Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain

- and Backbone Parameters from ff99SB. *J Chem Theory Comput.* **11**: 3696–3713 (2015).
- 147 Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. & Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys.* **79**: 926–935 (1983).
 - 148 Joung, I.S. & Cheatham, T.E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B.* **112**: 9020–9041 (2008).
 - 149 Li, P., Roberts, B.P., Chakravorty, D.K. & Merz, K.M. Rational design of particle mesh ewald compatible lennard-jones parameters for +2 metal cations in explicit solvent. *J Chem Theory Comput.* **9**: 2733–2748 (2013).
 - 150 Case, D.A., Belfon, K., Ben-Shalom, I.Y., Brozell, S.R., Cerutti, D.S., Cheatham III, T.E., *et al.* AMBER 20. .
 - 151 Olsson, M.H.M., Søndergaard, C.R., Rostkowski, M. & Jensen, J.H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput.* **7**: 525–537 (2011).
 - 152 Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys.* **132**: 154104 (2010).
 - 153 Hopkins, C.W., Le Grand, S., Walker, R.C. & Roitberg, A.E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J Chem Theory Comput.* **11**: 1864–1874 (2015).
 - 154 Cupellini L, L.F.C.E. Trespcoup - Software to compute TrEsp couplings. .. (2024).
 - 155 Wang, J., Cieplak, P., Li, J., Hou, T., Luo, R. & Duan, Y. Development of polarizable models for molecular mechanical calculations I: Parameterization of atomic polarizability. *Journal of Physical Chemistry B.* **115**: 3091–3099 (2011).
 - 156 Wang, J., Cieplak, P., Li, J., Wang, J., Cai, Q., Hsieh, M., *et al.* Development of polarizable models for molecular mechanical calculations II: Induced dipole models significantly improve accuracy of intermolecular interaction energies. *Journal of Physical Chemistry B.* **115**: 3100–3111 (2011).

- 157 Cupellini L, J.S.M.B. TraDA - Transition Density Analyzer. . . (2024).
- 158 Frisch, M.J., Trucks, G.W., Schlegel, H.B., Scuseria, G.E., Robb, M.A., Cheeseman, J.R., *et al.* Gaussian Development Version, Revision H.36. . (2010).
- 159 Muñoz-Losa, A., Curutchet, C., Galván, I.Fdez. & Mennucci, B. Quantum mechanical methods applied to excitation energy transfer: A comparative analysis on excitation energies and electronic couplings. *J Chem Phys.* **129**: 034104 (2008).
- 160 Senn, H.M. & Thiel, W. QM/MM Methods for Biomolecular Systems. *Angewandte Chemie-International Edition.* **48**: 1198–1229 (2009).
- 161 Corbella, M., Cupellini, L., Lipparini, F., Scholes, G.D. & Curutchet, C. Spectral Variability in Phycocyanin Cryptophyte Antenna Complexes is Controlled by Changes in the α -Polypeptide Chains. *ChemPhotoChem.* **3**: 945–956 (2019).
- 162 Caprasecca, S., Curutchet, C., Jurinovich, S. & Mennucci, B. PolChat: A polarisation-consistent charge fitting tool. .
- 163 Jordanides, X.J., Lang, M.J., Song, X. & Fleming, G.R. Solvation dynamics in protein environments studied by photon echo spectroscopy. *Journal of Physical Chemistry B.* **103**: 7995–8005 (1999).
- 164 Slaughter, B.D., Allen, M.W., Unruh, J.R., Bieber Urbauer, R.J. & Johnson, C.K. Single-molecule resonance energy transfer and fluorescence correlation spectroscopy of calmodulin in solution. *J Phys Chem B.* **108**: 10388–10397 (2004).
- 165 Westerlund, A.M. & Delemotte, L. Effect of Ca²⁺ on the promiscuous target-protein binding of calmodulin. *PLoS Comput Biol.* **14**: e1006072 (2018).
- 166 Wriggers, W., Mehler, E., Pitici, F., Weinstein, H. & Schulten, K. Structure and Dynamics of Calmodulin in Solution. *Biophys J.* **74**: 1622–1639 (1998).
- 167 Komeiji, Y., Ueno, Y. & Uebayasi, M. Molecular dynamics simulations revealed Ca²⁺-dependent conformational change of Calmodulin. *FEBS Lett.* **521**: 133–139 (2002).

- 168 Shepherd, C.M. & Vogel, H.J. A Molecular Dynamics Study of Ca²⁺-Calmodulin: Evidence of Interdomain Coupling and Structural Collapse on the Nanosecond Timescale. *Biophys J.* **87**: 780–791 (2004).
- 169 Fiorin, G., Biekofsky, R.R., Pastore, A. & Carloni, P. Unwinding the helical linker of calcium-loaded calmodulin: A molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics.* **61**: 829–839 (2005).
- 170 Shukla, D., Peck, A. & Pande, V.S. Conformational heterogeneity of the calmodulin binding interface. *Nat Commun.* **7**: 10910 (2016).
- 171 Wang, K., Sachdeva, A., Cox, D.J., Wilf, N.W., Lang, K., Wallace, S., *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat Chem.* **6**: 393–403 (2014).
- 172 Izumi, Y., Kuwamoto, S., Jinbo, Y. & Yoshino, H. Increase of the molecular weight and radius of gyration of apocalmodulin induced by binding of target peptide: evidence for complex formation. *FEBS letters* 495(1-2), 126-130. (2001).
- 173 Renger, T. & Müh, F. Theory of excitonic couplings in dielectric media : foundation of Poisson-TrEsp method and application to photosystem I trimers. *Photosynth Res.* **111**: 47–52 (2012).
- 174 Sobakinskaya, E., Schmidt am Busch, M. & Renger, T. Theory of FRET “Spectroscopic Ruler” for Short Distances: Application to Polyproline. *J Phys Chem B.* **122**: 54–67 (2018).
- 175 VanBeek, D.B., Zwier, M.C., Shorb, J.M. & Krueger, B.P. Fretting about FRET: correlation between kappa and R. *Biophys J.* **92**: 4168–4178 (2007).
- 176 Slaughter, B.D., Unruh, J.R., Allen, M.W., Urbauer, R.J.B. & Johnson, C.K. Conformational substates of calmodulin revealed by single-pair fluorescence resonance energy transfer: Influence of solution conditions and oxidative modification. *Biochemistry.* **44**: 3694–3707 (2005).

- 177 Robustelli, P., Piana, S. & Shaw, D.E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc Natl Acad Sci U S A*. **115**: E4758–E4766 (2018).
- 178 Best, R.B., Zheng, W. & Mittal, J. Balanced protein-water interactions improve properties of disordered proteins and non-specific protein association. *J Chem Theory Comput*. **10**: 5113–5124 (2014).
- 179 Muñoz-Losa, A., Curutchet, C., Krueger, B.P., Hartsell, L.R. & Mennucci, B. Fretting about FRET: Failure of the Ideal Dipole Approximation. *Biophys J*. **96**: 4779–4788 (2009).
- 180 Weiser, J., Shenkin, P.S. & Still, W.C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem*. **20**: 217–230 (1999).