Dynamics of self-assembled monolayers of molecular machines on the water/air interface

Matasović, Lujo

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Science / Sveučilište u Zagrebu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: https://urn.nsk.hr/urn:nbn:hr:217:196352

Rights / Prava: In copyright/Zaštićeno autorskim pravom.

Download date / Datum preuzimanja: **2024-11-29**

University of Zagreb
FACULTY OF SCIENCE
Department of Chemistry

Lujo Matasović

DYNAMICS OF SELF-ASSEMBLED MONOLAYERS OF MOLECULAR MACHINES ON THE WATER/AIR INTERFACE

Diploma Thesis

submitted to the Department of Chemistry,

Faculty of Science, University of Zagreb

for the academic degree of Master in Chemistry

Zagreb, 2022.

Acknowledgements

First and foremost, I want to thank my brother Mavro, my girlfriend Margareta and other members of my family, who were standing with me, with their unconditional love, patience, and support. Without them, I would have never had the strength to push through.

Many thanks to Marin and Jerko, my Simba and Pumba, in whose company I *Hakuna Matataed* through the University. It was quite a journey.

I also want to thank Dr Kodrin, Dr Kovačević and Dr Stilinović, who went well beyond their call of duty reading, revising, and organising my defence at the last minute. Without them, it would have been all topsy turvy.

Furthermore, I would like to thank all the fine scientists in Kaleta group, whose hard work tremendously contributed to this thesis, enriching it well beyond anyone's imagination.

I am particularly thankful for my mentor and an astounding group leader, Dr Kaleta, who welcomed me into his group with open arms, introduced me to the fine art of creating molecular machinery and kept believing in me when I doubted myself the most. I will never forget that.

Finally, I would like to wholeheartedly thank my dear friend Dr Rončević, who recognised my love for research, encouraged me to aim high and whose incisive guidance over the years has helped me realise my academic potential. His kindness, honesty, and merit have shown me how noble being ambitious can be.

Table of Contents

University of Zagreb

Faculty of Science

Department of Chemistry

Diploma Thesis

ABSTRACT

DYNAMICS OF SELF-ASSEMBLED MONOLAYERS OF MOLECULAR MACHINES ON THE WATER/AIR INTERFACE

Lujo Matasović

We present a computational study of self-assembled monolayers (SAMs) of several functional molecules which share a triptycene core and differ in the number of anchoring groups and functional mounts (photoswitches or light-driven motors). Structural properties of different SAMs on the water-air interface are explored by nanoscale molecular dynamics carried out in the tight-binding DFT formalism, as well as by static meta-GGA DFT calculations. Both approaches give similar results, revealing that tilting and bending in SAMs is significant in all investigated systems. Molecular dynamics shows that SAMs tilt collectively, although the motion of individual machines is not strongly correlated. Functional heads, triptycenes and anchoring groups are found to strongly affect SAM properties, and a structure bearing two triptycene units and three anchoring groups is identified as the most promising future synthetic target. Finally, the computational results are validated by comparing them with previously performed experiments (anisotropic spectroscopy and ellipsometry).

Sveučilište u Zagrebu

Prirodoslovno-matematički fakultet

Kemijski odsjek

SAŽETAK

DINAMIKA SAMOUDRUŽENIH MONOSLOJEVA MOLEKULARNIH STROJEVA NA MEĐUPOVRŠINI VODA/ZRAK

Lujo Matasović

U ovom radu računalnim su metodama istraženi samoudruženi monoslojevi (SAM) niza funkcionalnih molekula koje se razlikuju po funkcionalnoj jedinici (sklopka ili svjetlom upogonjeni motor) i broju veznih skupina, a dijele tripticensku jezgru. Strukturna svojstva različitih SAM-ova na međupovršini voda-zrak istražena su nanosekundnom molekulskom dinamikom provedenom u formalizmu DFT modela čvrste veze, te meta-GGA DFT statičkim izračunima. Oba pristupa daju slične rezultate i otkrivaju da su, u svim proučavanim sustavima, molekule značajno nagnute i savinute. Rezultati molekulske dinamike pokazuju da je nagibanje SAM-ova kolektivan proces, iako gibanje individualnih strojeva nije korelirano. Također, potvrđeno je da funkcionalne glave, tripticeni i vezne skupine značajno utječu na svojstva monoslojeva, te je kao glavni kandidat za sintezu novih funkcionalnih materijala identificiran molekulski sustav s dvije tripticenske jedinice i tri vezne skupine. Usporedbom s prethodno provedenim eksperimentima (anizotropnom spektroskopijom i elipsometrijom) potvrđeni su računalni rezultati.

(108 stranice, 42 slika, 16 tablica, 98 literaturnih navoda, jezik izvornika: Engleski)

Rad je pohranjen u Središnjoj kemijskoj knjižnici Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu, Horvatovac 102a, Zagreb i Repozitoriju Prirodoslovno-matematičkog fakulteta Sveučilišta u Zagrebu

Ključne riječi: DFT model čvrste veze, IR refleksijska adsorpcijska spektroskopija, meta-GGA DFT, molekulska dinamika, molekulski strojevi, samoudruženi monoslojevi,
Mentor: dr. sc. Jiří Kaleta
Neposredni voditelj: dr. sc. Igor Rončević
Nastavnik (imenovan od strane Kemijskog odsjeka): izv. prof. dr. sc. Ivana Biljan

Ocjenitelji:
1. izv. prof. dr. sc. Ivana Biljan.
2. izv. prof. dr. sc. Vladimir Stilinović.
3. prof. dr. sc. Davor Kovačević.
   Zamjena: doc. dr.sc. Ivan Kodrin

Datum diplomskog ispita: 6. Srpnja 2022

PROŠIRENI SAŽETAK

Dizajn novih materijala temeljenih na samoudruženim monoslojevima molekulskih strojeva, poput svjetlom aktiviranih sklopki i jednosmjernih motora, područje je sve većeg znanstvenog interesa, s primjenom u senzorici, katalizi i nanoelektronici.[I] U samoudruženim monoslojevima (SAM), pravilni slojevi molekulskih strojeva nastaju spontanom adsorpcijom na ravnim površinama. Znanstveni rad Kaletove grupe fokusiran je na dizajn pravilnih 2D redova molekulskih strojeva baziranih na tripticenu Langmuir-Bodgettovom metodom.[II]

Nedavno je pokazano da se SAM-ovi molekulskih strojeva usidreni na (111) površini zlata savijaju[III] otvarajući pitanja poput: utječe li savijanje i nagib molekula na stabilnost monoslojeva te mogu li se strukturna svojstva takvih sustava usmjeriti mijenjanjem građevnih jedinki molekulskih strojeva?

Ovaj rad primarno je usmjeren na dinamička svojstva, poput kolektivnog savijanja i naginjanja uređenih monoslojeva molekulskih strojeva temeljenih na tripticenu na međupovršini voda-zrak. Osnovni građevni motiv (tripticenska jedinica s funkcionalnom glavom i promjenjivim brojem veznih skupina) istraženih spojeva u ovome radu istaknut je crnim kvadratom na Dijagramu I. Tronožni sustavi (koji nose tri karboksilne skupine) temeljeni na tripticenu, razlikuju se po funkcionalnoj glavi, koje mogu biti sklopke (**Azo**, **Dia**, **Thio**), motor (**Fer**) ili još jedna tripticenska jedinica (**Tri**). Također, u radu su razmatrani i tzv. *PHE* sustavi (**Phe0**, **Phe1**, **Phe**) gdje molekule umjesto funkcionalne jedinice sadrže fenilni prsten, a razlikuju se po broju veznih skupina (Dijagram I) Vezne skupine u svim proučavanim spojevima su karboksilne kiseline, osim prilikom modeliranja spojeva u vakuumu, gdje su kiseline zamijenjene metilnim esterima.

[I] J. Kaleta, P. I. Dron, K. Zhao, Y. Shen, I. Císařová, C. T. Rogers, J. Michl, *J. Org. Chem.,* **80** (2015) 6173–6192.

[II] J. Kaleta, E. Kaletova, I. Cisarova, S. J. Teat, J. Michl, *J. Org. Chem.* **80** (2015) 10134–10150.

[III] I. Rončević, E Kaletová, K. Varga, I. Císařová, Z Bastl, J. Jiang, J. Kaleta, *J. Phys. Chem. C* **126** (2022) 7193–7207
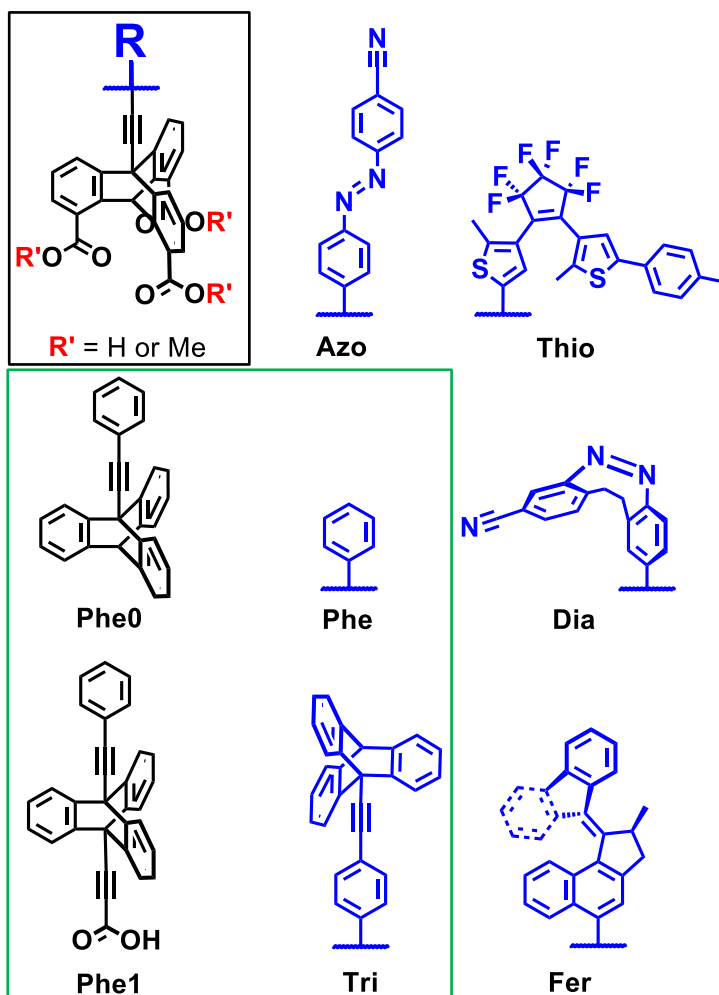
Diagram I. Strukture istraženih spojeva. Opća struktura razmatranog spoja prikazana je u crnom pravokutniku. Funkcijska skupina **R** može varirati (plave skupine). **Azo**, **Thia**, **Dia** i **Fer** sustavi sintetizirani su prije izrade ovog rada te su računalni rezultati za ove sustave uspoređeni s eksperimentalnima,[IV] dok su sustavi istaknuti zelenim pravokutnikom (**Phe0**, **Phe1** i **Phe**, **Tri)** razmatrani isključivo računalno.

Centralna veličina u ovome radu je prosječni kut nagiba Θ molekulskih strojeva u SAM-u, definiran kao kut odstupanja pravca koji prolazi kroz premošćujući C atom tripticenskog podnožja i aromatski prsten funkcionalne glave od normale ravnine, definirane preko

---

[IV] G. Bastien, C. Santos Hurtado, I. Rončević, L. Severa, J. Štoček, M. Dračínský, Z. Bastl, I. Císařová, J. Kaleta, *In preparation*

premoštenih tripticenskih C atoma strojeva u jediničnoj ćeliji monosloja (Slika I (a)). U radu je pretpostavljeno da se SAM-ovi udružuju prema uzorku prikazanom na Slici I (b).
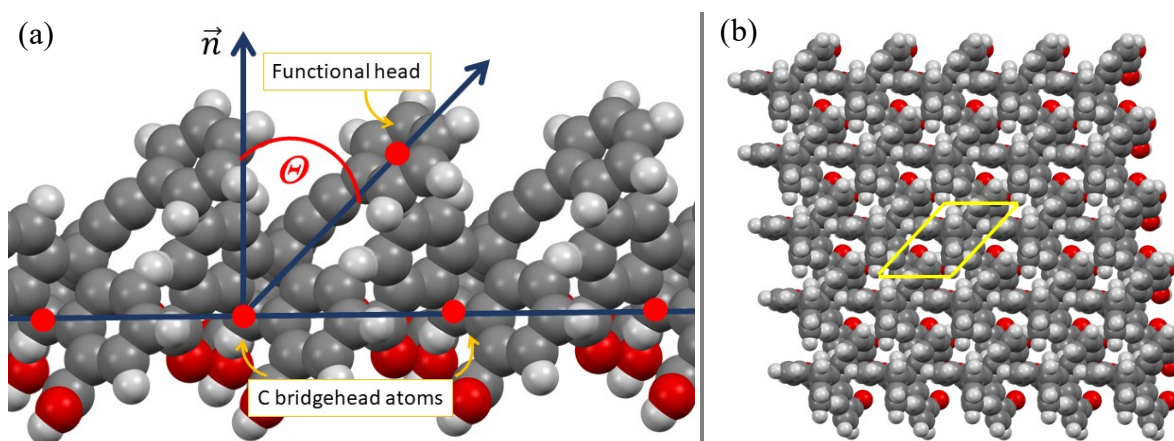
Slika I. (a) Definicija prosječnog nagibnog kuta Θ prikazanog na općenitoj strukturi samoudruženog monosloja molekulskih strojeva temeljenih na tripticenu. (b) Pretpostavljeni uzorak pakiranja samoudruženih monoslojeva razmatranih u radu. Žute linije omeđuju jediničnu ćeliju monosloja.

U ovom radu, kombinirane su tri različite računalne metode:

(i) nanosekundna molekulska dinamika na međupovršini voda-zrak DFT pristupom čvrste veze (DFTB)[V] s periodičnim rubnim uvjetima (PBC)

(i) klaster molekulska dinamika s nedavno predstavljenim GFN-FF poljem sila.[VI]

(ii) iscrpna pretraga plohe potencijalne energije SAM-ova u vakuumu meta-GGA DFT pristupom.

Također, računalni rezultati uspoređeni su s eksperimentalnim podatcima prikupljenim u Kaletovoj grupi upotrebom Langmuir Blodgettove tehnike, elipsometrije i anizotropne spektroskopije. Valja pritom napomenuti da, iako su obrada podataka, analiza i mjerenja napravljeni u sklopu ovog rada, sama mjerenja nisu provedena od strane autora te će metodologija vezana za eksperimente stoga biti naknadno objavljena.

[V] B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, et al., *J. Chem. Phys.,* **152** (2020) 124101–124120.

[VI] S. Spicher, S. Grimme, *Angew. Chem. Int. Ed.,* **2020**, *59*, 15665–15671

I.   Molekulska dinamika s periodičnim rubnim uvjetima

Monoslojevi na međupovršini voda-zrak modelirani su prema Slici II, s oko 15 Å vode i 20 Å vakuuma između svakog monosloja. Gornji i donji monosloj ne komuniciraju tijekom simulacije te se njihova svojstva mogu promatrati zasebno. Tablica I. prikazuje neke od mjerenih observabli promatranih u ovome radu. Θ_{bot} označava nagib donjeg, tripticenskog dijela molekule, Θ_{top} nagib gornjeg dijela molekule, dok B označava savijanje funkcionalne glave. Veličina d(top-top) definirana je kao prosječna udaljenost susjednih funkcijskih skupina, a CDF je kombinirana distribucijska funkcija između B i d(top-top).
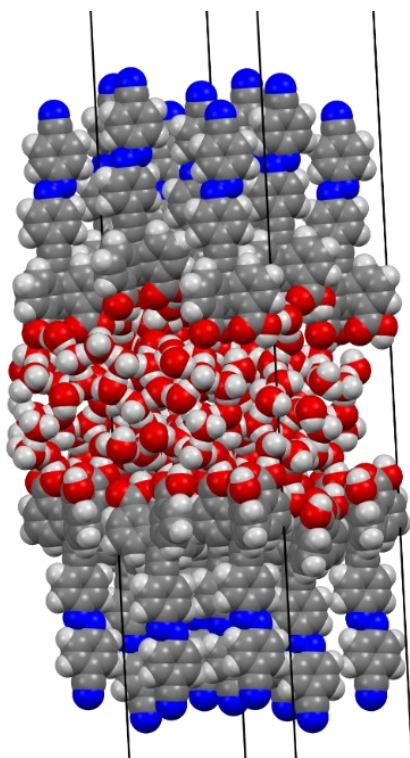
Tablica I. Primjeri observabli uzorkovanih metodama molekulske dinamike i statičkim izračunima.

Observabla	Opis
Θ_{bot} [°]	Nagibni kut tripticenskog postolja
Θ_{top} [°]	Nagibni kut funkcionalne glave
B [°]	Savijanje funkcijske glave, relativno prema tripticenskom postolju
d(top-top) [Å]	Minimalna prosječna udaljenost između funkcijskih skupina
CDF [°, Å]	Ovisnost B o d(top-top)
Korelacija između strojeva	Korelacija modova gibanja između individualnih molekulskih strojeva

Slika II. Bočni prikaz jedinične ćelije sustava simuliranog PBC molekulskom dinamikom. Za primjer je uzet *trans*-**Azo** sustav.

Nakon ekvilibracije, proučavane observable simuliranih sustava ne evoluiraju u vremenu. U produkcijskoj fazi, analiza observabli konvergiranih u simuliranom vremenu postaje invarijantna s obzirom na permutacije koraka molekulske dinamike što omogućava procjenu svojstava filma preko uzorkovanih podataka. Slika III, na primjeru *trans*-**Azo** sustava, pokazuje oscilaciju nagibnog kuta tripticena, te kuta savijanja funkcionalne glave u vremenu tijekom produkcijske faze simulacije.
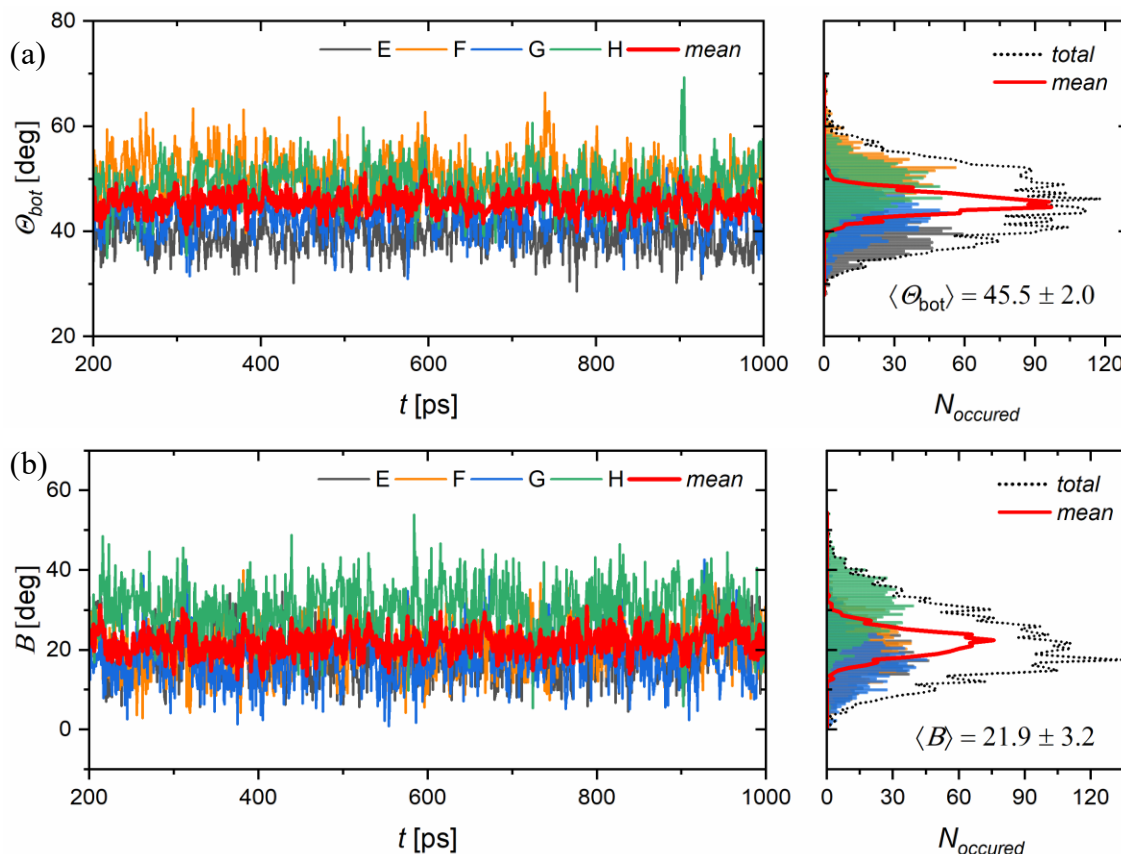
Slika III. Evolucija strukturnih svojstava *trans*-**Azo** sustava u vremenu. (a) nagibni kut tripticenskog postolja Θ_{bot}, (b) kut savijanja B. Tanke linije predstavljaju promjenu kutova u vremenu individualnih strojeva, a deblja crvena linija označava prosječni kut u svakom vremenskom koraku.

Phe sustav, simulirane sklopke (**Azo**, **Dia**, **Thio**), motor (**Fer**) te **Tri** tvore stabilne uređene filmove, iako variraju u prosječnom nagibu. Kod monoslojeva s manje od tri vezne skupine (**Phe0** i **Phe1**) granica između organske i vodene faze se nakon ekvilibracije gubi i sustavi su izuzetno nagnuti. To ukazuje na važnost triju veznih skupina u povećanju stabilnosti monoslojeva. Slika IV pokazuje z koordinate tripticena i molekula vode tijekom simulacijskog vremena na primjeru **Phe** i **Phe0** sustava.
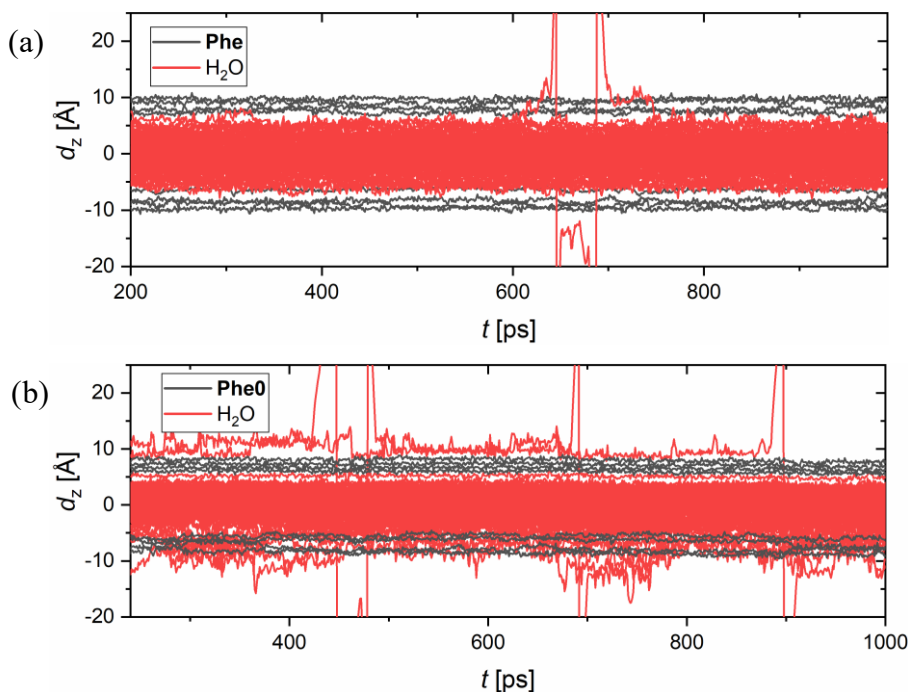
Slika IV. Vremenska evolucija *z* koordinata (okomito na površinu) vode (crveno) i tripticenskih postolja (siva) za (a) **Phe** i (b) **Phe0** sustav.

Jedan od najizraženijih faktora koji utječu na stabilnost SAM-ova je sklonost nekih monoslojeva naginjanju i savijanju više od drugih. Slika V prikazuje prosječne nagibne kutove tripticenskih postolja i funkcionalnih glava molekulskih strojeva u proučavanim sustavima. Prikazane vrijednosti nagibnih kutova dobivene su statičkom pretragom lokalnih minimuma i PBC molekulskom dinamikom. Sustavi s voluminoznim funkcijskim skupinama uglavnom su manje nagnuti dok najveće nagibe imaju monoslojevi u *PHE* obitelji, osobito molekule s manje od 3 vezne skupine (**Phe0** i **Phe1**). Međutim, svi se sustavi savijaju pod približno jednakim kutem (oko 20°), a sklonost savijanju raste kod sustava s funkcijskim skupinama kod kojih je dostupnost stabilizacije nekovalentnim interakcijama, poput $\pi-\pi^*$ slaganja, veća.

Normalna distribucija observabli dobivenih molekulskom dinamikom upućuje da, iako su simulirani sustavi mali (4 stroja u jediničnoj ćeliji) i simulacijsko vrijeme je kratko (1 ns), odabrana metoda dobro uzorkuje prostor mogućih svojstava monoslojeva. Slaganje između statičkih i dinamičkih rezultata je dobro, no statički izračuni redovito predviđaju manje nagibne kutove (Slika V), što je očekivano, budući da ne uključuju entropijske efekte ni interakciju s vodom. Trendovi ostaju očuvani u različitim metodama.
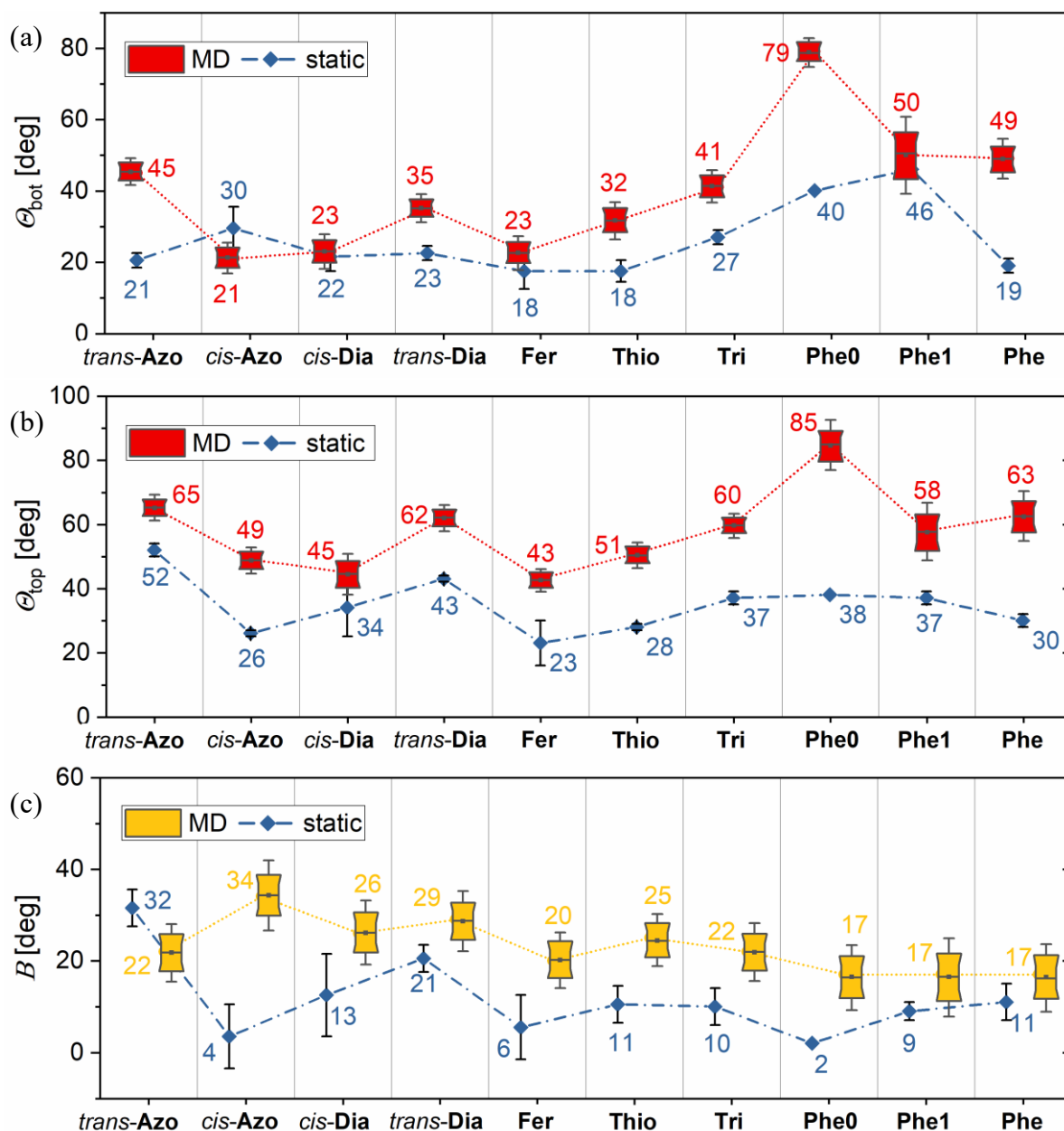
Slika V. Nagib donjeg (a) i gornjeg (b) dijela molekula u monosloju te kutevi savijanja (c) za sve simulirane sustave metodom PBC molekulske dinamike (dijagram kutije) i statičnim pretraživanjem (plave točke). Brojevi ispod i iznad pojedinog dijagrama opisuju srednje vrijednosti kuteva. Isprekidane linije su vizualni pomagači, bez fizičkog značaja.

Za razliku od 1D histograma, multidimenzijski korelacijski histogrami mogu ponuditi informaciju o kombinaciji parametara u prostoru mogućih konformacija molekula u monosloju koje konformeri mogu istovremeno imati, te odgovoriti na pitanje je li savijanje potaknuto stabilizirajućim nekovalentnim interakcijama između funkcijskih skupina ili težnjom da se minimiziraju steričke smetnje krupnih strukturnih motiva u filmu.

Geometrijska kompleksnost strojeva koji grade monoslojeve rezultira visoko koreliranim modovima gibanja unutar iste molekule dajući simetrične histograme koji opisuju pojavnost određenog parametra u monosloju. Međutim, svaka individualna konformacija može, u nekom vremenskom koraku, posjedovati samo podskup svojstava koji definiraju strukturu filma.

Slika VI prikazuje primjer dva različita slučaja korelacijskih grafova koji pokazuju zavisnost između savijanja molekula i udaljenosti susjednih funkcionalnih glava. U prvom slučaju, **Azo** sustav se značajno savija tijekom simulacije kako bi maksimizirao $\pi - \pi^*$ naslagivanje susjednih azobenzenskih skupina. Sustav je značajno savinut kada su mu funkcijske skupine blizu (crvena boja) i uspravan kada su udaljene (plava boja). S druge strane, **Fer** je značajno savinut u konformacijama kada su funkcijske skupine daleko jer monosloj nastoji smanjiti steričke smetnje između krupnih motora.
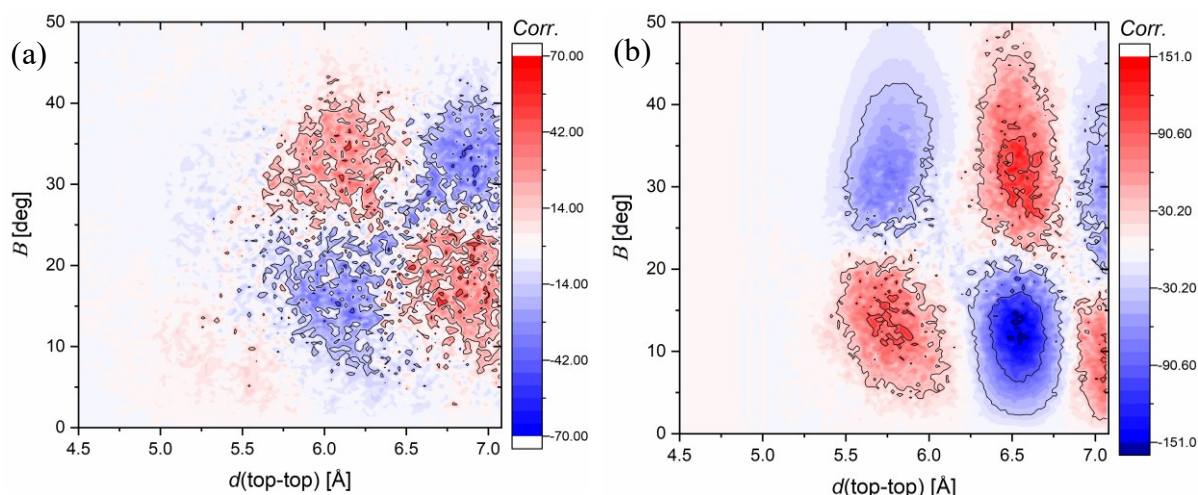
Slika VI. Korelacijski histogram ovisnosti molekulskog savijanja (B) i udaljenosti susjednih funkcionalnih glava (d(top-top)) na primjerima (a) *trans*-**Azo** i (b) **Fer** sustava. Crvena boja označava pozitivnu korelaciju, a plava negativnu korelaciju između varijabli

Tijekom simulacije, nagib molekulskih strojeva u monosloju je uglavnom kolektivan, s funkcijskim skupinama usmjerenim u istom pravcu tijekom cijele simulacije. Ova asimetričnost monoslojeva pokazuje da se strukturne promjene u monoslojevima zbivaju kolektivno te da je supramolekulska komunikacija u monosloju značajna. Međutim, gibanje individualnih strojeva zbog termičkih fluktuacija nije značajno korelirano.

II.   Klaster molekulska dinamika

Sustavi su simulirani i GFN-FF metodom implementiranom u xTB programu s otvorenim pristupom.[VII] Budući da korištena inačica GFN-FF polja sila ne posjeduje mogućnost simulacije s periodičkim rubnim uvjetima, za preciznu kontrolu prosječne površine po molekuli (APM), ručno je uveden elipsoidni potencijal oko simuliranog prostora. **Azo**, **Dia**, **Thio** i **Fer** sustavi simulirani su na međupovršini voda-zrak pri različitim APM-ovima (između 55 i 80 Å) između 1 i 2 ns.

Nepremostivi izazov prilikom upotrebe GFN-FF polja sila bio je neizbježni efekt prodiranja vode kroz monosloj na sučelju ruba simulacijskog sustava i ručno zadanog vanjskog potencijala. Naime, tripticenski strojevi na rubovima sustava ne mogu se ispravno pakirati zbog dodirivanja vanjskog potencijala. To ostavlja šupljine u strukturi monosloja kroz koje molekule vode mogu prodrijeti i uništiti organski film na površini. Energija stoga kontinuirano raste tijekom simulacije, sve dok prodiranje vodene faze ne prevlada i ne uništi strukturu organskog monosloja (Slika VII).
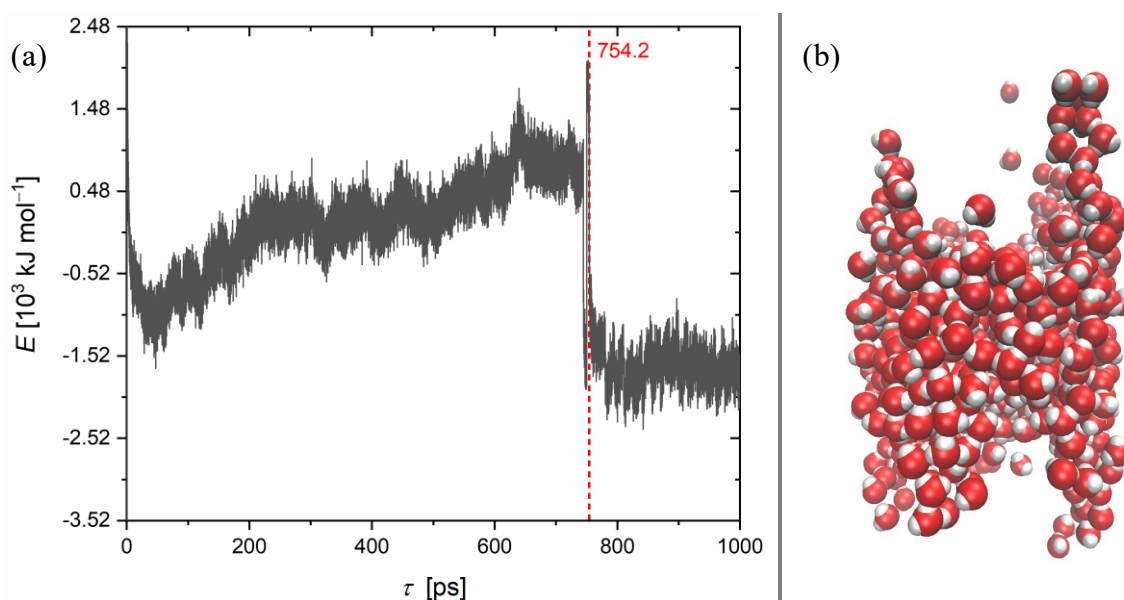
Slika VII. (a) primjer simulacije (**Azo** sustav pri eksperimentalnom APM-u) gdje, zbog nedostatka periodičkih rubnih uvjeta, energija sustava ne može ekvilibrirati, dok se monosloj ne razori protruzijama molekula vode. (b) Molekule vode simuliranog **Azo** sustava na cca 750 ps simulacijskog vremena. Probijanje vode kroz monosloj događa se

[VII] E. Bannwarth, S. Caldeweyher, A. Ehlert, P. Hansen, P. J. Pracht, S. Seibert, S. Spicher, S. Grimme, *WIREs Comput. Mol. Sci.* **11** (2020) 1493–1502.

samo na rubovima simulacijskog prostora. Ovo implicira da uništenje monosloja ne reflektira intrinzičnu nestabilnost filma, već je rezultat rubnih uvjeta

Međutim, u nekolicini sustava u kojima se uništavanje organskog sloja zbivalo osobito sporo, bilo je moguće odrediti (približno) konvergirane strukturne parametre. Tablica II daje primjer prosječnog naginjanja donjeg i gornjeg dijela molekule za **Fer** sustav simuliran pri eksperimentalnoj prosječnoj površini po molekuli (71 Å^2), dobiven metodama klaster (GFN-FF) i PBC (DFTB) molekulske dinamike. Porast energije tijekom simulacije u **Fer** slučaju bio je usporen zbog volumena funkcionalne glave.

Tablica II. Usporedba nagibnih kutova i kuta savijanja **Fer** sustava pri eksperimentalnom APM-u, dobivenih GFN-FF i PBC DFTB metodama molekulske dinamike.

	GFN-FF	DFTB
$\langle \Theta_{bot} \rangle$	21.6	22.7
$\langle \Theta_{top} \rangle$	32.5	37.4
$\langle B \rangle$	18.4	20.2

Nedavno je razvijeno GFN-FF polja sila koja podržava periodičke rubne uvjete.[VIII] Razmjerno dobro slaganje između DFTB i GFN-FF metoda ukazuje da je GFN-FF potencijalno dobar alternativni alat modelima DFT čvrste veze u predviđanju strukture ovakvih monoslojeva.

III.   Pretraživanje lokalnih minimuma

Statično pretraživanje lokalnih minimuma kroz raspon nagibnih kutova provedeno je upotrebom van der Wallsova funkcionala, za kojeg je pokazano da dobro opisuje nekovalentne interakcije.[IX] Interakcijska energija filma E_{film}, definirana kao:

$$E_{film} = E_{pbc} - E_{iso} \qquad \text{(I)}$$

[VIII] J. D. Gale, L. M. LeBlanc, P. R. Spackman, A. Silvestri, P. Raiteri *J. Chem. Theory Comput.*, **17** (2021) 7827–7849

[IX] K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, D. C. Langreth, *Phys. Rev. B* **82** (2010) 81101–81106.

gdje je E_{pbc} ukupna energija modelirane ćelije monosloja, a E_{iso} energija molekule stroja izoliranog u vakuumu. Općenito, E_{film} opada s povećanjem nagiba ili do minimuma određenog nagibnim kutom, ili se prestaje značajno mijenjati promjenom visine filma. Slika VIII pokazuje ovisnost nagiba gornjeg i donjeg dijela molekule te interakcijske energije filma o visini monosloja na primjeru **Azo** i **Fer** sustava. Tablica III daje pregled nagiba molekule za minimalne interakcijske energije filma pronađene za svaki sustav.
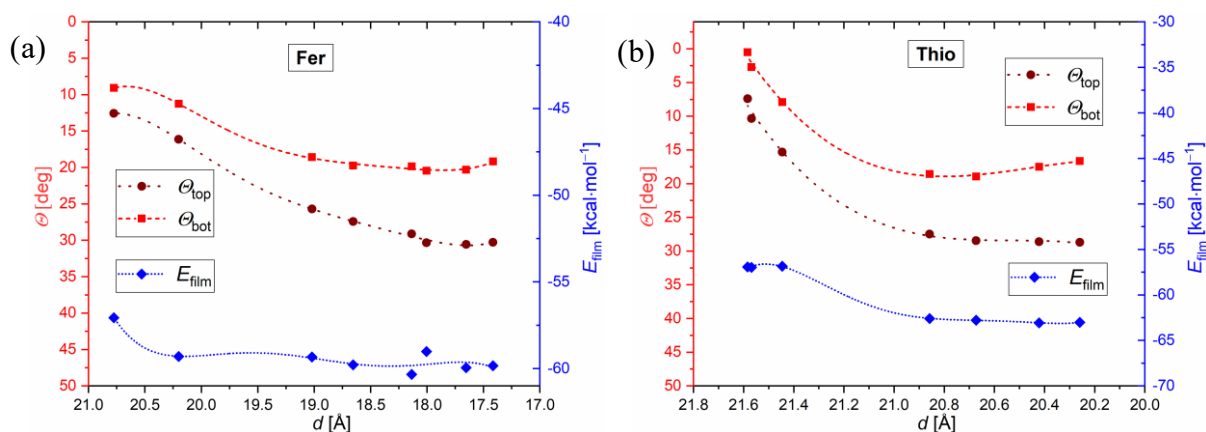
Slika VIII. Nagib i interakcijska energija filma kao funkcije debljine monosloja na primjerima: (a) **Fer** i (b) **Thio** sustava. Točke na grafovima prikazuju lokalne minimume na plohi potencijalne energije.

Tablica III. Konformeri s najpovoljnijim nagibom gornjeg i donjeg dijela molekule s pripadnim interakcijskim energijama filma.

	trans-**Azo**	*cis*-**Azo**	*cis*-**Dia**	*trans*-**Dia**	**Fer**	**Thio**
Θ_{bot} [deg]	20.5	23-36	17-26	22-23	11-20	16-19
Θ_{top} [deg]	52	25-27	25-43	43-44	16-30	27-29
$-E_{film}$ [kcal mol^{-1}]	54	44-45-	48-49	53-54	59-60	62-63

	Tri	**Phe0**	**Phe1**	**Phe**
Θ_{bot} [deg]	26-28	40	46	18-20
Θ_{top}, [deg]	36-37	38	36-38	28-31
$-E_{film}$ [kcal mol^{-1}]	52.5 - 53.5	25	28	38

Slika IX pokazuje promjenu interakcijske energije filma molekula *PHE* obitelji s nagibom. Minimumi na plohi potencijalne energije sustava bez ili sa samo jednom karboksilnom skupinom međusobno se značajno razlikuju u nagibu i stabilizacijskoj energiji (Slika IX (a)).

Također, s porastom broja veznih skupina, lokalni minimumi konformera s većim nagibom postaju plići, što pokazuje da uvođenje triju veznih skupina znatno stabilizira strukturu filma.
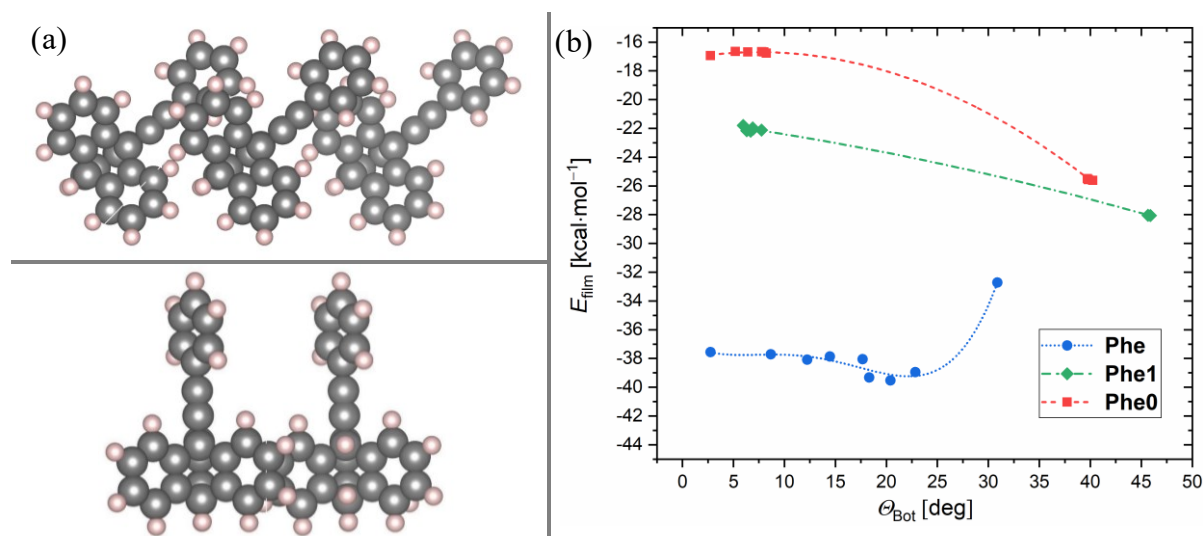
Slika IX. (a) Dva minimuma **Phe0** sustava s drastično različitim nagibom. C i H atomi prikazani su redom kao sive i bijele kuglice. (b) Ovisnost interakcijske energije filma o nagibnim kutovima *PHE* obitelji sustava bez (crvena), s jednom (zelena) te s tri (plava) vezne skupine. Točke na grafu prikazuju minimume na plohi potencijalne energije

IV.   Anizotropna IR spektroskopija i elipsometrija

Osim karakterizacije samoudruženih monoslojeva, polarizacijom modulirana infracrvena refleksijska adsorpcijska spektroskopija (PM-IRRAS) može se koristiti za procjenu prosječnog kuta nagiba molekula u monoslojevima usporedbom relativnih intenziteta vibracija molekula u izotropnom mediju i vibracija monoslojeva na vodljivoj površini.[X]

Za to su potrebne tri komponente: (i) PM-IRRAS spektar monosloja adsorbiranog na metalu, (ii) IR spektar izotropnog uzorka, primjerice praškastog uzorka u KBr paleti i (iii) teorijski izračun spektara kojim se asigniraju polarizacije vibracija. Da bi određivanje nagiba bilo moguće, moraju se pronaći vibracije čiji vektori prijelaznih momenata posjeduju isključivo okomitu ili paralelnu komponenu. Procjena kuta nagiba tada se može izračunati po:

$$\langle\theta\rangle = \arccos\left(\frac{1}{1 + 2\frac{I_{\mathrm{T}}(\mathrm{iso})}{I_{\mathrm{T}}(\mathrm{LB})}\cdot\frac{I_{\mathrm{L}}(\mathrm{iso})}{I_{\mathrm{L}}(\mathrm{LB})}}\right)^{0.5} \tag{II}$$

gdje je $\langle\theta\rangle$ prosječni kut nagiba molekula u monosloju, $I(iso)$ intenzitet vibracije u izotropnom mediju, $I(\mathrm{LB})$ intenzitet vibracije na metalnoj površini, a T i L u indeksu označavaju redom transverzalnu i longitudinalnu polarizaciju vibracije. Slika X prikazuje primjer vibracija s vertikalnim i okomitim vektorom prijelaznog momenta i primjer anizotropnog, izotropnog i izračunatog vibracijskog spektra **Azo** molekule. Prosječni nagibni kutevi na navedeni su način izračunati za eksperimentalno dostupne **Azo**, **Dia**, **Fer** i **Thio** sustave. Rezultati su sažeti u Tablici III.
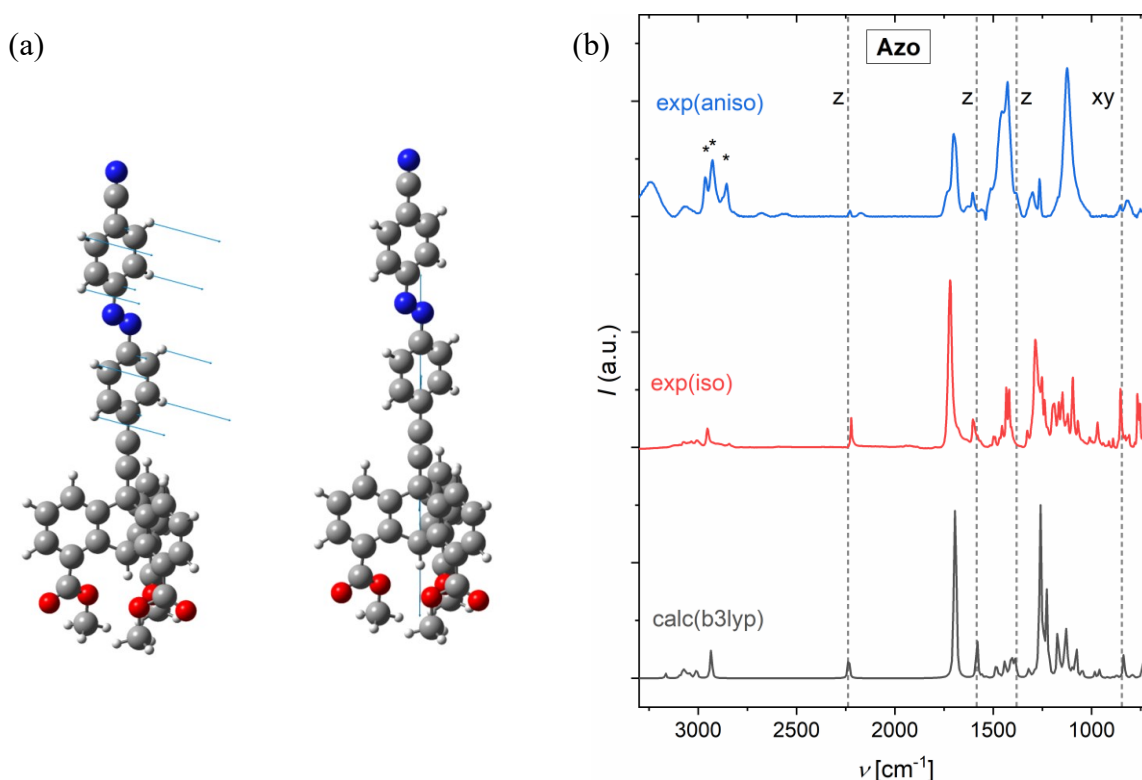
Slika X. (a) Vibracije s isključivo paralelnim (lijevo) i okomitim vektorima (desno) momentima prijelaza (plave strelice) s obzirom na metalnu površinu, na primjeru **Azo** molekule. (c) PM-IRRAS (plavo), izotropni (crveno) i izračunati (sivo) IR spektar **Azo** sustava s naznačenim signalima koji odgovaraju okomitim (xy) te paralelnim vibracijama s obzirom na normalu površine.

Tablica III. Prosječni nagibni kutovi tronožnih molekulskih strojeva adsorbiranih na (111) površinu zlata procijenjeni usporedbom izo- i anizotropnih IR spektara

Sustav	**Azo**	**Dia**	**Thio**	**Fer**
	43.1	26.5	36.7	28.3
	36.5	29.2	34.3	26.5
Θ_{est}	39.0	23.2	43.2	34.4
	-	38.9	45.7	20.4
$\langle\Theta\rangle$	34 ± 4	29 ± 8	40 ± 6	27 ± 7

Za sintetizirane sustave, prosječni nagibni kut je određen i usporedbom elipsometrijski određene debljine monosloja s maksimalnom teorijskom visinom uspravnih molekula. Prosječni nagibni kutovi tada su izračunati kao *arccos* kvocijenta dviju debljina. Rezultati su sažeti u Tablici IV.

Tablica IV. Maksimalna teorijska i izmjerena elipsometrijska debljina i prosječni nagibni kut eksperimentalno dostupnih sustava molekulskih monoslojeva adsorbiranih na zlatu

Sustav	$d(\text{izračunato})_{\text{max}}$	$d(\text{elipsometrija})$	Θ_{avg}
Azo	22.5	19 ± 2	32 ± 9
Dia	17.8	19 ± 2	6 (donji limit)
Fer	18.5	19 ± 1	3 (donji limit)
Thio	20.8	17 ± 1	35 ± 5

Slika XI ilustrira razliku u prosječnim nagibima monoslojeva 4 razlilčita stroja. Rezultati pokazuju kako se trendovi u nagibnim kutovima dobiveni različitim tehnikama mijenjaju s različitim strojevima. Nagibi određeni PM-IRRAS tehnikom nalaze se između računalnim metodama (statičko skeniranje i molekulska dinamika) procijenjenih nagiba, što je očekivano, budući da interakcija između zlata i veznih skupina stabilizira uspravne strukture, a termičko gibanje molekula, koje statički računi ne mogu opisati, povećava nagibni kut.

Mjerenje elipsometrijske debljine redovito precjenjuje debljinu monosloja,[XI] što objašnjava zašto, dok su općeniti trendovi uglavnom dobro reproducirani sa sve četiri metode, nagibni

---

[XI] H. Fujiwara in H. Fujiwara, R. Collins (Eds.) *Spectroscopic Ellipsometry for Photovoltaics*, Vol 212 Springer: Cham, 2018, pages 155−172.

kutevi procijenjen elipsometrijski (**Azo** i **Thio** imaju velike kuteve nagiba, **Dia** i **Fer** male) su niži u odnosu na ostale metode.
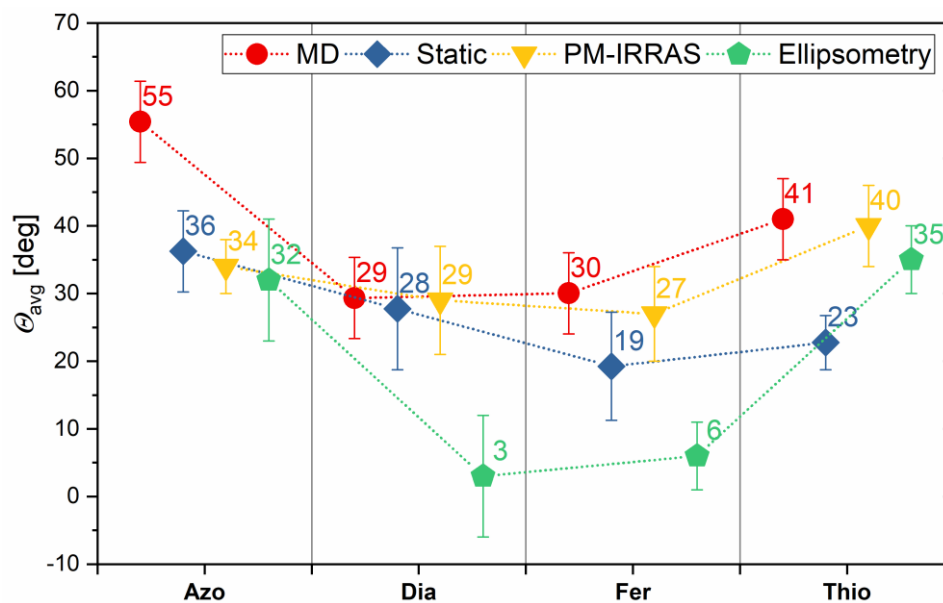
Slika XI. Izračunati prosječni kutevi nagiba metodom molekulske dinamike (crveno) i statičnim izračunima (plavo) te procijenjeni eksperimentalno, PM-IRRAS tehnikom (žuto) i preko elipsometrijske debljine monosloja (zeleno).

§ 1. INTRODUCTION

Designing novel sophisticated molecular-level machines is a clear way forward in the development of the next-generation smart materials, with applications ranging from information storage[1,2] and photovoltaics[3–6] to directed drug delivery.[7] In context of this thesis, the term molecular machine refers to various light-activated switches and motors which can repetitively respond to the external stimuli without significant loss of efficiency.

In self-assembled monolayers (SAMs), regular arrays of molecular machines are formed by their spontaneous adsorption on various surfaces, such as metallic surfaces, silica, or crystal faces, and are usually characterized by large and ordered domains.[8–10] As SAMs present the possibility of achieving both the unrestricted switching characteristic of molecules in solution and the structural order typical for crystalline solids, furthering the understanding of the structure and properties of these systems is pertinent for improving the design of novel materials for real-world applications. The approach which has been pursued in Kaleta group utilises the Langmuir-Blodgett technique, enabling smooth production of large surfaces covered by ordered arrays of molecular machines based on triptycene[10–12] (Figure 1(a)). As these assemblies are stabilised by van der Waals interactions, the final film structure is determined by molecular building blocks which allows for finetuning of the film properties.

It has recently been suggested that SAMs anchored to a (111) gold surface show tendency to collectively bend[10]. This insight opened a window to a series of questions such as: do triptycene-based SAMs prone to bending form 2D or 3D structures? Can the packing of these films be directed by supramolecular engineering? How do the bending and tilting of individual molecules affect the packing? How significant are the $\pi-\pi^*$ interactions in determining the structural properties of SAMs? Can the interplay between the triptycene interactions and the anchor−surface interactions be accurately described?

In presented study we attempt to answer some of these questions, focusing primarily on the collective tilting and bending of regular arrays of molecular machines. We also explore how different structural motifs influence the film structure, and further investigate the dynamic properties of these self-assemblies, such as the collectivity of their modes of motion and the role of entropy in their stability. The average tilt angle Θ, central to this thesis, can be measured as an inclination of the line connecting the C-bridgehead atom of triptycene stand and the

aromatic centre of functional group from the normal of a plane defined by the triptycene C bridgehead atoms of the neighbouring machines. Figure 1 (a) shows the general definition of the tilt angle of a triptycene-based machine, while the assumed packing pattern, which was observed experimentally on similar systems deposited on the water surface,[12] is illustrated in (b).
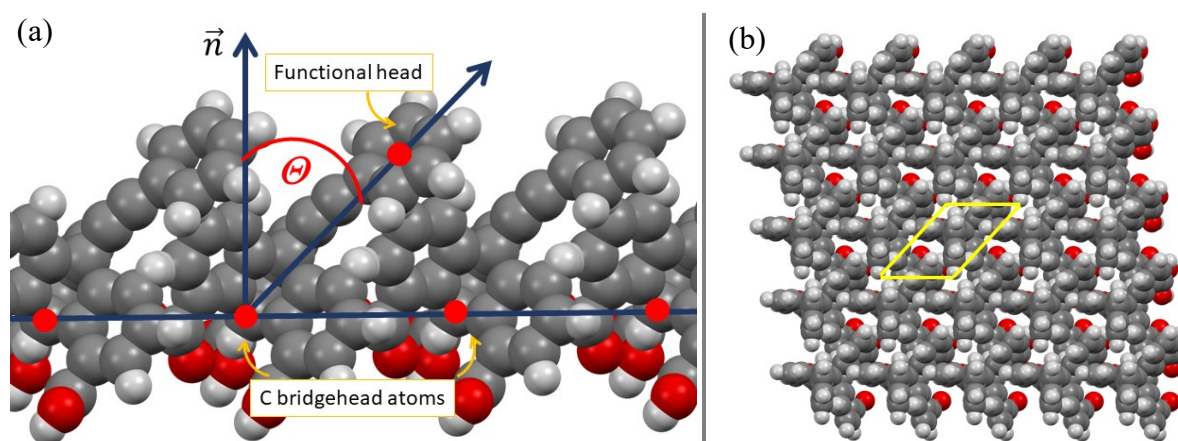
Figure 1. (a) definition of the average tilt angle θ illustrated on a general SAM structure of triptycene-based machines, (b) top-down view of the monolayer structure of triptycene based machines and the assumed packing pattern. Unit cell vectors are denoted with yellow lines.

In effort to further stabilise the SAMs of molecular machines on surfaces, the most recent focus of Kaleta group[13] has been on the colloquially called *tripodal systems*, which share the same type of tripodal stand containing one carboxylic function on each phenyl blade of the triptycene core and differ in their functional heads (upper right corner of Chart 1). There are several common features in all the investigated structures, namely:

(i)   A core made of the triptycene unit whose phenyl blades organize and stabilize the monolayer by intermolecular $\pi-\pi^*$ interactions

(ii)   Functional head defining the functionality of a molecular machine

(iii)   A triple bond linker providing sufficient separation between the functional head and the triptycene stand

(iv)   A varying number of carboxylic anchoring groups connected to the triptycene unit that force the terminal bond in a perpendicular perpendicular orientation in respect to the metallic or aqueous subphase surface.

Chart 1 summarises all the investigated systems in the presented work. **Azo**, **Dia** and **Thio** are tripodal systems bearing azobenzene, diazocine and thiophene photoswitches, respectively. In case of **Azo** and **Dia** systems, their stable and metastable isomers will be differentiated with *trans*- and *cis*- prefixes throughout the thesis, referring to the isomers of their mounted functional units. **Fer** is the tripodal molecular motor, bearing a second-generation Feringa's functional head, based on an overcrowded alkene,[14] and **Tri** represents a model system with two triptycene units. Furthermore, *PHE* series of systems, namely **Phe0**, **Phe1** and **Phe** all bear a phenyl ring as a functional head and differ only in the number of anchoring groups, with zero, one or three carboxylic anchors, respectively. Concerning anchoring groups (denoted in red in Chart 1), carboxylic acids should be assumed, except the calculations performed in vacuum, where the methylated analogues are modelled instead.
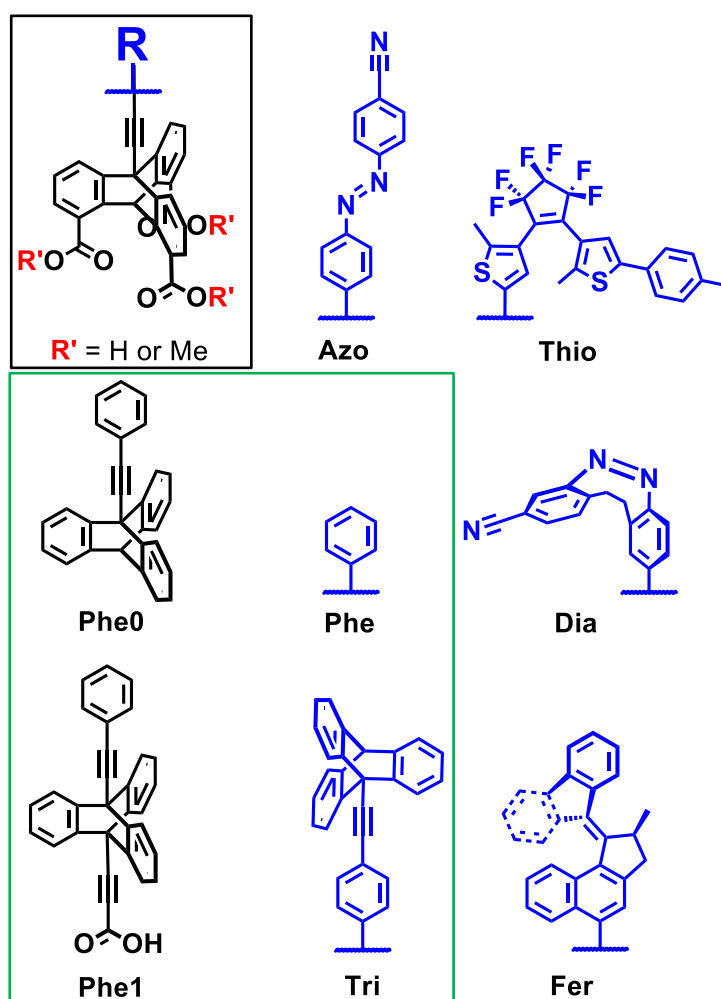
Chart 1. Structures of investigated compounds. General structure of a tripodal system is denoted in the black rectangle with a varying **R** (blue

moieties). **Azo**, **Thio**, **Dia** and **Fer** were synthesized prior to this thesis[13] while the *PHE* series (**Phe0**, **Phe1**, **Phe**) and **Tri** were investigated only computationally and are highlighted by the green rectangle.

This thesis presents an extensive computational and spectroscopic investigation of SAMs properties of molecular machines and can be divided into three major parts:

(i) Molecular dynamics investigation of SAMs properties on the water-air interface

(ii) Extensive scanning of potential energy surfaces of SAMs in vacuum

(iii) The evaluation of the SAMs structure by means of iso- and anisotropic IR spectroscopy.

As such, the thesis heavily relies on the extensive collaborative work that was performed in Kaleta group. Synthesis and characterisation of **Azo**, **Dia**, **Thio** and **Fer** systems were done by other people,[13] but the available experimental data is readily used throughout the thesis. Moreover, although processing and the analysis of spectroscopy measurements were done by the author, and are important part of the thesis, the author did not perform the measurements.[13]

§ 2. LITERATURE REVIEW

Study of molecular machines, from synthesis and characterisation to potential applications has developed into a full-fledged interdisciplinary area of research. In this chapter we define molecular machines and describe differences in their modes of operation, concentrating mostly on switches and motors investigated in this thesis. In subsequent sections, we discuss the achievability of collective behaviour of regular arrays of molecular machines and their self-assembly. We mention different experimental techniques used in structure determination of self-assembled monolayers and we point the readers to papers which use the mentioned techniques in context of SAMs characterisation. Infrared spectroscopy techniques we describe in more detail, as IR reflection adsorption spectroscopy was utilised for tilt determination in this thesis.

2.1. Molecular machines

Generally, when we think of classic, macroscopic mechanical machines, we tend to think of some apparatus composed of multiple parts, each having a definite function and being able to collectively perform a particular task. Following this logic, molecular machines can be defined as an assembly of molecular components designed to perform specific tasks of mechanical nature in response to appropriate external stimuli. Alternatively, to collectively include motors and switches in the definition, one can think of molecular machines as the sets of functional molecular components that can use the transformation of energy in a predictive manner and whose externally controlled movements of nuclei (conformational, configurational, translational, or rotational movement) either produce work or store the information. For a molecular machine to maintain operation within an environment where thermal motion exceeds the actual motions of the machines, it must be able to use or rectify such fluctuations as an operational principle.[15,16]

With the rapid advancements in molecular biology, a wide spectrum of diverse molecular machines with high complexity has been found in nature.[17,18] Cytoplasmic proteins capable of movement, such as myosins and kinesins, are supramolecular conjugates that use adenosine triphosphate (ATP) molecules as fuel to slide along linear actin filaments and microtubules

transporting different substrates[18]. In response to certain biochemical stimuli, these proteins perform directed or programmed motions. Maybe the most representative example of a biomolecular machine is the ATP synthase, which synthesises and hydrolyses ATP through rotary motion.[19] Considering the variety and complexity of these examples, it is not surprising that the inspiration for the intelligent design of artificial molecular machines has mostly been found in nature. Nanomachines of the biological world serve as proven examples of the potential utility and sophistication of nanotechnology and constitute a goal yet to be achieved by the synthetic approach. In recent years, a great deal of effort has been put into the design of highly functional artificial molecular machines.[20]

For the molecular machines to be able to operate, they need a source of energy as an input, so it is a valid thing to ask oneself: what is the optimal energy source that should be used to power them? Not surprisingly, most of the biological molecular motors are powered by chemical means, either by varying the concentration gradients of chemical species in the microenvironment of a molecular machine (e.g., proton pumps) or in a cascade of exergonic reactions (e.g., ATP hydrolysis).[19] However, when designing an artificial molecular machine, inputs of chemical energy generate an array of problems, such as the need for the constant addition of new reactants to the system or dealing with the generation of waste products that compromise the operation of molecular devices. These are formidable challenges, as they destroy the autonomy and reusability of molecular machines as functioning complex materials. Luckily, besides chemical energy, there are other sources of energy that can be utilised to power a molecular machine. Light as an energy input is especially interesting since, besides not generating waste products, has several further advantages compared to other methods, such as:

(i)   Switching of photoswitches or rotary motion of molecular motors can be (de)activated by controlling the wavelength of incident light in relation to the absorption spectrum of molecular machines, and the amount of energy is easily controlled by its intensity. Stability and performance tests are practical with photobleaching experiments.

(ii)   Energy transfer can be done without a physical connection to the molecular machines

(iii)   Utilizing lasers as an energy source provides an opportunity of working both in small spaces in very short time intervals and irradiating large areas of material.

(iv)   Monitoring the operation of molecular machines is usually straightforward.[XII]

For these reasons, the molecular machines designed in Kaleta group and described in the later chapters of this thesis are all light driven.[10–13,22–25]

Besides the energy input, the motion of a machine needs to be directed. If there is no organised direction during the mechanical motion of a molecular machine, either no work can be done on average, or the system cannot transform the energy to store information. At room temperature, the thermal noise is not negligible, and every easily rotatable bond in an organic molecule is going to rotate both in a clockwise and anticlockwise direction, governed by the Brownian motion. For example, the symmetry of an acetonitrile molecule at room temperature is better described as $C_{\infty v}$ instead of C_{3v} because what is being observed is the average of rotation minima on the potential energy surface. If freely rotatable functional groups possess strong dipole moments, they can be considered as molecular rotors, because their motion is correlated and can be strongly coupled with the environment. Apart from molecular rotors, molecular machines can generally be divided into two classes: if the controlled movement of a molecular system can be used to perform work, the machine is considered to be a *motor*, while systems that cannot do so, but can be used for storing information, are distinguished as *switches*.[26]

2.2. Molecular switches and motors

2.2.1. Switches.

The defining feature of a molecular switch is its ability to reversibly shift between two or more (meta)stable states. The input of energy in the form of a photon can lead to a population of an excited state that has conical intersections which lead to relaxation to the other isomer.[27] Switches can in general be categorised as either active or passive (in which case they are typically called sensors). Plethora of different switches were designed over the years, such as

---

[XII] In some cases, monitoring the operation of fuel-driven molecular machines is also straightforward, for example, by measuring the H_2O_2 concentration which is used as a fuel is exceedingly simple.[21] However, these examples are quite rare.

hydrazones, spiropyranes, fulgides and fulgimides, and have mostly been studied in solution.[26] In this work, three types of active switches based on different chromophores serve as functional heads mounted on triptycenes: (i) azobenzene (ii) diazocine and (iii) thiophene switch, and their modes of operation are summarised in Figure 2.
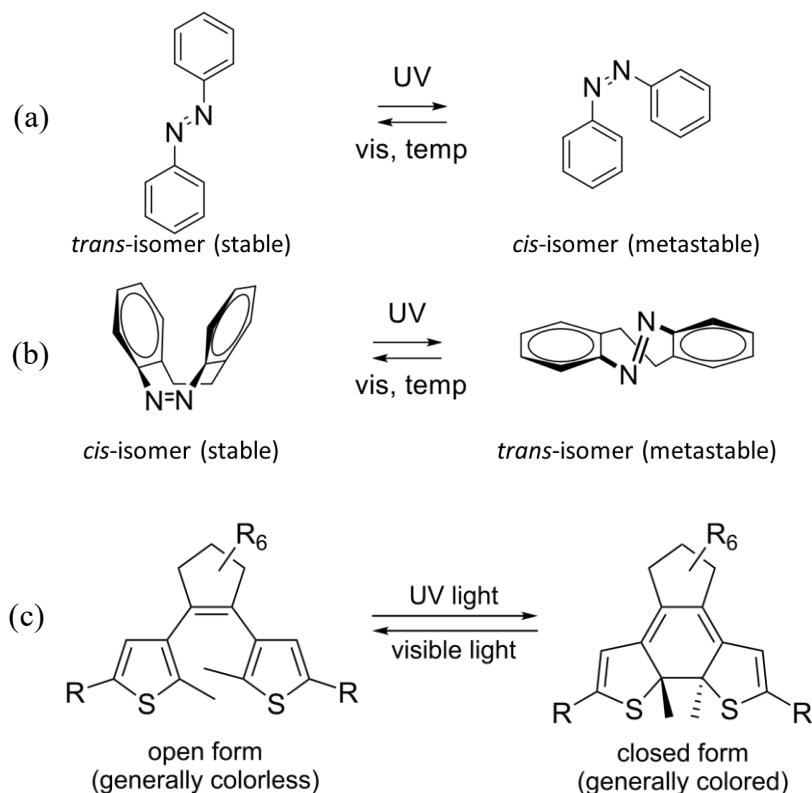
Figure 2. Switching process of (a) azobenzene (b) diazocine (c) thiophene switches, adapted from[26]

2.2.2. Molecular motors

Contrary to rotaxanes[28] and catenanes,[29] motors discussed in this thesis can be defined as molecules held together by chemical bonds, composed of two functional parts that can rotate relative to each other around a common axis. The part with a larger moment of inertia is defined as a stator, while the part with a smaller moment of inertia is called a rotor. In general, molecular motors can be studied either floating freely in an isotropic solution, attached to a surface, or embedded in a crystalline solid. When mounted onto a flat substrate, they are called azimuthal when their rotation axis is parallel, and altitudinal when it is perpendicular to the surface normal (Figure 3).[26]
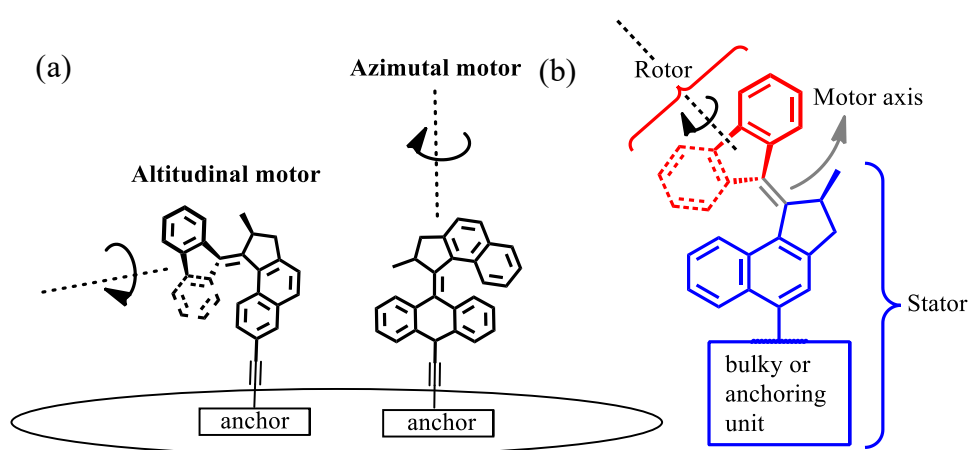
Figure 3. (a) Azimuthal vs altitudinal rotors (b) General components of a molecular motor.

Unsymmetrically substituted alkenes ($R_1R_2C = CR_3R_4$) have *E* and *Z* isomers which cannot interchange thermally. If the interconversion between the two forms can be induced by light, these molecules can be used as a basis for the previously described molecular switches. However, to make a motor out of it, one needs to introduce directionality into these systems which can sometimes be achieved by overcrowding the alkene around the double bond with the introduction of substituents.[14]

Generally, if the substituents are large enough, the planarity of the double bond is disrupted and a variety of twisted and folded conformations can be adopted, including helically shaped conformations. The direction in which the upper and lower halves of an overcrowded alkene rotate when passing through the transition state geometry is uncontrolled in overcrowded alkenes and hence, although absorption of a photon will lead to roughly a 180° rotation about the $C = C$ axis, the rotation will proceed both clockwise and anticlockwise in equal measure. However, augmenting the intrinsic helical chirality of the molecule with a second stereogenic centre imposes a bias for one conformation (pseudo-equatorial) over the other possible conformation (pseudo-axial) of a methyl group that is part of a five or six-membered ring. The introduction of stereogenic centres on these switches was key to the development of the family of rotary molecular motors based on overcrowded alkenes, usually called Feringa's type motors (Figure 4).[15]
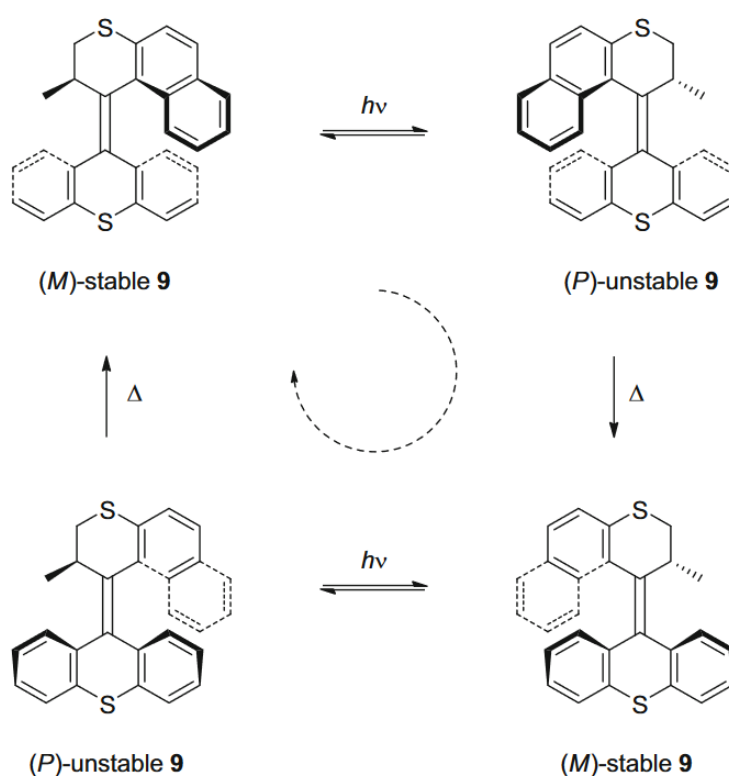
Figure 4. Interconversion process in a general type of Feringa's motor, based on the unsymmetrically substituted overcrowded alkene, adapted from[30]

2.3. Achieving collective behaviour

As chemistry is still mostly thought of as the "solution science", it is to be expected that the first efforts in synthesis, characterisation and functionality evaluation of molecular machines have been invested in molecular rotors and switches randomly oriented in isotropic solutions. However, for most real-world applications, for example, in next generation memory devices, or molecule-propelled nanostructures, switches and motors need to be attached to a macroscopic surface, either curved or flat, or imbedded in cavities of crystals.[26] Individual molecules also need to be arranged in regular 2- and 3D arrays and switches and motors should all be oriented in the same direction. In case of regular arrays of molecular motors, their rotational motion produces a net torque larger than zero, providing motion at macroscopic level.[31] For successful operation of these types of materials, it is desirable to maximise correlation between rotational motion of individual machines and minimize the collective behaviour of other degrees of freedom, such as thermal fluctuations, as those are bound to

destabilise the regularity of assembled arrays. Correlation in modes of motion can be achieved either through mechanical interference or by the inter-rotor interactions. In case of switches, when organised properly, the switching of specific zones[XIII] on the surface or within the crystal[XIV] can be controlled externally, for example by irradiation or with an STM tip, making these materials attractive candidates for memory storage units.[32]

There are multiple approaches in construction of regular arrays; in three dimensions, the most straightforward way is to grow crystals of suitably designed molecules,[33-35] embed the machines inside the metal-organic frameworks (MOFs),[33] or make a multi-layered structure by continuous deposition on surfaces. However, crystal engineering of such 3d materials with desired properties is a daunting task and remains a challenge in solid state community. Prepared crystals should be free of significant defects and should achieve desired intermolecular arrangements, distances, and orientations within the crystal. It is not surprising that only a handful of such arrays showed tendency for collective behaviour. This provides us with clear motivation to control the structure and properties of assemblies of molecular machines in regular 2D arrays.

2.4.   Self-assembled monolayers of molecular machines

Self-assembly can be defined as a process of spontaneous and reversible molecular organization into well-defined architectures through non-covalent interactions, such as hydrogen and halogen bonds and van der Waals interactions. Although these forces are weak, their collective interactions can produce structurally and chemically stable structures through an interplay between energetic end entropic contributions. For self-assembly to be achievable, the molecular components need complementary properties such as specific surface characteristics, surface charge, polarizability, and surface functionalities to self-assemble into different regular patterns.[36-37]

[XIII] Excitation wavelength for most known switches is longer than the intermolecular distance in typical regular arrays. This makes achieving the control of switching on molecular level improbable. However, excitation of whole zones remains within experimental reach in the future.

[XIV] The switching wavelength is influenced by the environment of the switch, so excitations of specific zones should also be achievable inside the 3D structures.

A functional material based on molecular machines should have two key features: the unrestricted modes of operation of switches and motors and a high degree of organisation necessary for achieving collective behaviour (for example, allowing molecular motors to all unrestricted rotate in the same direction upon irradiation). Self-assembly of molecular machines into organised two-dimensional arrays is thus a popular way of preparation of chemically and structurally well-defined organic architectures.[26]

2D-assemblies are particularly interesting since the properties of surfaces they are adsorbed on can be tailored by introducing different structural motifs to molecules building the monolayers. However, the robustness of the monolayers is very dependent on several factors, such as: the structure of the adsorbate, anchoring of the adsorbate onto the substrate (effect of anchoring groups, energetic and entropic effects governing the adsorption), and the lattice parameters of a metallic surface – difference in the lattice parameters will direct the self-assembly process and thus influence structure of the formed 2D films.[26]

There are three major techniques for preparation of these systems: (i) self-assembly on metallic surfaces, (ii) Langmuir−Blodgett technique, and (iii) anchoring of molecular machines on the flat facets of porous crystalline matrices. Each of these approaches has its strengths and weaknesses and has been successfully utilised in Kaleta group.[10−13,22−26,38,39]

2.4.1. Triptycene based SAMs

It is well-known that triptycenes and similar structures readily pack on a surface, producing regular 2D arrays by self-assembly.[40−42] This proclivity of triptycenes to form organised supramolecular films was utilized in Kaleta group by successfully designing several generations of amphiphilic triptycene-based rods for Langmuir-Blodgett applications and subsequently installing both switches and rotors onto them. Examples of different generations of triptycene based machines designed in Kaleta group are shown in Figure 5.
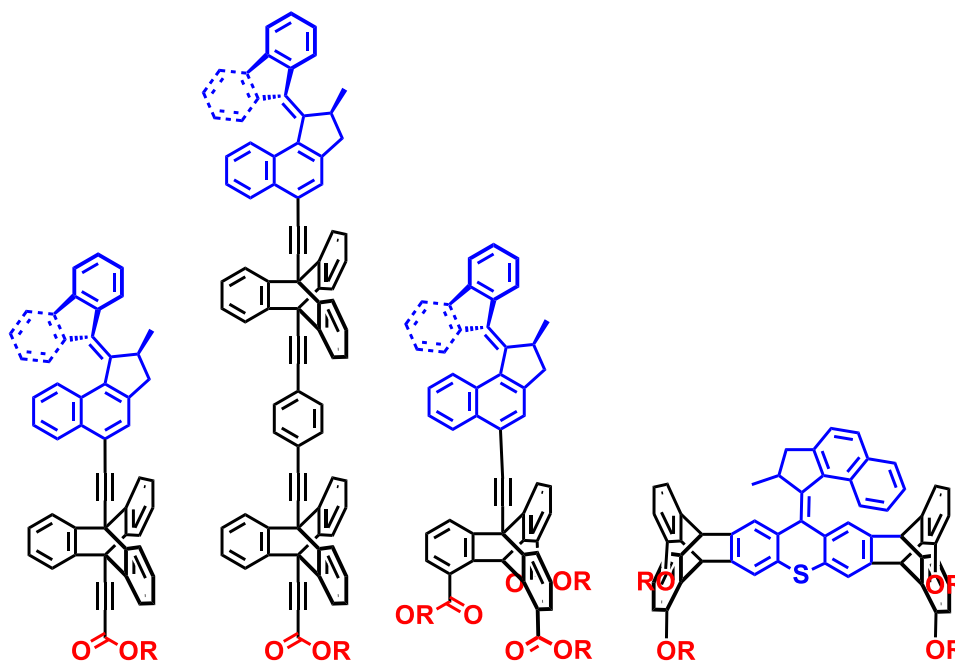
Figure 5. Examples of triptycene based molecular machines designed in Kaleta group: (a) monopods (1st generation),[11] (b) double-deckers (2nd generation),[24] (c) tripodal systems (3rd generation),[13] (d) tetrapods (4th generation).[43] Coloured motifs may vary; overcrowded alkene molecular motor is assumed in all examples for simplicity, although other functional mounts were also successfully prepared for each generation of SAMs. R denotes an arbitrary anchoring group.

2.5. Polymorphism in triptycene-based films.

Theoretical prediction of polymorphism is a very challenging task; kinetic effects frequently determine which polymorph will be dominant in the system, and evaluation of thermodynamics of different possible polymorphs is seldom enough. Moreover, the difference in stability between different phases is highly dependent on the local environment and can change drastically with nucleation conditions and slight variations of thermodynamic variables. That the polymorphism of triptycene based LB-films is still debated is thus hardly surprising. [12,42]

Das et al.[42] have recently recorded STM images of tripodal systems bearing no functional head and suggest simultaneous presence of both hexagonal and trigonal polymorphs on (111) Ag surface deposited on Au. Pronounced polymorphism was observed, with the formation of several very different packing patterns highly dependent on the preparation conditions, ranging from close-packed hexagonal and honeycomb molecular arrangements to trigonal shapes.

In case of complex triptycene-based molecular machines, situation is even more complicated, as effects of molecular heads can influence the relative stability of different polymorphs as well as the type of substrate, number of anchoring groups and, in case of the assemblies at the water/air interface, the prominent effects of complexation with different ions.[11] A recent computational study done by Rončević et al. indicated that the hexagonal polymorph is the most stable one out of three investigated in case of arrays of triptycene-based machines bearing a single anchoring group in vacuum (Figure 6).[24]

Although many different polymorphs are thus plausible, in this work we assume all tripods are packing in patterns represented as polymorph 1 in Figure 6, as this is the polymorph observed experimentally on similar systems deposited on the surface of water subphase,[12] and a thorough investigation of every possible polymorph would be outside the scope of this thesis.
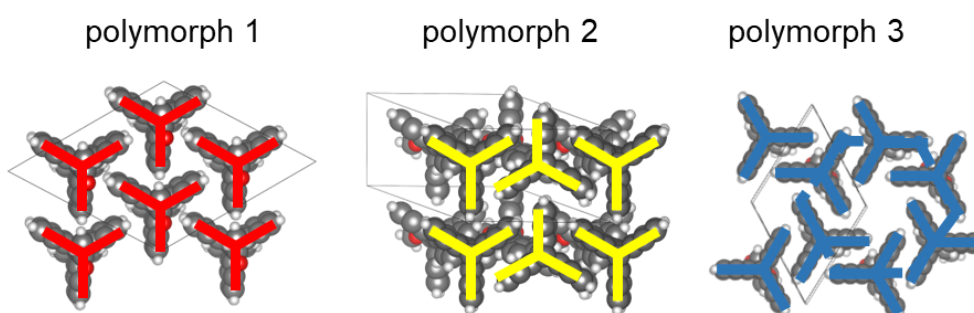
Figure 6. Some of the possible polymorphs of triptycene based monolayers, adapted from[24]

2.6. Langmuir-Blodgett technique

The Langmuir-Blodgett (LB) technique is one of the most frequently used methods of preparing organized molecular assemblies as it provides a level of control over the orientation and placement of molecules in monolayers that are difficult to otherwise achieve. Generally, LB technique consists of two main steps: (a) Preparation of floating monolayer film at air-water interface (b) Deposition of the Langmuir film on solid substrate (Figure 7 (a)). The layer of molecules on a liquid surface is usually termed a Langmuir monolayer and after transferring it is called a Langmuir–Blodgett film.[44]

To form a Langmuir monolayer film, the molecule of interest is dissolved in a volatile organic solvent that will not dissolve or react with the subphase. The dilute solution is then placed on the subphase of the LB trough. The solvents evaporate and the surfactant molecules

spread over the subphase surface in the LB trough. When organic molecules arrive at the air/water interface with their hydrophobic parts pointing towards the air and the hydrophilic groups towards water, the initial high energy interface tends to relax by exchanging the surrounding interace with the hydrophilic – hydrophilic and hydrophobic – hydrophobic pairings of molecular parts. Hence, the molecules at the interface are non-covalently anchored to water surface, oriented parallel to its normal.[44]

Stability of the Langmuir monolayer is estimated by monitoring the surface tension which is affected to a large extent by the presence of molecules at the water/air interface. Initially, the effect of molecules on the surface tension is relatively low since the molecules float freely in the LB-through. When the barrier is compressed, the available surface area of the floating film is reduced. As a result, inter molecular distance between the adjacent molecules shortens and the surface tension decreases, and floating film-building molecules start to interact with each other.

Typically, this surface pressure (π) (force exerted by surface film per unit length) is equal to the decrease in surface tension of the air/water interface due to the presence of the monolayer and can be expressed as:

$$\pi = \gamma_0 - \gamma \qquad (1)$$

where γ_0 is the surface tension of the air/water interface and γ is the surface tension in the presence of a monolayer. During the compression of the organic phase, self-organization of the molecules occur and the floating monolayer at the interface undergoes several phase changes. These changes during the compression can be described by the surface pressure / area per molecule (π / APM) isotherm. A typical π / APM isotherm curve is shown in Figure 7 (b). Shape of the curve is largely dependent on both the condition of the water subphase (pH, ionic strength, type of ions in the solution, temperature) and on the properties of molecules building the organic phase. π / APM isothermes provide information about thermodynamic behaviour and the stability of the floating film, the orientation of the molecules and the existence of the phase transition as well as of the conformational transformation of air-water interface.[45]
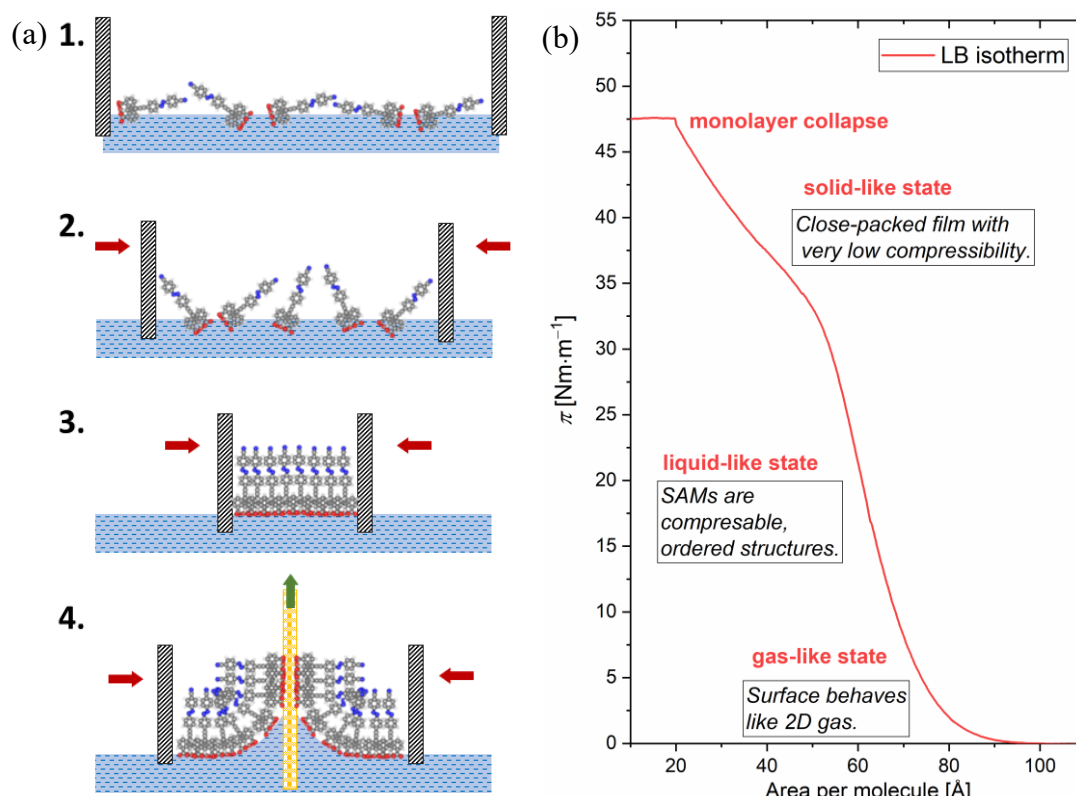
Figure 7. (a) Schematic representation of the LB deposition process (b) An example of the LB isotherm.

The surface pressure for the film deposition is usually chosen to be in the solid-like region. However, the film can be deposited on any point of the π / APM isotherm and, in Kaleta group, it is usually deposited at the point where π just starts to rise since it corresponds to the APM at which the monolayers have the minimum energy. The transfer of monolayer film occurs via hydrophobic interactions between the organic phase and the substrate or by the hydrophilic interaction between the head groups of the molecules and the hydrophilic surface. By consequtive repetition of dipping the substrate in the LB trough, the desired number of layers can theoretically be deposited.

2.7. Characterisation of Langmuir-Blodgett films

Experiments can provide a lot of information about the SAM structure on different surfaces, and many innovative techniques have been successfully utilized over the years in order to characterise self-assembled films. When possible, structure determination by grazing incidence X-ray diffraction (GIXD) is an obvious choice.[12] However, it is an expensive method and therefore rarely practical when other, cheaper techniques are available. When films are

deposited on a metallic surface, various analytical techniques can be easily performed, ranging from spectroscopies (for example, X-ray photoelectron spectroscopy (XPS) polarization modulation-infrared reflection-adsorption spectroscopy (PM-IRRAS), Raman), through surface plasmon resonance (SPR) and ellipsometry to various microscopies (such as atomic force microscopy (AFM), scanning tunnelling microscopy (STM)).[26] Moreover, Langmuir-Blodgett technique, described in the previous section, besides being a classic method for preparation of such films, can also provide information about the relative density of the self-assembled molecules on the water surface by estimating the mean molecular area. **Table 1** summarises the most frequently applied techniques for LB film characterisation, and the information they usually provide.[44]

Table 1. Techniques commonly used for LB film characterisation and description of properties they evaluate.

Experimental technique	Extracted information
Infrared reflection absorption spectroscopy (IRRAS)	SAM packing patterns and conformation, degree of ionization, H-bonding, chemical and structural changes, molecular orientation.
Surface potential measurements	Polarization, orientation
Ellipsometry	Refractive index and thickness measurement
X-ray diffraction / reflection	Inter layer spacing, in-plane lattice structure, in-plane / out-of-plane orientation
Grazing incidence and small angle – X-ray diffraction	Inter layer spacing, in-plane lattice structure, in-plane / out-of-plane orientation
Neutron diffraction	Inter-layer spacing
UV-Vis absorption spectroscopy	Electronic transition and orientation
Raman Spectroscopy	Identification and orientation, conformation of molecules, molecular interactions within LB films
Scanning electron microscopy (SEM)	Surface morphology, domain structure, patterns, holes, and defects (resolution of 50 nm).
Transmission electron microscopy (TEM)	Surface morphology, domain structure, patterns, holes, and defects (resolution of 50 nm).
Scanning tunnelling microscopy (STM)	Imaging surface at the atomic level, visualisation of defects, sub-molecular packing (lateral resolution of 0.2 nm).
Atomic force microscopy (AFM)	Imaging surface at the atomic level, visualisation of defects, sub-molecular packing (lateral resolution of 0.2 nm).
X-ray photoelectron spectroscopy (XPS)	Quantitative analysis of chemical composition, molecular orientation estimate
Electron spectroscopy for chemical analysis (ESCA)	Quantitative analysis of chemical composition, molecular orientation estimate
Electrochemical measurements	Electrical behaviour.

Brewster Angle Microscopy (BAM)	In-situ study of various SAM phases (sensitive to the surface density and monolayer domains anisotropy)

In Kaleta group, PM-IRRAS has been routinely used in SAM characterisation, as it is a simple, cheap, and non-destructive method of acquiring the molecular information such as conformation, chemical bonds formation with the substrate, orientation of functional groups of SAMs and estimation of average tilting. Basic theory behind the operation of PM-IRRAS will be described subsequently in more detail, as the PM-IRRAS results are important for this thesis. Ellipsometry, although usually used to determine optical properties of materials, is another frequently applied technique in Kaleta group used for the estimation of the average optical thickness of organic films adsorbed on conductive surfaces.

§ 3. THEORETICAL BASIS

3.1. Iso- vs Anisotropic IR spectroscopy

The importance of polarisation modulation IR reflection adsorption spectroscopy in the context of SAMs is threefold, as it: (i) enables the characterisation of organic film structure by providing information about characteristic functional groups present within SAMs, (ii) provides information about the degree of their organisation on conductive surfaces since modulated reflectivity is mostly independent from isotropic adsorption of the material and (iii) enables determination of average tilting of the molecules of which the films are composed of.[11]

This is done by comparing the change in relative intensities of vibrations with either planar or coplanar polarisations with respect to the conductive surface the molecules are adsorbed on with the same vibrations in an isotropic medium. This works since only the normal component of the transition moment vector of the molecular vibration mounted on the conductive surface is active, so the relative intensities of IR signals of monolayers adsorbed on metals compared to the same signals in the the solution or in the KBr pellets provide the information about the orientation of the vibrational transition moment vectors on the surface.

The orientations of the adsorbed molecular groups can be easily calculated from the IRRAS (or corrected PM-IRRAS) spectra because of the anisotropy of the surface electric field (only the component of the electric field normal to the surface E_Z is different from zero in the vicinity of the metallic surface). Knowing this, we can write for the absorbance I_Z of light polarised along z direction:

$$I_Z = \langle \cos^2 \theta_x \rangle A_x + \langle \cos^2 \theta_y \rangle A_y + \langle \cos^2 \theta_z \rangle A_z \tag{2}$$

where A_i are the components of the i-polarised absorbance of the molecule. The absorbance I_{iso} measured in KBr pellet under isotropic conditions can be approximated as:

$$I_{iso} = \frac{1}{3} A_x + \frac{1}{3} A_y + \frac{1}{3} A_z \tag{3}$$

The ratios K between z-polarised and isotropic absorbances are different for longitudinal transitions (polarised along z) and transverse transitions (polarised in xy-plane):

$$longitudinal\ transition: \qquad K_\mathrm{L} = \frac{I_\mathrm{Z}}{I_\mathrm{iso}}, A_x, A_y = 0, A_z \neq 0 \tag{4}$$

$$transverse\ transition: \qquad K_\mathrm{T} = \frac{I_\mathrm{Z}}{I_\mathrm{iso}}, A_x, A_y \neq 0, A_z = 0 \tag{5}$$

With the approximation that the plane of the carboxylate group is oriented randomly with respect to the molecular z axis, molecular orientation distribution is isotropic in xy plane, so $\langle \cos^2 \theta_x \rangle = \langle \cos^2 \theta_y \rangle$. If we assume that the number of molecules in the KBr pallet is the same as the number of molecules in the LB monolayer, for purely polarised transitions, with only z or xy component, measured ratios between z-polarised ($I_z = I_\mathrm{LB}$) and isotropic absorbances (I_iso) can take only a single value for longitudinal and transverse transitions. For longitudinal transitions:

$$K_\mathrm{L}(\mathrm{LB/iso}) = \frac{\langle \cos^2 \theta_x \rangle A_x + \langle \cos^2 \theta_y \rangle A_y + \langle \cos^2 \theta_z \rangle A_z}{\frac{1}{3} A_x + \frac{1}{3} A_y + \frac{1}{3} A_z} = \frac{\langle \cos^2 \theta_z \rangle A_z}{\frac{1}{3} A_z} = 3 \langle \cos^2 \theta_z \rangle \tag{6}$$

while for transverse transitions we can write[XV]:

$$K_\mathrm{T}(\mathrm{LB/iso}) = 3 \langle \cos^2 \theta_{XY} \rangle = \frac{3}{2} (1 - \langle \cos^2 \theta_z \rangle) \tag{7}$$

With this, we can use simple arithmetic to determine the average tilt angle of the molecule:

$$\frac{K_\mathrm{T}(\mathrm{LB/iso})}{K_\mathrm{L}(\mathrm{LB/iso})} = \frac{\frac{3}{2}(1 - \langle \cos^2 \theta_z \rangle)}{3 \langle \cos^2 \theta_z \rangle} = \frac{1}{2 \langle \cos^2 \theta_z \rangle} - 1 \tag{8}$$

$$\Longrightarrow \qquad \langle \cos^2 \theta_z \rangle = \frac{1}{1 + 2 \dfrac{K_\mathrm{T}(\mathrm{LB/iso})}{K_\mathrm{L}(\mathrm{LB/iso})}} \tag{9}$$

If all molecular axes were tilted from the surface normal equally, their average tilt angle would be given by the equation:

[XV] Here we used: $\cos^2 \theta_X + \cos^2 \theta_Y + \cos^2 \theta_Z = 1$

$$\langle\theta\rangle = \arccos\left(\cfrac{1}{1 + 2\cfrac{I_{\mathrm{T}}(iso)}{I_{\mathrm{T}}(\mathrm{LB})} \cdot \cfrac{I_{\mathrm{L}}(iso)}{I_{\mathrm{L}}(\mathrm{LB})}}\right)^{0.5} \tag{10}$$

Figure 8 shows an example of theoretically evaluated purely longitudinal (transition moment only along z-axis) and purely transverse (transition moment only along xy-axes) polarised vibrations of **Azo**.

(a)   (b)
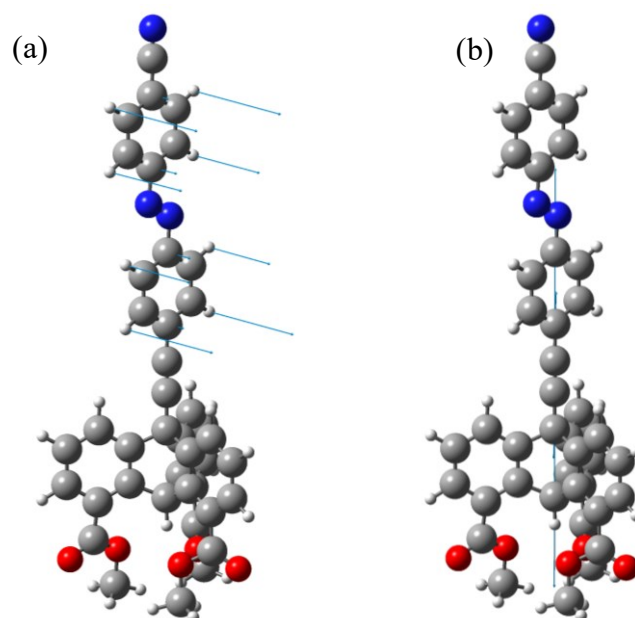
Figure 8. Examples of vibrations with either only (a) parallel or (b) perpendicular transition moment vector components (denoted by the blue arrows).

3.2. When do we need molecular dynamics?

Since, at the room temperature, a molecular system probabilistically samples a huge range of different states, results of experimental measurements are usually observable averages over many configurations at a finite time. In contrast, advanced electronic structure methods provide the theoretical framework within which the properties of individual microstates of different systems can be accurately predicted. The bridge between the two, linking microstates to macroscopic properties is the field of statistical mechanics. The macroscopic average $\langle O\rangle$ of some observable defined as:

$$\langle O \rangle = \frac{M}{Q} \int\limits_{-\infty}^{\infty} o(p, q) e^{-H(p,q)/k_B T} dq \tag{11}$$

where p and q are momentum and position variables, o is the microstate at the specific (p, q), H is the Hamiltonian, T is the thermodynamic temperature, k_B is the Boltzmann constant, and the canonical partition function Q is given by:

$$Q(N, V, T) = M \int\limits_{-\infty}^{\infty} dp \int\limits_{-\infty}^{\infty} e^{-H(p,q)/k_B T} dq \tag{12}$$

with M being the pre-factor that accounts for the indistinguishability of the particles of the same type. The ergodic hypothesis postulates that the ensemble and time averages are equivalent:

$$\langle O \rangle = \frac{M}{Q} \int\limits_{-\infty}^{\infty} dp \int\limits_{-\infty}^{\infty} e^{-H(p,q)/k_B T} dq = \lim_{\tau \to \infty} \int\limits_{0}^{\infty} o(p(t), q(t)) dt \tag{13}$$

This fundamental proposition of statistical mechanics is at the heart of the molecular dynamics. In MD, instead of observing many microscopic realizations of the given state, one can observe the time evolution of pre-equilibrated single state over a sufficiently long time. Macroscopic properties are insensitive to detailed starting conditions of microstates since they depend on them in some averaged way.

Many properties of systems can be properly evaluated just by relaxing the geometry of the investigated system and by finding the observables expected values from the resulting Schrodinger equation, or some of its analogues. It is worth asking whether investigating the properties of the system dynamically is a worthy endeavour. In single-point calculations, estimating a position-dependent equilibrium macroscopic observable from an individual configuration q' obtained by geometric relaxation means approximating the physical ensemble by a single microstate:

$$\langle O \rangle \approx o(q') \tag{14}$$

Physical properties of many systems can be accurately calculated assuming the approximation (14) is valid, where an ensemble average is approximated with a phase space function obtained for a single configuration. In short, estimating a position dependent

equilibrium macroscopic observable from an individual configuration is reasonable only if the associated Boltzmann factor:

$$e^{-V(q')/k_B T} \tag{15}$$

where V is the potential energy of configuration q', is much larger than that for any other ensemble member. So, for the static approach to give meaningful results, the weight of the single sampled configuration needs to overweigh the contribution of all other possible configurations. This is plausible in systems with well-defined deep minima at room conditions and there are many examples where this works fine.[46]

Estimating whether this approximation is valid is far from trivial. The failures of single point approach are well known in the AIMD community and have been demonstrated in the bulk state.[47] For example, thermal motion of atoms distorts the structure of highly symmetric crystals, so the perfect symmetric arrangements between the atoms which single point calculations predict for all volumes are very rare at the room temperature. As a result, the bulk modulus is frequently overestimated using *ab initio* static approaches, no matter the rigour of applied technique, as the predicted function describing the hardness/volume relationship has stronger curvature. MD, on other hand readily predicts bulk modulus comparable to the experimental value, as thermally induced symmetry breaking softens the material.[48]

These observations make for a strong argument why molecular dynamics should be used to investigate structural properties of SAMs.

3.3. **Choosing the right method**

Force-field methods are a natural choice when one wants to simulate large systems at long time scales and have been a go-to approach for the biology community over the last 50 years. With many advances in the field, multiple pre-parametrised force fields have been developed, varying in their complexity and specificity.[49]

On the other hand, in the realm of electronic structure methods ever-so-rising in complexity DFT variants are dominating the computational chemistry and material science by providing good accuracy with reasonable computational cost.[50] Between the two extremes lies the realm of semi-empirical methods. Figure 9 summarises size and time scale limits of the most popular classes of simulation methods used in computational chemistry and material science.[51]
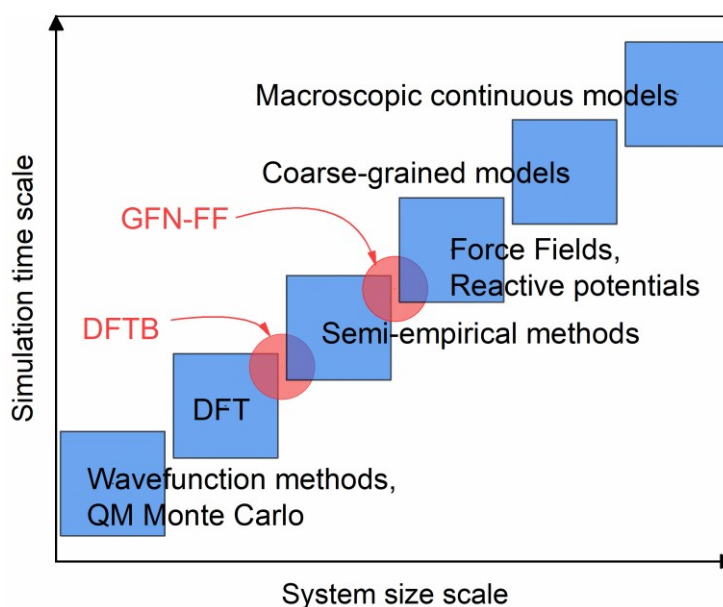
Figure 9. Schematic representation of computational complexity scaling in families of different theoretical frameworks frequently used in computational chemistry and materials science. Adapted from[51]

Choosing the method for SAMs promerties simulations on the air-water interface was a very difficult decision. On one hand, the self-assembly, packing and stability of SAMs are heavily driven by dispersion interactions which are difficult to describe using a force field, especially due to the fact that in complex films of molecular machines that we describe, the dispersion interactions stem from rather exotic triptycene units and aromatic functional heads which highly differ in their structure.

Moreover, the investigated geometries would look like they are far from the minimum due to their proclivity to bending already observed in similar systems[10] which is particularly hard to accurately describe with polynomial terms usually found in FFs, since these functions are robust only close to the equilibrium. In our systems, bending and tilting are likely a result of the surface effects and a complex interplay between noncovalent interactions governed by the structure of molecular machines. Both are very hard to describe with a generic force field.

To address these challenges properly, applying electronic structure methods may seem like a natural choice. However, capturing collective behaviour of SAMs, describing the dynamics of non-trivial modes of motion of individual machines, and estimating the degree at which individual bending and tilting drive the structural changes of the whole film requires simulation of a large system.

In this thesis we have thus investigated the suitability of simulating meso-scale systems with semi-empirical (tight binding DFT, implemented in open-source DFTB+ program[52]) and a next-generation FF method (GFN-FF, recently developed by Grimme et al.[53]), in hope to get the best of both worlds by correctly addressing the non-covalent interactions while simulating large enough systems on a long enough time scale. Both methods are briefly discussed in the following sections.

3.4. Tight binding density functional (DFTB) approach.

The cornerstone of tight binding DFT models are the Kohn-Sham equations. Starting from the reference density $n_0(r)$, its total energy functional is expanded as a Taylor series, up to some term:

$$\mathcal{E}^{DFTB}[n_0 + \delta n] = \mathcal{E}^0[n_0] + \mathcal{E}^1[n_0, \delta n] + \mathcal{E}^2[n_0, (\delta n)^2] + \dots \tag{16}$$

The reference density is just a composition of neutral atomic densities of a modelled system – a density of non-interacting atoms with zero net charge. Therefore, $n_0(r)$ artificially contains no charge density transfer between atoms. $\mathcal{E}^1$ describes charge density fluctuation, and $\mathcal{E}^2$ term can intuitively be understood as the change in the chemical hardness of an atom. The accuracy of the DFTB model implementation, among other things, depends on the order at which the energy functional is being expanded. In recent implementations, the standard expansion is up to a third term.[52]

One of the defining features of DFTB approach is the cost-effective usage of minimal basis set, providing better accuracy than minimal basis pure DFT without significant computational cost. In tight binding approach, the orbitals of valence electrons are expanded in minimal local basis. This means that a single radial function is used for each angular momentum state: s-state needs one radial function, p needs three, d- needs five etc. Orbitals are calculated by solving the atomic Kohn-Sham equations with an additional confining, usually harmonic potential. Confining the atomic orbitals creates compressed atomic-like orbitals which can describe bonding situations since otherwise, the orbitals of free atoms would be too diffuse. Confinement potential is spherically symmetric basis for wave function expansion. This pseudo-atom description is calculated with pure DFT only once, and the resulting localised orbitals are later

used in the wavefunction construction, represented as the linear combination of atomic orbitals.[54]

3.4.1. Limitations of the method.

DFTB is an approximate model, so it will have abundancy of limitations setting the bar on the maximal possible accuracy. General tight-binding framework involves several general approximations: (i) restricting Hamiltonian to include only valence electrons, which are directly involved in the electronic properties of interest (ii) valence Hamiltonians are defined using the minimal basis sets, which means that each valence orbital is defined by a single atomic function. (iii) the monopole approximation is necessary, which does not allow on-atom polarization (interaction energies of hydrogen bonds are thus severely underestimated), (iv) many-body geometric parameters, such as dihedral angles, cannot be parametrised, as the fitting of pair potentials in repulsive energy term is a two-centre problem, (v) DFTB is built on a DFT foundation, so it inherits all the well-known weaknesses of the local DFT formalism.[54]

3.4.2. When to use DFTB?

DFTB is obviously not an *ab initio* method, because it contains parameters. However, most of the parameters have a theoretically solid foundation. Since the starting point of DFTB are tightly bound electrons and perturbative treatment of their interactions, DFTB is particularly suited for covalent systems[54,55] and performs well with modelling valence electrons with significant amount of delocalisation.[56] Although less robust compared to *ab initio* approaches, fitting of pair potentials in repulsive energy term readily alleviates some of the DFTB inaccuracies. As the nature of repulsive potential is two-centred, bond lengths, stretch frequencies, and bond energies can be targeted efficiently. This is the reason DFTB performs better than minimal basis set DFT. In some cases, DFTB can even provide better results than double-ζ DFT using the GGA functionals, which is impressive, but can be traced back to the parametrisation.[57]

   Concerning the practical aspects, calculation of large systems becomes available since TB models easily achieve better scaling with the system size compared to standard DFT methods. DFTB is also relatively fast, so study of dynamical properties by accessing longer time scales becomes possible. Moreover, with a solid foundation in DFT, method development and improvements can be tested with DFTB and later imported into DFT, and investigating trends

in structural families with the DFTB approach is frequently being demonstrated in publications.[54]

DFTB method is especially suitable for studying the behaviour of complex organic systems and their interaction with different interfaces, especially when comparison of properties governed by the structural motifs differing in the amount of delocalisation, triple bond flexibility or tendency for π- π^* stacking are of interest. Furthermore, we wanted to evaluate the robustness of the method which would allow for the study of structural changes that occur upon switching, which necessitates breaking and making of covalent bonds. Simulating those requires semi-empirical approach at least.

3.5. GFN-FF method.

Going a step further from the DFTB approach, one can imagine a force field, with parametrisation being done automatically using DFT or semi-empirical methods. In this case, generation and breaking of covalent bonds would be impossible to simulate, but the parametrisation problems would be alleviated, provided the method used for parametrisation can be trusted. The recently developed GFN-FF forcefield is an example of such an approach.

Although considerably faster compared to tight binding approach, the current iteration of GFN-FF is still considerably slower compared to classical force fields. Furthermore, xTB implementation of GFN-FF[58] is not transparent, and it is difficult to tell whether the automatic topology generation module estimates the right topology.[59]

3.6. Modelling the non-covalent interactions.

Making high quality xc-DFT functionals is exceptionally hard as they should accurately model the r^{-6} attractive long-range interactions. At first, it would seem that one could just add the van der Waals interactions by hand, but it would inevitably modify the short-range interactions which DFT accurately models without corrections. This includes the lack of long-range electron correlation that translates to underestimated or missing London dispersion. An accurate account of vdW forces is essential for achieving the reliable description of intra-molecular comunication in SAMs. DFTB van der Waals forces are modelled by changing the repulsive potential.[54]

The Tkatchenko–Scheffler (TS) correction[60] includes vdW interactions as London-type atom-pairwise C_6/R^6-potentials, which is an aprroach similar to other dispersion correction

methods, for example, to DFT-DF2. However, in TS, dispersion coefficients and damping function are charge-density dependent, enabling the method to consider variations in vdW contributions of atoms due to their local chemical environment. TS corrections are thus comparable to many-body correction approach, but with significantly smaller computational cost as the polarizability, dispersion coefficients, and atomic radii of atoms in a molecule or a solid are computed from their free-atomic values. At short distances between the atoms, potentials are damped because DFT(B) already captures the electron correlation.[52]

Furthermore, TS dispersion correction method is cheap and already implemented in DFTB+ code, does not require any external libraries nor it involves any special, system-specific additional parametrisation, and was therefore a practical choice in this thesis.

3.7. Correlation statistics of time series

The subtleties of estimating the correlation between two or more different timeseries are the topic of an ongoing debate.[61] General consensus is that there is no ideal technique for estimating the timeseries cross-correlation, and the challenges can be roughly divided into three categories: (i) sensitivity of correlation estimate approaches to properties of investigated data (ii) frequent presence of spurious correlation (iii) the unavoidable presence of autocorrelation within the timeseries.

3.7.1. Spurious correlation

In signal processing, if different events are associated but not casually related either due to coincidence or due to presence of some hidden factor (common response variable) the mathematical relationship behind it is referred to as spurious relationship. Spurious correlation[XVI] between time series can occur due to many reasons and is a well-known problem in statistics.[62]

Figure 10. shows a more subtle example of spurious correlation between three 1D random walkers. Custom python code snippet used for simulation can be found on GitHub address.[63] It can be seen that, by simple Pearson analysis, the walkers w1 and w3 seem significantly

---

[XVI] Although strictly speaking there is a difference between dependence and correlation, in this thesis the terms will be used interchangeably.

correlated although the data sets are obviously completely independent. This is c aused exclusively by the within-series dependence because (i) random walks are not ergodic processes as they are non-stationary, so their time averages do not converge to any value at $\tau \to \infty$, (ii) step of a random walker is essentially a small perturbation of a previous step, so within series dependence introduces spurious correlation. This is further accented by smoothing, as the variability in the series is reduced and introduced into correlation calculation, increasing the apparent correlation.[64]
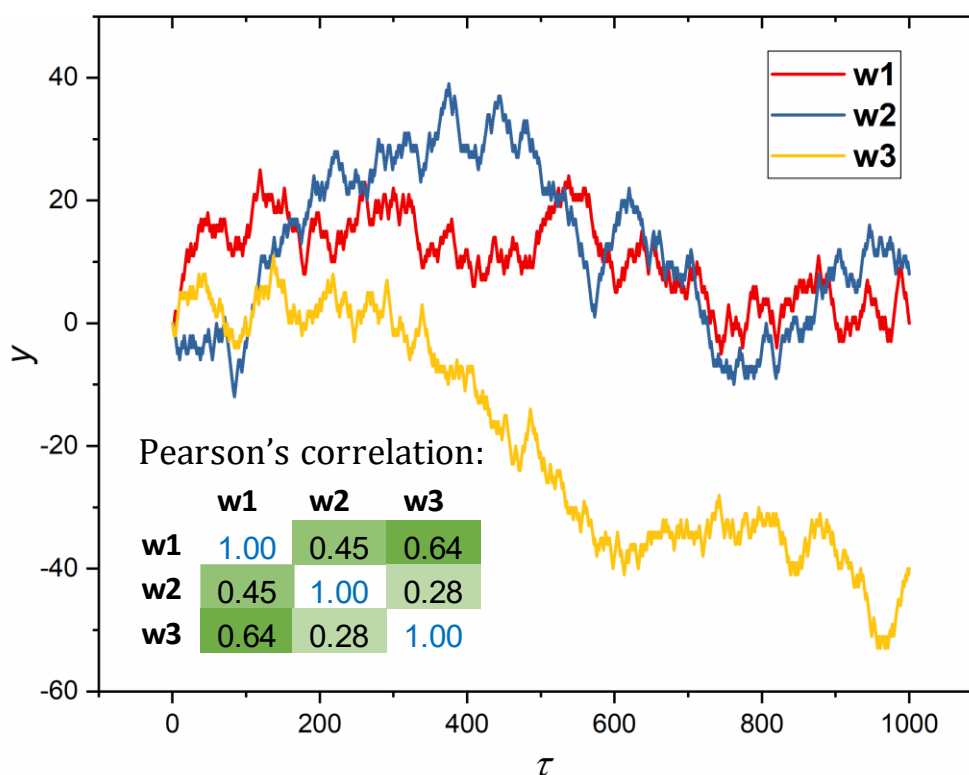
Figure 10. Movement of three independent 1D random walkers. Table embedded in the figure shows Pearson correlation coefficient matrix of their movement. Movement between walkers **w1** (red) and **w3** (yellow) seems highly correlated but is in fact completely spurious. Pearson coefficients of 0.45 (**w1**/**w2**) and 0.64 (**w1**/**w3**) are of medium and high significance, respectively.

3.7.2. Autocorrelation

Non-parametric tests for trend determination in time series, frequently require the data in the model to be autocorrelation-free and normally distributed. Of course, sampled data does not always meet this requirement. For instance, in case of molecular dynamics trajectory analyses, the successive observations in the time series are not stochastically independent. If the

autocorrelation is present within the time series, the correlation within the system will stay prone to giving arbitrary results. Furthermore, in MD analyses, removing the autocorrelation effects is also generally desirable for calculation of the standard error[XVII]:

$$\sigma_{\bar{o}} = \sqrt{\frac{\langle o^2 \rangle - \langle o \rangle^2}{N}} \tag{17}$$

where the number of independent measurements (N) is needed. N generally differs from the number of MD steps, as every MD step is dependent on the previous steps. Evaluating standard error is pertinent in the analysis of the MD runs at short time scales, as the error can determine whether the simulation was run long enough.

In this thesis, we apply the simplest approach of block averaging,[65] which works as follows: the block averaging algorithm starts with grouping the data into a series of block sizes: $1, 2, 4. ...$, together with the standard deviations obtained by computing the variance of the block averages. For each block size, an estimate of the standard deviation of the mean is given by:

$$\sigma_B = \sqrt{\frac{1}{N_B - 1} \sum_{i=1}^{N_B} (\langle X_i^2 \rangle - \langle X \rangle_B^2)} \tag{18}$$

where X_i is an average over block i, $\langle X \rangle_B$ is the average over all blocks and N_B is the total number of blocks. Assuming that a sufficiently large dataset and the number of blocks has been included in the analysis, the standard deviation of the mean is then proportional to the square root of the mean squared deviation in block averages. Generally, σ_B increases with the block length B asymptomatically to some limiting value, after which the deviation starts to significantly oscillate around that constant. This constant can be taken as the true value of the standard deviation. This observation is hard to prove, but it is worth noting that other methods,

---

[XVII] At first, one could think that standard deviation is a good descriptor of unreliability of estimated observable averages. However, since σ describes the spread of the sampled distribution, σ of some observable has a defined physical meaning independent of the quality of sampling. For example, $\sigma(E_{NVT}) = \sqrt{\langle E_{NVT}^2 \rangle - \langle E_{NVT} \rangle^2} = c_V k_B T^2$, shows that, standard deviation of NVT sampled total energy is closely related to the volume heat capacity. Terefore, to determine whether the observable has been properly sampled, using standard error is preferred.

such as bootstrap procedures regularly give similar results.[66] Figure 11 (a) schematically summarises the algorithm, while (b) illustrates the change of standard deviation with the block length.
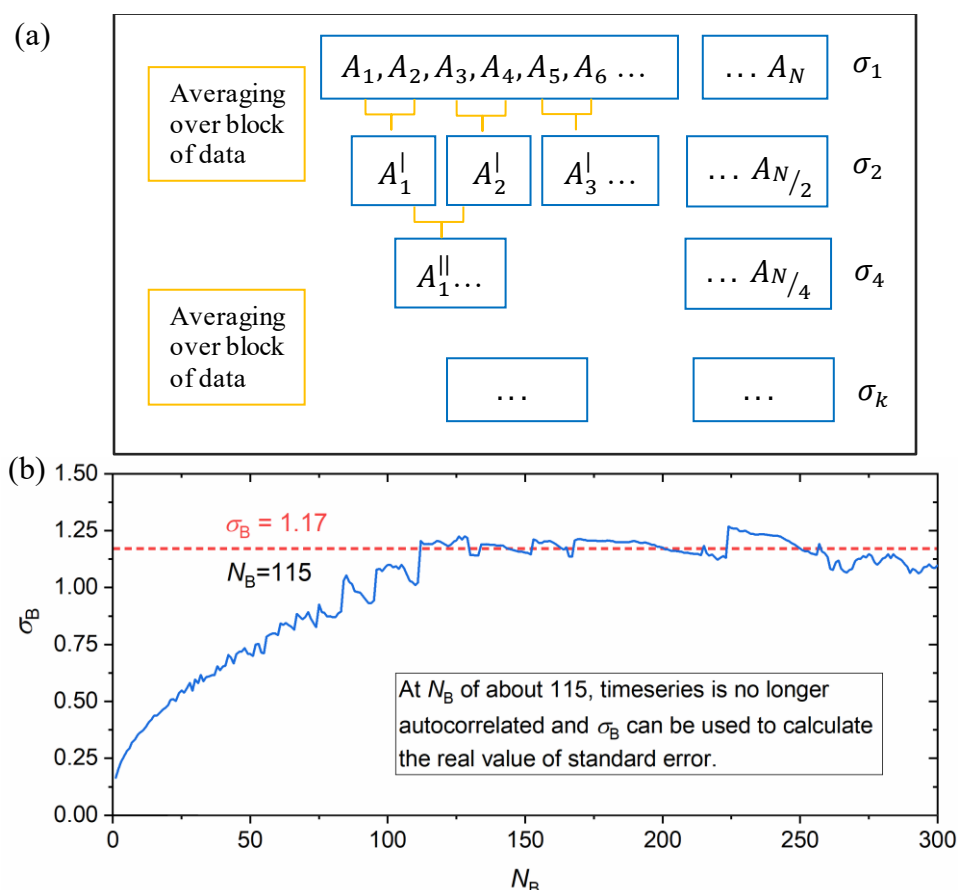
Figure 11. (a) Scheme depicting the block averaging procedure. (b) Change in the standard deviation with the block size for a generic MD sampled observable. At the plateau denoted by the red line, the error estimate is independent of the time steps.

3.8. Statistical measures of correlation

3.8.1. Pearson product-moment correlation

This is probably one of the simplest and most used correlations in statistics. It is a measure of the strength and the direction of a linear relationship between two continuous variables. However, using Pearson correlation in a meaningful way between time series is difficult due to the already mentioned temporal autocorrelation problem which is prone to producing the

coefficients with non-meaningful significance. This can be tackled by calculating the cross-correlation function with different lags. However, in molecular dynamics simulations, sampling of equilibrated observable averages necessarily means the sampled properties oscillate around some mean value. As the time series are trendless and periodic, there will always be a lag (a phase shift) at which the correlation will be statistically significant. Furthermore, Pearson correlation coefficient gives misleading results for the interpretation of the dependence structure, unless the following assumptions about the investigated data are met:

(i)   Two variables are continuous (measured at the interval or ratio level).

(ii)   Pearson's correlation coefficient is sensitive to outliers, which can have a very large effect on the line of the best fit and the Pearson correlation coefficient. Therefore, including outliers poses a risk of getting misleading results.

(iii)   Variables should be approximately normally distributed and homoscedastic. To assess the statistical significance of the Pearson correlation, bivariate normality is required. This assumption is generally difficult to assess

Finally, it cannot be stressed enough that Pearson's correlation determines the degree at which a relationship is linear. Put another way, it determines whether there is a linear component of association between two continuous variables. As such, linearity is not actually an assumption of Pearson's correlation. If the relationship between two timeseries is not linear, the relationship between them might be better described by other statistical measures, for example, with Kendall statistics.[67]

3.8.2. Kendall's tau b (τ_b) correlation coefficient

Kendall's Rank Correlation Coefficient is a measure of nonlinear dependence between two random variables. There are only two assumptions that the data must satisfy: (i) used data must be continuous or ordinal, (ii) two variables should have a monotonic relationship. If these are satisfied, the Kendall correlation coefficient can be calculated using:

$$\tau = \frac{p_C - p_D}{N} \tag{19}$$

where p_C is number of concordant pairs, p_D is the number of discordant pairs, and:

$$N = \binom{n}{k} = \frac{1}{2}n(n-1) \tag{20}$$

is the binomial coefficient denoting total number of pair combinations, e.g., number of ways to choose two items from an n-item sample size. Two observations (X_i, Y_i) and (X_j, Y_j) are concordant if:

$$
\begin{aligned}
&\text{1. } X_i < X_j \text{ and } Y_i < Y_j \text{ or} \\
&\text{2. } X_i > X_j \text{ and } Y_i > Y_j
\end{aligned}
\tag{21}
$$

and discordant if:

$$
\begin{aligned}
&\text{1. } X_i < X_j \text{ and } Y_i > Y_j \text{ or} \\
&\text{2. } X_i > X_j \text{ and } Y_i < Y_j.
\end{aligned}
\tag{22}
$$

Finally, two observations can also be tied if:

$$
X_i = X_j \text{ and/or } Y_i = Y_j.
\tag{23}
$$

To correct for the tied pairs, Kendall's τ_b rank coefficient is usually used instead:

$$
\tau_b = \frac{p_C - p_D}{\sqrt{(p_C + p_D + X_0)(p_C + p_D + Y_0)}}
\tag{24}
$$

Where X_0 is number of pairs tied to X variable, and Y_0 number of pairs tied to Y variable. In case of the perfect agreement between the two timeseries, the coefficient τ will have the value of 1, and when two timesereis are in a perfect disagreement (if the rankings are perfectly reversed) the coefficient will have the value of -1. For independent values of X and Y, the coefficient should be approximately 0.[68]

Again, autocorrelation should be either removed prior to analysing the dependence between the two variables, or a modified version of Mann-Kendall test (for example, a pre-whitening approach suggested by Yue et al.[69]) should be performed. In this thesis, the former was used due to its simplicity and because removing the autocorrelation was useful for error estimates in trend analysis.

It is easy to note that Kendall τ_b coefficient serves the exact same purpose as the well-known Spearman rank correlation as both indicate to which extent 2 timeseries are monotonously related.[70] Moreover, both correlations measure monotonicity relationships and have an intuitive interpretation. However, between the two, Kendall's approach is usually preferred since the statistical properties (sampling distribution and standard error) are better defined for Kendall's

τ_b than for Spearman-correlations. Kendall's τ_b also converges to a normal distribution faster. The consequence is that significance levels and confidence intervals for Kendall's Tau tend to be more reliable than for Spearman correlations. Moreover, Kendall correlation is preferred to the Spearman correlation since it is generally more robust (has a smaller gross error sensitivity) and more efficient (has a smaller asymptotic variance).[70]

It is also worth noting that generally, absolute values for Kendall's Tau tend to be smaller compared to Spearman correlation coefficient. When both coefficients are calculated using the same data set, the $|\tau b| \approx 0.7{\cdot}|Rs|$ relationship is usually observed. Also, Kendall's Tau typically has smaller standard errors compared toa Spearman correlations.[70]

In terms of computation, Kendall correlation has a $O(n^2)$ computation complexity compared with $O(n \log n)$ of Spearman correlation, where n is the sample size.[71] The other issue is the presumed monotonicity which does not necessarily apply for every set of paired observables one could find in an MD simulation. Nonetheless, in this thesis, we chose the Kendall formalism for description of correlation between different observables. Our decision is a result of several reasons:

(i)   We wanted a single approach to describe the dependence between all sorts of observables, and we could not guarantee that the sampling of all the observables would follow the normal distribution.

(ii)   Calculating Kendall rank correlation is much simpler compared to obtaining the cross-correlation functions across different lag times. Furthermore, the significance of dependence between the two observables can be described by a single number.

(iii)   Kendall rank correlation has already been used in the MD community.[72]

(iv)   Mann-Kendall trend analysis, which we used to determine the equilibration period in *ab initio* MD using a custom python library,[63] is one of the most robust methods for estimating trends in time series. Expanding these methods to calculate the correlation was very practical.

(v)   Although the main issue with the Kendall approach is the unfavourable scaling, calculations were well within our computational reach and vectorization processes under python *scipy* package made it reasonably fast.

3.9. Mann-Kendall trend analysis

Many useful quantities, such as internal energy, geometric parameters, work-function, probability distributions etc, can be expressed as simple temporal averages.[XVIII] Since averages are meant to represent the equilibrium properties, these results should be stationary. Therefore, the trajectory from which the observables are calculated must be trendless in order to extract the relevant predictions. In any case, the initial part of the trajectory is affected by initial conditions and should naturally be discarded before calculating the averages. The equilibration period can often be visually determined and is frequently practiced when large systems are studied in long timeframes (tens of ns). However, when *ab initio* or semi-empirical molecular dynamics methods are used, acquiring trajectory is very expensive and, while skewing the calculated observables due to inclusion of initial conditions is to be avoided, excluding large portions of the trajectory due to crude visual inspection is very unremunerative. Short available timeframes and introduction of rigorous numerical criteria is therefore necessary.[73]

The Mann–Kendall trend test (MK)[74,75] is a widely used non-parametric test for trend detection in timeseries which does not require the data to be normally distributed nor linear. It is thus a good numerical tool for estimating the length of equilibration period. Moreover, MK is conceptually exceptionally simple and easy to implement in trajectory analyses codes.

The intuition behind MK directly builds on the previously discussed Kendall rank correlation framework: if data is completely trendless, the number of concordant pairs of points should be equal to the number of discordant pairs within the same timeseries. We can define the MK statistic S such that:

$$S = 2I - \binom{n}{2} \tag{25}$$

where I is the number of distinct increasing pairs of data points. It is assumed that the S statistic follows the normal distribution, and the mean value is 0 when the data is completely trendless. The null hypothesis, stating there is no trend in data, is rejected if S deviates significantly from 0. There are several values which can be taken as a threshold for determining

---

[XVIII] Gibbs free energy is, of course, an exception here.

the statistical significance; in this thesis quantile is drawn from normal distribution and standard deviation given by:

$$\sigma_S = \left[\frac{1}{18} n(n-1)(2n+5) \right]^{0.5} \tag{26}$$

where n is the number of data points. In practice, the trend evaluation is started with raw data and pieces of data are removed until no trend can be observed in the rest of the trajectory, and Mann-Kendall test is passed. Passing the MK test ensures that the trajectory is free of trend and also of discontinuous abrupt changes of data trends, such as polymorph interconversions during the production run.

It is clear that the just-discussed original Mann-Kendall test considers neither the serial correlation nor the seasonality effects. When analysing MD results, however, the observed data are usually autocorrelated and can also contain seasonal changes, which will result in misinterpretation of trend tests results. Over the years, different versions of MK test have been proposed, which account for these effects.[76]

3.10. Computational cost reduction strategies

One of the main challenges with DFT based MD is the computational cost and unfavourable scaling due to system's size e.g., with the increasing number of molecules in the system. With modern supercluster infrastructures available to scientific communities which usually have thousands of processing units per user, natural approach to dealing with the issue is to utilise the rising availability of computer power. However, no matter the availability of computational resources, the bottleneck in the performance of a parallel program is always going to be the maximum parallelisability of the performed routines. In modern computational chemistry codes, not all operations can be parallelised with equal success. Amdahl's law[77] describes the expected speed up for operations which are parallelised only partially. The expected speed-up is:

$$S(N) = \frac{1}{1 - p + \frac{p}{N}} \tag{27}$$

where N is the number of available processing units, and p is the fraction of operations performed in parallel. We can see that, for very large numbers of processors, the limiting speed up goes to:

$$\lim_{N \to \infty} S(N) = \frac{1}{1 - p} \tag{28}$$

Figure 12 illustrates the problem. Optimal speedup depends on the parallel fraction parameter; using a large number of processors becomes efficient only when the performed routine is highly parallelisable.
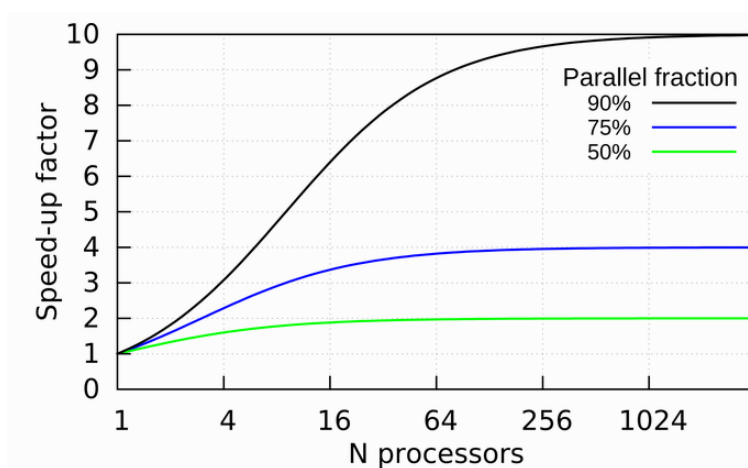
Figure 12. Dependence of maximal theoretical computational speed-up on the number of available processing units for three arbitrary algorithms varying in parallelisability. Adapted from[78]

Amdahl's law can provide a reasonable indication of the performance of a problem on increasing numbers of processors, depending on the parallelism of the task. In the case of DFTB, the computational effort for conventional diagonalisation (the dominant part for most calculations) grows as the cube of the number of atoms with the optimal parallel fraction of about 60%. Practically, this means that, even with the full MPI parallelisation, using more than 36 processors in simulations speeds the calculations up by less than 4%.[52]

Another possibility of achieving the necessary speed-up is by utilizing the extended Lagrangian approach (ExL).[79] Compared to the standard Born-Oppenheimer molecular dynamics, ExL explicitly introduces the electronic degrees of freedom as (fictitious) dynamical variables, writing an extended Lagrangian for the simulated degrees of freedom. This leads to a system of coupled equations of motion for both ions and electrons. As a result, the explicit

minimization of charge density at each ionic step of the simulation, as done in Born–Oppenheimer MD, is not needed. After the initial convergence of wavefunction, the extra degrees of freedom keep electrons in their ground state which corresponds to each new ionic configuration visited along the phase space path, thus yielding accurate ionic forces. Convergence of the wavefunction is not repeated every ionic step, but every n-th timestep instead. Unfortunately, this approach becomes unreliable in case of large nearly degenerate systems, such as the molecular machines simulated in this thesis, bearing many aromatic units which are very close in energy.[79]

Charge evolution without minimizations between time steps necessarily leads to instabilities that break the simulation. The integration step $d\tau$ must be chosen carefully since a too large step does not allow for a correct sampling of high-frequency modes and causes SCF convergence issues. However, too small step is terribly inefficient as it produces useless time-correlated data. To reduce the computational cost, choosing the right time step between ionic steps is thus crucial. The safe approach is to just set the $d\tau$ to be about 10 smaller than the period of oscillation of fastest degree of freedom.

To speed up the simulation, one can, however, consider increasing the mass of the lightest atoms in the system. This is viable to do, because the position dependent quantities are independent of mass, so the modification should not affect the outcome of simulation when the static quantities are of interest:

$$\langle O \rangle = \frac{\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} o(q) e^{-H(p,q)/k_BT} dq}{\int_{-\infty}^{\infty} dp \int_{-\infty}^{\infty} e^{-H(p,q)/k_BT} dq} = \tag{29}$$

$$= \frac{\int_{-\infty}^{\infty} e^{-\sum p_i^2/2m_i k_BT/2k_BT} dp \int_{-\infty}^{\infty} o(q) e^{-V(q)/k_BT} dq}{\int_{-\infty}^{\infty} e^{-\sum p_i^2/2m_i k_BT/2k_BT} dp \int_{-\infty}^{\infty} e^{-V(q)/k_BT} dq} \tag{30}$$

From the equation 30, it can be seen that, when calculating a static average property from phase space sampling, the mass bearing terms cancel each other out. This choice was utilised in this thesis. However, one must be aware of the limitations of this approach. Study of dynamic observables, such as velocity distribution functions, diffusion coefficient or reconstruction of vibrational spectra from the trajectories becomes impossible, because time evolving properties are mass dependant.

§ 4. METHODS

4.1. Molecular Dynamics with Periodic Boundary Conditions

4.1.1. System preparation

Starting geometries were prepared in the following way: both the molecular geometries and unit cell parameters of 2x2 films were relaxed in vacuum as described in Section 4.4. Bilayers were then assembled using the unit cell parameters of relaxed films placed in the xy-plane, with the z-direction extended to ensure at least 20 Å of vacuum. The water layer was independently prepared in Gromacs.[80] This procedure was adopted due to DFTB+ code[52] not having a practical implementation of the unit cell optimization algorithms. Each system contained 4 machines in each layer, resulting in 8 machines in total, encapsulating 89 water molecules between them. For the subsequent trajectory analysis, machines in one layer were denoted as A, B, C and D, while the machines in the other were named as E, F, G and H (Figure 13, Chart 2 and Chart 3).

4.1.2. Optimisation

Geometries of assembled systems were relaxed with the DFTB+ code,[52] using the conjugated gradients method, until the change in total energy between two consecutive steps dropped below $1e^{-6}$ Ha. Throughout optimisation and subsequent dynamics simulation, the Third-Order Parametrisation for Organic and Biological systems (3OBS) Slater-Koster Hamiltonians were used.[81]

Chart 2. Molecular bilayer assembly scheme
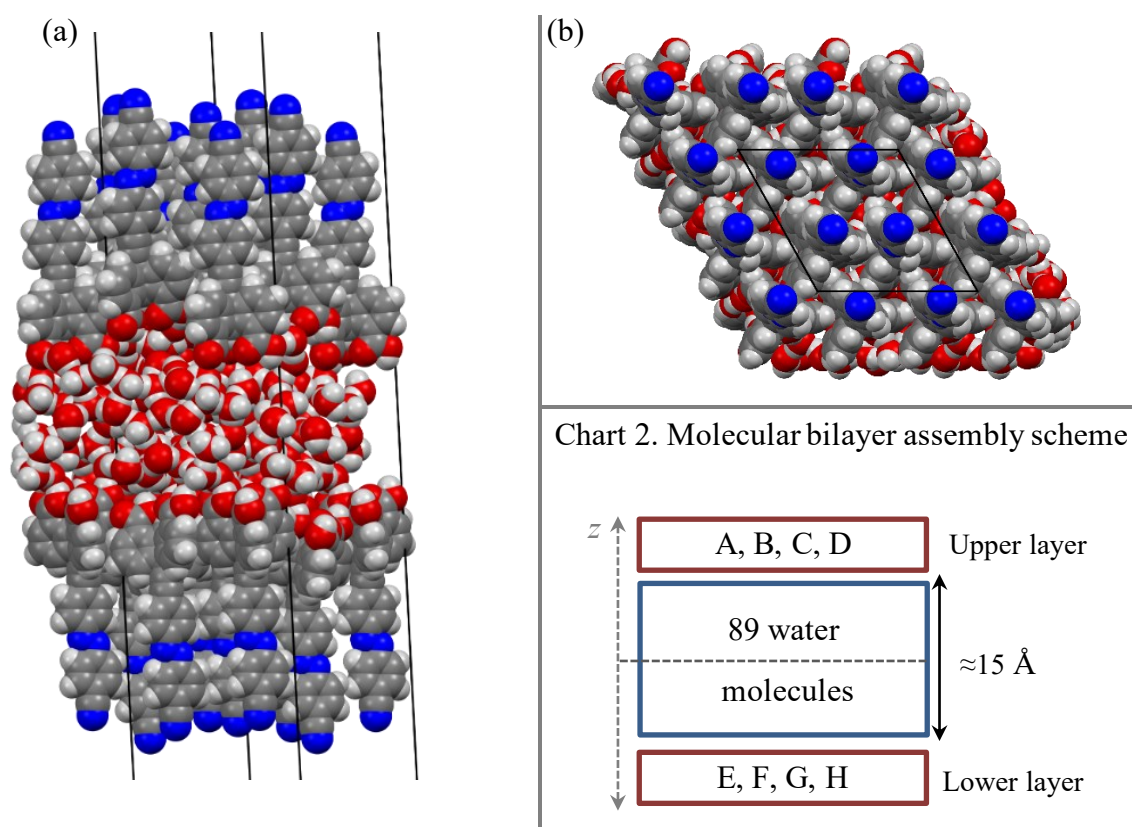
Figure 13. (a) Side-by-side and (b) top view of prepared *trans*-**Azo** system.
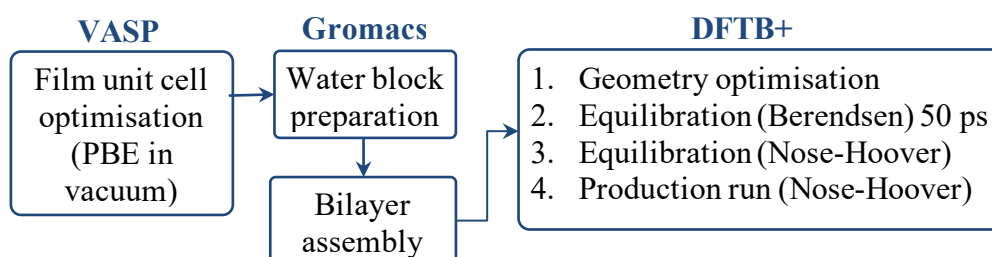
Chart 3. System preparation workflow.

4.1.3. Molecular dynamics simulations

All simulations were run for at least 1 ns in the canonical ensemble at 300 K, with timestep of 2 fs. To keep the simulation stable, masses of all H atoms were increased to 4 amu. The self-consistent cycle tolerance was set to $1e^{-6}$ Ha and Tkatchenko-Scheffler model[60] was used to correct for dispersion interactions. The *Divide-And-Conquer* eigensolver[82] algorithm was used, as other recommended eigensolvers available with LAPACK 3.0[83] led to instabilities and sporadic simulation behaviour. Equilibration was started using the Berendsen thermostat at 300

K for the first 50 ps. Subsequently, the thermostat was changed to Nosé-Hoover, with coupling strength of 3200 cm^{-1} and was used for the rest of the simulation. An exemplary input file can be found on GitHub[63] and thermostat benchmark details are presented in Section S6 of the SI.

4.1.4. Determining the equilibration period

The equilibration period was chosen by evaluating the time derivative and amount of trend of smoothed total energy curves along the full trajectory path. The strength of the trend during the production run was estimated by dividing the temporal evolution of each observable into 20 equally sized parts and performing the Mann-Kendall statistics on increasing time intervals, as was described in Section 3.9 in Theoretical Basis.

4.2. Trajectory Analysis

Topology reconstruction via domain decomposition and the unwrapping of film geometries were done in Travis MD processing tool.[84,85] This led to the possibility of (i) extracting only the structural motifs necessary for further analyses, (ii) reducing the autocorrelation effects by saving every n-th trajectory frame. These 'purified' trajectories were then further analysed, mostly using custom Python code (available at a GitHub repository free of charge)[63], with the exception of distance determination between different ring centres and their representative combined distribution functions (CDFs), for which we used a combination of Travis and our custom Python code. Chart 4 shows the general workflow used during the analysis, while input file examples and code used in this thesis can be found on.[63] To describe dynamics of film structure, a set of descriptors was defined. Definitions are summarized in Figure 14 with brief explanations provided in the accompanying text.
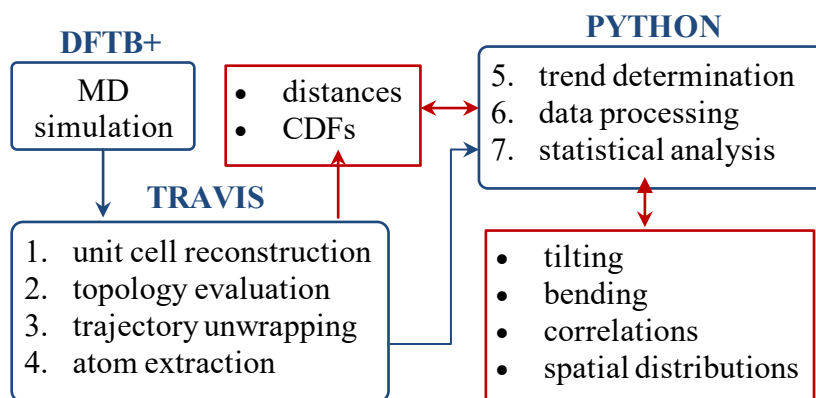
Chart 4. Trajectory analysis flowchart. CDFs stand for combined distribution functions.
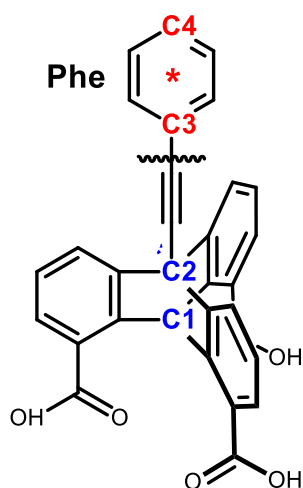
Table 2. Defining SAM features used in film structure determination

Obs.	Definition
Θ_{bot}	angle between C2–C1 and normal to surface
Θ_{top}	angle between C4–C3 and normal to surface
Θ_{top2}	angle between C5–C6 and normal to surface
B	angle between C3–C4 and C2–C1[XIX]
d(top-bot)	neighbouring distance between * and C2
d(top-top)	neighbouring distance between ring centres (*)
d(top-top2)	neigbouring distance between lower (red *) and upper (violet *) ring centres (**Azo**)
d_z	z-coordinate of C2
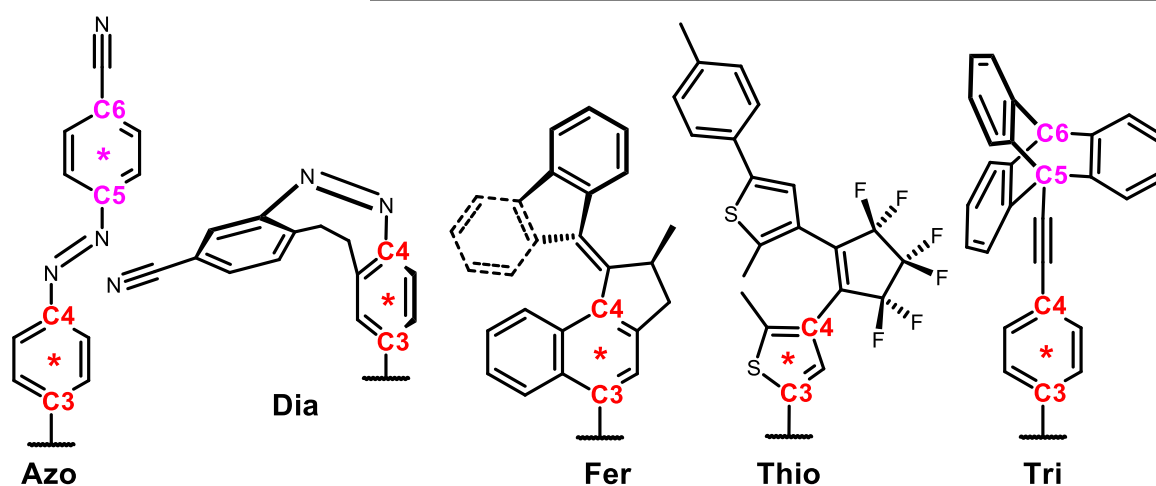
Figure 14. Topology graphs of investigated systems, highlighting the motives used in the analysis of film structure. Asteriks denote ring centroids, and colours highlight molecular motifs at different heights: *blue* triptycene unit, *red* lower and *violet* (where applicable) topmost parts of functional heads.

[XIX] Bending needs to be monitored separately from Θ_{top} and Θ_{bot} since the molecules can bend in direction opposite to the tilt of their functional heads (See subsequent Sections).

4.2.1. Molecular tilting

Time evolution of molecular tilt was determined by tracking the vector $v12$ = C2(x,y,z) – C1(x,y,z) of every individual machine, as defined in Figure 14, and calculating the angle between them and the surface normal vector, defined as the normal to the xy surface.

4.2.2. Bending of the functional head

Functional head bending was defined as the angle between vectors $v34$ = C4(x,y,z) – C3(x,y,z) and vector $v12$ of individual machines, defined in Figure 14. For **Azo** and **Tri** systems, an extra vector was defined as $v56$ = C6 – C5, to provide further information about the structure and the behaviour of the topmost part of these machines. Tilting ($\Theta_{top2} \equiv \sphericalangle(\vec{v}_{56}, \vec{n})$) and bending $B_2 \equiv \sphericalangle(\vec{v}_{56}, \vec{v}_{12})$ of topmost parts of these molecules were calculated in respect to the triptycene stand in these cases. It is worth noting that definition of B as general bending is not equivalent to $\Theta_{bot} + \Theta_{top}$ defined in the previous paragraph – on the contrary, comparing the B to the sum of tilts leads to further information about the directionality of tilting and bending with respect to the monolayer.

4.2.3. Average minimal neighbouring distance between functional heads.

To estimate the tendency of functional groups to stack, their rotational freedom and tendency to interact with triptycene stands due to tilting, radial distribution functions (RDF) of different parts of molecules building the films can be calculated and compared. However, RDFs are not reliable metrics for small unit cells.[85] To solve this issue, we implemented the following procedure: we (i) neglected the periodic boundary conditions and (ii) built a 4x4 topologically unwrapped supercell, (iii) calculated all distance combinations of interest, (iv) removed duplicates, (v) calculated the temporal averages of each distance to (vi) finally extract the minimal distance value for every pair of machines based on those temporal averages. As the machines are ordered in regular arrays, every machine is necessarily surrounded by other machines, and, as the films are on average collectively tilted in a single direction, there has to be a unique minimal distance for 2 out of 4 machines in the primitive unit cell. Figure 15 illustrates the chosen neighbouring distances. By taking the average of those two, quantity of interest can be sampled. Minimal neighbouring distance between the functional heads and triptycene stands was calculated the same way, with the minor change of tracking the triptycene stands (C2 atom in Figure 14) as well as the functional heads. Note that, in this case, tracking

distances seizes to be a symmetric problem, so more combinations had to be considered. General workflow of this procedure is summarised in Chart 5.
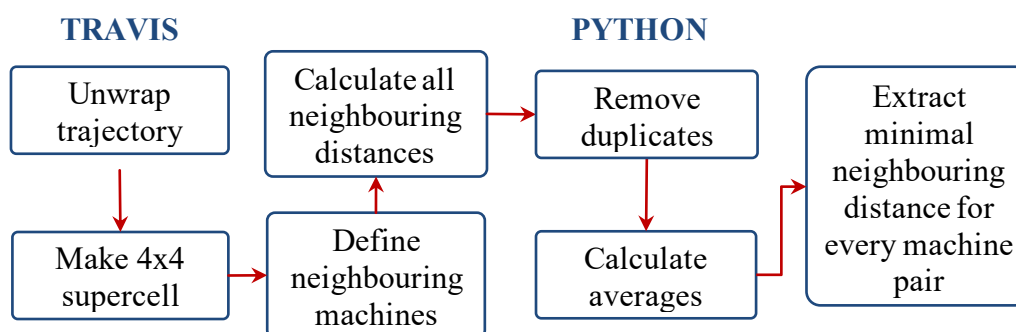
Chart 5. Minimal average neighbouring distance calculation workflow.
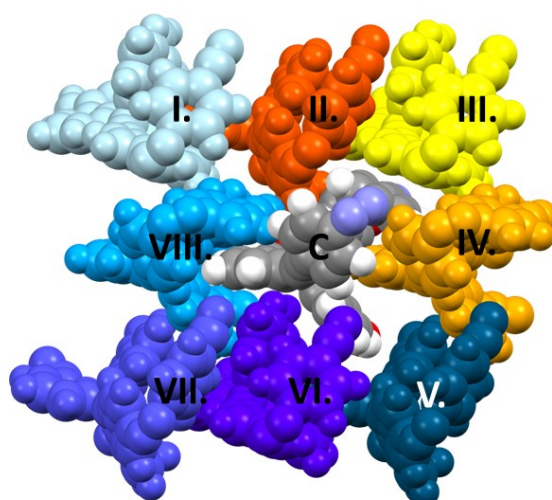
Figure 15. Stylised representation of the monolayer structure. For the central molecule (**C**) candidates for the minimal average neighbouring distance calculation are molecules **II**. **III**. and **IV**, denoted in warm colours.

4.2.4. Combined distribution functions

Generally, different types of distribution functions can be visualized as 1 D histograms taken over scalar data. By simultaneously tracking two (or more) distribution functions over all molecules in the system and over all timesteps, one can combine these histograms into higher dimension heat plots. For each individual configuration, an n-element tuple is created for scalars that are being tracked, and data is represented in an n-dimensional tensor form. Distribution of values in the tensor reveals the relationship between the distribution functions, as every tuple

entry stems from the unique spatial and temporal calculated configuration. In this thesis, only 2-element tuples were combined. This was done for radial distribution functions (RDFs) between functional heads and two other functions: (i) previously defined angular distribution functions (ADFs) corresponding to bending (B) and RDFs corresponding to distances between functional heads and triptycene stands, as was defined in Table 2.[XX]

4.2.5. Correlation plots of combined distribution functions

If the CDF domain elements were uncorrelated, their resulting combined tensor would be the Cartesian product of its 1D distribution functions. So, by internally creating the Cartesian product of the two histograms and subtracting it from the precomputed CDF, one can reveal the dependence between multiple distribution functions. It is worth noting that this procedure requires the distribution size of the sample to approach the size of population. This procedure was previously used for hydrogen bond analysis,[86,87] and it can be done using the Travis code.

4.2.6. Cross-correlation of individual machines

Tilting and bending correlations between individual machines in the film were estimated by calculating the Kendall rank (τ_b) correlation coefficients matrices, as defined in the Theoretical basis chapter, Section 3.8.2. Pairwise correlation between mean tilting and bending was calculated using the same method (code on GitHub address[63]).

4.3. Cluster-based Molecular Dynamics

4.3.1. System preparation

Molecular systems were prepared as bilayers, with two regular arrays of 3x3 molecular machines oriented in the opposite direction and a water layer between them. Systems were optimized using the GFN-FF[53] method as implemented in the xTB code.[58] For precise control of the mean molecular area during the simulation, an ellipsoidal (egg-shaped) external potential

---

[XX] As previously mentioned, RDFs are not well defined for small unit cells, however, as we were interested only in smallest distances, we monitored the RDFs only at less-than-half of the shortest unit cell vector distances.

was introduced (Figure 16 (a)). This provided the opportunity to simulate all the experimentally available systems, as well as the basic **Phe** system at 5 different APMs each, ranging from 55 A^2 to 85 A^2. Two different external potential shapes were tested, namely the $V_{\log}$ and V_{pol} (Figure 16) with different parameters. V_{pol} with a large exponent ($a = 30$) provided best results, as the constant gradient beyond the boundaries of $V_{\log}$ potential resulted in unstable simulations with larger ($\Delta\tau > 1.5$ fs) timesteps.

$$V_{\log} = \sum_{A} k_B T \log\left(1 + e^{\beta(|R_A - O| - R)}\right) \quad (31)$$

$$V_{\text{pol}} = \sum_{A} \left(\frac{|R_A - O|}{R}\right)^{\alpha} \quad (32)$$
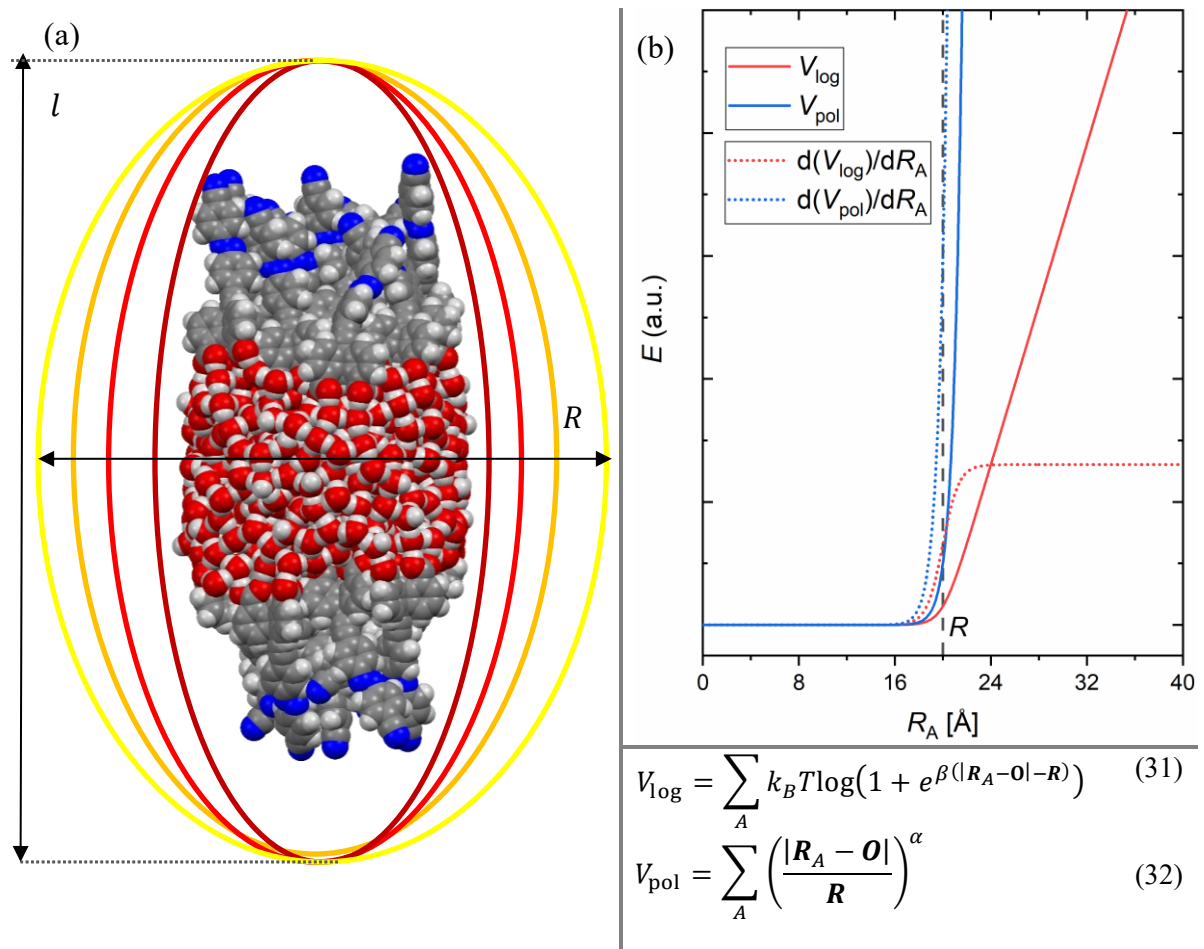
Figure 16. (a) Bilayer after the optimisation step with the GFN-FF method (b) Two types of the APM dependant custom potential functions ($V_{\log}$ and V_{pol}) and their gradients. By changing the value of R parameter, film area of the bilayer system can be tuned.

4.3.2. Optimisation

To ensure a proper cylindric shape during the simulation, external potential was introduced incrementally during the optimisation step, by reducing the R value and increasing the power of the polynomial. The optimisation criterion between each step was set to crude in the xtb code

$(5 \times 10^{-4}$ Ha$)$. After reaching the desired APM value, system was relaxed so the total energy change between two ionic steps was less than 1×10^{-6} Ha.

4.3.3. MD simulation

All 20 systems (5 types of tripodal machines at 5 different APMs) were run between 1 and 2 ns at 300 K (Berendsen thermostat), with 2 fs timesteps. The external potentials used to simulate different APMs can be found in Table S12 (Section S9 in the SI).

4.4. Scanning of local minima

4.4.1. Film relaxation

Based on our previous results,[10,24] we constructed arrays of molecular machines with unit cells containing one machine, resulting in a symmetric SAM in the xy plane. To avoid self-interaction, at least 15 Å of vacuum was included in the z direction. A series of geometries was prepared, differing in the starting tilt angles Θ [55°, 90°] with 5° increments (Figure 1 in the Introduction chapter). Film geometries were relaxed by sampling only the Γ-point, using the VDW-DF2[88] (optB86b-vdW variant[89]) functional, a plane wave basis with a 600 eV energy cutoff, with the projector-augmented wave method as implemented in VASP.[90-92] Dipolar corrections of the total energy up to second term were included in the z-direction. A convergence criterion of Helmann-Feyman forces equal to 0.01 eV Å^{-1} was set, both atomic positions and lattice parameters were allowed to relax, but the unit cell volume was kept fixed to prevent the errors due to Pulay stress.

4.4.2. Film interaction energy as a function of molecular tilt

We investigated how the interaction energy and tilt angles of the top and bottom parts of the molecule change as a function of film height. Film height was simply defined as the difference between atoms with highest and lowest z-coordinate values, hydrogens excluded while the film interaction energy E_{film} was defined as:

$$E_{film} = E_{pbc} - E_{iso} \tag{33}$$

Where E_{pbc} is the (dipole corrected) total energy of molecular films calculated in vacuum, and E_{iso} is the total energy of an isolated machine. Note that, the more negative the E_{film} is, the greater the film stabilization effect.

4.5. IR and Raman spectroscopy

4.5.1. Computational details

Optimisation of isolated geometries (methylated analogues of experimental tripodal systems) and the calculations of IR and Raman spectra were carried out with the Gaussian16 package.[93] IR spectra were modelled at the B3LYP/6-311+G(d,p) level of theory using the Grimme's long-range corrections with Becke-Johnson damping (D3BJ). Raman spectra were found to be more sensitive to functional choice. Several methods were tested, with the M06-2X/6-311+G(d,p) level of theory providing the best agreement to experiment. Solvent effects were described by performing optimizations as well as frequency and polarizability calculations with the polarisable continuum implicit solvent model. Calculated spectra were subsequently scaled by the factor of 0.9710, which was previously determined by a simple single-parameter linear regression model.[10]

4.5.2. Data analysis.

The average tilting and bending of molecules in SAMs were determined according to the equation 10 in the Section 3.1 of Theoretical Basis. The experimental techniques used to acquire spectra are described elsewhere.[13] Due to their readiness to make hydrogen bond networks as well as coordinates with zinc ions, the IR spectra of triacids were unsuitable for the analysis, so methyl esters were measured instead.

The PM-IRRAS baseline was corrected using a high-pass filter in order to remove the Fourier components with low frequencies, corresponding to the baseline undulations. When necessary, deconvolutions of interesting peaks were done using Gaussian fitting (for an example, see Figure S16 in Section S7).

§ 5. RESULTS

5.1. Molecular Dynamics with Periodic Boundary Conditions

5.1.1. Thermodynamics

Figure 17 summarises the thermodynamics of DFTB MD simulations during 1 ns for few systems, while the rest can be found in Section S6 of the SI. Mann-Kendall statistics revealed that 200 ps of equilibration period completely removes the total energy trend in the later part of trajectory. An exception is the **Tri** system, which took another 100 ps due to bulkier structure and subsequently longer relaxation time. During the production run, the temperature is well converged at 300 K among all the investigated systems justifying the choice of thermostat parameters (Section 4.1.3 in Methods). The total energy is mostly well converged too, except for the *trans*-**Azo** system which shows a slight, but continuous drop during the last 50 ps of simulation time, due to the destruction of one of the layers.
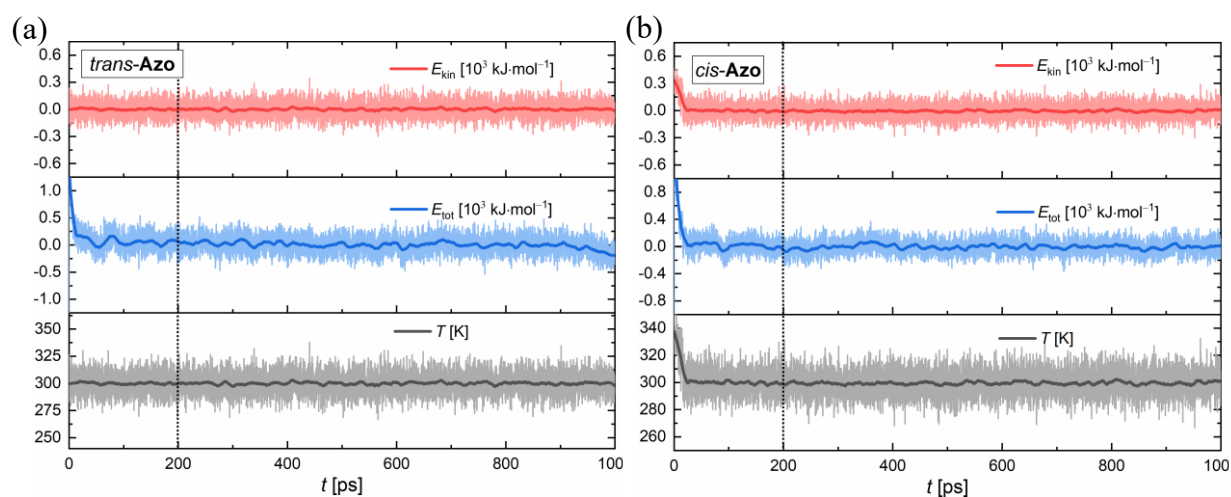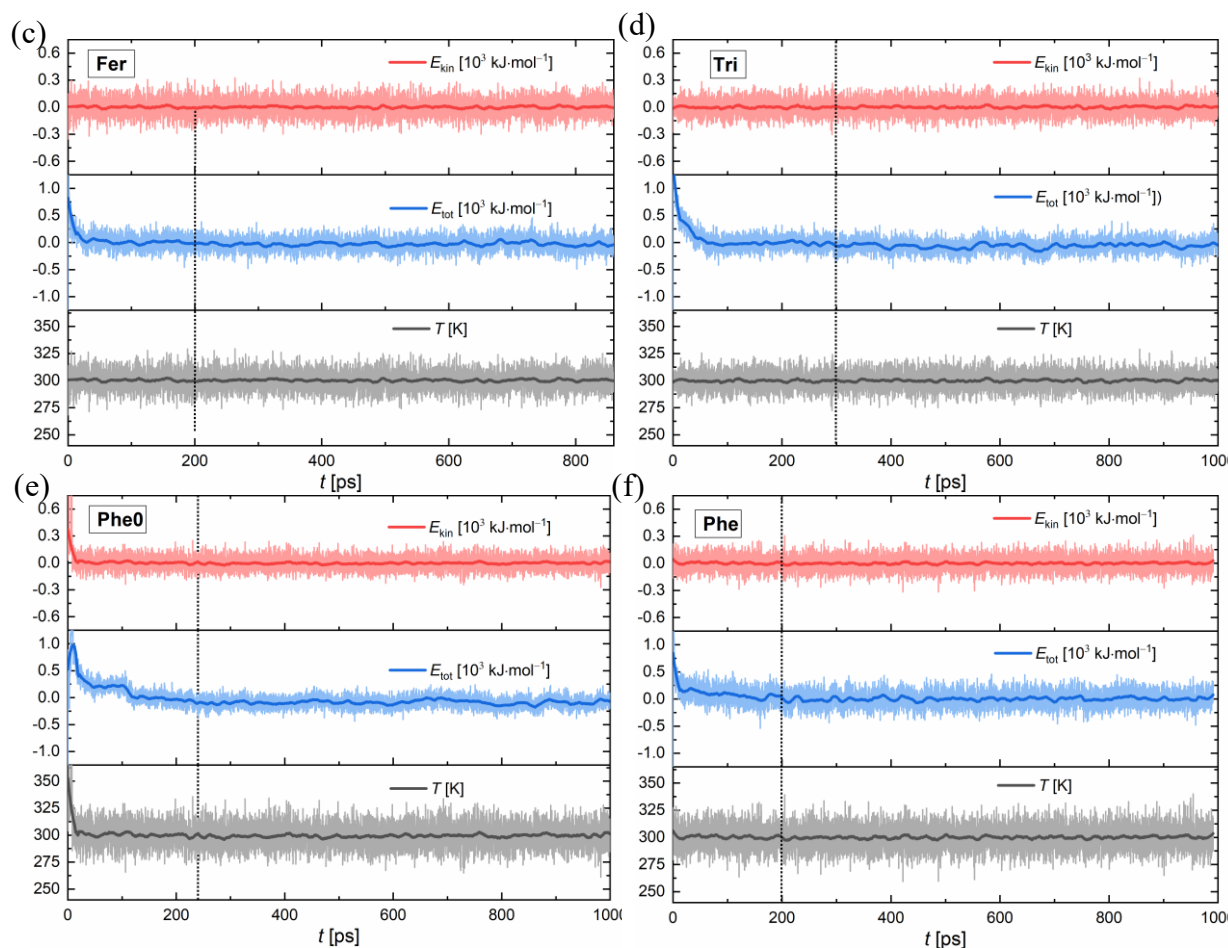
Figure 17. Kinetic energy *red*, total energy *blue* and temperature *grey* for (a) *trans-* **Azo** (b) *cis-***Azo** (c) **Fer** (d) **Tri** (e) **Phe0** and (f) **Phe** systems. The chosen equilibration cut-off is denoted by a dotted vertical line. Energy mean is shifted to zero.

5.1.2. Film stability

With around 10 Å of water molecules between the monolayers, our expectation was that the motors in the top layer will not be affected by the motion of the bottom layer and vice versa. A pairwise correlation analysis shows that this is indeed the case (Table S8 − Table S10 in the SI), meaning that the trajectories of individual SAMs can be analysed separately. Figure 18 shows the density profiles of water molecules and triptycene stands of **Azo**, **Fer**, **Phe0** and **Phe** systems (as defined in Table 2 in Methods) projected onto the surface normal. *z*-coordinate of every molecule in the system was tracked individually, allowing us to monitor: (i) the change of the distance between the two monolayers in time, (ii) the variation in the heights of all the molecules in each layer, (ii) the number of water molecules that escape the confinement of the double-layer during the simulation, and (v), if this is happening, the effect of these phenomena on the stability of SAMs on water surface.
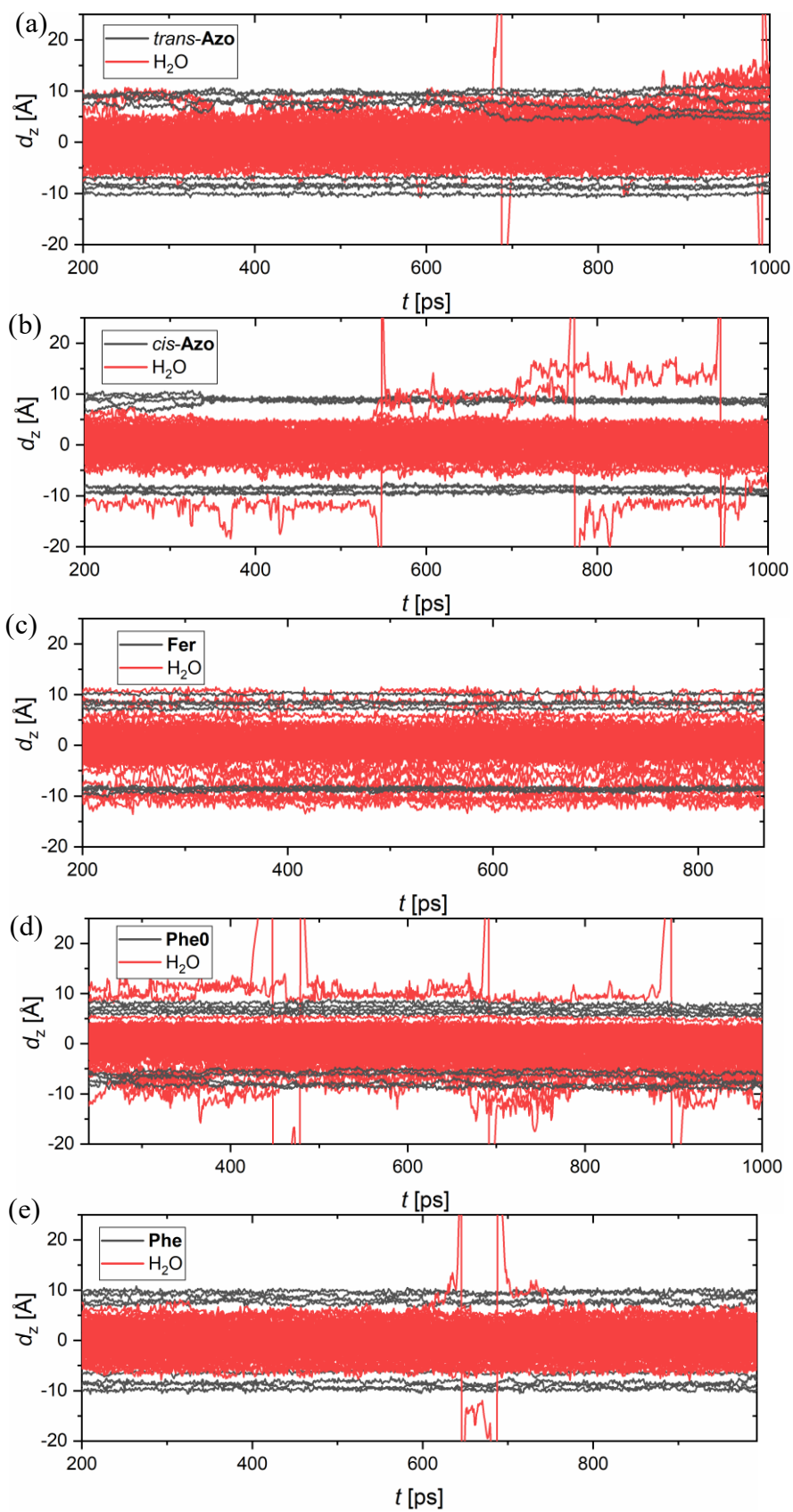
Figure 18. Temporal evolution of water (*red*) and triptycene stand (*grey*) densities in the *z*-direction (perpendicular to the film surface) for (a) *trans*-**Azo**, (b) *cis*-**Azo** (c) **Fer** (d) **Phe0** and (e) **Phe** systems. *z*-coordinates of water oxygen and upper bridgehead carbon atoms represent the density distribution of water molecules and individual triptycene stands, respectively. See Figure 14 for definition and details.

5.1.3. Structural properties of SAMs

As was shown in the previous section, for most systems at least one of the two layers remains stable during the whole 1 ns run, without significant water protrusions. Structural analysis of the stable layer is presented in following sections, while the exemplary results for the less stable one can be found in S6 Section in the SI.

All monitored observables (tilting of the bottom and top molecular parts, bending of the functional heads and distances between different molecules making SAMs) for each individual machine oscillate around a constant value, i.e., mean values of monitored observables do not evolve in time during the production run (for example, thin lines in (a), (b) and (c) in Figure 19). This is important since in this case, the analysis becomes invariant under the permutations of trajectory frames (provided the autocorrelation effects have been taken care of) allowing us to draw conclusions about the film structure based on the sampled data.

Overall, although well converged, the time averages of tilt and bend angles of individual machines are mostly dispersed across a wide range of angles. This results in a narrow and highly symmetric distribution of their means, with small standard deviations, but broad and frequently multimodal cumulative distributions (full red vs grey dotted histograms in Figure 19). This is calculated by summing up the monitored distributions of individual switches. The film, being confined by the fixed film density, is thus composed of a range of stable tilted configurations differing in the amount of tilt. Analyses of molecular motors are divided into two sections, based on the structural motifs which influence their monolayer behaviour the most: (i) effects of functional heads and (ii) effects of anchoring groups on SAMs. Due to an overwhelming amount of analysed data, in our effort to make this thesis enjoyable to read, the analyses of some systems were repositioned from the main text to the SI. For (i), detailed trajectory analyses are presented for **Azo** and **Fer** systems, while the results for **Dia**, **Thio** and **Tri** are in Sections S1-S4 of the SI. In section (ii), **Phe0** and **Phe** results are summarised in the main text, and **Phe1**

system analysis can be found in Section S5 in the SI. Table 3 summarises observables used in film structure analysis as were defined in the Methods section.

**Table 3.** Calculated observables used in trajectory analysis

Observable	Description
Θ_{bot} [deg]	Triptycene stand tilt angle
Θ_{top} [deg]	Functional head tilt angle
B [deg]	Bend of functional head, relative to the triptycene base
d(top-bot)	Distance between functional heads and triptycene stands
d(top-top)	Distance between functional heads of individual machines
d(top-top2)	Distance between upper and lower part of the neighbouring functional heads
RDF-RDF CDFs	Correlation between d(top-bot) and d(top-top)
RDF-ADF CDFs	Dependence of B on d(top-top)
MK correlation	Correlation of different modes of motion between individual machines

5.1.4. Effects of functional heads: trans- Azo system

On average, tilting of *trans*-**Azo** molecules is significant and increases with the molecular height, due to the flexible linker connecting the functional head to the triptycene stand. Furthermore, the tilting of the upper part of the molecule (Θ_{top} = 65.3°) is roughly equal to the sum of the triptycene stand tilt (Θ_{bot} = 45.5°) and the functional head bend (B = 21.9°, sum = 43.4°), suggesting that both the tilting of the lower and the upper part of the molecule on average occur in the same direction (Figure 20 (b)). Projecting bend defining vectors onto unit cell base vectors confirms this truly is the case (Figure S13 in the SI).

Minimal average neighbouring distances between triptycene stands and functional heads are consistently smaller compared to minimal distances between functional heads (Figure 19 (d)). Moreover, although the functional heads have a significant amount of rotational freedom during the simulation, resulting in a wide distribution peak (d_{min} = 5.12 Å to d_{max} = 8.03 Å), the two distribution peaks have almost no overlap. This suggests the proximity of functional heads to stands during the simulation is not merely a result of Brownian motion of flexible functional heads but is also governed by energetic contribution. This was further explored by tracking the distances between both phenyl rings of the **Azo** switch. REF _Ref107261110 \r \h Figure 20indicates that the lower and upper phenyl rings of neighbouring switches are in proximity

most of the simulation, thus explaining the significance and directionality of tilting in the lower and upper part of the switch.

(d)
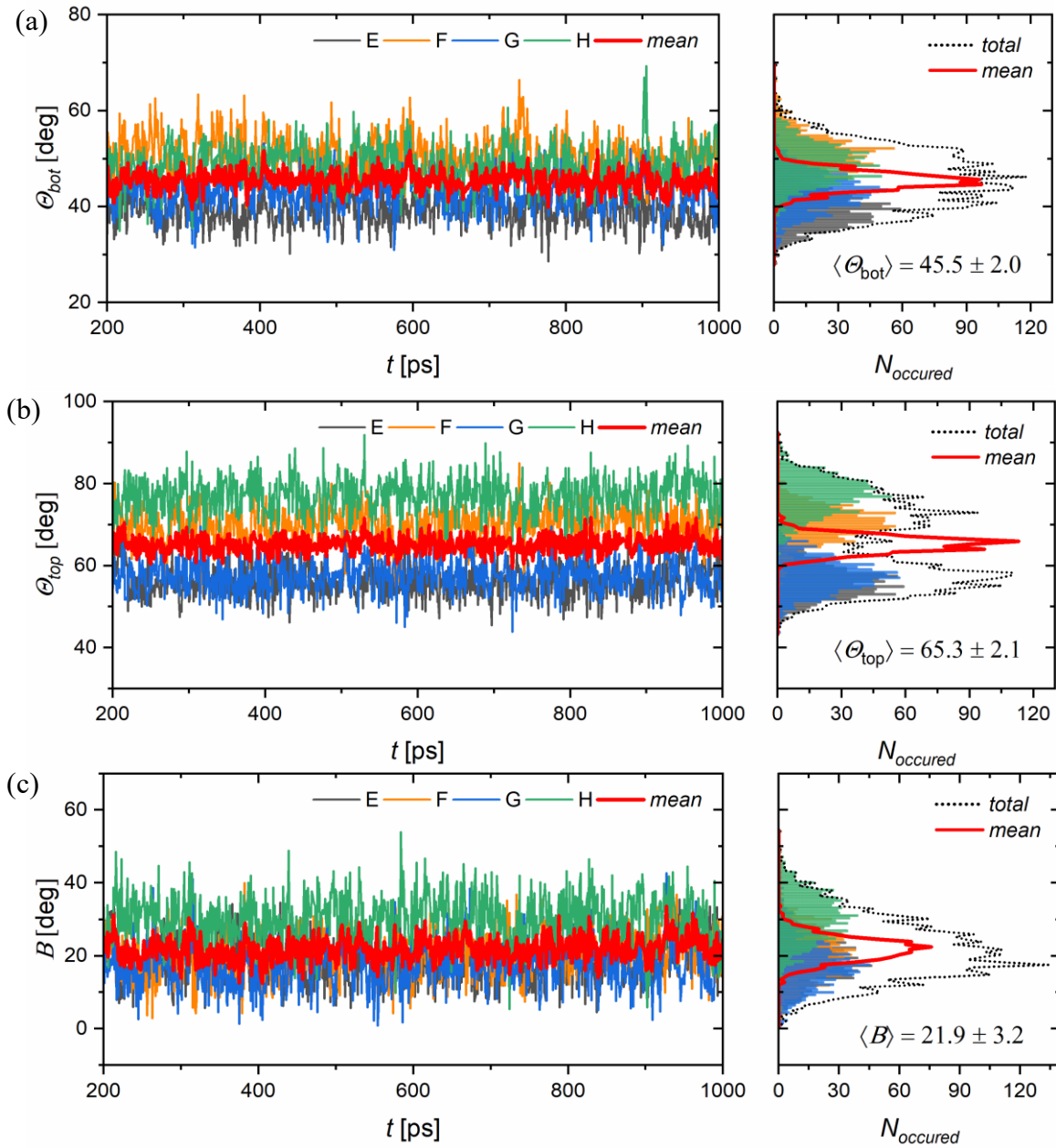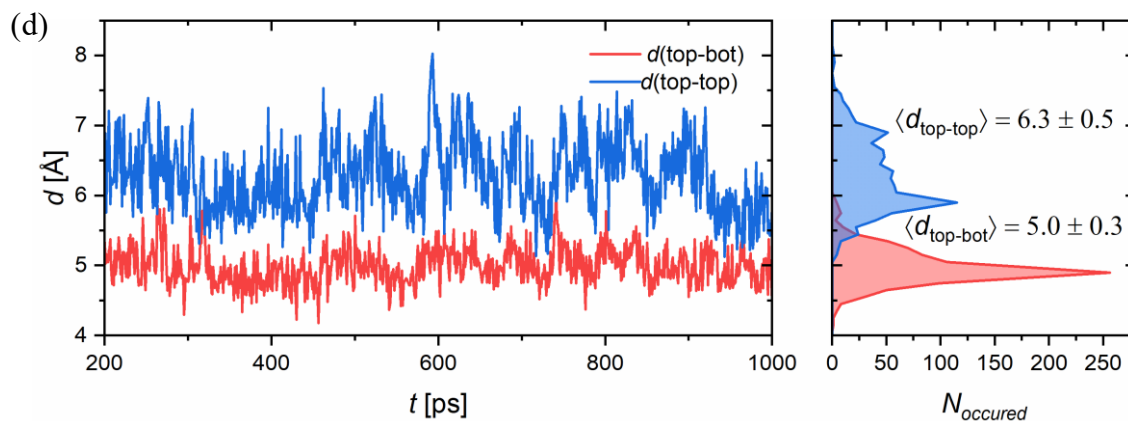
Figure 19. Evolution of *trans*-**Azo** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2.

(a)

(b)
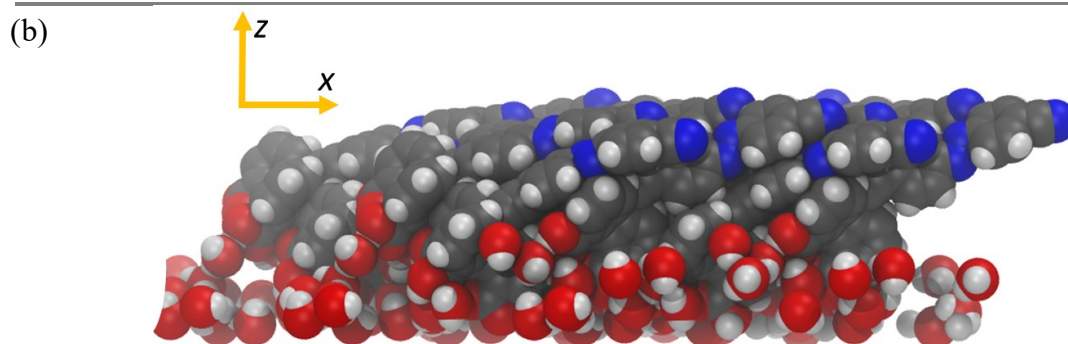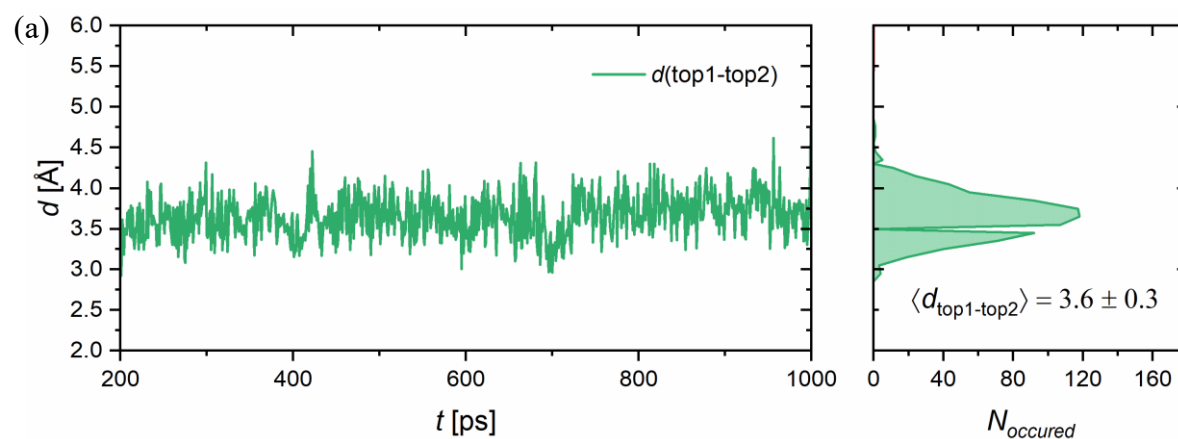
Figure 20. (a) Minimal average observed distance between upper and lower phenyl rings of the neighbouring azobenzene moieties (b) Snapshot of *trans*-**Azo** system at cca 620 ps. Collective tilting in single direction is clearly visible

This trend was further investigated by the simultaneous calculation of both the molecular bending and the distance between functional heads, revealing a highly symmetric 2D histogram, with a single maximum at around [$d = 6.6$ Å, $B = 20°$]. Furthermore, the combined distribution composed of radial distribution function (RDF) between different functional heads and the RDF between functional heads and triptycene stands was calculated, showing two distinct minima, with the higher density at lower values of stand / functional head distance. Their respective correlation graphs, in the right panel of Figure 21, have distinct X-shape patterns, with a strong negative correlation in the y = − x direction. This confirms that, for the configurations where the bending is significant, functional heads tend to be relatively close to each other with significant gravitation towards the triptycene stands.
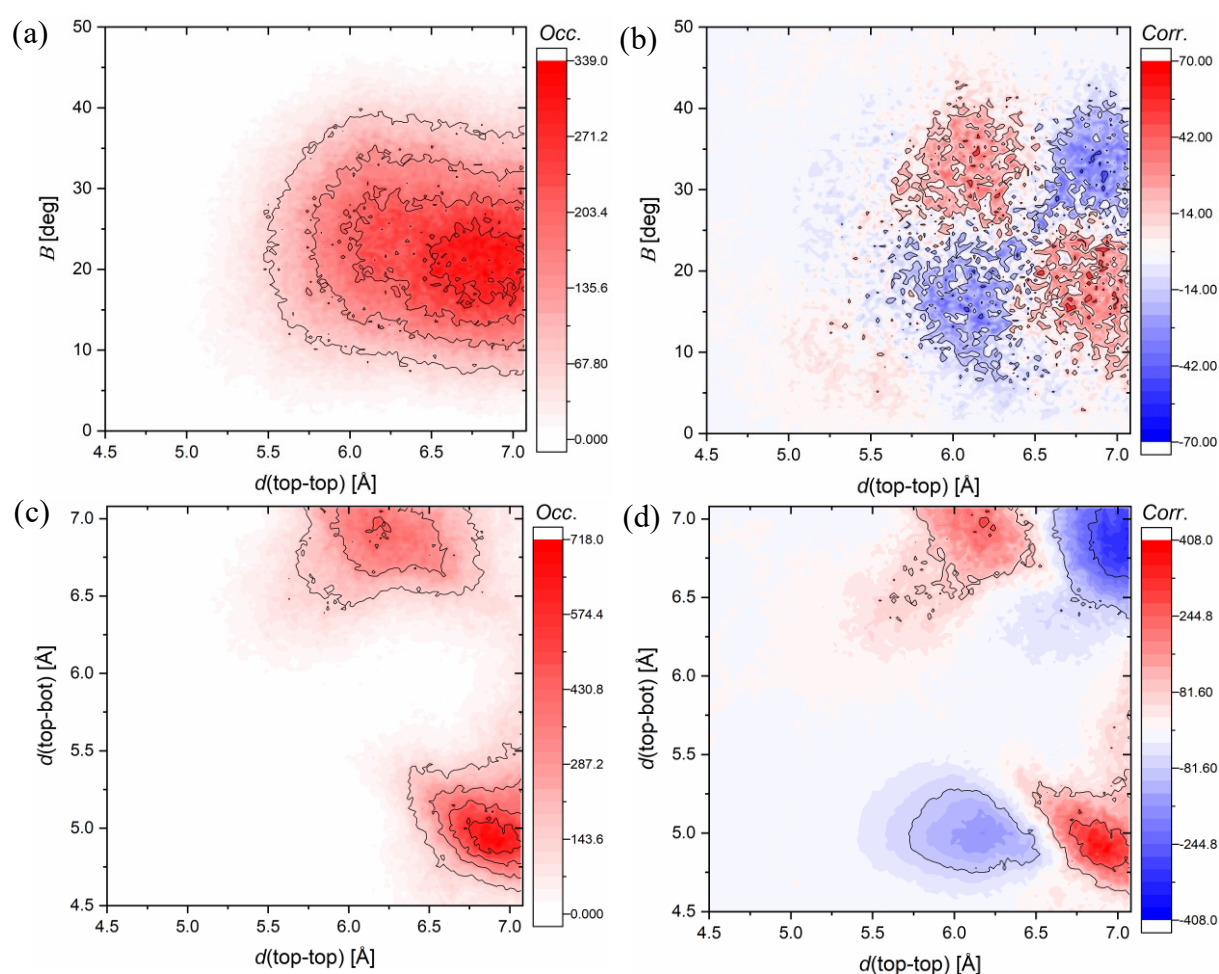
Figure 21. *trans*-**Azo** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. The red colour depicts positive correlation,

white depicts uncorrelated events, and blue depicts configurations less probable compared
to the completely uncorrelated case (see Methods, Section 4.2.5).

No significant correlation between the switches is present in *trans*-**Azo** system, with the most
significant Kendall rank coefficients having both slightly positive and negative values in a
similar range (Table 4). However, comparing the correlations of different observables (i.e., top
and bottom molecular tilt), we see that, the positive and negative coefficients alike, are slightly
more pronounced between the same pairs of individual machines, suggesting the caught trend
to be more than a mere noise artefact produced by the random motion of sampled observables
(see E-H and F-G pairs in Table 4). On the other hand, major changes in the upper layer film
structure produce highly correlated motions between neighbouring switches (Table S8 Table
S10 in the SI).

Table 4. *trans*-**Azo** Kendall correlation matrices for tilting and bending of individual
machines

	Θ(top) [deg]				Θ(bot) [deg]				B [deg]			
	E	F	G	H	E	F	G	H	E	F	G	H
E	1.00	-0.01	-0.01	-0.13	1.00	0.03	-0.01	-0.09	1.00	0.04	0.05	-0.11
F	-0.01	1.00	0.14	-0.03	0.03	1.00	-0.07	-0.05	0.04	1.00	0.03	-0.03
G	-0.01	0.14	1.00	0.05	-0.01	-0.07	1.00	0.07	0.05	0.03	1.00	0.05
H	-0.13	-0.03	0.05	1.00	-0.09	-0.05	0.07	1.00	-0.11	-0.03	0.05	1.00

5.1.5. Effects of functional heads: cis-Azo system

Like its more stable *trans*-**Azo** analogue, the metastable *cis*-**Azo** system has well converged
observables with narrow distributions of their mean, but wide range of individual tilt and bend
angles. However, mean tilt angle value of the top and bottom parts of the molecule is smaller
(21°), while bending angle mean is larger (34°) compared to *trans*-**Azo**, suggesting there is some
tendency for bending in the opposite direction from the film tilt. The way the functional group
centres were defined (centre of a phenyl ring connecting the C≡C linker), average neighbouring
distance between functional head centres and stands becomes greater compared to the distance
between neighbouring centres. This is, again, reasonable since the geometry of switched
azobenzene moieties prevents the rings from stacking too close to one another (note how the
peak positions of average distances in Figure 22 (d) changed places due to switching).

Figure 22. Evolution of *cis*-**Azo** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) average observed neighbouring distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2.

Functional head geometry of *cis*-**Azo** also leads to the less clear histograms of combined distribution functions and their correlation plots, compared to *trans*-**Azo** system. Instead of one distinct maximum, we can now see a plethora of superimposed extrema in case of RDF/ADF (Figure 23 (a)), and a single symmetric maximum in case of RDF/RDF (Figure 23 (c)) 2D histograms. Concerning correlation plots, while a rough outline of X-pattern in Figure 23 (d) can still be observed, graphs are more dispersed and distinction between areas of positive and negative correlation is less clear.

Table 5 shows that the motion correlation between the machines is negligible, except for the lower tilting of G and H neighbouring machines, whose tilt curves mimic each other reasonably well, but are shifted by about 20 degrees, producing a τ_b coefficient of ≈ 0.25.
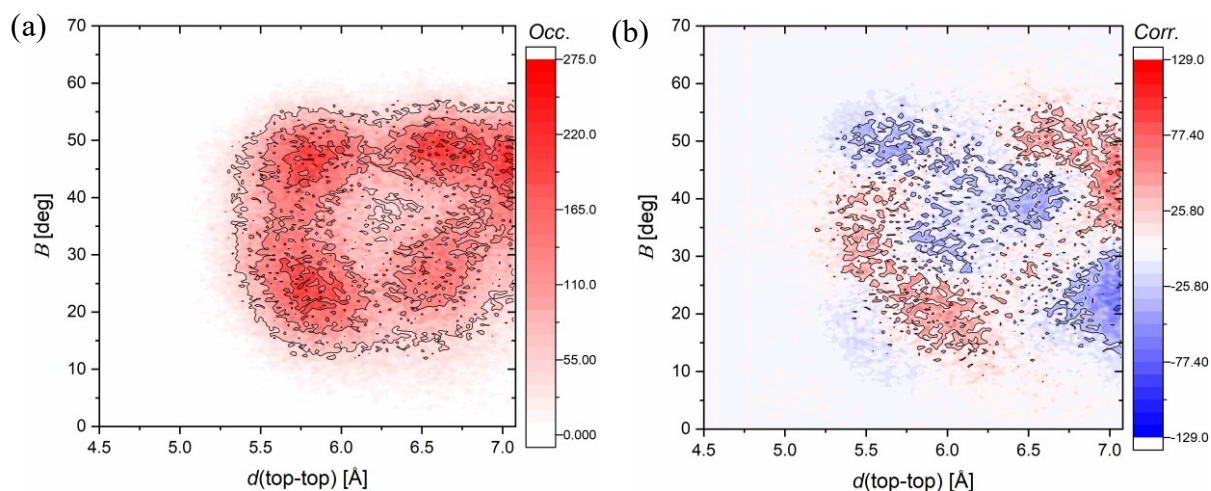
Figure 23. *cis*-**Azo** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. The red colour depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable compared to the uncorrelated case (see Methods, Section 4.2.5).

Table 5. *cis*-**Azo** Kendall correlation matrices for tilting and bending of individual machines

	Θ(bot) [deg]				Θ(top) [deg]				B [deg]			
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	0.08	0.00	-0.02	1.00	0.00	-0.10	0.01	1.00	0.04	0.00	0.05
F	0.08	1.00	-0.09	-0.07	0.00	1.00	0.02	-0.07	0.04	1.00	0.03	0.01
G	0.00	-0.09	1.00	0.26	-0.10	0.02	1.00	0.02	0.00	0.03	1.00	0.06
H	-0.02	-0.07	0.26	1.00	0.01	-0.07	0.02	1.00	0.05	0.01	0.06	1.00

5.1.6. Effects of functional heads: FER system

Triptycene stands of **Fer** system tilt at 22.7° while its motor heads tilt at 37.4° and bend at 20.2° on average. Tilting of the **Fer** systems core does not induce tilting of functional heads in the same direction. Overall, angular observables are well converged, with the exception of Θ_{bot} angle, that shows a slight increasement in mean ($\approx 7°$) at about 380 ps. It subsequently drops again ($\approx 5°$) approximately at 560 ps resulting in a somewhat rougher mean angle curve. Similar trend can be observed in bending but is far less marked in Θ_{top} time evolution. This roughness is a result of the tilt angle change in all 4 machines, instead in a single one. Looking at the neighbouring distances between the motor heads and their distance from closest triptycene stands, one can again observe a high amount of overlap at 5.8 Å. Neighbouring distance between motor heads is bimodal instead of following the normal distribution which is a result

of sudden rise at 380 ps (to 7 Å) and subsequent slow drop in distance until the minimum at about 560 ps (d = 5.4 Å) is reached. It can be seen that the change in distances closely mimicks the Θ_{bot} and B tilt curves (Figure 24).
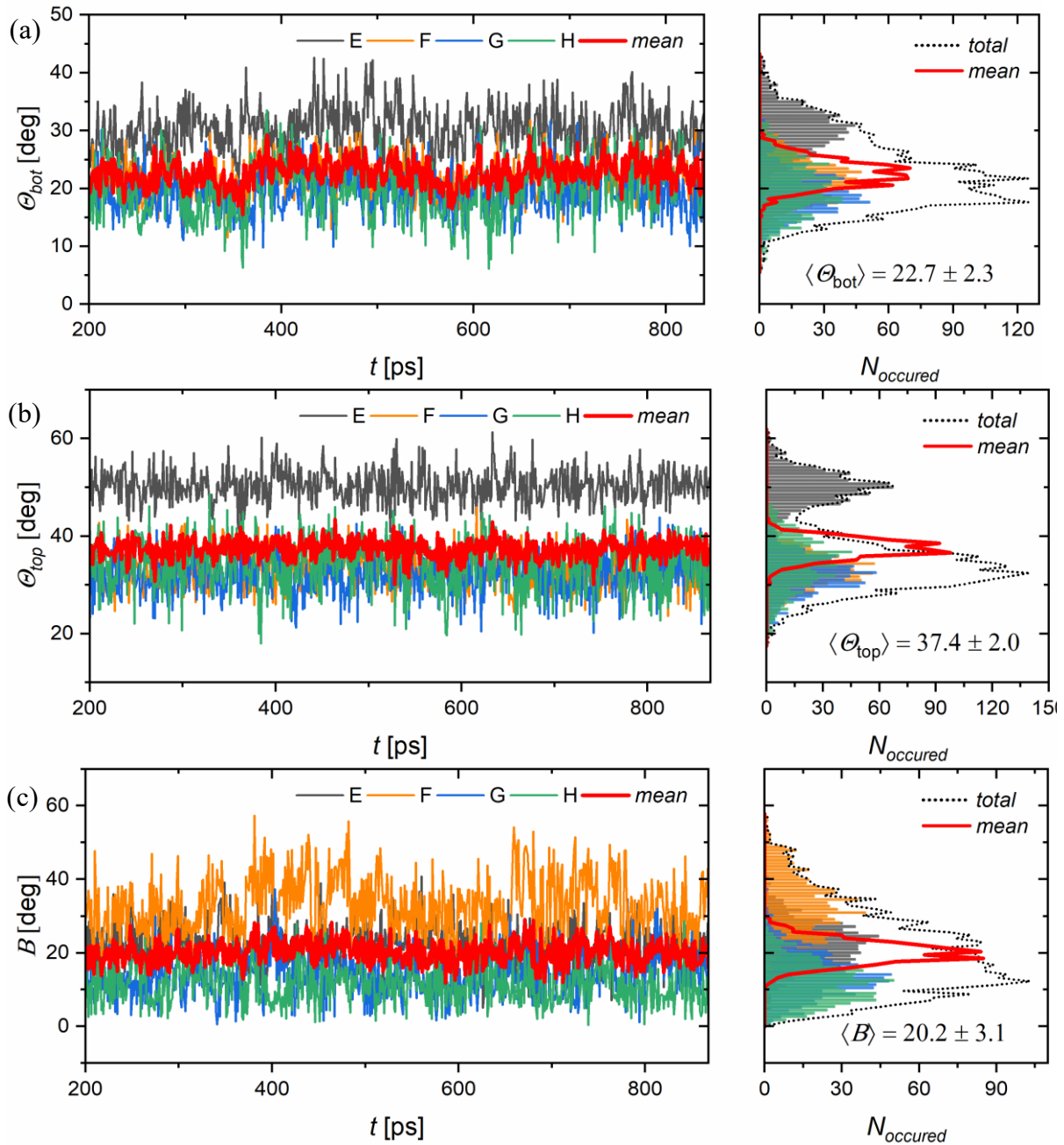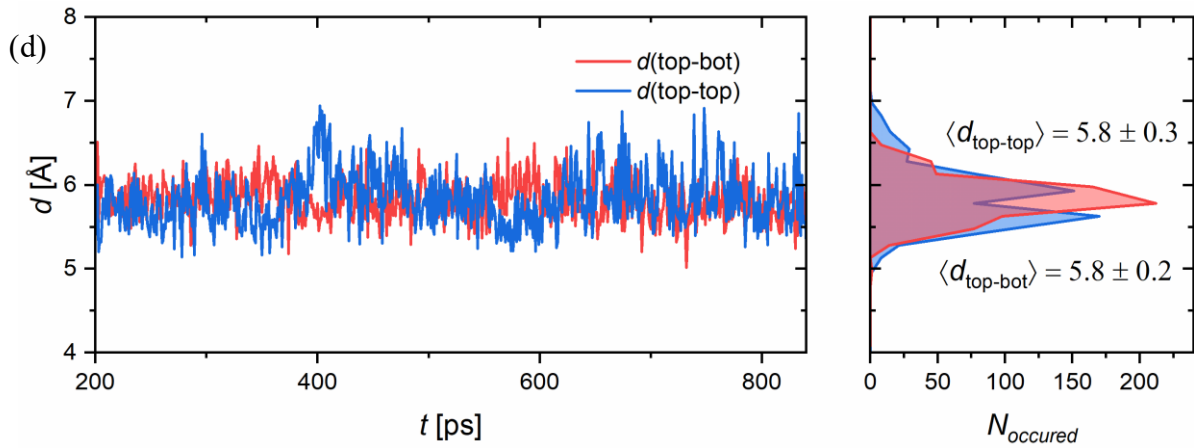
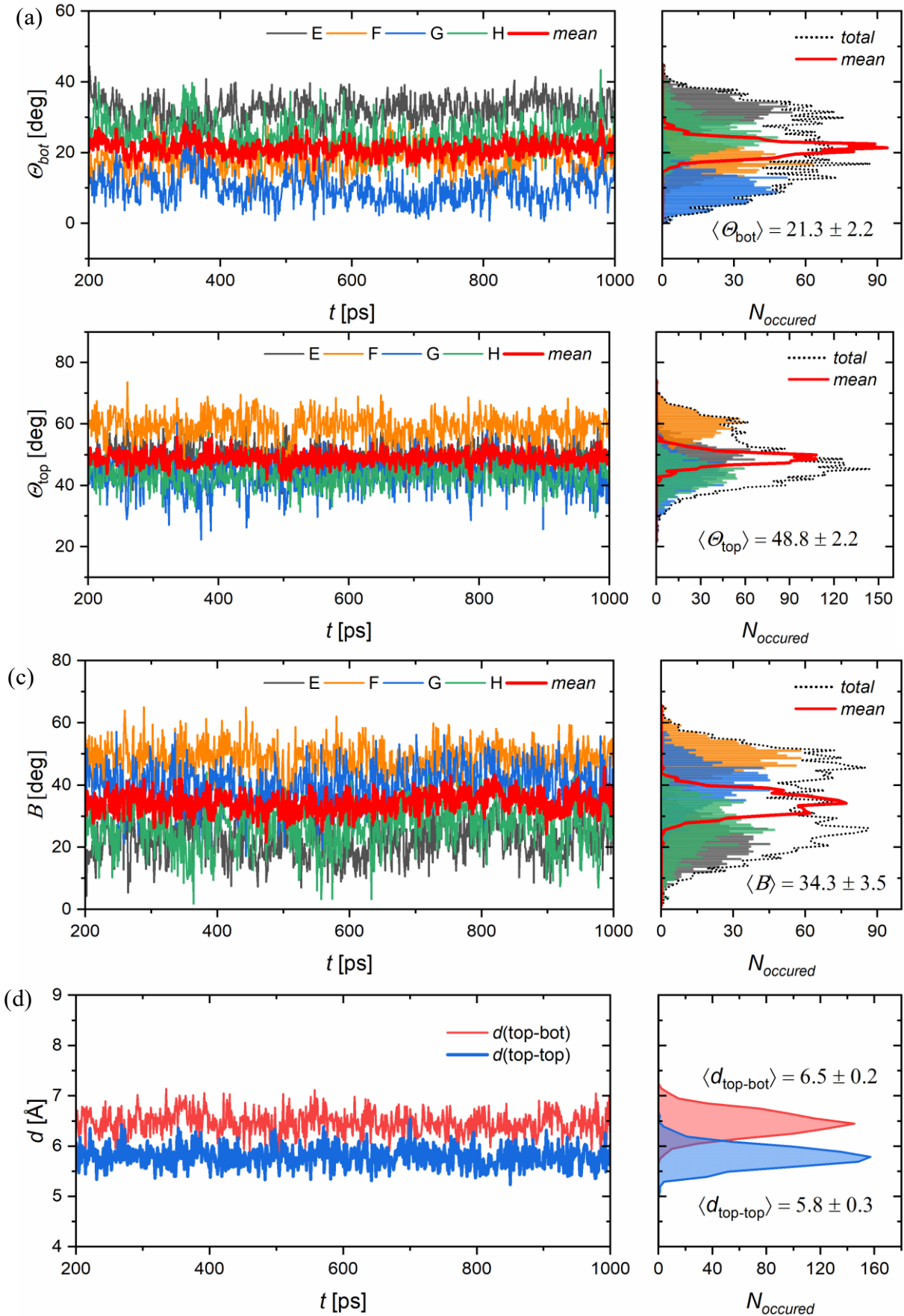Figure 24. Evolution of **Fer** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2.

In **Fer** case, ADF-CDFs show two distinct, symmetric maxima with a well-defined X-pattern type correlation plot, showing a strong positive correlation between small bend angles of motors in proximity to each other ($B = 14°$, $d = 5.7$ Å) and at larger angles and distances ($B = 32°$, $d = 6.5$ Å). Surprisingly, RDF-RDF CDF histogram shows a single, wide maximum (d(top-top) = 5.7 Å, d(top-bot) = 6.8 Å) (Figure 25). However, its respective correlation plot has a dispersed range of positively and negatively correlated distance pairs, reflecting the observed interchange between two sets of distances during the simulation run. Correlation plots do not reveal a significant amount of correlation besides a single medium sized Kendall coefficient representing correlated tilts between F and G triptycene stands (Table 6).
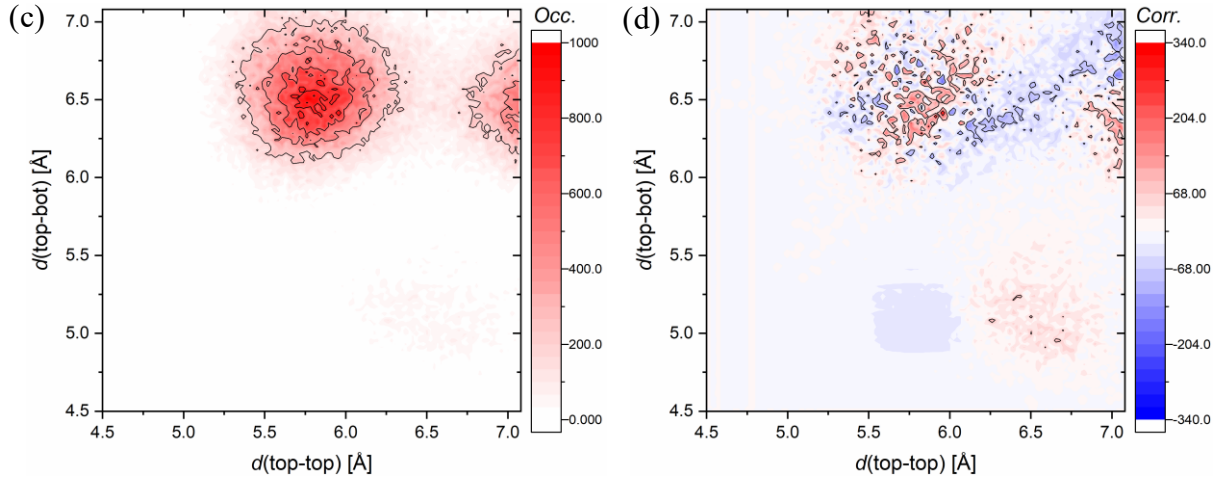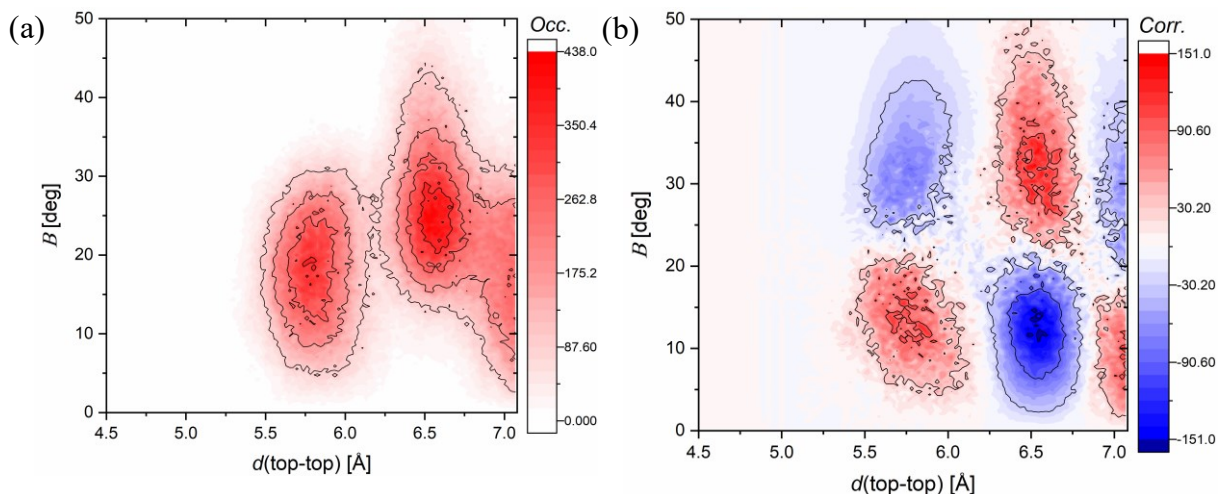
Figure 25. **Fer** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable compared to the uncorrelated case (see Methods, Section 4.2.5).

Table 6. Fer Kendall correlation matrices for tilting and bending of individual machines

	Θ(stand) [deg]				Θ(fh) [deg]				B [deg]			
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	-0.01	0.02	-0.02	1.00	-0.10	0.17	0.02	1.00	0.05	0.00	-0.01
F	-0.01	1.00	0.13	0.23	-0.10	1.00	-0.13	-0.06	0.05	1.00	0.01	-0.05
G	0.02	0.13	1.00	0.07	0.17	-0.13	1.00	-0.01	0.00	0.01	1.00	-0.03
H	-0.02	0.23	0.07	1.00	0.02	-0.06	-0.01	1.00	-0.01	-0.05	-0.03	1.00

5.1.7. Effects of anchoring groups: Phe0 system

To decouple effects of anchoring groups from those of functional heads on the structural properties of SAMs, we simulated a series of systems bearing only a phenyl ring as a functional group, with 0, 1 and 3 carboxylic groups. The results for 1 ns simulation runs of **Phe0** and **Phe** systems are given in the following sections, while the results for **Phe1** system can be found in the S5 section in the SI.

Although well converged, the **Phe0** SAMs are exceptionally tilted during the the whole simulation, with triptycene core tilts of around 79° on average, and top parts of the molecule's tilts of 85°. It i s interesting to note that the *PHE* family of systems in general on average bends less markedly compared to systems with mounted switches or a motor. Neighbouring distances

between top and bottom molecular parts are well converged, but highly overlapped during the whole simulation run (Figure 26). Correlation plots are highly dispersed, both in ADF/RDF and RDF/RDF cases; their respective CDF histograms have one ($B = 18°$, $d = 5,2$ Å), and two (5.5 Å, 6.5 Å; 5.5 Å, 5.2 Å) distinct symmetric maxima, respectively. Correlation between individual machines is not pronounced, and the Kendall tau correlation matrices can be found in the SI (Table S6).
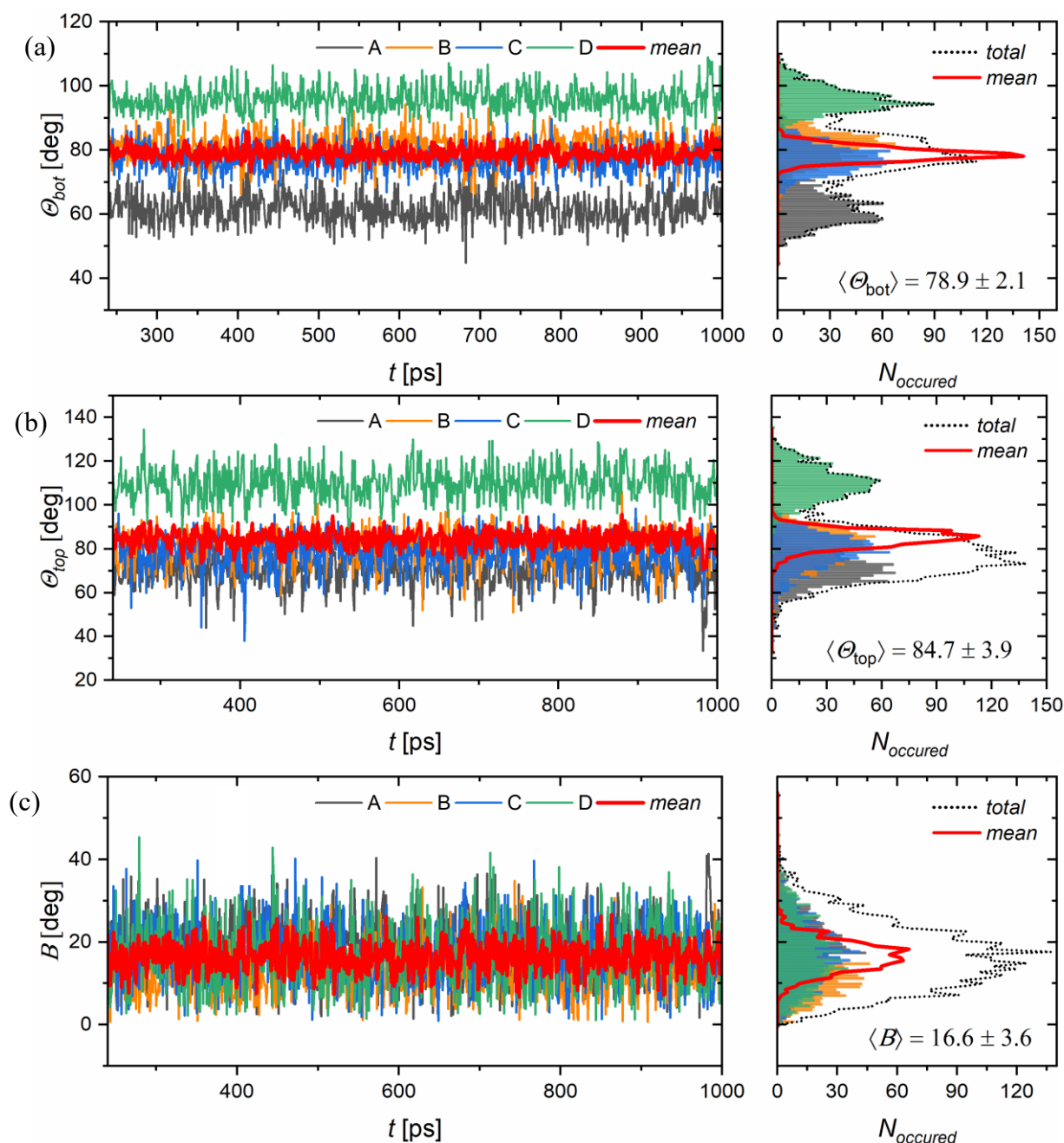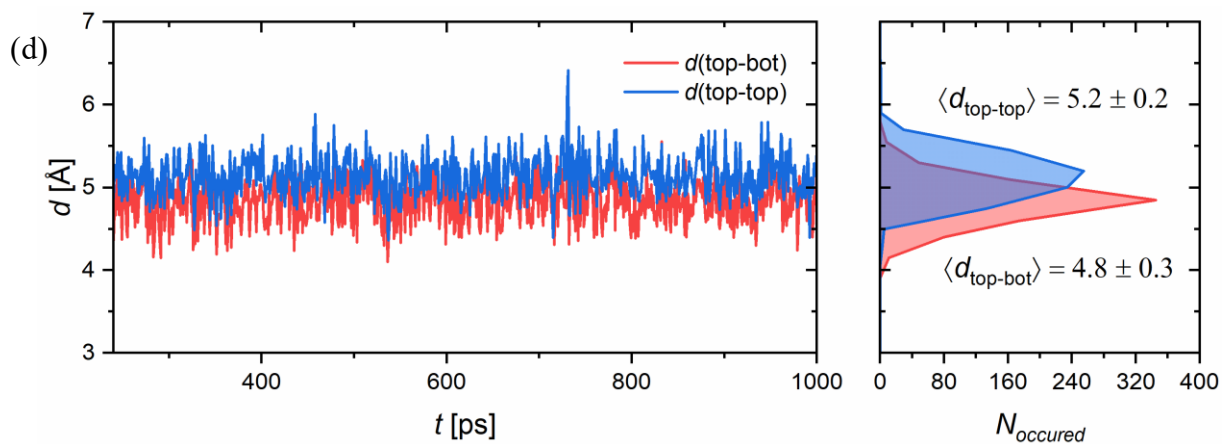
Figure 26. Evolution of **Phe0** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (A − D), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2.
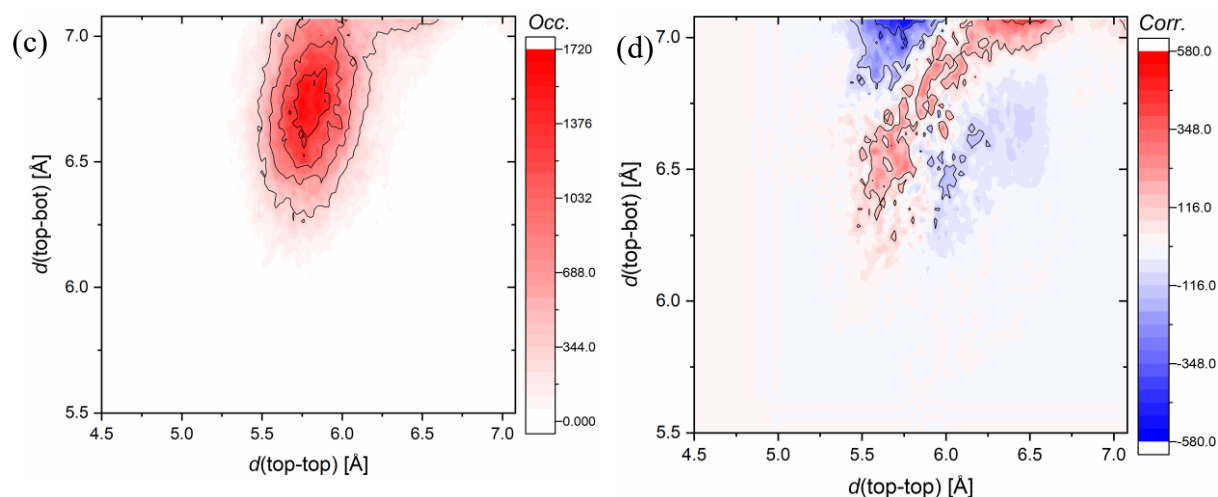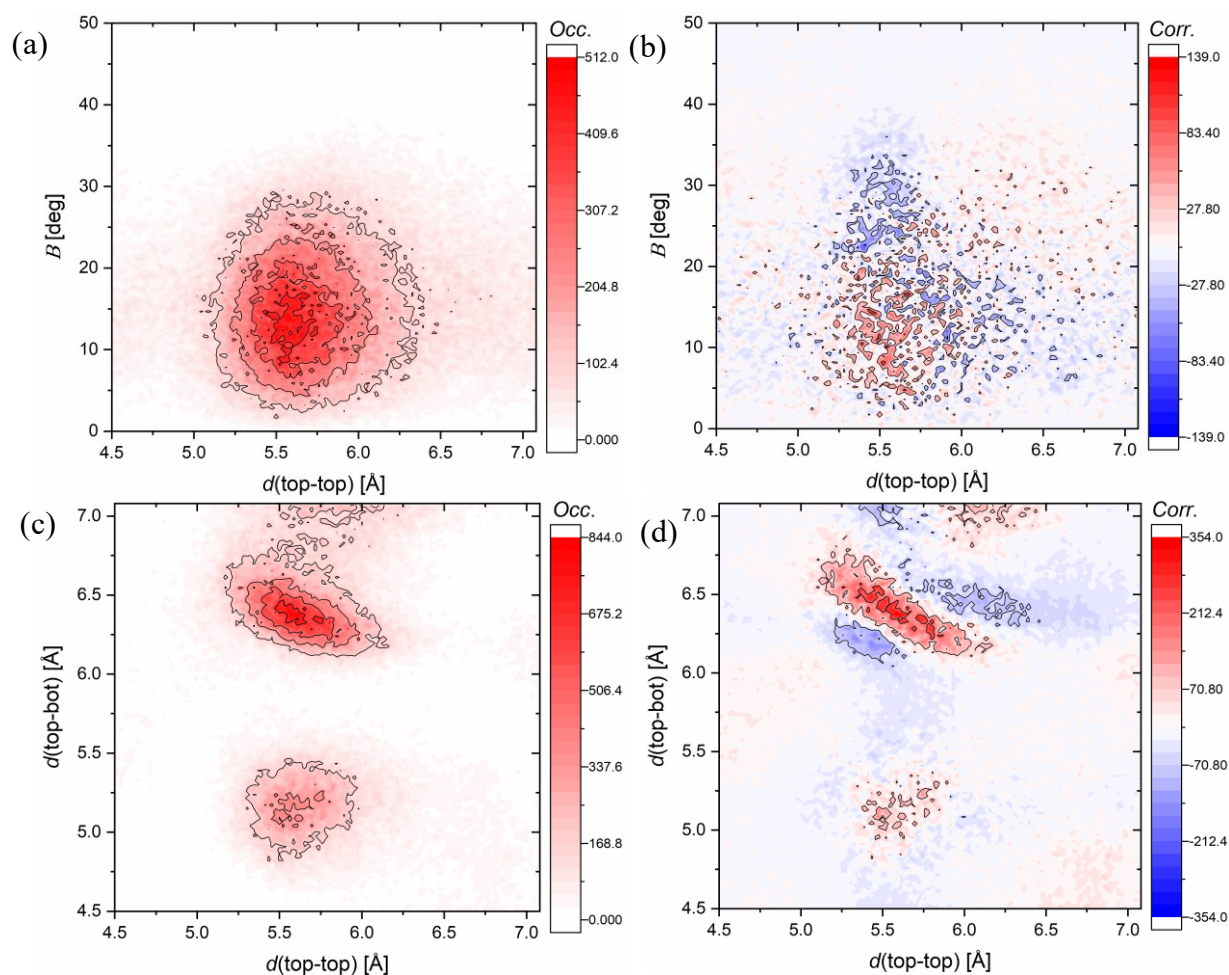
Figure 27. **Phe0** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable than in the uncorrelated case (see Methods, Section 4.2.5).

5.1.8. Effects of anchoring groups: Phe system

Phe system's tilt angle is on average 49.0° for the lower part, and around 60° for the phenyl head. Average bending of the functional head is around 16.5°, and is, as was already pointed out, similar to the other *PHE* family members. Looking into Figure 28 (d), one can note the phenyl rings are generally closer to neighbouring triptycene bases (d(top-bot) = 5.2 Å) compared to the neighbouring rings (d(top-top) = 6.6 Å). This is well conserved during the production run, without abrupt fluctuations, producing well defined normal-like distributions without significant overlap. Combined distribution functions show very broad symmetric peaks. ADF-CDFs single maximum is at about (d = 6.6 Å, B = 16°). Although not particularly well defined, the correlation graph shows a slightly positive correlation between large functional had distances and smaller bends, and configurations with smaller functional head distances having higher bends. RDF-RDF CDF is similar to its ADF analogue (Figure 29).
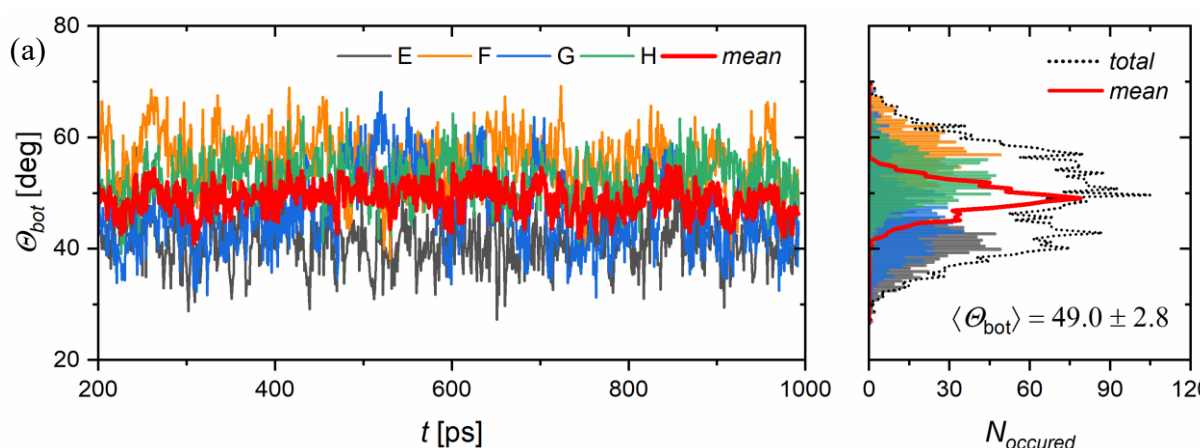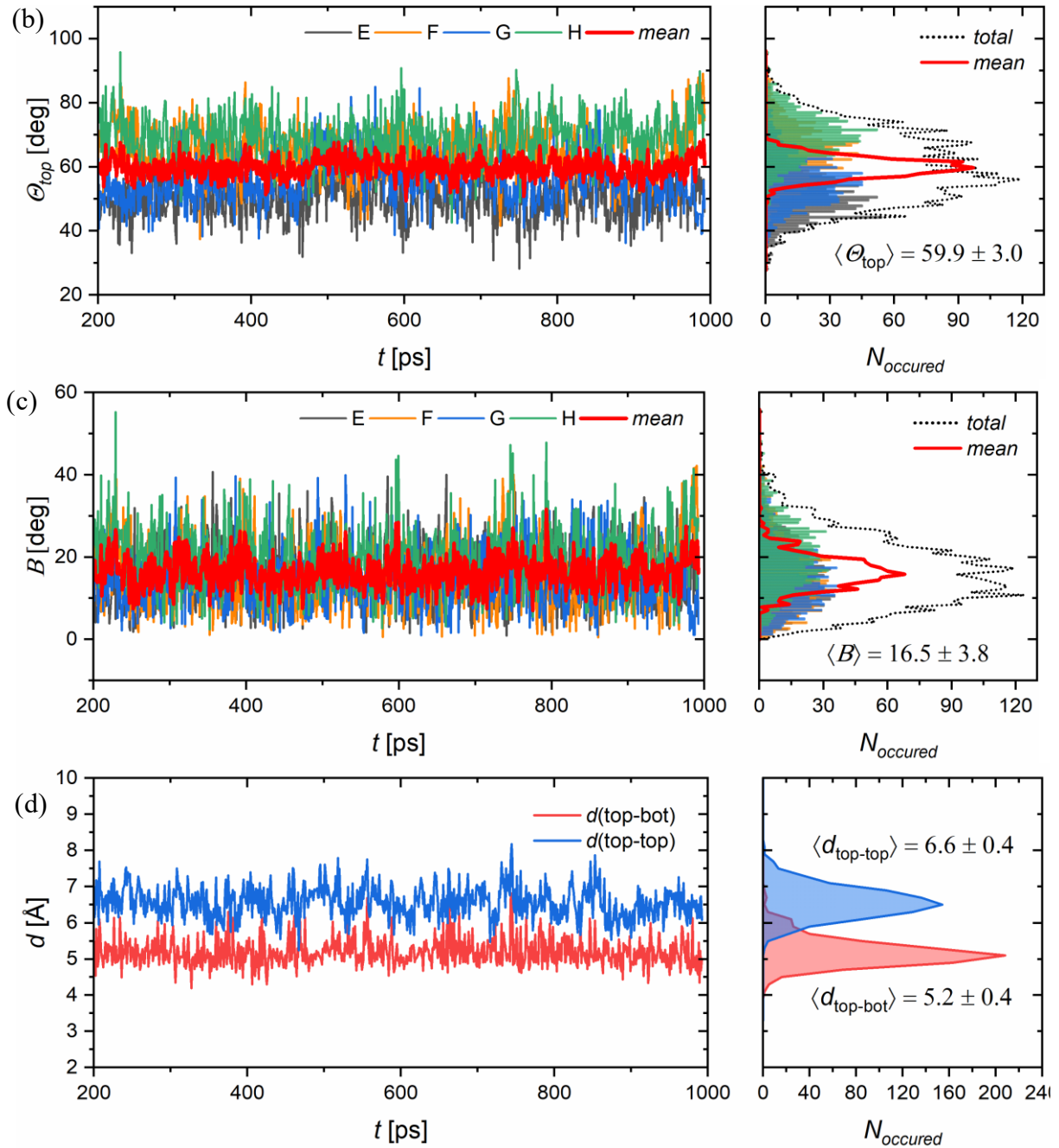
Figure 28. Evolution of **Phe** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2..
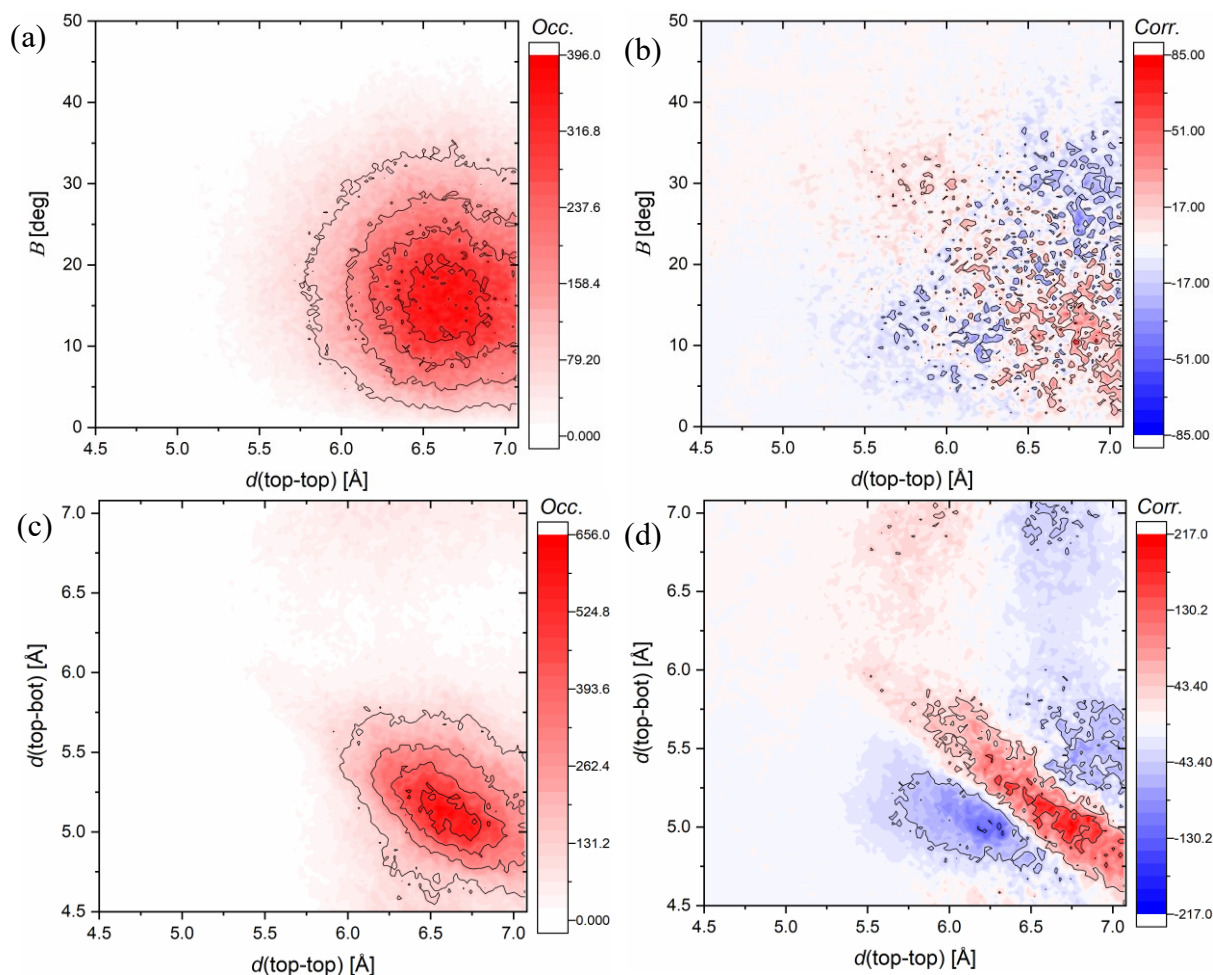
Figure 29. **Phe** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable compared to the uncorrelated case (see Methods, Section 4.2.5).

5.2. Cluster-based Molecular Dynamics

5.2.1. Issues with the method

Lack of the periodic boundary conditions invariably leads to a protrusion of the water molecules through film layers at the system edges. As the systems have finite dimensions and the collisions of molecules with the external potential wall are perfectly elastic, the net force of water layer is going to be 0 in the xy plane, but > 0 in the z direction. As a consequence, after enough collisions, the film inevitably gets destroyed at the interface between elliptic potential and the

organic layers – inability of triptycenes to properly stack due to their position at the edges of the system leaves cavities in the film for the water to escape through into the vacuum. This leads to unphysical results, as is shown in the Figure 30: the energy steadily rises throughout the simulation, unable to equilibrate, until the protrusion of water and subsequent mixing of organic and solvent phases destroy the layers, lowering the energy of the system. This happens with all the systems, differing only in simulation timescale necessary for the bilayer to get destroyed.

It is worth noting that this issue is not present in systems including periodic boundary conditions. Further GFN-FF energy profile examples of investigated systems can be found in S9 Section of SI (Table S13).
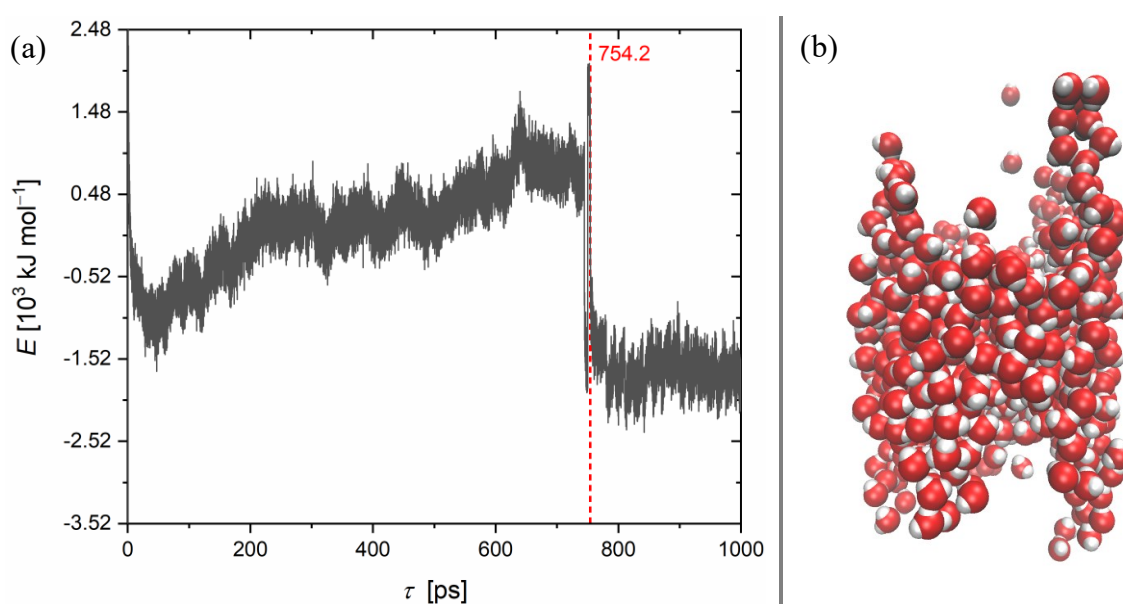
Figure 30. (a) An example of an energy profile of a cluster simulated with the GFN-FF method. The system cannot equilibrate due to the lack of periodic boundary conditions. (b) Water molecules of **Azo** system at 750 ps of simulation time. The water protrusions happen almost exclusively on the edges of the system, indicating that the phenomenon is not a result of intrinsic film instability, but occurs due to the presence of an external potential.

However, in rare cases where the destruction of the monolayer was particularly slow, basic structural parameters, including tilting and bending, were analysed. Figure 31. illustrates such an example: in case of **Fer** system simulated at the experimental APM, monolayer behaves reasonably well due to the slow increase in energy. Although the convergence of analysed

observables is not particularly good, results are still comparable with the PBC MD results, and are summarised in **Table 7**.

Table 7. Tilt angle comparisons of Fer system obtained by GFN-FF and PBC DFTB+ MD methods

	GFN-FF	DFTB+
$\langle \Theta_{bot} \rangle$	21.6	22.7
$\langle \Theta_{top} \rangle$	32.5	37.4
$\langle B \rangle$	18.4	20.2
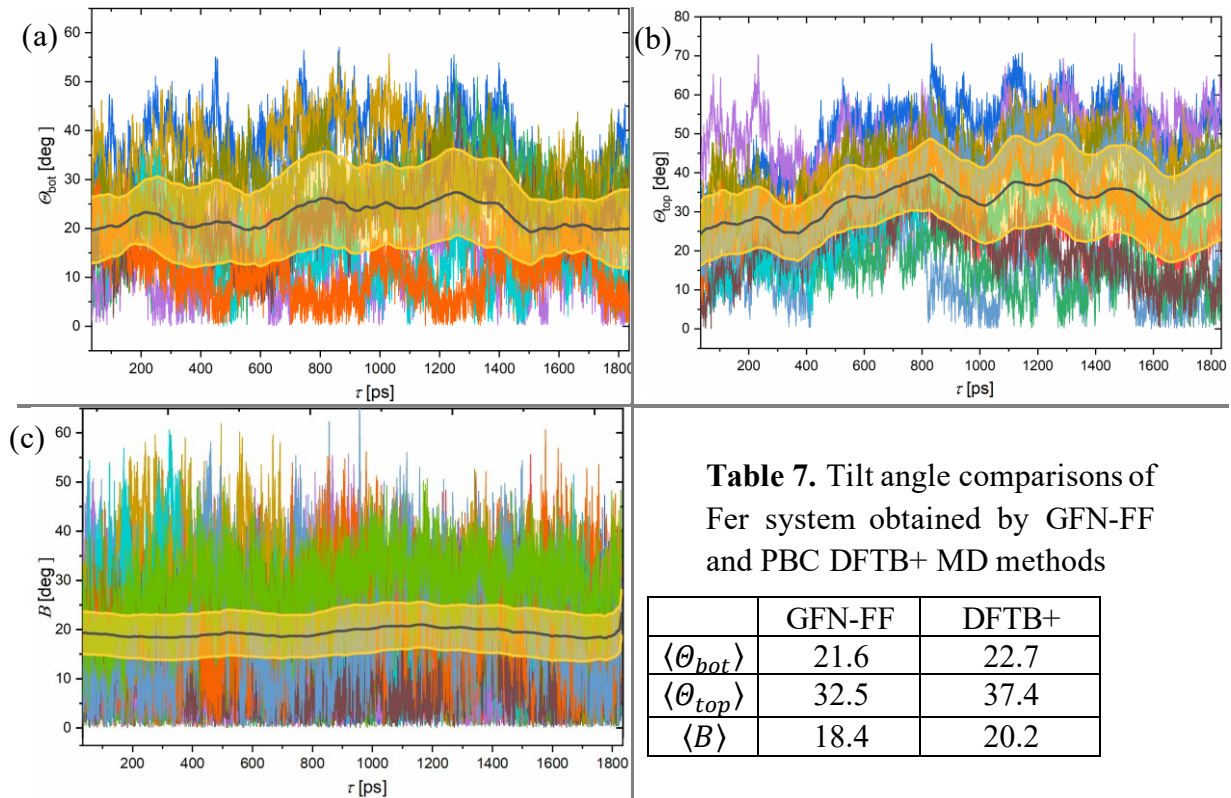
Figure 31. (a) Evolution of average bottom (a), top (b) and bending (c) angles for **Fer** system obtained from the GFN-FF MD trajectory. Temporal evolution of angles for each machine in the system is denoted with thin coloured lines, thick black line is the mean angle in every timestep, and yellow zone represents the 95% confidence intervals. The roughness of the mean angle curves can be contributed to the endge effects.

5.3.   Scanning of local minima

Scanning of periodic structures of all investigated systems across different tilt angles was performed using the van der Waals density functional,[80] which was shown to provide good description of weak interactions at a reasonable computational cost. Overall, film interaction energy drops with increasing tilt angles to either a distinct minimum or until flattening of the energy curve occurs. At low angles, triptycene interactions are dominant because the functional heads are too far apart to interact. **Fer** and **Thio** curves strongly resemble each other. *cis*-**Dia**, *trans*-**Dia** and *cis*-**Azo** systems are also somewhat similar, each showing a minimum between 25°-35° for the stand, and 30°-40° for the switch tilt (**Table 8**). The shape of their curves also

mimics the **Phe** system, suggesting that the interactions between triptycene stands are more dominant compared to the interactions between functional heads (Figure 32 and (c) in Figure 33). **Azo** is somewhat specific, with a huge drop in stabilization energy (around 20 kJ·mol[1] in total) at higher angles, where the interactions between the functional heads become a dominating stabilizing factor.

It is worth noting that the top and bottom tilt angle curves generally do not cross each other. In fact, they tend to diverge at higher angles. This means that the larger the tilt angels are, the more molecules tend to bend. *cis*-**Azo** is an exception here, with the lines crossing each other at about 30°, suggesting that at this point, the stabilization from triptycene stand interactions becomes more important. Furthermore, during the optimization of different geometries, most of the geometries ended up gravitating towards the point where the lines cross, suggesting there is a local minimum, hidden by the rich and shallow potential energy surface of the film (note the negligible change in interaction energy for a wide range of tilt angles)
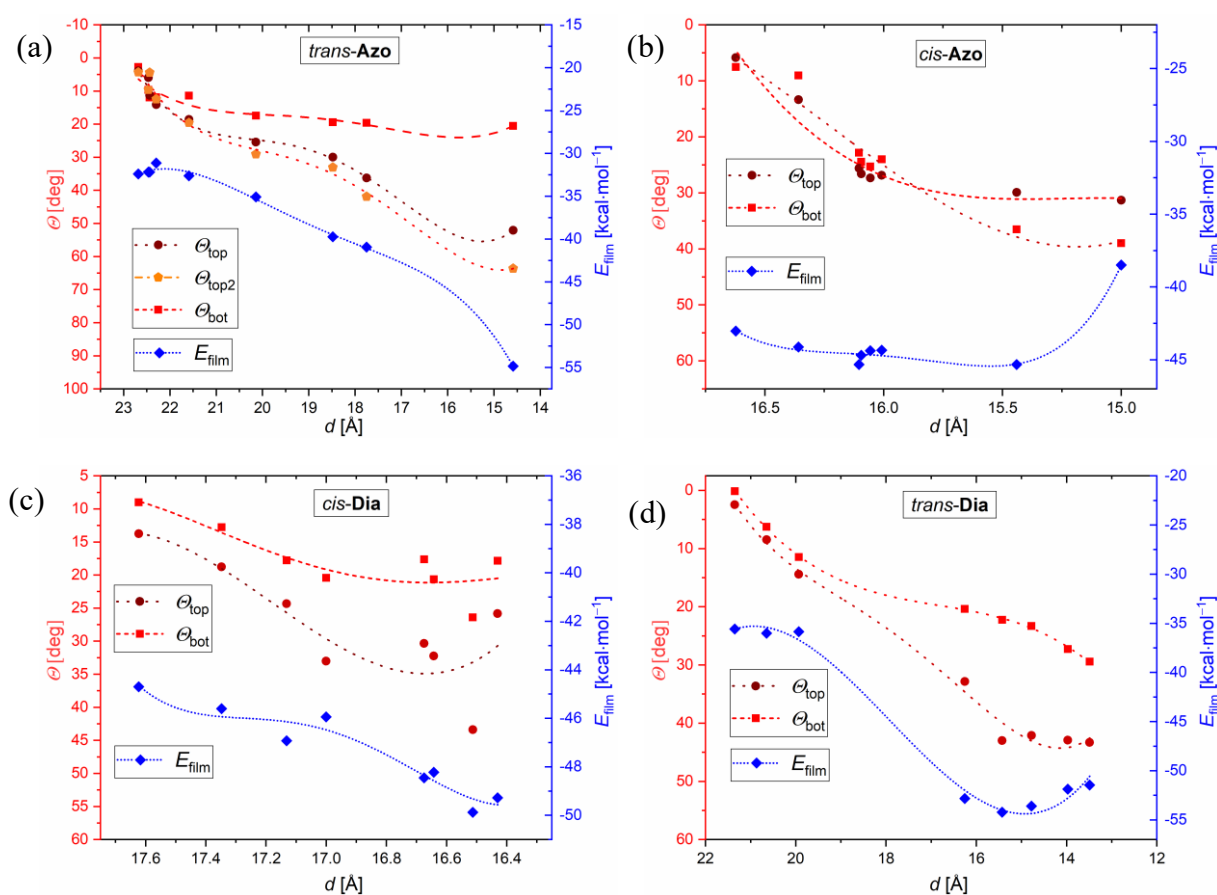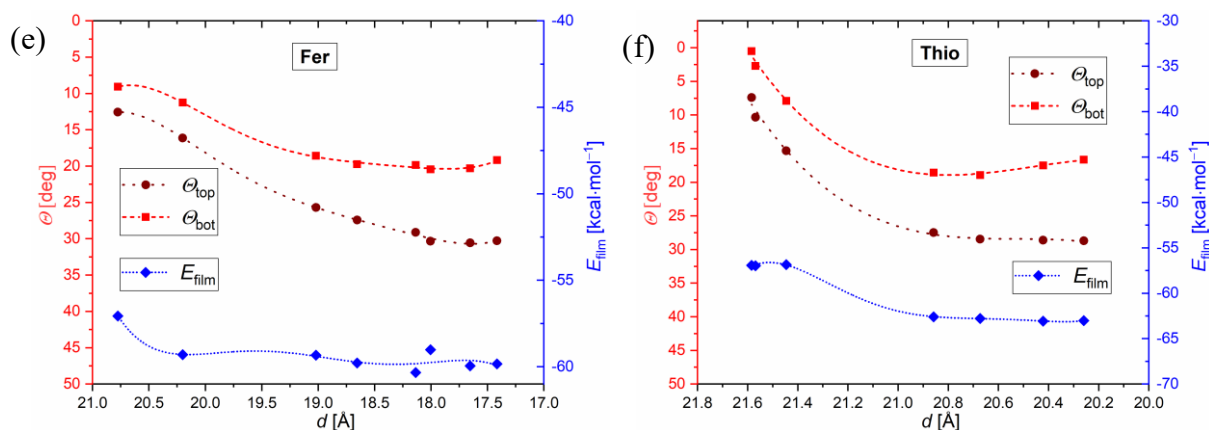
Figure 32. Molecular tilt and film interaction energy as functions of film thickness for (a) *trans*-**Azo**, (b) *cis*-**Azo**, (c) *cis*-**Dia**, (d) *trans*-**Dia**, (e) *closed* **Thio** and (f) **Fer** systems. Top and bot subscripts are denoting tilting of the triptycene stands and the functional groups, respectively.

Concerning the **Phe** family of systems, the situation is somewhat different. In the case of **Phe0** and **Phe1**, all structures agglomerate around two different angle groups: the first one is the less stable straight structure with no interaction between the heads, while the second group forms at high tilting angles where the dominating interaction becomes the one between phenyl moieties and triptycenes. The latter group is lower in energy by about 5 kJ·mol[1] in both systems (**Table 8**). The change in the hierarchy of interactions is confirmed with the crossing of tilt angle curves between the groups (Figure 33). On the other hand, in the case of the **Phe** system, angle curves diverge at higher tilt angles with a well-defined film interaction energy minimum at around (Θ_{bot} = 20°; Θ_{top} = 30°). Figure 33 (d) shows the complex effect of two triptycenes and the middle phenyl linker of the **Tri** system. Two local minima are observed, with the deeper one being at higher tilt angles. It is interesting to note how Θ_{top} and Θ_{bot} completely mimic each other.
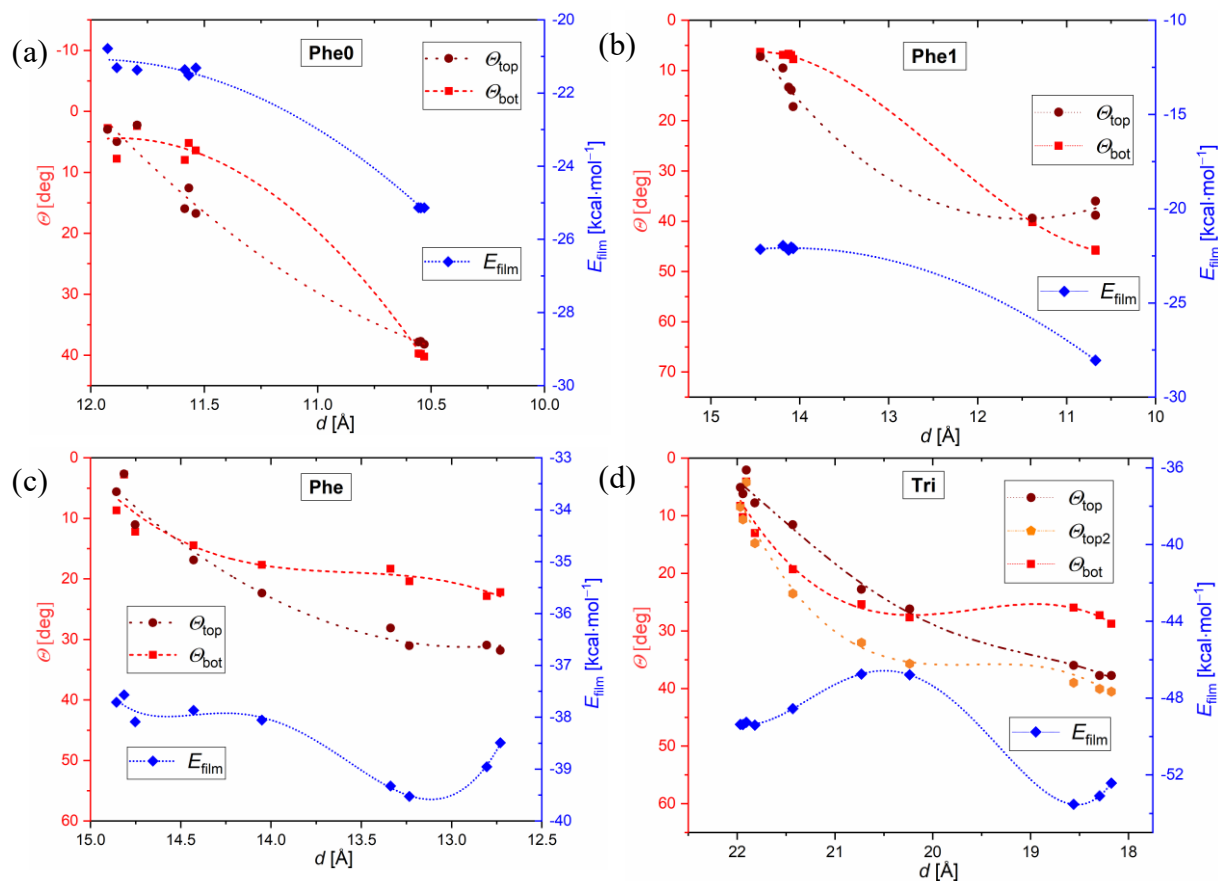
Figure 33. Molecular tilt and film interaction energy as functions of film thickness for: (a) **Phe0**, (b) **Phe1**, (c) **Phe** and (d) **Tri** systems.

Table 8. Energetically most favourable top and bottom molecular tilt angles with their respective film interaction energies. Where applicable, Θ_{top2} is denoted in brackets

	trans-**Azo**	*cis*-**Azo**	*cis*-**Dia**	*trans*-**Dia**	**Fer**	**Thio**
Θ_{bot} [deg]	20.5	23-36	17-26	22-23	11-20	16-19
Θ_{top}, (Θ_{top2}) [deg]	52 (64)	25-27	25-43	43-44	16-30	27-29
$-E_{film}$ [kcal mol^{-1}]	54	44-45	48-49	53-54	59-60	62-63

	Tri	**Phe0**	**Phe1**	**Phe**
Θ_{bot} [deg]	26-28	40	46	18-20
Θ_{top}, (Θ_{top2}) [deg]	36-37 (39-41)	38	36-38	28-31
$-E_{film}$ [kcal mol^{-1}]	52.5 - 53.5	25	28	38

5.4. **Spectroscopy**

For an experimental estimate of average SAM tilting, three components are necessary: (i) polarisation modulation infrared reflection-adsorption spectroscopy (PM-IRRAS) spectra, isotropic IR spectra of samples measured in KBr pellet, and the theoretical IR spectra (grey curves in Figure 34), which are used for determination of the monitored vibrational modes' polarization direction.

5.4.1. Isotropic and anisotropic infrared spectroscopy.

The PM-IRRAS technique in context of SAM characterisation is frequently used to determine whether the reasonably well organised SAMs[XXI] are present on a metallic surface by analysing the characteristic vibrations of functional groups present in the sample.

Figure 34 summarises the vibration spectra of investigated systems. Starting from the highest monitored frequencies, at about 3500 cm^{-1}, C−H triptycene bridgehead stretch can be identified. The intensity of this peak is rather weak in case of the isotropic IR, while it is strongly overlapped, producing wide PM-IRRAS peaks in cases of **Azo**, **Fer** and **Thio** systems. **Dia** is an exception, and this peak was therefore used as a purely z-polarised peak in the tilt angle determination. In the 3000-3400 cm^{-1} range, peaks corresponding to the atmospheric impurities can be observed (highlighted by asterisk in Figure 34), overlapping the peaks corresponding to the stretches of triptycene stands (note the comparison with the IR spectra in the same frequency range). These signals are known to be regularly present in PM-IRRAS spectra of triptycene based SAMs.[10,11] At around 2250 cm^{-1} one can identify a weak C≡C bond stretch, again more pronounced in isotropic spectra. In cases of **Azo** and **Dia** switches, these signals are stronger, as the carbon C≡C stretch is accompanied by C≡N stretch, similar in frequency. The most prominent characteristic signals of tripodal machines are probably the high intensity peaks in 1700 cm^{-1} region, corresponding to the C=O stretches of carboxyl groups. There are three of these (one for every carboxylic group) and are basically degenerate, producing a single clean sharp peak. Intensities of these signals are very pronounced in FT-IR spectra while being significantly lower in the PM-IRRAS case (respectively red and blue curves in Figure 34)

[XXI] Polarisation modulation evens out vibrations with randomly oriented transition moments, recording only the signals of monolayers with collective orientation, e.g., reasonably well-organized films.

suggesting pronounced tilting of triptycene stands. Below 1600 cm^{-1} several high intensity peaks can be observed: (i) multiple skeletal stretching and bending vibrations of triptycene stand and aromatic rings of functional heads (1420 cm^{-1}, 1580 cm^{-1}, 1540 cm^{-1}), (ii) δ (stand aromatic breathing) at around 1380 cm^{-1} and (iii) the strongly overlapped pair of peaks at 1250 cm^{-1}, identified as C–H scissoring of triptycene units.
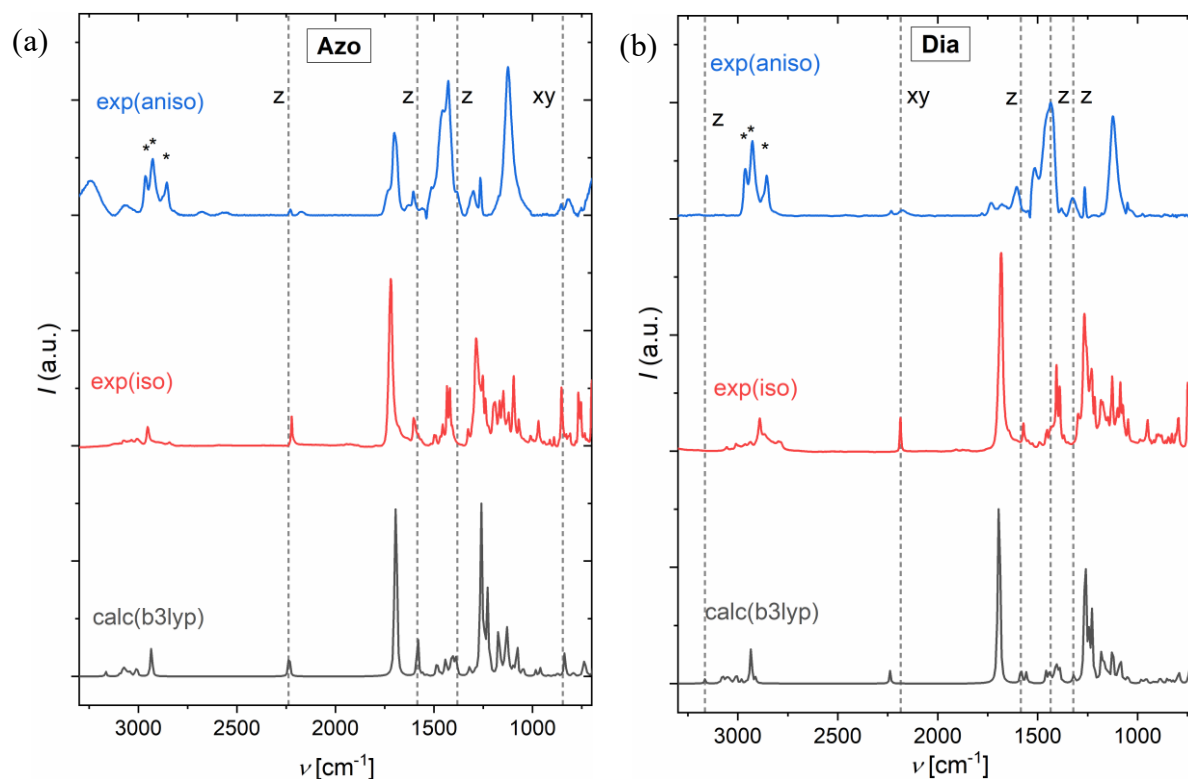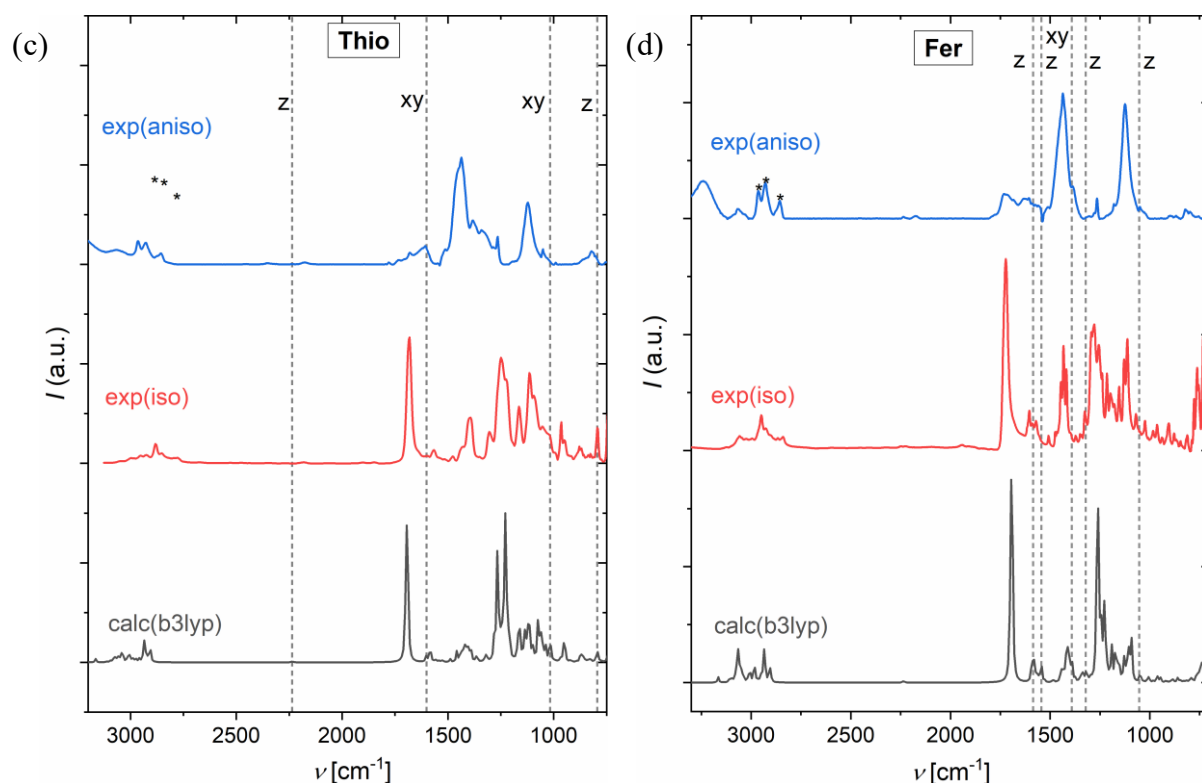
Figure 34. Experimental anisotropic (*blue*) and isotropic (*red*) vs theoretical spectra (*gray*) of (a) **Azo**, (b) **Dia** (c) **Thio** and (d) **Fer** tripodal machines.

5.4.2. Average tilt angle

Besides SAMs characterisation, PM-IRRAS can also be utilised to estimate the average tilt angles of molecules composing the monolayers. Referring to Section 3.1 in Theoretical Basis chapter, it can easily be shown that the useful vibrations should have either a purely parallel *(xy)*, or purely perpendicular (*z*) polarisations of **Azo**, **Dia**, **Fer,** and **Thio** corresponding transition dipole moment vectors.

Table 9 summarises the characteristic vibrations used to determine the tilting. Aromatic breathing stretches of triptycene stands and functional heads were generally used for the *z*-polarised vibrations, as their respective signals were quite easy to deconvolute. Although low in intensity, C≡C stretches were also used as they are easy to identify and do not overlap with other peaks. Finding suitable *xy*-polarised vibrations is a more difficult task, as those are heavily overlapped in regions of larger wavelengths. Assigned suitable peaks were mostly the skeletal stretches and wagging modes of triptycene stands.

For each combination of *z* and *xy* polarised vibrations, a tilt angle was estimated. This yielded three angle estimates for **Azo**, and four for the other systems. With a limited number of

signals with only a single transition moment component and suffering from high signal-to-noise ratio, it was impossible to differentiate between the tilts of upper and lower parts of the molecules. Calculated angels for all experimentally investigated systems can be found in Table 10.

Table 9. Vibrations with transition moment vectors polarised exclusively in z or xy direction, used in experimental determination of tilt angles for **Azo**, **Dia**, **Fer** and **Thio** tripodal systems. I_{iso} / I_{LB} represents the ratio between FT-IR and PM-IRRAS peak intensities.

System	vibration	polarisation	ν [cm^{-1}]	I_{iso} / I_{LB}
Azo	δ(C–H stand)	xy	845–855	0.73
	δ (stand aromatic breathing)	z	1382–1385	0.32
	ν(C–C stand)	z	1582–1587	0.20
	ν(C≡C) + (C≡N)	z	2234–2238	0.24
Dia	δ (stand aromatic breathing)	z	1285-1350	0.10
	δ (**Dia** aromatic breathing)	z	1422-1464	0.12
	ν(C–C stand)	z	1565-1618	0.07
	ν(C≡N)	xy	2151-2216	0.77
	ν(C–H stand)	z	3160-3180	0.25
Fer	δ (**Fer** aromatic breathing)	z	1018-1053	0.21
	δ (stand aromatic breathing)	z	1322-1338	0.18
	w(C–C stand)	xy	1369-1392	1.45
	ν(C–C Fer)	z	1544-1557	0.34
	ν(C–C stand)	z	1584-1595	0.10
Thio	w(C–C Thio)	z	778-805	1.12
	δ (**Thio** aromatic breathing)	xy	1017-1029	0.59
	ν(C–C **Thio**)	xy	1595-1601-	0.71
	ν(C–C)	z	2223-2236	0.31

Table 10. Tilt angles of tripodal machine films

System	**Azo**	**Dia**	**Thio**	**Fer**
Θ_{est}	43.1	26.5	36.7	28.3
	36.5	29.2	34.3	26.5
	39.0	23.2	43.2	34.4
	-	38.9	45.7	20.4
$\langle \Theta \rangle$	34 ± 4	29 ± 8	40 ± 6	27 ± 7

Concerning the accuracy of the described experimental determination, several challenges need to be considered: firstly, finding well-resolved signals with either *z*- or *xy*-polarisation is more challenging because, in contrast to previous work done in Kaleta group, [10,11] tripodal systems bear three carboxylic groups, instead of one. Electron-rich oxygen atoms dissymmetrise most C=C skeletal vibrations, limiting the number of frequencies that can be used in determining the tilt. This made the tilt angle determination particularly challenging in case of **Fer** system, as Feringa's motor, being an overcrowded alkene with several fused aromatic rings, has most of the useful vibrations with the similar energy to the stretches of the triptycene base, making the peaks of those vibrations highly overlapped. Secondly, considering that the theoretically estimated relative energies of interesting peaks are not very reliable, especially due to the possibility of small energy shifts resulting from the surface and packing effects in case of PM-IRRAS measurements, there is a large possibility of integrating the wrong peak. Finally, one has to be aware that the peak intensities are proportional to the cosine of an angle squared. This means that small errors in area under the peaks result in large errors in tilt angle estimates.

5.4.3. Raman spectroscopy

Comparing the off-resonance Raman spectroscopy measurements of switches before and after the irradiation of switch-inducing wavelengths, one can track the progress of the switching photoreaction by looking directly at the chromophore. Figure 35 and **Table 11** show the resulting spectra of **Azo** and **Dia** switches in solution before and after the irradiation.[13] Comparison between calculated and monitored Raman spectra for the investigated systems in powder (calculated in vacuum) can be found in SI (Figure S15, Section S7). The presented results confirm this technique is a promising route for direct measurements of the kinetics of machines operation.
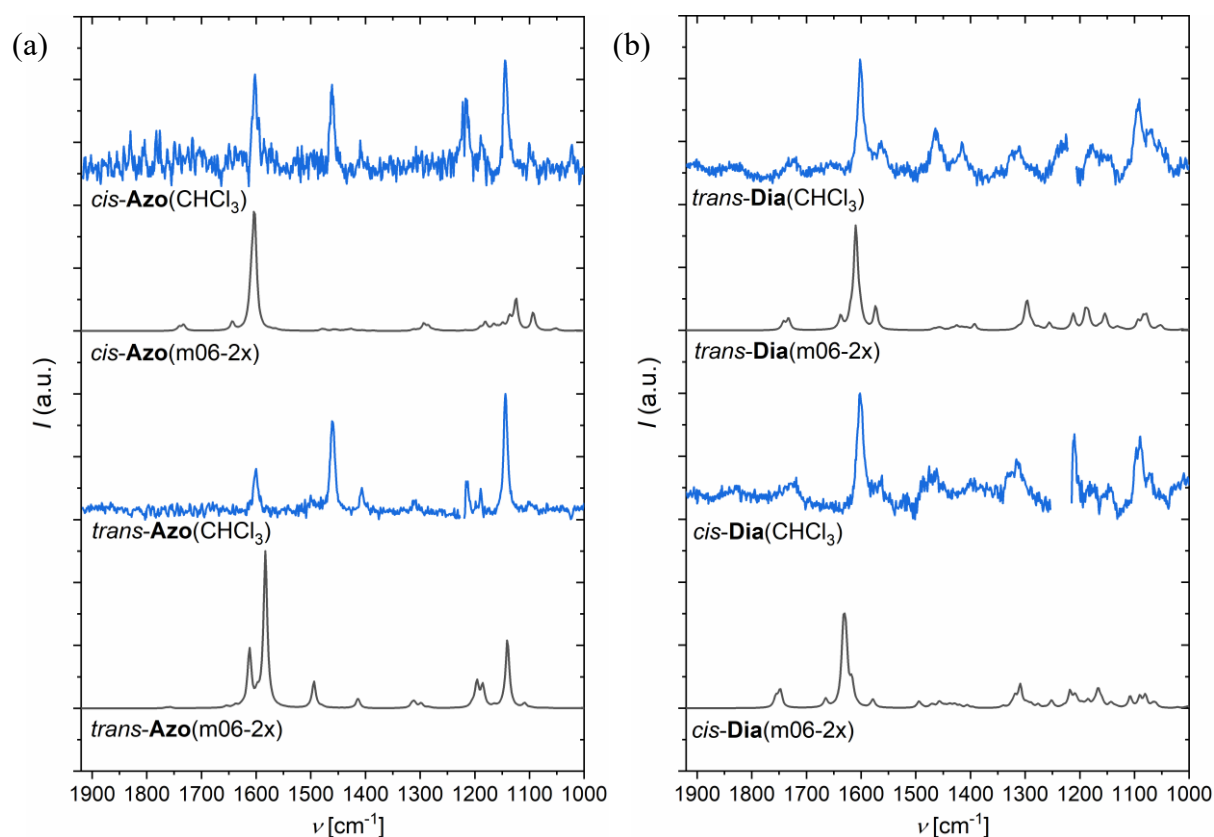
Figure 35. *Off-resonance* Raman spectra of **Azo** (a) and (b) **Dia** system monitored in solution before and after irradiation (blue curves) and their calculated *cis-* and *trans-* analogues (gray curves).

Table 11. Characteristic Raman signals of **Azo** and **Dia** systems.

System	Vibration	v [cm^{-1}]
Azo	σ(**Azo** aromatic scissoring)	1620–1650
	v(N=N)	1570–1580
	ρ(**Azo** aromatic rocking)	1490–1510
	v(C=C, Azo aromatic)	1190–1200
	σ(triptycene scissoring)	1090–1100
Dia	v(C=O, carboxyl)	1780–1790
	v(N=N)	1650
	σ(**Dia** aromatic scissoring)	1620–1640
	ρ(aromatic rocking)	1300–1320

5.5. Ellipsometric thickness and mean molecular area

For synthesised tripodal systems, tilt angle was also determined by comparing the measured monolayer thickness using the variable-angle Stokes ellipsometer[13] with maximal theoretical

molecule lengths in case of perfectly strainght molecules. Average angle was determined by the simple trigonometry, as the *arccos* of the ratio between the two thicknesses. Results are summarised in Table 12.

Naturally, ellipsometric thickness determination suffers from plethora of issues too. The tendency of ellipsometry to overestimate the thickness is a well-known problem.[98] The unreliability of the estimate is a result of substrate roughness. Moreover, using a single empirical value for the index of refraction during the measurement necessarily limits the precision of the thickness estimate. Since the index of refraction changes with the optical density of organic layer, it will depend on the average tilt and on the type of the molecules building the SAMs.

Table 12. Maximal theoretical vs measured ellipsometric thickness and determined average tilt angle of synthesised tripodal systems.[XXII]

System	$d(\text{calc})_{max}$	$d(\text{ellipsometry})$	Θ_{avg}
Azo	22.5	19 ± 2	32 ± 9
Dia	17.8	19 ± 2	6 (lower limit)
Fer	18.5	19 ± 1	3 (lower limit)
Thio	20.8	17 ± 1	35 ± 5

Moreover, mean molecular areas for the synthetically available systems were estimated during the Langmuir-Blodgett deposition procedure. Figure 36 shows the π / APM isotherms, defined in Literature Review (Section 2.6), for the synthesised systems (**Azo**, **Dia**, **Thio** and **Fer**) used for the adsorption of Langmuir films on the gold surface. Table 13 summarises the experimentally estimated APMs of the investigated systems. As neither the ellipsometry nor the LB deposition were performed by the author of this thesis, the measurement details will subsequently be published elsewhere.[13]

[XXII] **Dia** and **Fer** estimates of tilt angle using ellipsometric thickness is not straightforward since, due to geometry of diazocine and motor moieties, the tilting of the molecule can increase the molecular height.[13]
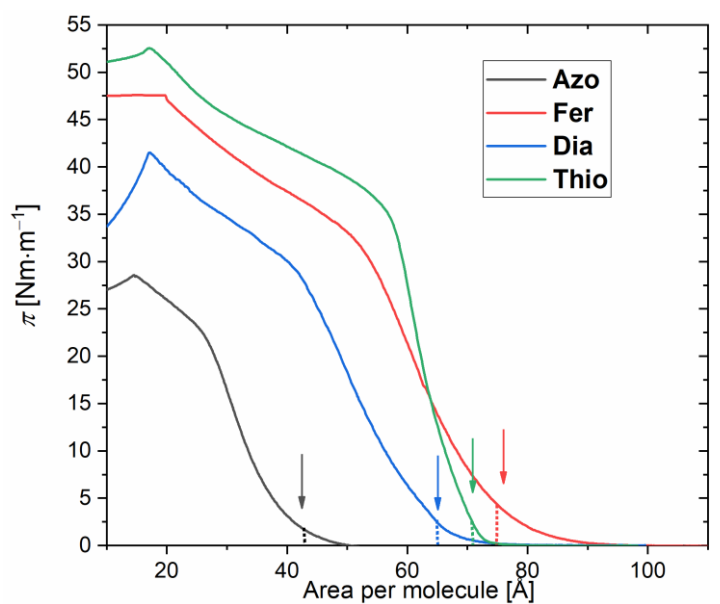
Table 13. Average APMs of investigated systems, as determined by the LB technique

System	APM (Å^2 / molecule)
Azo	40 ± 3
Dia	65 ± 1
Fer	71 ± 3
Thio	75 ± 2

Figure 36. LB isotherms of **Azo** (grey) **Fer** (red), **Dia** (blue) and **Thio** (green) systems. Arrows denote the point on the isotherm with the APMs at which the deposition was performed.[13]

§ 6. DISCUSSION

6.1. Structure and stability of SAMs

The monolayer structure is governed by an interplay between several factors, which can mostly be clustered into: (i) enthalpic (interactions of anchoring groups with the metallic or water subphase and non-covalent interactions between the molecular machines), (ii) entropic contributions (thermal motion of the individual building blocks) and, in case of the water-air interface simulations (iii) the effects of SAM interactions with the water surface (solvation effects, hydrophobic forces between the water and the organic phase, protrusion of the water molecules through the monolayer due to the drop in surface tension).

On the experimental side, both PM-IRRAS data clearly indicates that films are observed on the gold surface. Specifically, the presence of z-polarized carboxylic groups and triple bond stretches in the PM-IRRAS spectra strongly suggest that that tripodal systems form organized SAMs. We can conclude that with fair certainty as modulated reflectivity is independent of the isotropic adsorption of material which ensures that the observed characteristic peaks are not randomly oriented on the conductive surface. Moreover, the measured ellipsometric thickness of synthesised tripodal systems (**Azo**, **Dia**, **Fer** and **Thio**) is exclusively in range between 16 and 21 Å in all cases, suggesting that the films are not 3D assemblies, but monolayers. (**Table 12**, Section 5.5 in Results).

6.2. Tilting and bending of SAMs

Probably the most striking factor influencing the SAM stability is the propensity of some monolayers to tilt and bend more than the others. In this thesis, we were able to estimate the tilt angles of already synthesised **Azo**, **Dia**, **Fer** and **Thio** tripodal systems using the anisotropic IR spectroscopy. A combination of factors, outlined in the Section 5.4, resulted in considerable errors of isotropic/anisotropic signal intensity ratios which limited the accuracy of the average monolayer tilt estimate. However, the estimated angles can be compared to their calculated analogues providing the opportunity to discuss the differences in average tilting measured by two completely different experimental approaches and calculated independently by scanning of the local minima and by molecular dynamics.

Figure 37 and **Table 14** illustrate the change in average tilt angles across four different types of machine monolayers, showing that the overal trend in tilting in the systems is replicated by all four methods. MD estimated tilts are the largest in all cases, while static calculation results follow the same trend as the MD but the calculated angles have considerably smaller values. It is crucial to note that the tilts experimentally determined with the PM-IRRAS are generally somewhere between the static and the dynamic theoretical picture. This is expected since the effect of interactions between gold atoms and anchoring groups stabilise the less tilted conformers, while the effect of the molecular thermal motion, which the static calculations cannot describe, induces the tilting. As was already explained, the ellipsometric measurements overestimate the film thickness which explains how, while the general trends are observed in all four techniques, tilt angles estimated with ellipsometry (**Dia** and **Fer** have particularly low estimates) are much lower.
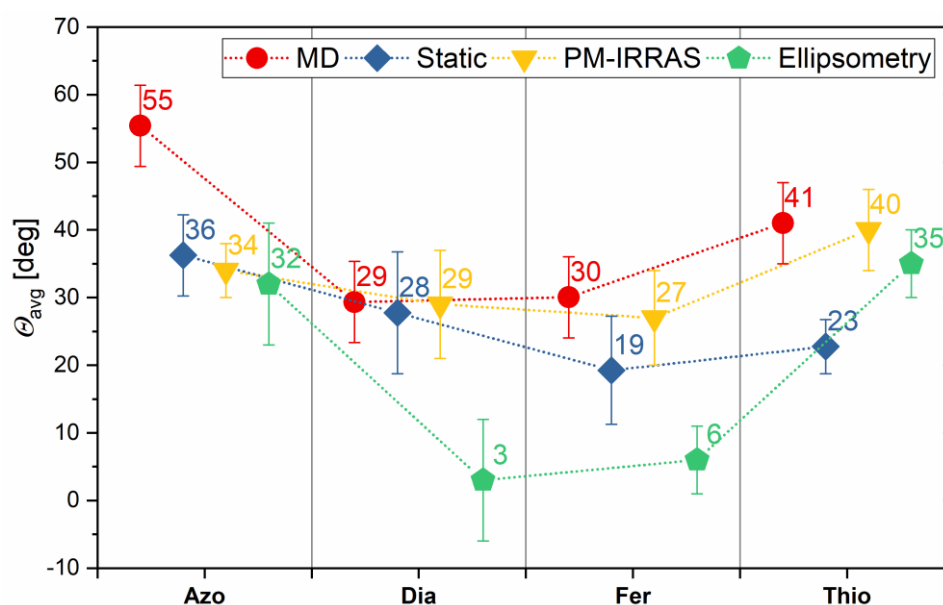
Figure 37. Comparison between calculated (MD – *red*, Static – *blue*) and experimentally determined (PM-IRRAS – *yellow*, ellipsometry – *green*) average tilt angles for Azo, Dia Fer and Thio systems.

Table 14. Comparison between average tilt angles for **Azo**, **Dia**, **Thio** and **Fer** systems, determined with two different experimental (ellipsometry, PM-IRRAS) and two computational approaches (static scanning, PBC molecular dynamics).

System/Technique	Ellipsometry	PM-IRRAS	Static scanning	MD
Azo	32 ± 9	34 ± 4	36 ± 6	55 ± 6

Dia	6 (lower limit)	29 ± 8	28 ± 9	29± 5
Fer	3 (lower limit)	27 ± 7	19 ± 8	34± 4
Thio	35 ± 5	40 ± 6	23 ± 4	41± 5

To further illuminate the difference between the effects of molecular motion and water surface with the static, energetic contributions on the SAM structure, detailed comparison of the computational results is necessary. We have shown that, when calculated in vacuum, molecules do not bend significantly, since the conformations where tilting of the bottom part of the molecule is pronounced is readily realised. Resulting conformers have anchoring groups in the proximity of neighbouring triptycene centres, since no anchor-substrate stabilisation is available. This type of packing is observed in vacuum across most of the investigated systems (Figure 32 in Results) suggesting that the pattern is driven by aromatic interactions. On surfaces, however, anchoring groups interact with the substrate and achieving an equilibrium between non-covalent machine – machine and anchoring group – surface interactions leads to significant bending. This is also prominent on the water-air interface since the hydrophobic effects and presence of H-bond networks of carboxylic acids and surface water molecules are considerable. Nonetheless, one still has to be aware that the major contributing factor to bending and the average tilt in the dynamic picture is likely the presence of thermal motion e.g., entropy. Since the average tilt is also a result of thermal motion, the bend angle as an average over many MD snapshots will have larger values when no negative angles have been defined.[XXIII]

Figure 38 illustrates tilting and bending angles of all investigated systems obtained by static scanning and PBC simulations. Looking at the decoupled top and bottom tilt angles (defined in Figure 14, Methods section 4.2) in all simulated systems, both statically and dynamically, few interesting things can be noted. Firstly, in case of all measured angles (Θ_{bot}, Θ_{top} and B), static and MD results closely mimic each other, but the static picture systematically predicts lower angles. This is crucial because it shows that already discussed reproduction of trends between different methods remains preserved when properties of individual building blocks are

---

[XXIII] For example, average computed angle of inclination of a simple 1D mathematical pendulum, with a defined space of possible angles in $\left[0, \frac{\pi}{2}\right]$ instead of $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, will result in average angle of $\frac{\pi}{4}$ instead of 0. The same reasoning applies here. This is, of course, completely fine as experimental evaluation necessitates absence of negative angles anyway (see definition of average tilt in Section xx in Literature Review).

investigated, making the otherwise unavailable investigation of specific phenomena by the MD method more credible. Secondly, although the MD calculated angles cover a significant range, their distributions are highly symmetric in *all cases* (note how the mean (gray dot in boxes) and median (gray line in boxes) values are almost in perfect alignment). This is important since it indicates that, although simulated systems were small, their size is sufficient for sampling the behaviour of SAMs. At first glance, one could argue that such a close resemblance between static and dynamic calculations may be a result of the optimisation procedure – optimisation of monolayers had first been performed in vacuum and transfered to water surface for further optimisation afterwards. However, we made sure that the optimised films are straight molecules, with tilt angles close to 0°, so all the tilting had to spontaneously occur during the equilibration process which, in our opinion, eliminates this potential concern.
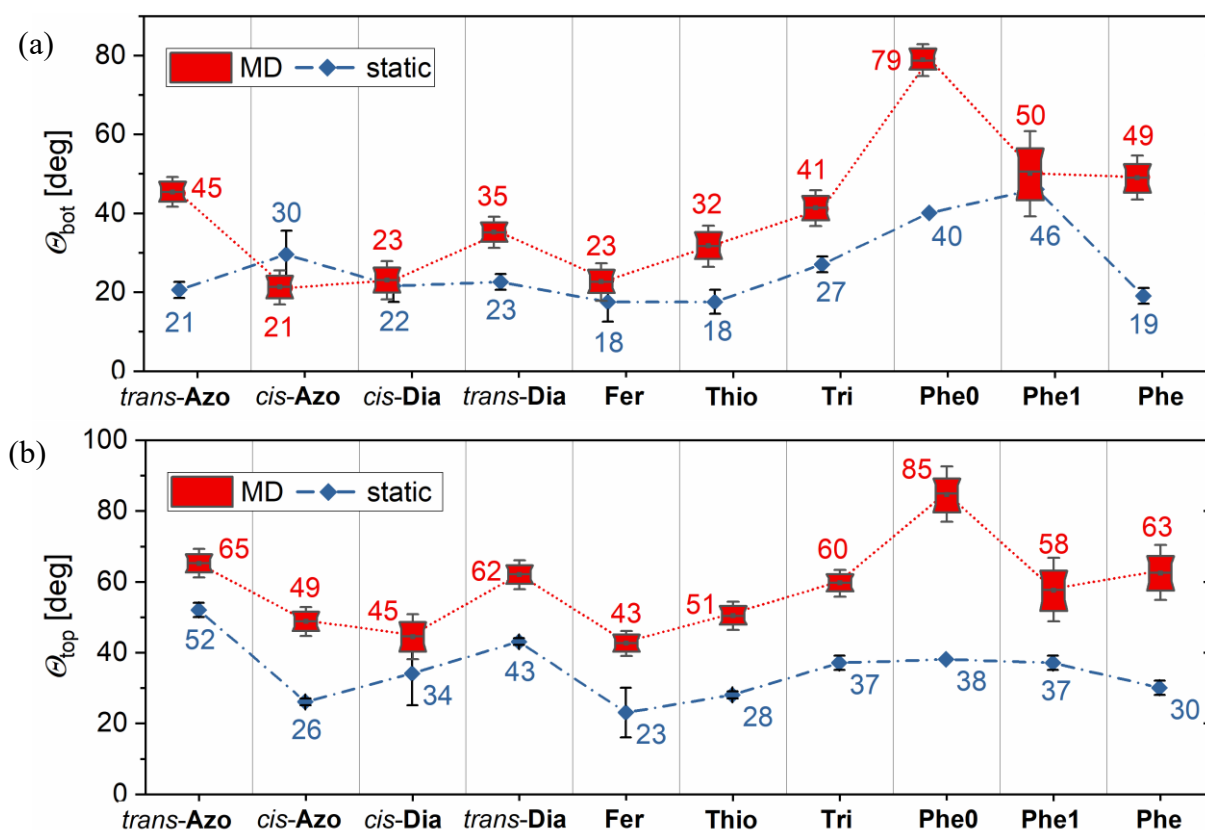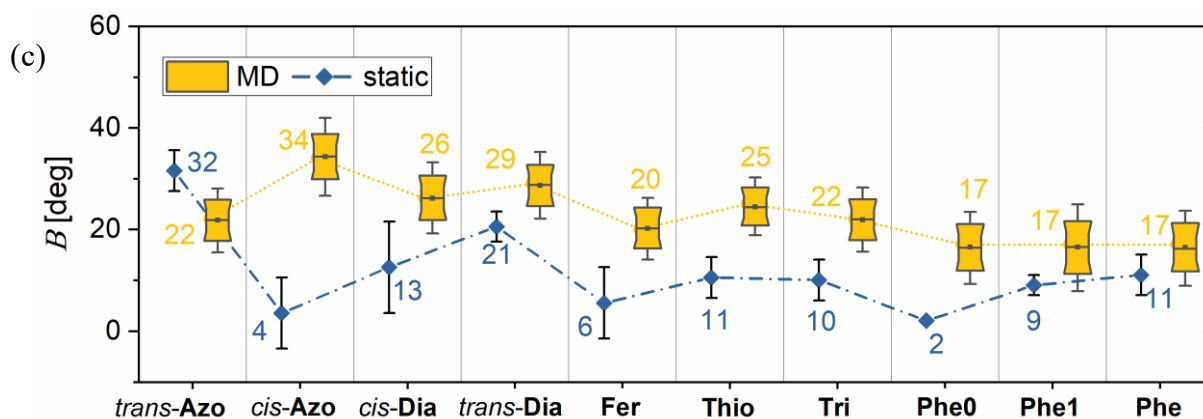
Figure 38. Tilting of the lower part (a), upper part (b) and bending angles (c) summaries of all simulated monolayers with PBC MD (box plots) and with static scanning (diamond scatters). Lower and upper tilting angles and bending angles were defined in Table 2 in Methods, Section 4.2. Numbers above and below the box plots indicate mean values of the calculated angles, points inside the boxes reflect median values, box inflections are 5th, 25th, 75th and 95th percentiles of distribution respectively, and short horizontal lines indicate the position of outliners. Dashed connections between boxes are visual helpers, without any physical meaning.

6.2.1. Why is MD titling of trans-Azo system so overestimated?

The largest discrepancy between the calculations and the experiment are observed in case of *trans*-**Azo** system. This is a long, narrow, and flexible molecule with comparatively small APM (Table 13 in Results). This difference also translates to static vs dynamic calculation, where MD evaluated Θ_{bot} is much larger compared to the static case, but B evaluated statically is much larger compared to MD calculated bending. This is interesting, as it points that, in static picture, **Azo** system readily bends in order to maximise π–π* stacking of functional heads while, in the dynamic picture, pronounced tilting of the triptycene stands ensures the same stacking pattern albeit with less bending. This is surprising and two explanations for this observation are possible. Firstly, the crude static scanning was possibly unable to find the energy minimum corresponding to the conformer with less pronounced bending while in contrast, with the presence of thermal motion, system reached this minimum after the equilibration period. The second possibility is that the available APM during the simulation run was too large for the proper description of the dynamic behaviour of **Azo** system (NVT ensemble had to be simulated due to the limitations of DFTB+ program). This resulted in more available space between neighbouring machines making the tilting easier to achieve. However, since the **Phe** system stays perfectly stable and does not tilt (event the bottom part) as significantly as the **Azo**

molecules, the somewhat confusing behaviour of **Azo** system is more likely to be explained with the former argument: the π–π* stacking of azobenzene units truly is the driving force behind tilting.

6.2.2. What drives molecular bending?

Bending within the molecule is usually destabilising phenomenon, and the system being prone to it means bending probably enables other supramolecular motifs to form stronger stabilising interactions. Since self-assembled monolayers on dynamic substrate are in quasi-liquid state, when molecular machines on average bend their functional heads in the same direction as they tilt, the distance between their functional heads generally gets smaller. This way, bending can be justified by enabling more film-stabilising interactions between neighbouring functional units; *trans*-**Azo**, *trans*-**Dia** and **Tri** are good examples of this behaviour. However, some systems tend to bend in the opposite direction from the tilting of functional heads (**Table 15**). *cis*-**Dia** and *cis*-**Azo** are the most prominent examples of this, which can be explained by their asymmetric functional head geometries. This trend is observed only in MD; the most stable conformer of *cis*-**Dia** found with static scanning bends in the same direction as *trans*-**Azo**, but the bending angle is less marked (Figure 39). *PHE* family also shows the lack of bending directionality, but this is most likely due to significant tilting and interactions of functional heads with water surface.

However, since similar phenomena can be observed in the static picture (**Table 8**, Section 5.3), it can be concluded that the proclivity for bending is in large part a consequence of the monolayers trying to self-organize into packing pattern observed in vacuum despite the presence of a dynamic surface. This confirms that the computational evaluation of monolayers in vacuum remains a valuable tool for the structure / energy relationship investigation in extended systems.

Table 15. Tendency of simulated systems for bending in the opposite direction from film tilt.

System	$\langle\Theta_{bot}+\Theta_{bend}-\Theta_{top}\rangle$
trans-**Azo**	2.1
cis-**Azo**	6.8
cis-**Dia**	13.7
trans-**Dia**	1.8

(a)
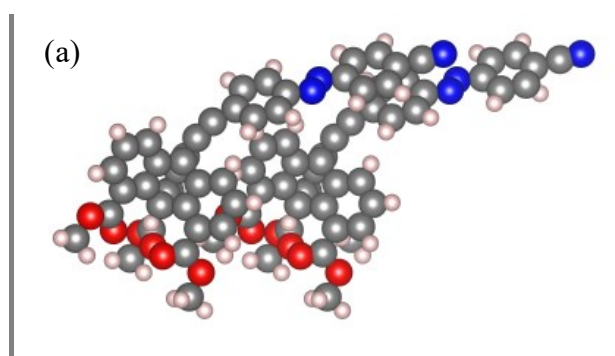
Fer		5.5
Thio		6.0
Tri		3.0
Phe0		10.8
Phe1		9.0
Phe		5.6
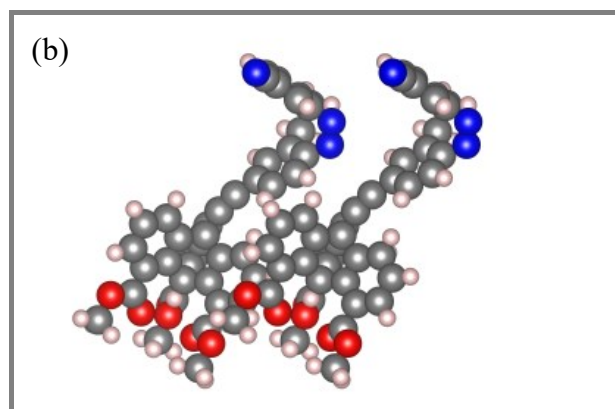
Figure 39. (a) *trans*-**Azo** and (b) *cis*-**Dia** conformations, obtained by scanning of the local minima, with maximum film stabilization energies, indicating the bending direction relative to the overall tilt of the molecule

6.3. Influence of surface on SAM stability.

We can expect that the extent of tilting and bending will be dependent not only on the nature of the adsorbed molecules but also on the crystal lattice of the metal, as was suggested recently.[10,42] This makes decoupling the effects of molecular structure on the packing from the surface effects challenging. However, simulating the systems on water does not suffer from these problems, as the surface is dynamically disordered during the simulation run.

Besides the long-range interactions between aromatic parts of molecular machines, the general stability of monolayers on the water-air interface can further be contributed to the dipole-dipole interactions between carboxylic acids and molecules of water and their proclivity to make H-bond networks. It is worth noting that, although the carboxylic acids were simulated in this thesis, in a real experimental setup, the anchoring groups most likely form complexes with Zn ions.[11,12] In fact, without the presence of ions in the LB troughs, the monolayers on the water surface do not self-assemble successfully but aggregate in a zig-zag pattern.[11] When coordinated with carboxylic groups from several proximal triptycene units, Zn ions serve as an extra locking mechanism which prevents stacking of the organic phase in 3D. For practical reasons, however, pure water was used for as the subphase during the MD simulations. Looking at the results, two things can be observed: (i) systems stay preserved during the simulation time because regular organic monolayers and water do not mix and (ii) interactions between water and anchoring groups clearly stabilise the system. This indicates that finetuning the mean molecular area before the simulation in vacuum probably resulted in SAMs being stuck in the

shallow range of local minima far away from the aggregation point, without the need for extra stabilisation by coordination with Zn. This prevents the aggregation in direction of the surface normal by keeping the molecules at larger distances.

Nonetheless, it is worth noting that the stability of systems in water is going to be highly dependent on where on the LB isotherm the system is being simulated (Figure 36 in Section 5.5). To study the effects of structural motives on APMs of SAMs, such as their tendency to form high density islands on water surface, polymorph interchange kinetics within the layer or the velocity of lateral diffusion of molecules, much larger systems would need to be simulated, on much greater timescales. This was well beyond the scope of this thesis.

It is interesting to note that, even though the structure of all SAMs stays preserved during the production runs, in some cases (**Tri**, **Dia**, **Fer**, compare Figure 18 in Section 5.1.2, and Figure S12 in the SI) there is a significant amount of water present at the hight of triptycene units. The presence of water at the triptycene level was already observed experimentally on similar systems with grazing incidence X-ray diffraction analysis.[12] In the case of *cis*-**Dia** and **Fer** systems, the geometry of their functional heads tends to prevent the molecules from stacking close together, thus allowing the water molecules to occupy the volume between triptycene stands. At the same time, the same configurations prevent water from further diffusing through the film. Nonetheless, both *cis*-**Dia** and **Fer** triptycenes stay at approximately the same height and with similar orientation during the simulation and do not lose energy. This shows that, although certain degree of water protrusion can be expected when triptycene units are further apart, stabilizing interactions between functional heads preserve the film structure.

6.4. Influence of structural motifs on the monolayer structure

6.4.1. Functional heads.

In static picture, the effect of steric hindrance due to bulkiness of functional heads is straightforward. The bulkier the functional head is, the flatter the film interaction energy curve at higher tilts because the volume of the heads limits the amount of interaction between the triptycenes. For this reason, as the angle increases, the film stability starts to increase as well. Comparison between *trans*-**Dia** and *trans*-**Azo** systems is particularly interesting, since both systems have approximately the same shape and aromatic parts prone to $\pi-\pi^*$ stacking at the topmost level of molecule. However, compared to *trans*-**Azo** system, tilting of *trans*-**Dia** is

somewhat lower, which can be attributed to two different factors: (i) *trans*-diazocine group is slightly bulkier and (ii) the presence of N≡N linker of *trans*-diazocine moiety closes an 8-membered ring whose geometry limits the available distance for stacking of the top and the bottom phenyl rings which is so prominent in *trans*-**Azo** case. The difference in energetic and steric effects on the film structure is also indicated by the less defined ADF / RDF correlation plot of *trans*-**Dia** (Figure S4 in the SI).

However, although the dependence of the tilt on the bulkiness of functional heads is intuitive, tilting is not linearly dependant on the volume of functional heads, because the readiness of molecules to pack in a way that maximises the triptycene interactions is also highly dependent on the geometry of the functional head. The lack of symmetry in *cis* isomer of diazocine moiety can, for example, explain the relatively small average tilt angle and large neighbouring distance between functional heads, while thiophene, which is more voluminous, tilts more compared to *cis*-**Dia**.

The averaged minimal neighbouring distances between individual functional heads and their distances from triptycene stands differ due to multiple highly coupled effects related either to packing (bulkiness of functional heads and their geometric complexity) or dynamics (freedom in different modes of motion). The fact that the neighbouring functional head distances are further apart in most systems compared to the neighbouring distance between triptycene stands and functional heads implies the interactions between aromatic parts of functional heads and triptycenes may also play an important role in promoting the overall tilt. Furthermore, while the distances between functional heads tend to differ between the systems, there is much less variability in shapes of their distributions. *d*(top-bot) descriptors have sharp, well defined normal distributions in all of the investigated systems (Figure 40). The triptycenes thus serve as an extra locking agent for molecular conformations with significant tilting of lower parts, which can further be confirmed by looking at the RDF/RDF plots. They indicate whether the frequency of interactions occurring between functional heads surpasses the frequency of interactions between functional heads and triptycene stands (**Table 16**).
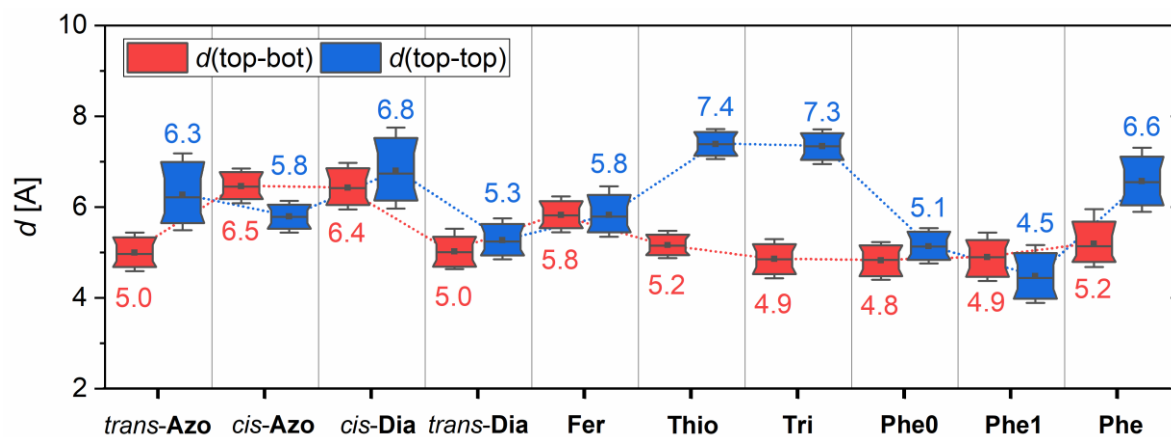
Figure 40. Summary of average minimal neighbouring functional heads (d(top-top)) and triptycene stand distances (d(top-bot)), as was defined in Table 2 in Methods, Section 4.2., for all investigated systems. Numbers above and below the box plots indicate mean values of distance distributions, points inside the boxes are median values, box inflections are 5th, 25th, 75th and 95th percentiles of distribution, and short horizontal lines indicate the outliners. Dashed connections between boxes are visual helpers, without physical meaning.

However, when looking at the Figure 40, one has to bear in mind that the averaged minimal neighbouring distances cannot be directly tied to the experiment, unless X-ray diffraction is possible, since we simulated the NVT ensemble with a small unit cell size.

It has already been discussed how the aromatic systems tend to endure large bending angles due to their ability to stack (*trans*-**Azo**, *cis*-**Azo**, **Phe**), while bulky systems bend out of necessity, due to voluminous functional heads (**Fer**, **Thio**). Compared to 1D distribution functions, tracking the multidimensional correlation plots provides illuminating information on the space of possible conformations that can be sampled[XXIV] and indicate whether the bending is induced by the tendency to stabilize the monolayer with the non-covalent interactions, or to minimize the steric hindrance of bulky structural motives. The lack of symmetric maxima in some correlation graphs can thus be explained as a result of greater degree of freedom in modes of motion of systems lacking extra stabilisation from functional groups while the high complexity of correlation plots explains why a wide range of angles within the same SAM

[XXIV] However, it is worth noting that the general shape of correlation plot will still be highly dependent on the number of interchanging phases that are being sampled during the simulation runs. Therefore, exploring the temporal convergence of every CDF dimension is advisable.

structure is observed – topological complexity of film-building machines necessarily produces highly correlated modes of their individual motion, necessitating an interchange between the mutually exclusive conformations of SAM building molecules. Each individual conformation can, at a certain time, have only a subset of properties that on average define the film structure.

Table 16 summarises this observation, dividing the molecular machines in three general categories: (i) *trans*-switches and **Phe** system with significant amount of bending due to π–π^* stabilisation. They bend to achieve proximity to other functional heads leaving a negative X pattern in correlation plots. (ii) systems bearing voluminous heads which tend to bend less when close to other functional heads and maximise their bending when far apart. from them to This trend is a result of the tendency to minimise the collisions between the heads which arise due to steric hinderance. (iii) Systems with either large degree of rotational freedom or non-symmetric geometrically complex heads have highly dispersed or complex plots, respectively. It is hard to generalise in these cases.

Table 16. Summary of ADF/RDF and RDF/RDF correlation plots for the systems simulated with the PBC molecular dynamics

system	ADF-RDF CDF	RDF-RDF CDF
trans-**Azo**	Strong bending due to π-π^* stacking ($-$ X pattern)	Structural proximity driven by interactions ($-$ X pattern)
cis-**Azo**	Bending due to steric hinderance ($+$ X pattern)	(complex plot)
cis-**Dia**	Complex modes of motion due to functional head geometry (dispersed plot)	Structural proximity driven by interactions ($-$ X pattern)
trans-**Dia**	Strong bending due to π - π^* stacking ($-$ X pattern)	(complex plot)
Fer	Bending due to steric hinderance ($+$ X pattern)	Structural proximity driven by steric effects ($+$ X pattern)
Thio	Bending due to steric hinderance ($+$ X pattern)	Structural proximity driven by steric effects ($+$ X pattern)
Tri	$g(r)_{max} > 7.08$ A	$g(r)_{max} > 7.08$ A
Phe0	High degree of motional freedom (dispersed plot)	High degree of motional freedom (dispersed plot)
Phe1	High degree of motional freedom (dispersed plot)	High degree of motional freedom (dispersed plot)
Phe	Strong bending due to π - π^* stacking ($-$ X pattern)	Structural proximity driven by interactions ($-$ X pattern)

6.4.2. Anchoring groups

Figure 38 shows that the lower parts of tripods are significantly tilted. This is surprising, as our expectation was that the presence of three carboxylic groups will minimize the tilting, resulting in almost perfectly straight lower parts of the investigated molecules. Tripodal triptycene core is identical for all machines, so it can be argued that the effect of interactions between the anchoring groups and the surface is probably similar in all systems, provided the film packing between different types of machines is comparable. However, it is worth noting that the nature of the interactions between carboxylic functions and the (111) gold substrate has not been thoroughly understood yet. Nonetheless, it is likely non-covalent since the LB films get strongly perturbed by the STM tip upon STM characterisation.[13] This limits the applicability of spectroscopy techniques in context of tripodal systems, but it does indicate that the SAMs are not chemisorbed on the gold surface. This observation also explains the relatively large average tilt angles estimated with PM-IRRAS; if three carboxylic groups were covalently bound, tilting of the lower parts of tripodal systems would be minimal, resulting in almost perfectly straight molecules. Further computational investigation of tripods-gold interaction would provide better insight into the matter.

Theoretical stability of **Phe** system is similar both on the water surface and in vacuum. Comparing the interaction energy curves between the systems in *PHE* family reveals, quite surprisingly, that even without the surface to interact with, the **Phe** system is the most stable of three. Stabilising effects of H-bond networks between the anchoring groups can, at least in case of vacuum calculations, be eliminated, as the methylated analogues of tripodal systems were used in the scanning of local minima. This strongly points toward the idea that the drastic difference in stability between **Phe1** and **Phe** systems in vacuum is a result of anchoring groups preventing SAMs from considerable tilting due to steric effects.

This trend also translates well to the water surface, where **Phe** system remains stable during the whole run, although the amount of accessible surface area is considerable, because the phenyl group is very small and allows individual molecules within the film a wide range of motion. The main factor preventing the SAM destruction by water is most likely the difference in polarity between the upper parts of molecular machines and the carboxylic anchors. Contrary to **Phe**, however, **Phe0** and **Phe1** show significant amount of water protrusion through the layer, due to the smaller number of anchoring groups.

Comparing the structures of **Phe0** and **Phe1** systems in vacuum, we can see that their top and bottom tilts are very similar, because the anchoring group on the triptycene's C bridgehead atom does not play any stabilising role without a surface. When placed on the air-water interface, however, the tilting of the **Phe1** is, although still large compared to other systems, much lower compared to **Phe0** since a single carboxylic group readily interacts with water. **Phe0** bears no anchoring groups, so the film inevitably gets destroyed.

Looking both at the energy profile of *PHE* family of systems (Figure 33) and their stability on water (Figure 18), the presence of two minima greatly differing in tilt is obvious. In fact, by comparing the stabilisation energies of their film conformations, the groups are so different, that one can think of them as of two different film polymorphs. As the number of anchoring groups increases, the energy minimum of the polymorph with large tilt becomes shallower, while, in the dynamic picture, the equilibrium state gets shifted towards the polymorph with less tilt. This is perfectly visible from the inspection of the production runs: **Phe0** system is in highly tilted phase all the time, **Phe1** looks unstable, as the two phases are approximately equally favourable, and the systems regularly shifts between the two, while **Phe** system remains stable in less titled phase. Even at higher angles, the anchoring groups are sterically preventing too much tilting to occur, ensuring the preservation of the straight polymorph. This confirms that introducing three carboxylic anchoring groups moves the machine design in the right direction.
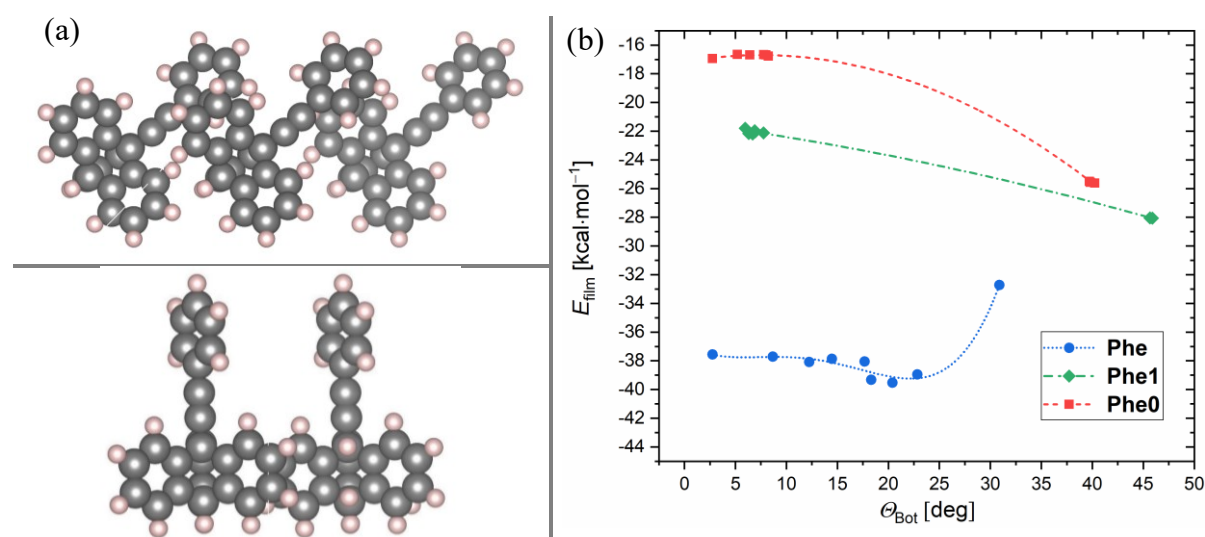
Figure 41. (a) Two minima of **Phe0** system highly differing in tilt angle. C and H atoms are represented as grey and white spheres, respectively. (b) Film interaction energies as a

function of tilting angles for *PHE* family of systems with zero (*red*), one (*green*) and three (*blue*) anchoring groups. Points represent local minima on the potential energy surface.

6.5. Can extra triptycenes stabilise the film structure?

The lower layer of the **Tri** system has some water escaping the bottom, but it immediately gets stuck below the top triptycenes (Figure 42 and Figure 18, in Section 5.1.2). The average tilt angle is approximately the same, both in **Tri** and **Phe** case. While the bottom tilt is somewhat less prominent in case of **Tri** system, the top tilt is larger, as an extra triptycene attached to phenyl bridge of **Tri** system tries to maximise the number of interactions. This is readily observed in the amount of bending, which is more pronounced in **Tri** system, compared to **Phe**. However, all the measured observables of **Tri** have significantly narrower distributions compared to **Phe**, clearly showing the effects of extra stabilisation ($E_{film} \approx -53 \ kcal \cdot mol^{-1}$). This provides further motivation for building already proposed double-decker systems.[24]
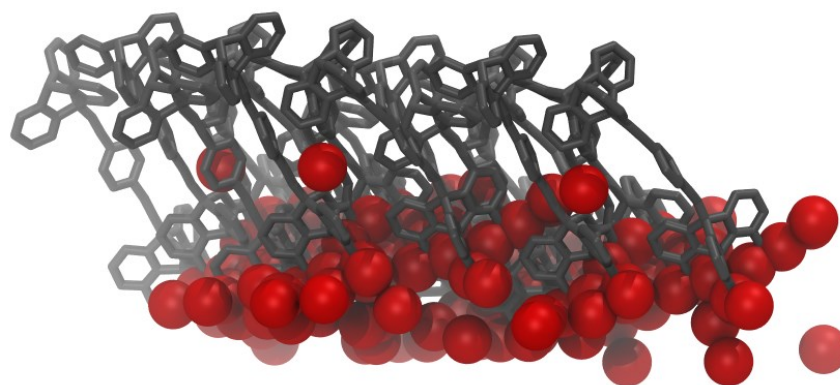
Figure 42. Snapshot of the stable layer of triptycene. Although water protrudes through cavities between triptycene stands, it gets stopped by the array of mounted triptycenes. Tri molecules are represented as grey tubes, and water molecules as red spheres. For clarity, H atoms are not shown.

6.6. Effect of switching

Noting the functional head and triptycene stand distances in cases of **Azo** and **Dia** systems, it is interesting to see how the interaction dynamic changes upon switching. Overall, the large

change in the photoswitch geometry upon switching introduces considerable changes in structural properties of their respective SAMs, which is reflected in differences in their average top and bottom tilt angles, bending properties, film stabilisation energies, complexity in their correlation plots and packing patterns. Furthermore, metastable isomer of **Azo** functional heads prevents stacking maximisation via bending, thus forcing moieties to bend in the opposite direction from their collective tilts. **Dia** system shows the opposite behaviour since *trans* isomer of diazocine is metastable. However, one has to be aware that in physical systems, switching occurs with lower than 100% quantum yields[96] resulting in less pronounced structural changes. Unfortunately, simulations of systems with combined stable and meta stable conformers of photoswitches was, while possible in principle, out of the scope of this thesis.

Raman spectra (Figure 35) clearly show that switching is successful in solution. Reasonably good resolutions of acquired spectra, both for powder and solution samples the cases of all synthesised motors, interpreted with the help of performed calculations are a step forward in measuring the switching kinetics of SAMs on a gold surface using SERS. Further possibility of using SERS to measure switching of trypticene based machines on a gold surface will be explored in the future.

6.7. Are molecular machines moving collectively?

In Theoretical basis section, we mentioned the pitfalls of calculating the correlation between different timeseries. The major problem in cross-correlation evaluation between non-stationary processes, as was explained on the example of random walkers, is their non-ergodicity. However, modes of motion investigated in our case are likely both stationary and ergodic, and we have indicated this with a careful trend analysis and a successful estimation of the converged averages. Moreover, the within-series dependence was removed with the block-averaging technique, ensuring the sporadic correlation does not emerge due to small perturbations in consecutive measurements.

Generally, local modes of motion between the machines in stable layers are not correlated once the system, with its collective tilt and functional heads position, is aligned in a single direction. This is well captured with Kendall correlation matrices; the highest correlation coefficient is around 0.26 for **Fer** system (Table 6 in Section 5.1.6), which is reasonable due to an overcrowded alkene moiety limiting the molecular movement. However, during the equilibration period, correlation of motion between machines is extremely pronounced. This

suggests that, although the communication between neighbouring machines is prominent, and molecules are packed closely enough to move collectively, the motion produced by thermal fluctuations is not strong enough to destabilise the regularity of energetically favourable film structure. This is further confirmed by noting the correlation between the motion of metastable films (Table S8 −Table S10). In these systems, the average observables are hard to evaluate because the films tend to oscillate between two or more phases during the simulation. However, the modes of motion in these cases are still ergodic, as the average would not necessarily change indefinitely with $\tau \to \infty$. It is crucial to note that the degree of correlation in these cases is much more pronounced and includes several pairs of machines, usually the ones in closest contact due to tilting. This provides insight into how changing between different polymorphs can be induced due to collective rearrangements of molecules hopping between several shallow minima. This is also observed in static calculations in case of topologically more complex systems, for example **Dia**, **Thio** and **Fer** systems. Note how there are different conformations with wide range of tilt angles that are almost degenerate in energy ($E_{film} < 1$ kcal mol^{-1}).

Finally, it is worth mentioning that, while in cases of bend and tilt angles estimation, calculation of cross-correlation is plausible, there are cases when this technique would hardly be justified. For example, calculating the motion dependence of individual machines in the lateral diffusion process cannot be described by a single metric, since the process does not converge in finite amount of time. Here, more sophisticated techniques should be performed.[64]

6.8. Phase transitions and polymorphism in 2D films

Proper description of interphase conversion processes is a challenging problem in computational chemistry and material science. Phase interchange can be a very slow process (up to hundreds of ns) and catching the dynamics of a change can be hard to achieve even with highly parallelised force fields.[98] Futhermore, structure properties acquired by MD simulations are going to be highly dependent on starting conditions.

In this work, our main focus was on the description of static properties of the experimentally observed polymorph modelled by antecedent rigorous geometry relaxation. The observables were shown to be converged on the investigated time scale and the analysis was straightforward, while it still captured the general features of the investigated films.

6.8.1. Why do different layers of same system differ in stability?

It is obvious that, although part of the same simulation run, one layer is usually more prone to phase transitions (structural changes of the film promote water / triptycene stand mixing) compared to the other layer of the same system. This behaviour is a result of the difference in starting configurations – during the system optimization process, the lower- and upper-layer finish in different local minima. This means that there is a high probability that the difference in starting configurations between the layers will result in a different sampling of the phase space, yielding slightly different trajectories. However, although capturing different phases of SAMs was successful in metastable layers, our approach cannot provide any insight into the kinetics or rations of different film polymorphs present in the physical systems so focusing on stable layers with well converged properties was a rational decision.

6.9.   On choice of MD methods

6.9.1. Validity of chosen method.

When choosing the method, one of the main concerns was the imposed limit size. Four molecular machines make a questionably small system, and it does not seem plausible that ergodic hypothesis is being fulfilled here. Moreover, the computational cost dictates a relatively short timeframe, less than a nanosecond, and it is known that structural changes of self-assembled extended systems are regularly orders of magnitude slower processes.[98] Furthermore, the shallow minima between different tilts in most systems provide a wide range of angles and, with only four individual machines in the unit cell, one does not have a system large enough to confirm which machines have the conformation that best represents the film structure. Considering these issues, it is somewhat to our surprise that, by just computing the average observable between the four well converged machines, the method works so well. This is clearly shown both by reflecting the trends observed with other methods and overestimating the calculated values by the amount which was expected due to the interaction with water.

6.9.2. Was going with the DFT approach a fair trade off?

On one hand, by deciding to go with a tight binding approach, we are willingly sacrificing the possibility of simulating numerous SAM building machines at large timescales and thus catching processes like interconversion between different phases or lateral motion of SAM

islands. However, accurate manual parametrisation of topologically complex systems, such as triptycene based machines is not straightforward, and fine-tuning of the geometrical parameters for the whole family of structurally different molecular machines investigated in this thesis to account for both the flexibility of the triple bond bridges and the effects of dispersion interactions between different aromatic units would be a very challenging task. Changing the structure of molecular machines would necessitate repeating the fitting procedure. In contrast, once the methodological framework was established, our approach works out-of-the-box. This can prove useful to groups studying the films and interfaces with stronger experimental component to their work in the future. Moreover, although this was beyond the scope of this thesis, our approach enables simulation of processes that require breaking and reattachment of covalent bonds, such as the switching processes within SAMs or the formation of complexes with different ions on the water/air interface. Dynamics of these phenomena is well worth exploring and is now feasible on the meso-scale level. Finally, we hope that this thesis demonstrates that our method performs surprisingly well within the imposed limitations of the chosen model, and that the provided experimental evidence confirms our claim.

6.9.3. How reliable is Grimme's force field?

The unsurpassable challenge with the GFN-FF method involved the unavoidable edge effects - water protrusions throughout the layer at the interface between system edges and custom assigned external potential. However, in few systems in which destruction of layers was delayed, some useful information could still be extracted: Table 7 in Results clearly shows that the agreement in average tilting is reasonable between the two MD methods.

It is worth mentioning that, just recently, the PBC version of GFN-FF was developed.[59] Fairly good agreement in case of **Fer** suggests GFN-FF would be a feasible alternative to tight binding DFT in modelling these types of systems and can serve as a valuable tool in determination of SAM structure. Unfortunately, the paper was published after the final steps in writing this thesis had already been made, and further exploration of the strengths and drawbacks of this approach was, unfortunately, impossible.

§ 7. CONCLUSION

In the presented study, structural properties of several molecular machines based on triptycene were investigated by extensive computational and experimental investigation. Main goal was to determine how different building blocks of molecular machines influence the average tilting and bending of the molecules in the SAMs.

Two meso-scale MD approaches of simulating SAMs of molecular machines on the water-air interface on a nanosecond scale were performed. Lack of the periodic boundary conditions implementation in GFN-FF approach proved fatal since water protrusions at the edges of simulated system led to inevitable destruction of the layers. On the other hand, PBC molecular dynamics approach based on DFTB formalism generated stable nanosecond trajectories with well converged static observables, such as tilting of different parts of molecular machines, bending, and neighbouring distances of different structural motifs.

Overall, all stable films on the air-water interface collectively tilt in a single direction during the production run. The exception here is **Phe0** whose lack of anchoring groups led to the destruction of the film structure. Moreover, thermal motion of individual machines is generally not cross-correlated which indicates that the already proven significant degree of communication between the neighbouring machines does not influence the overall film stability.

The lower parts of the molecule on average tilt between 20 and 45 degrees, with a varying distance between functional units and their neighbours (5.3–7.8 Å between other functional heads and $5.0 - 6.5$ Å between neighbouring triptycenes). Functional mounts bending generally oscillates around 20°. The exceptions are **Phe0** and **Phe1** systems whose lack of stabilisation from extra anchoring groups results in high tilting (65–75°) and lower bending (16–17°) angles.

However, we have shown that asymmetric collective tilting of SAMs coupled with geometric complexity of their building blocks regularly produces highly correlated modes of their individual motion. Individual conformations thus describe only a subset of properties that on average define the film structure. This indicates the dangers of making conclusions based on unidimensional analysis of parameters. Properties within individual machines are highly correlated and their correlation is more pronounced in bulky (**Fer**, **Thio**, **Tri**) and π-π* stabilised (**Azo**) systems.

By extensive scanning of local minima, PES of all the machines spanning a wide range of tilting angles was investigated statically using the vdw DFT. Sampled structural motifs pair well with the results of molecular dynamics, providing deeper insight into energetic contribution differences. Changes in tilting and bending across different systems closely follow the molecular dynamics trend, but the values are smaller due to lack of entropic contributions and the effects of surface.

Finally, calculated average tilt angles of **Azo**, **Dia**, **Fer** and **Thio** were compared with the PM-IRRAS and ellipsometry measurements of the same systems adsorbed on (111) gold surface. Determined average angles reflect computationally determined trends well. PM-IRRAS results are between MD and static calculations due to gold surface – carboxylic groups interactions contrasted against a more dynamic water surface. Angles calculated based on the ellipsometric thickness measurements follow the trend observed by other techniques but are lower due to the known problem of thickness overestimation.

§ 8. LIST OF ABBREVIATIONS AND SYMBOLS

ADF	Angular distribution function
AFM	Atomic force microscopy
AIMD	*ab initio* molecular dynamics
APM	Average area per molecule
CDF	Combined distribution function
DFTB	Tight-binding density functional theory
ExL	Extended Lagrangian approach
FF	Force field
FT-IR	Fourier transfrom infrared spectroscopy
GIXD	Grazing incidence X-ray diffraction
LB	Langmuir Blodgett films
MK	Mann Kendall statistics
PBC	Periodic boundary conditions
PM-IRRAS	Polarization modulation infrared reflection adsorption spectroscopy
RDF	Radial distribution function
SERS	Surface enhanced Raman spectroscopy
STM	Scanning tunneling microscopy
TS	Tkatchenko–Scheffler dispersion correction
XPS	X-ray photoelectron spectroscopy

§ 9. REFERENCES

1. C. P. Collier, J. O. Jeppesen, Y. Luo, J. Perkins, E. W. Wong, J. R. Heath, J. F. Stoddart, *J. Am. Chem. Soc.* **123** (2001) 12632–12641.

2. V. Balzani, A. Credi, M. Venturi (Eds.), *Molecular Devices and Machines: A Journey into the Nanoworld*, Wiley-VCH Verlag, 2006, pages 267–277.

3. M. Suda, Y. Thathong, V. Promarak, H. Kojima, M. Nakamura, T. Shiraogawa, M. Ehara, H. M. Yamamoto, *Nat. Commun.* **10** (2019) 2455.

4. S. Nachimuthu, L. T. Wu, J. Kaleta, H. Y. Yu, P. Wu, J. C. Jiang, *Mater. Chem. Phys.* **277** (2022) 125563.

5. S. Kumar, M Merelli, W. Danowski, P. Rudolf, B. L. Feringa, R. C. Chiechi, *Adv. Mater.* **31** (2019) 1807831.

6. J. E. Green, J. W. Choi, A. Boukai, Y. Bunimovich, E. Johnston-Halperin, E. DeIonno, Y. Luo, B. A. Sheriff, K. Xu, Y .S. Shin, H. R. Tseng, J. F. Stoddart J. R. Heath, *Nature* **445** (2007) 414–417.

7. G. Alachouzos, A. M. Schulte, A. Mondal W Szymanski, B. Feringa, *Angew. Chem. Int. Ed.* **61** (2022) e202201308.

8. J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, G. M. Whitesides, *Chem. Rev.* **105** (2005) 1103–1170.

9. J. A. A. W. Elemans, S. Lei, S. De Feyter, *Angew. Chem. Int. Ed.* **48** (2009) 7298–7332.

10. I. Rončević, E Kaletová, K. Varga, I. Císařová, Z Bastl, J. Jiang, J. Kaleta, *J. Phys. Chem. C* **126** (2022) 7193–7207.

11. J. Kaleta, E. Kaletova, I. Cisarova, S. J. Teat, J. Michl, *J. Org. Chem.* **80** (2015) 10134–10150.

12. J. Kaleta, J. Wen, T. F. Magnera, P. I. Dron, C. Zhu J. Michl, *PNAS.* **115** (2018) 9373–9378.

13. G. Bastien, C. Santos Hurtado, I. Rončević, L. Severa, J. Štoček, M. Dračínský, Z. Bastl, I. Císařová, J. Kaleta, *In preparation*.

14. N. Koumura, E. M. Geertsema, A. Meetsma, B. L. Feringa, *J. Am. Chem. Soc.* **122** (2000) 12005–12006.

15. J. Michl in V. Balzani, A. Credi, M. Venturi (eds), *Topics in Current Chemistry*, Vol. 354, Springer, Heidelberg, 2014, pages 163–213

16. C. Joachim, J.K. Gimzewski in J. P. Sauvage (ed), *Structure and bonding*, Vol 99, Springer Verlag Berlin 2001, pages 1–18.

17. P. D. Boyer, *Annu. Rev. Biochem.* **66** (1997) 717–749

18. W. Kühlbrandt, *Annu. Rev. Biochem.* **88** (2019) 515–549.

19. D. L. Nelson, M. M. Cox, *Lehninger Principles of Biochemistry*, Springer Nature, 2017, pages 380–389.

20. I. Aprahamian *ACS Central Science* **6** (2020) 347–358.

21. V. Richards, *Nat. Rev. Chem.* **1** (2017) 0018.

22. D. Loncaric, C. Santos Hurtado; M. Dracinsky; I. Roncevic; E. Masson; J. Kaleta. *In Preparation*.

23. J. Kaleta, G. Bastien, J. Wen, M. Dračínský, E. Tortorici, I. Císařová, P. D. Beale, C. T. Rogers, J. Michl, *J. Org. Chem.* **84** (2019) 8449–8467.

24. L. Severa, C. Santos Hurtado, E. Kaletová, I. Rončević, M. Mašát, J. Štoček, M. Dračínský, I. Císařová, Z. Bastl, J. Kaleta, *In Preparation*.

25. C. Santos Hurtado, G. Bastien, M. Mašát, J. R. Štoček, M. Dračínský, I. Rončević, I. Císařová, C. T. Rogers, J. Kaleta, *J. Am. Chem. Soc.* **140** (2020) 9337–9351.

26. J. Kaleta in Z.L. Pianowski (Ed.), *Molecular Photoswitches. Chemistry, Properties, and Applications*, John Wiley & Sons, 2022, pages 711–733.

27. N. Koumura, R. W. Zijlstra, R. A. van Delden, N. Harada, B. L. Feringa, *Nature* **401** (1999) 152–155.

28. J. A. Bravo, F. M. Raymo, J. F. Stoddart, A. J. P. White, D. J. Williams, *Eur. J. Org. Chem.* **11** (1998) 2565–2571.

29. P.R. Ashton, C. L. Brown, E. J. T. Chrystal, T. T. Goodnow, A. E. Kaifer, K. P. Parry, D. Philp, A. M. Z. Slawin, N. Spencer, J. F. Stoddart, D. J. Williams, *J. Chem. Soc. Chem. Commun.* **9** (1991) 634–639.

30. B. L. Feringa in V. Balzani, A. Credi, M. Venturi (eds), *Topics in Current Chemistry*, Vol. 354, Springer, Heidelberg, 2014, pages 139–163.

31. C. Lemouchi, K. Iliopoulos, L. Zorina, S. Simonov, P. Wzietek, T. Cauchy, A. Rodríguez-Fortea, E. Canadell, J. Kaleta, J. Michl, D. Gindre, M Chrysos, P. Batail, *J. Am. Chem. Soc.* **135** (2013) 9366–9376.

32. J. Lin Zhang, J. Qiang Zhong, J. Dan Lin, W. Ping Hu, K. Wu, G. S. Qin Xu, A. T. Wee, W. Chen, W. *Chem. Soc. Rev.* **44** (2015), 2998–3022.

33. E. B. Winston, P. J. Lowell, J. Vacek, J. Chocholoušová, J. Michl, J. C. Price, *Phys. Chem. Chem. Phys.* **10** (2008), 5188–5191.

34. B. Rodríguez-Molina, S. Pérez-Estrada, M. A. Garcia-Garibay, *J. Am. Chem. Soc.* **135** (2013), 10388–10395.

35. C. Lemouchi, C. S. Vogelsberg, L. Zorina, S. Simonov, P. Batail, S. Brown, M. A. Garcia-Garibay, *J. Am. Chem. Soc.* **133** (2011) 6371–6379.

36. S. Zhang in K. H. J. Buschov (Ed.), *Encyclopedia of Materials: Science & Technology,* Elsevier Science, Oxford, UK, 2001, pages 822–834.

37. S. Zhang, *Nat. Biotechnol.,* **21** (2003) 1171–1178.

38. J. Kaleta, P. I. Dron, K. Zhao, Y. Shen, I. Císařová, C. T. Rogers, J. Michl, *J. Org. Chem.,* **80** (2015) 6173–6192.

39. J. Kaleta, J. Chen, G. Bastien, M. Dračínský, M. Mašát, C. T. Rogers, B. L. Feringa, J. Michl, *J. Am. Chem. Soc.,* **139** (2017) 10486–10498.

40. J. H. Chong, M. J. MacLachlan, *J. Org. Chem.,* **72** (2007) 8683–8690.

41. J. Liu, M Kind, B. Schüpbach, D. Käfer, S. Winkler, W. Zhang, A. Terfort, C. Wöll, *J. Nanotechnol.,* **8** (2017) 892–905.

42. S. Das, G. Nascimbeni, R. O. de la Morena, F. Ishiwari, Y. Shoji, T. Fukushima, M Buck, E. Zojer, M. Zharnikov, *ACS Nano* **15** (2021) 11168–11179.

43. K. Bezdekova, G. Bastien, I. Roncevic, J. Kaleta, *In preparation*.

44. S. A. Hussain, B. Dey, D. Bhattacharjee, N. Mehta, *Heliyon,* **4** (2018) e01038.

45. M. C. Petty, *Thin Solid Films*, **1** (1992) 417–426.

46. P. J. Hasnip, K. Refson, M. J. Probert, J. R. Yates, S. J. Clark, C. J. Pickard *Phil. Trans.R. Soc.,* **372** (2014) e20130270.

47. P. Lazar, T. Bučko, J. Hafner, *J. Phys. Rev. B*, **92** (2015) 224302–224308.

48. T. Bučko F. Šimko, *J. Chem. Phys.,* **144** (2016) e064502.

49. J. A. Harrison, J. D. Schall, S. Maskey, P. T. Mikulski, M. T. Knippenberg, B. H. Morrow, *Appl. Phys. Rev.*, **5**, (2018) e031104.

50. P. Makkar, N. N. Ghosh, *RSC Adv.,* **11** (2021) 27897–27924.

51. F. Spiegelman, N. Tarrat, J. Cuny, L. Dontot, E. Posenitskiy, C. Martí, A. Simon, M. Rapacioli, *Adv. Phys. X*, **5** (2020), 17102–17168.

52. B. Hourahine, B. Aradi, V. Blum, F. Bonafé, A. Buccheri, C. Camacho, C. Cevallos, M. Y. Deshaye, T. Dumitrică, A. Dominguez, et al., *J. Chem. Phys.,* **152** (2020) 124101–124120.

53. S. Spicher, S. Grimme, *Angew. Chem. Int. Ed.,* **2020**, *59*, 15665–15671.

54. P. Koskinen, V. Mäkinen, *Comput. Mater. Sci.,* **49** (2009) 237–253.

55. S. Malola, H. Häkkinen, P. Koskinen, *Appl. Phys. Lett.,* **94** (2009) e043106.

56. P. Koskinen, H. Häkkinen, G. Seifert, S. Sanna, T. Frauenheim, M. Moseler, *New J. Phys*., **8** (2006) 043039

57. M. Gaus, A. Goez, M. Elstner, *J. Chem. Theory Comput.,* **9** (2013) 338–354.

58. E. Bannwarth, S. Caldeweyher, A. Ehlert, P. Hansen, P. J. Pracht, S. Seibert, S. Spicher, S. Grimme, *WIREs Comput. Mol. Sci.* **11** (2020) 1493–1502.

59. J. D. Gale, L. M. LeBlanc, P. R. Spackman, A. Silvestri, P. Raiteri *J. Chem. Theory Comput*., **17** (2021) 7827–7849.

60. A. Tkatchenko, M. Scheffler, *Phys. Rev. Lett.* **102**, (2009) 073005.

61. E. van den Heuvel, Z. Zhan, *Am. Stat.* **76** (2022) 44–52.

62. H. A. Simon, *JASA* **49** (1954) 467–479.

63. https://github.com/SuncedDjever/DYNAMICS-OF-SELF-ASSEMBLED-MONOLAYERS-OF-MOLECULAR-MACHINES.git

64. J. D. Hamilton, *Time Series Analysis*. Princeton University Press, Princeton, 1995, pages 571–630.

65. H. Flyvbjerg, H. G. Petersen, *J. Chem. Phys.* **91** (1989) 461–466.

66. M. E. J. Newman, G. T. Barkema. *Monte Carlo Methods in Statistical Physics*, Oxford University Press, 1999, pages 210–258.

67. G. U. Yule, *J.Roy.Stat.Soc.*, **89** (1926) 1–63.

68. https://online.stat.psu.edu/stat509/lesson/18/18.3 (Accessed 20 April 2022)

69. S. Yue, C. Wang, *Water Resources Management* **18** (2004) 201–218.

70. C. Croux, C. Dehon, *Stat. Methods Appl.* **19** (2010) 497–515.

71. W. R. A. Knight, *J. Am. Stat. Assoc.* **61** (1966) 436–439.

72. K. F. Wong, T. Selzer, S. J. Benkovic, S. Hammes-Schiffer, 2005, *PNAS*, **102** 6807–6812.

73. S. K. Schiferl, D. C. Wallace, *J. Chem. Phys.* **83** (1985) 5203–5209.

74. H. B. Mann, *Econometrica* **13** (1945) 163–171.

75. M. G. Kendall, *Rank Correlation Methods*, Charles Griffin, London. 1975.

76. M. Hussain, I. Mahmud *J. Open Source Softw.*, **4** (2019) 1556–1558.

77. G. M. Amdahl, *Validity of the single processor approach to achieving large scale computing capabilities*, Spring joint computer conference (AFIPS '67 (Spring)). New York, 1967, pages 483–485.

78. https://github.com/dftbplus/recipes/blob/main/docs/parallel/amdahl.rst (Accessed on 12 June 2022).

79. B. Aradi, A. M. N. Niklasson, T. Frauenheim *J. Chem. Theory Comput.* **11** (2015) 3357–3363.

80. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, E. Lindahl, *SoftwareX*, **1–2** (2015) 19–25.

81. M. Kubillus, T. Kubař, M. Gaus, J. Řezáč, M. Elstner, *J. Chem. Theory Comput.* **11** (2015) 332–342.

82. J. J. M. Cuppen, *Numerische Mathematik.*, **36** (1981) 177–195.

83. E. Anderson, Z. Bai S. C. Bischof, I. Blackford, J. Demmel, J. J. Dongarra Du Croz, A. Greenbaum, et al. *LAPACK Users's Guide* (3rd Ed.), Society for Industrial and Applied Mathematics, Philadelphia, PA 1999.

84. M. Brehm, B. Kirchner *J. Chem. Inf. Model.* **51** (2011) 2007–2023.

85. M. Brehm, M. Thomas, S. Gehrke, B. Kirchner, *J. Chem. Phys.* **152** (2020) 164105.

86. M. Brehm, H. Weber, A. S. Pensado, A. Stark, B. Kirchner, *Phys. Chem. Chem. Phys.*, **14** (2012) 5030–5044.

87. M. Brehm, H. Weber, A. S. Pensado, A. Stark, B. Kirchner, *Z. Phys. Chem.*, **227** (2013) 177–204.

88. K. Lee, E. D. Murray, L. Kong, B. I. Lundqvist, D. C. Langreth, *Phys. Rev. B* **82** (2010) 81101–81106.

89. J. Klimeš, D. R. Bowler, A. Michaelides, *Phys. Rev. B* **83** (2011) 195131.

90. G. Kresse, J. Furthmüller. *Phys. Rev. B* **54** (1996) 11169–11186.

91. G. Kresse, J. Hafner. *Phys. Rev. B* **49** (1994) 14251–14269.

92. G. Kresse, J. Furthmüller, *Comput. Mat. Sci.* **6** (1996) 15–50.

93. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji et al., *Gaussian 16* (Revision A.03) Gaussian, Inc., Wallingford CT.

94. M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett*. **92** (2004) 246401−246405.

95. H. Fujiwara in H. Fujiwara, R. Collins (Eds.) *Spectroscopic Ellipsometry for Photovoltaics*, Vol 212 Springer: Cham, 2018, pages 155−172.

96. V. Ladányi, P. Dvořák, J. Al Anshori, L. Vetráková, J. Wirz, D. Heger, *Photochem. Photobiol. Sci.*, **16** (2017) 1757−1761.

97. Phillips, P.C.B, *J. Econom.*, **33** (1986), 311−340.

98. J. Matti, H. Henrik, M. Luca, J. Iae-Hyung, M. S. Markus, J. Jae-Hyung, M. Hector, M. Ralf, I. Vattulainen, *Faraday Discuss.*, **161** (2013) 397−417.

§ 10. SUPPLEMENTARY INFORMATION

S1. Structural properties of SAMs: *cis*-Dia system

Lower part of *cis*-**Dia** SAMs on average tilts at 23°. The discrepancy between the sum of average Θ_{bot} and B angles (49.3°) and the average functional head tilt (35.6°) is very pronounced, indicating that, contrary to the case of **Azo** system, diazocine head readily tilts in the opposite direction from the film tilting. The overlap between neighbouring distances between top/top and top/bot centres is also significant and slightly varies in time, suggesting either a similar strength between the stands and functional head interactions or a wide variety in different modes of functional head motion due to complex, unsymmetrical geometry of diazocine moiety. Correlation plot embedded in bottom right corner of Figure S1 shows a well-defined X-pattern, indicating that finding film motives where top and bottom parts of neighbouring switches are either both close together or far apart is extremely unlikely.
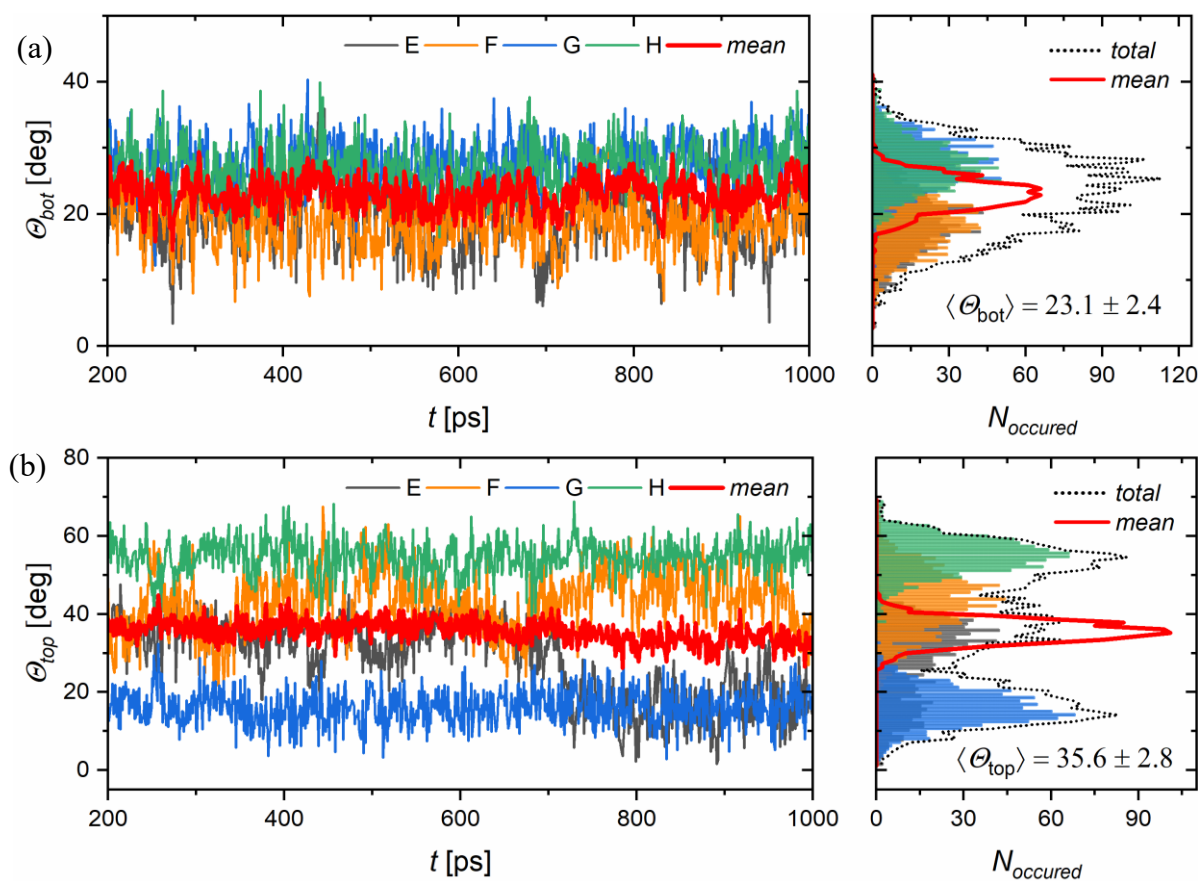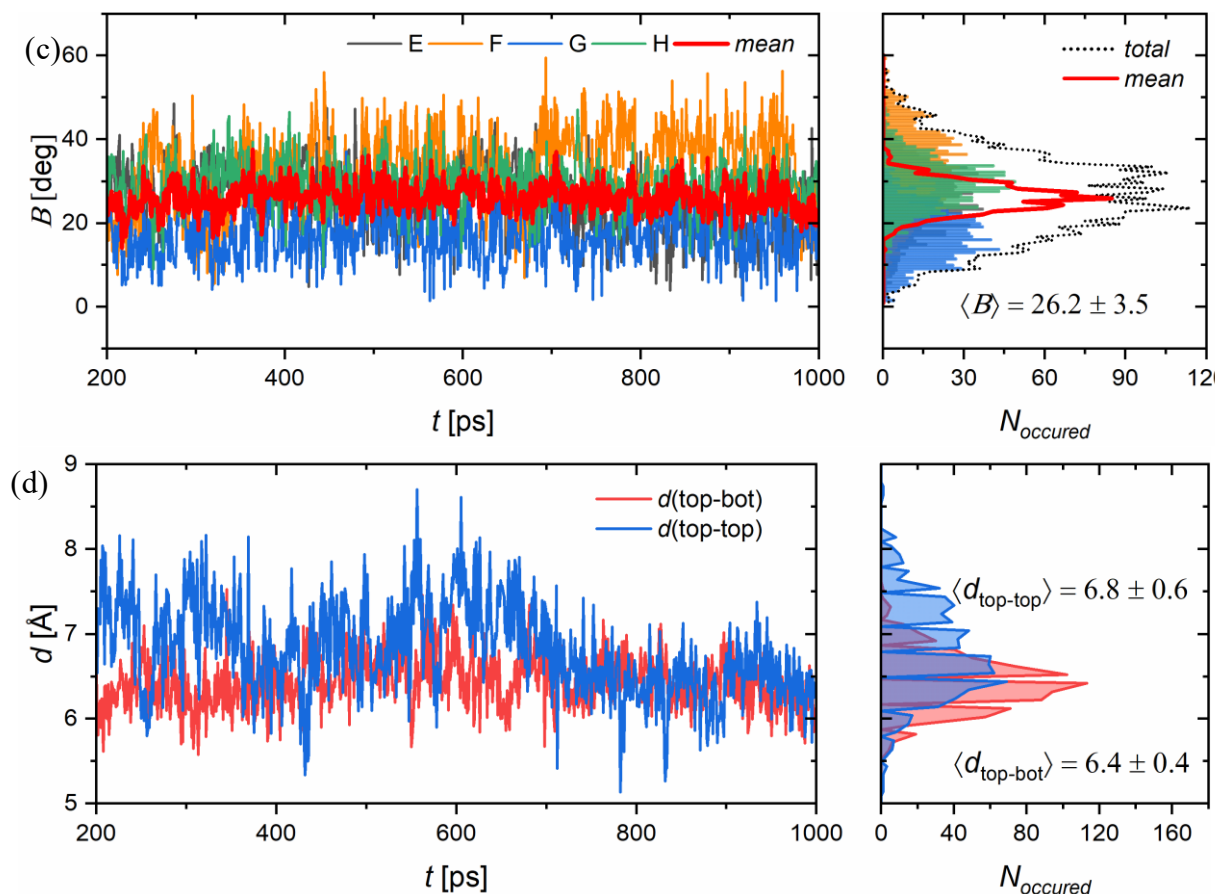
Figure S1. Evolution of *cis*-**Dia** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2.

Combined distribution function plot provides limited information as most interactions occur above the 7.08 Å distance and meaningful RDFs above that point are impossible to calculate. One can see a highly dispersed maximum at bend angles of 30° and functional head distance of around 7 Å, with a correlation plot showing only certain combination of bend/functional head distance are likely. Functional head tilt of F and E machines is not entirely converged; it is an interplay between two different angles instead, producing wide overlapped distribution histograms and the negative correlation of medium strength between the two (-0.21 for top tilt, -0.17 for the bend) machine motions.
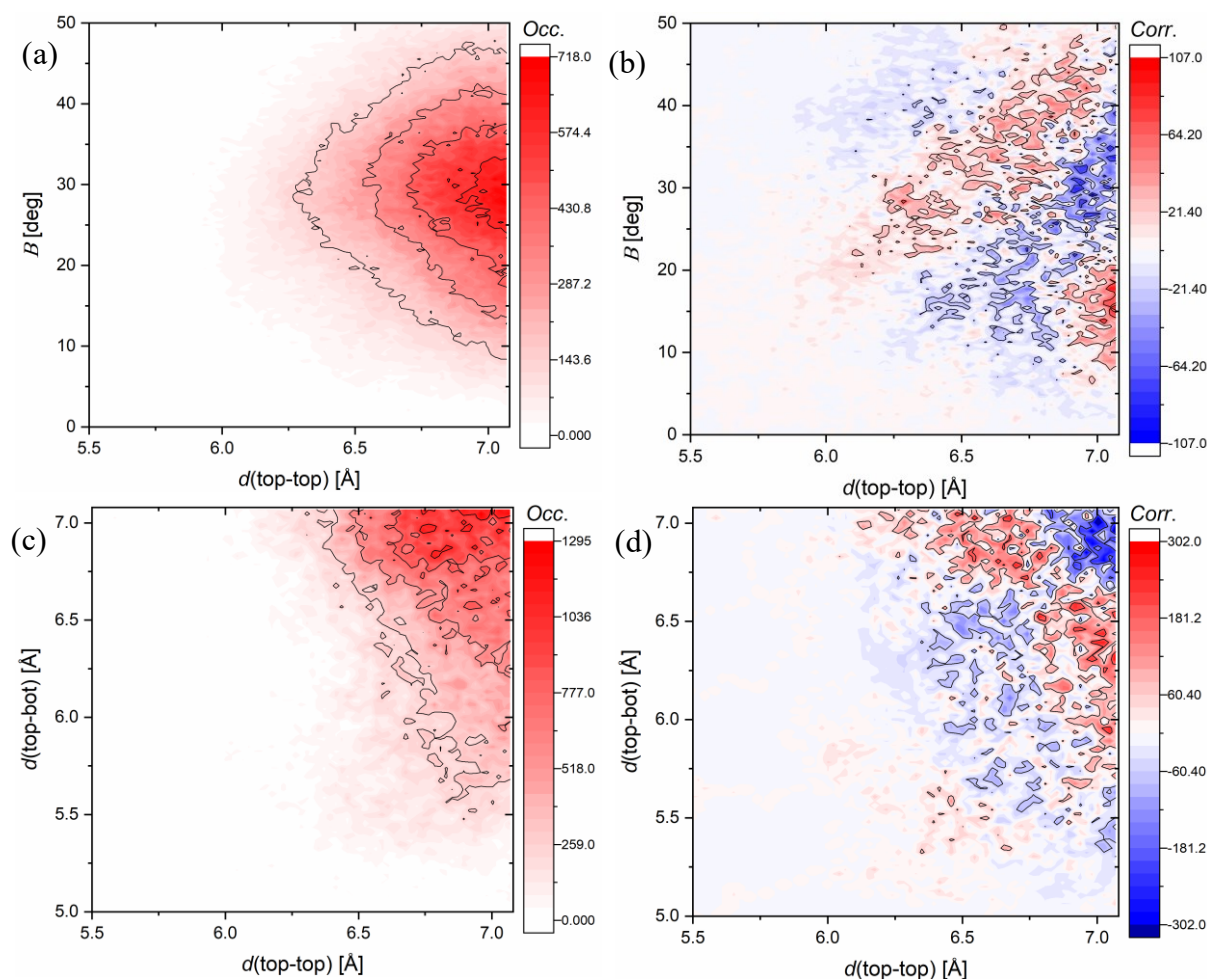
Figure S2. *cis*-**Dia** system: Combined distribution functions depicting: **(a)** distance between functional heads versus functional head bending angle and **(b)** its respective correlation plot. **(c)** distance between functional heads vs distance between functional heads and triptycene stands and **(d)** its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable than in the uncorrelated case (see Methods, Trajectory Analysis section).

Table S1. *cis*-**Dia** Kendall correlation matrices for tilting and bending of individual machines

	Θ(bot) [deg]				Θ(top) [deg]				B [deg]			
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	0.10	0.10	0.06	1.00	-0.21	-0.06	-0.03	1.00	-0.17	-0.05	-0.01
F	0.10	1.00	0.04	0.11	-0.21	1.00	-0.06	-0.12	-0.17	1.00	0.08	-0.02
G	0.10	0.04	1.00	0.06	-0.06	-0.06	1.00	0.00	-0.05	0.08	1.00	-0.03
H	0.06	0.11	0.06	1.00	-0.03	-0.12	0.00	1.00	-0.01	-0.02	-0.03	1.00

S2. Structural properties of SAMs: *trans*-Dia system

trans-**Dia** SAMs highly resemble those of *trans*-**Azo**, bases tilted at 35° on average, average bending of 29° and 62° of functional head tilts which, again, show propensity towards tilting in the same direction as the film structure (sum of Θ_{bot} and B is around 64°). However, contrary to *trans*-**Azo**, d(top-top) = 5.3 Å and d(top-bot) = 5.0 Å are almost perfectly overlapped, with two narrow, normal-like distributions.
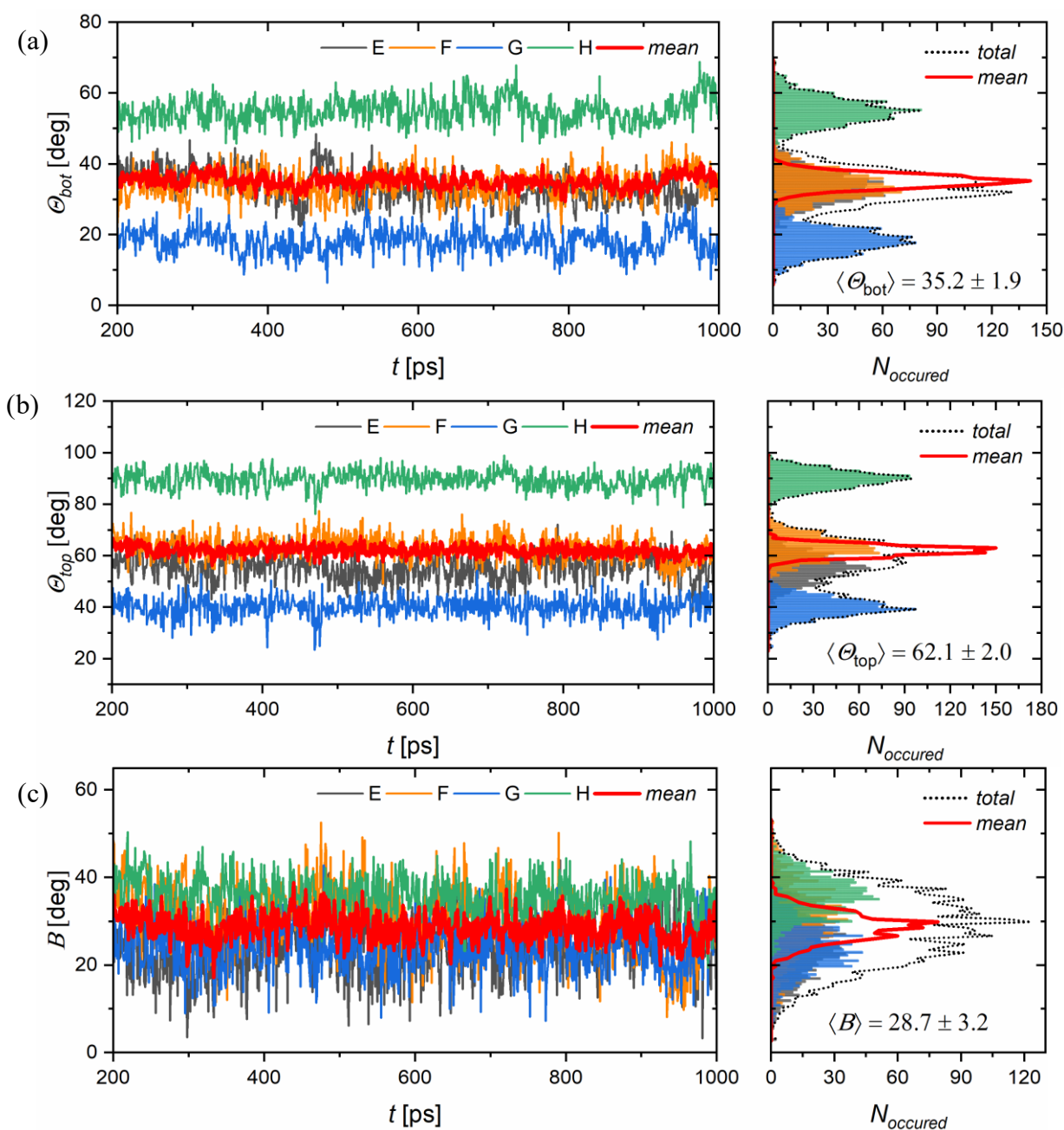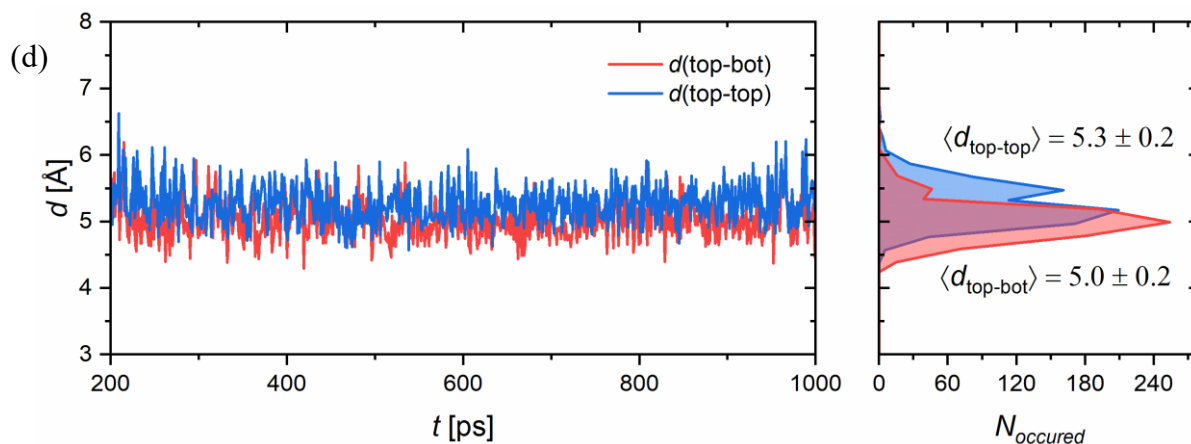
Figure S3. Evolution of *trans*-**Dia** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2..

Angular CDF plot (Figure S4 (a)) shows a clear overlap between functional head distances (d = 5.3 Å) and a range of bending angles (20° – 40°), producing an unsymmetric maximum on the 2D histogram. Several less pronounced maxima are also present ([d = 4.2 Å, B = 28°], [d = 5.7 A, B = 24°]) but are mostly negatively correlated. Overall, the angular correlation plot shows an X-pattern, similar to the trans-Azo system, with strong positive correlation at d= 5.8 Å and B = 22°, but is less defined, suggesting more rotational freedom between trans-Dia functional heads.

RDF-RDF combined distribution 2D histogram continues the trends of multiple weak maxima. However, a single, highly symmetric maximum at d = 5.2 d= 5.0 Å, also described in Figure S3 (d) dominates the landscape, showing a clear preference of *trans*-**Dia** films for configurations with lower and upper parts of neighbouring molecules at similar distances. Moreover, its respective correlation plot reveals a large dominating positive correlation of exactly these conformations.

Kendall correlation matrices show no significant collective trends in molecular modes of motion, with lowest correlation coefficient of -0.18 in case of upper tilt, which is not significant. Also, there is a complete lack of collective trends in bending between the machines.
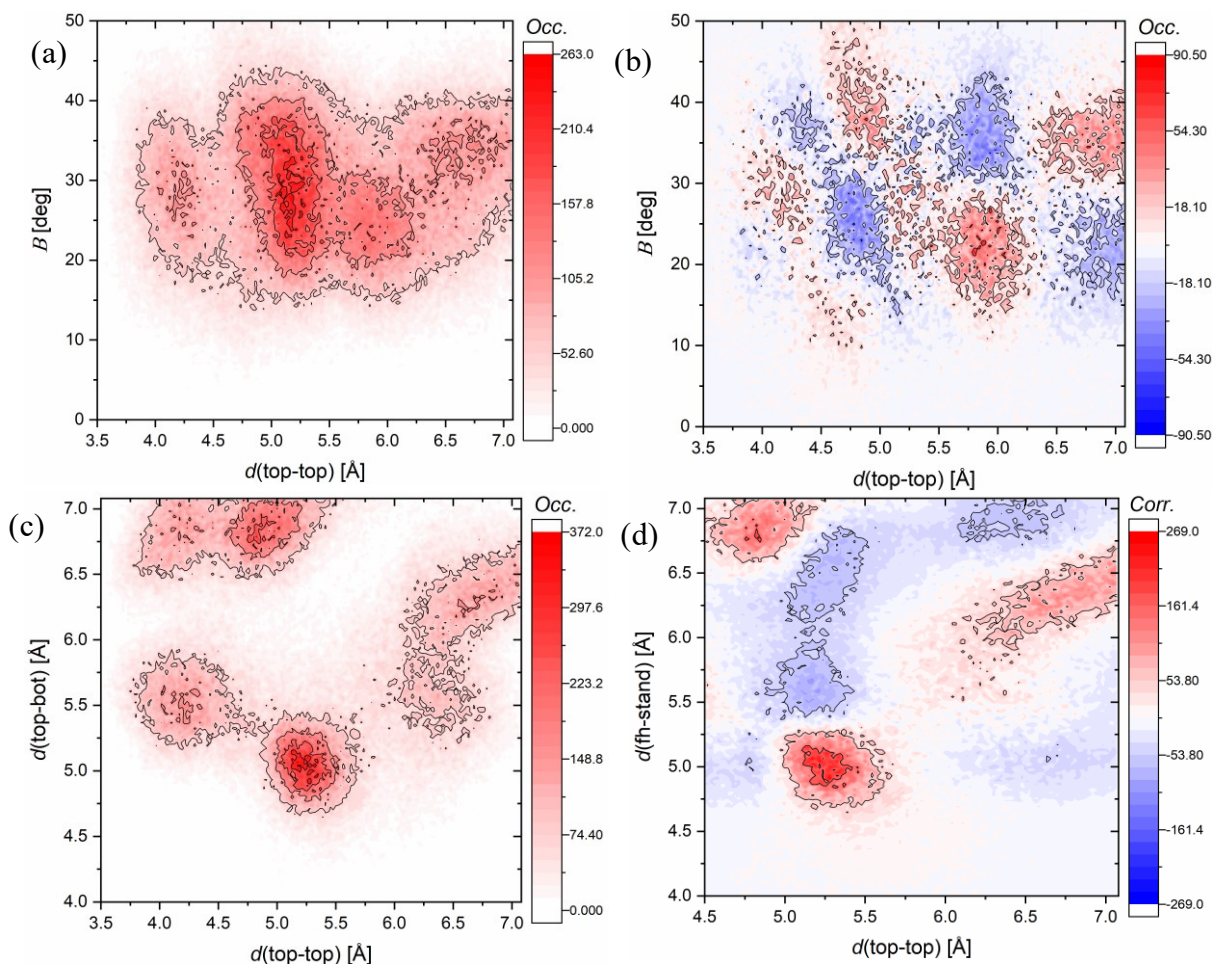
Figure S4. *trans*-**Dia** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable than in the uncorrelated case (see Methods, Trajectory Analysis section).

Table S2. *trans*-**Dia** Kendall correlation matrices for tilting and bending of individual machines

	Θ(bot) [deg]				Θ(top) [deg]				B [deg]			
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	-0.03	-0.05	-0.04	1.00	-0.03	-0.07	-0.18	1.00	0.03	-0.01	0.04
F	-0.03	1.00	0.08	-0.03	-0.03	1.00	-0.08	0.04	0.03	1.00	0.00	0.00
G	-0.05	0.08	1.00	0.07	-0.07	-0.08	1.00	0.13	-0.01	0.00	1.00	0.01
H	-0.04	-0.03	0.07	1.00	-0.18	0.04	0.13	1.00	0.04	0.00	0.01	1.00

S3. Structural properties of SAMs: Thio system

Figure S5. Evolution of **Thio** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2..
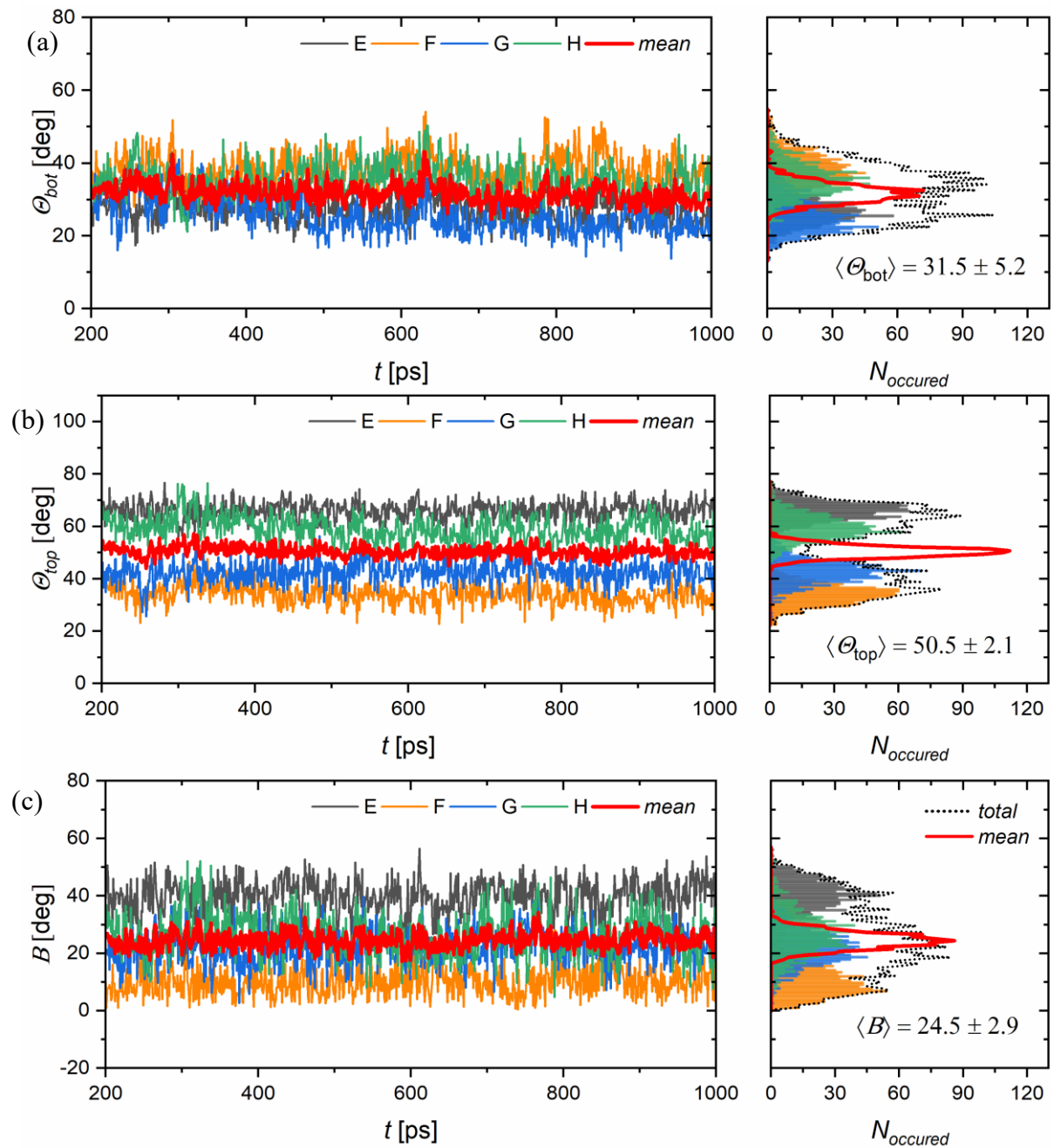

Figure S6. **Thio** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot.

(c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable than in the uncorrelated case (see Methods, Trajectory Analysis section).
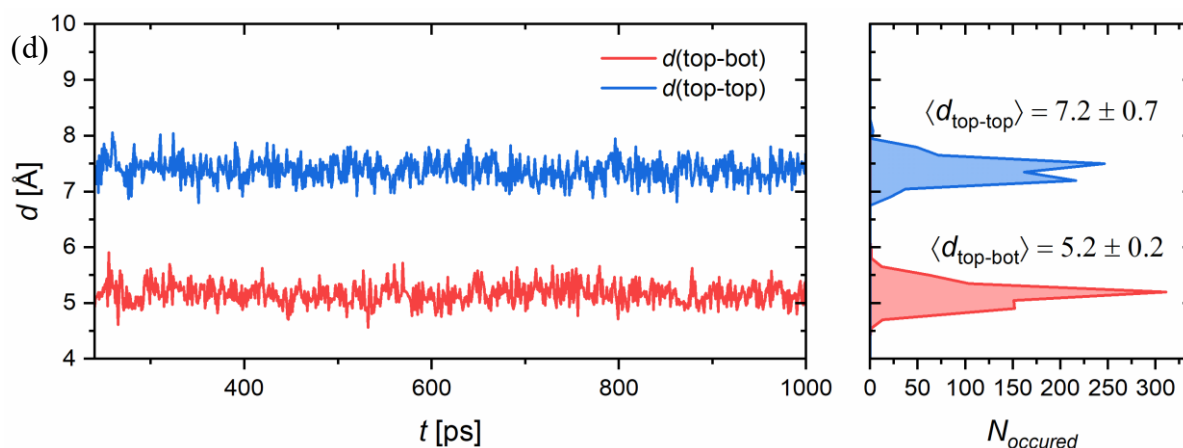
S4. Structural properties of SAMs: Tri system

Figure S7. Evolution of **Tri** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2..
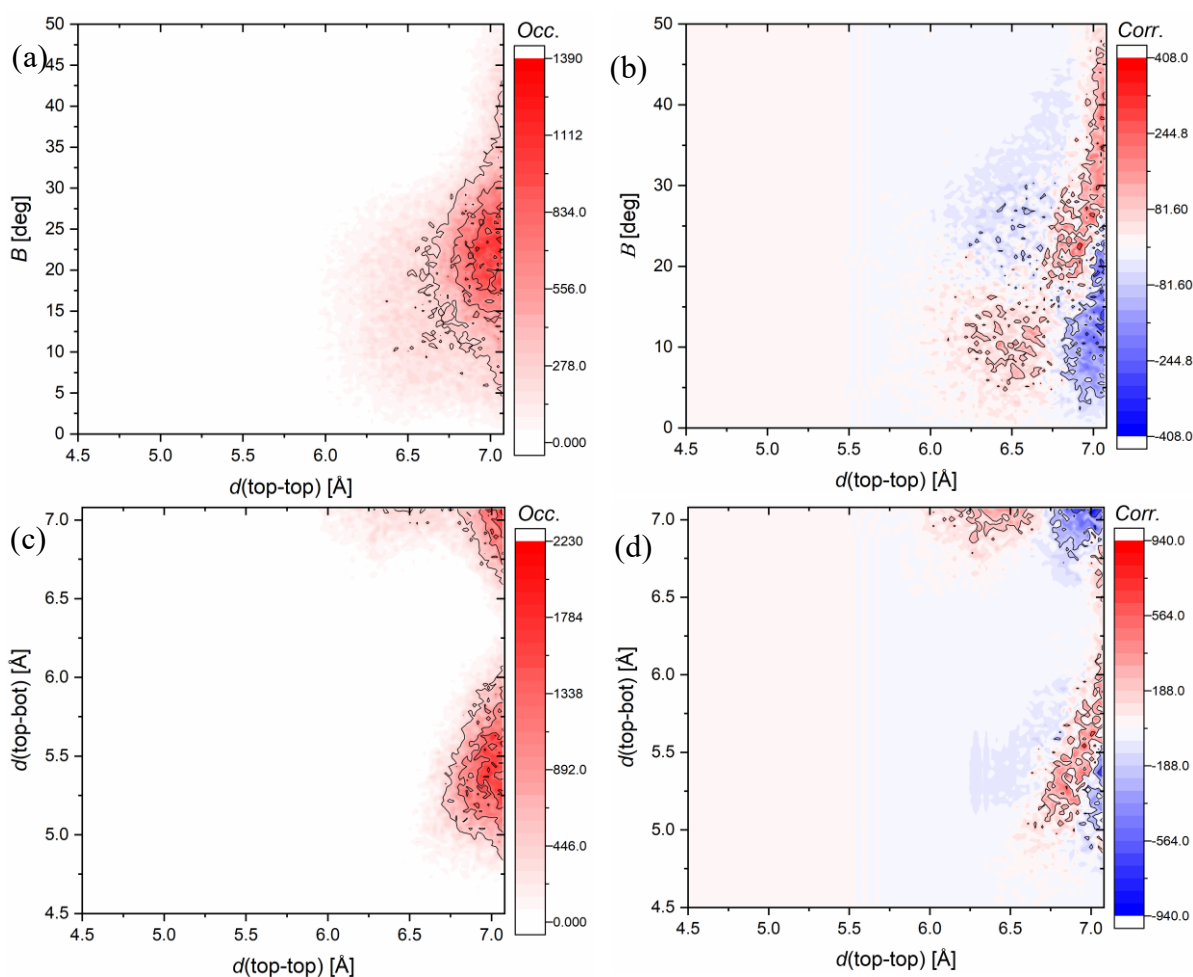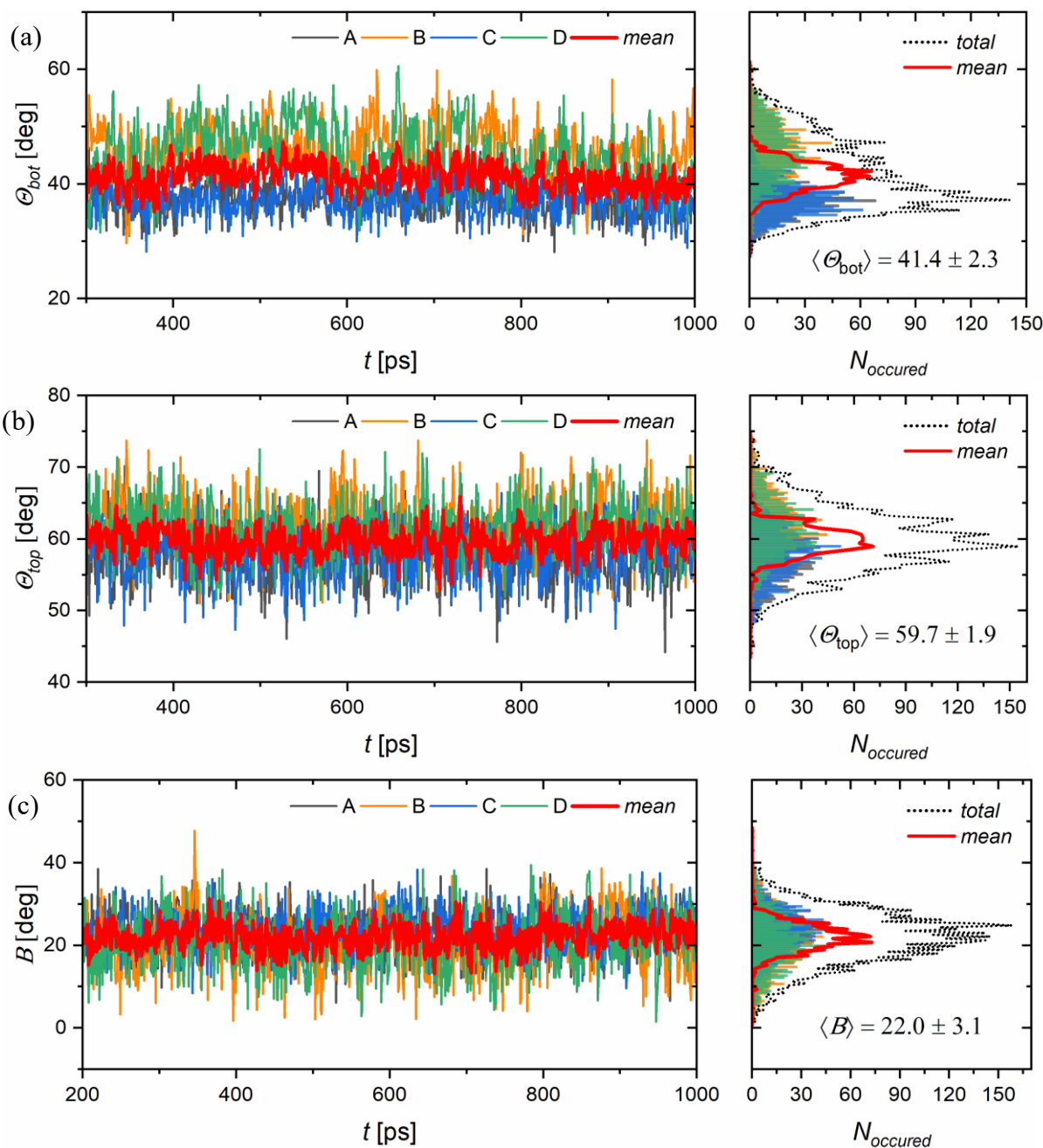

Figure S8. **Tri** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot.

(c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable than in the uncorrelated case (see Methods, Trajectory Analysis section).

Table S3. Tri Kendall rank correlation matrices for tilting and bending of individual machines

	Θ(bot) [deg]				Θ(top) [deg]				B [deg]			
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	0.16	0.06	0.10	1.00	-0.07	0.02	0.04	1.00	0.00	0.06	-0.06
F	0.16	1.00	0.01	-0.01	-0.07	1.00	0.01	0.00	0.00	1.00	-0.04	0.07
G	0.06	0.01	1.00	0.19	0.02	0.01	1.00	-0.06	0.06	-0.04	1.00	0.04
H	0.10	-0.01	0.19	1.00	0.04	0.00	-0.06	1.00	-0.06	0.07	0.04	1.00

S5. Structural properties of SAMs: Phe1 system
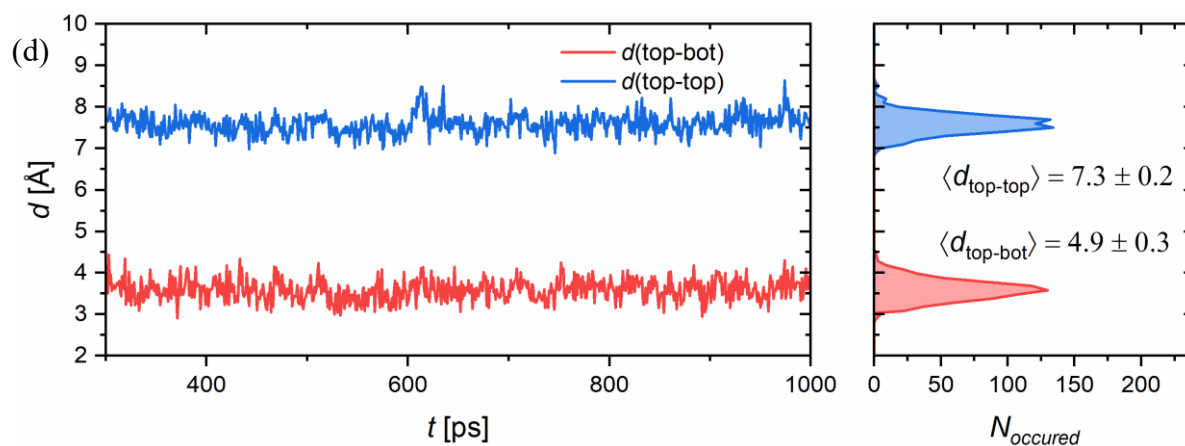

(a)

(b)

(c)
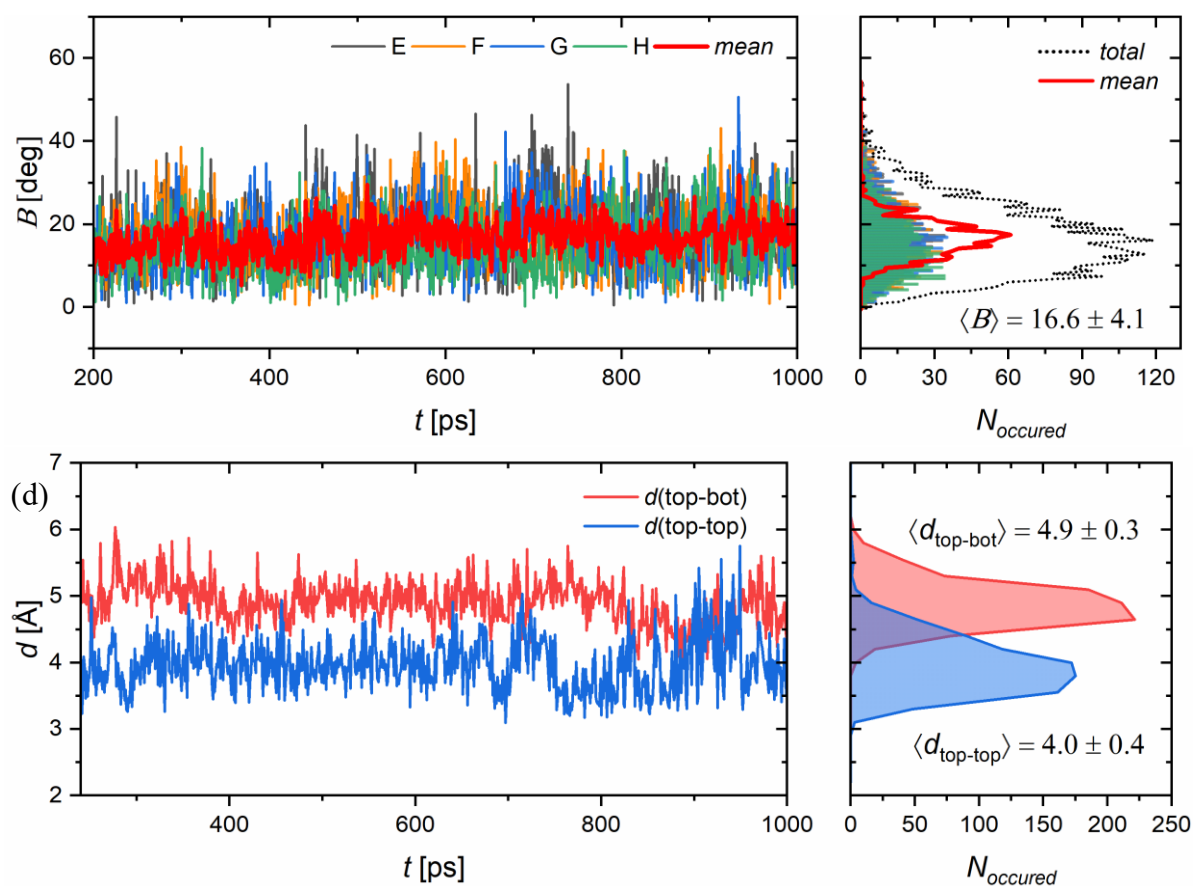
Figure S9. Evolution of **Phe1** film structural properties: (a) triptycene stand and (b) functional head tilt angles and (c) the bending angle. Thin coloured lines are monitored angles of individual machines (E − H), while thick red line denotes the mean angle in every timestep. (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2..
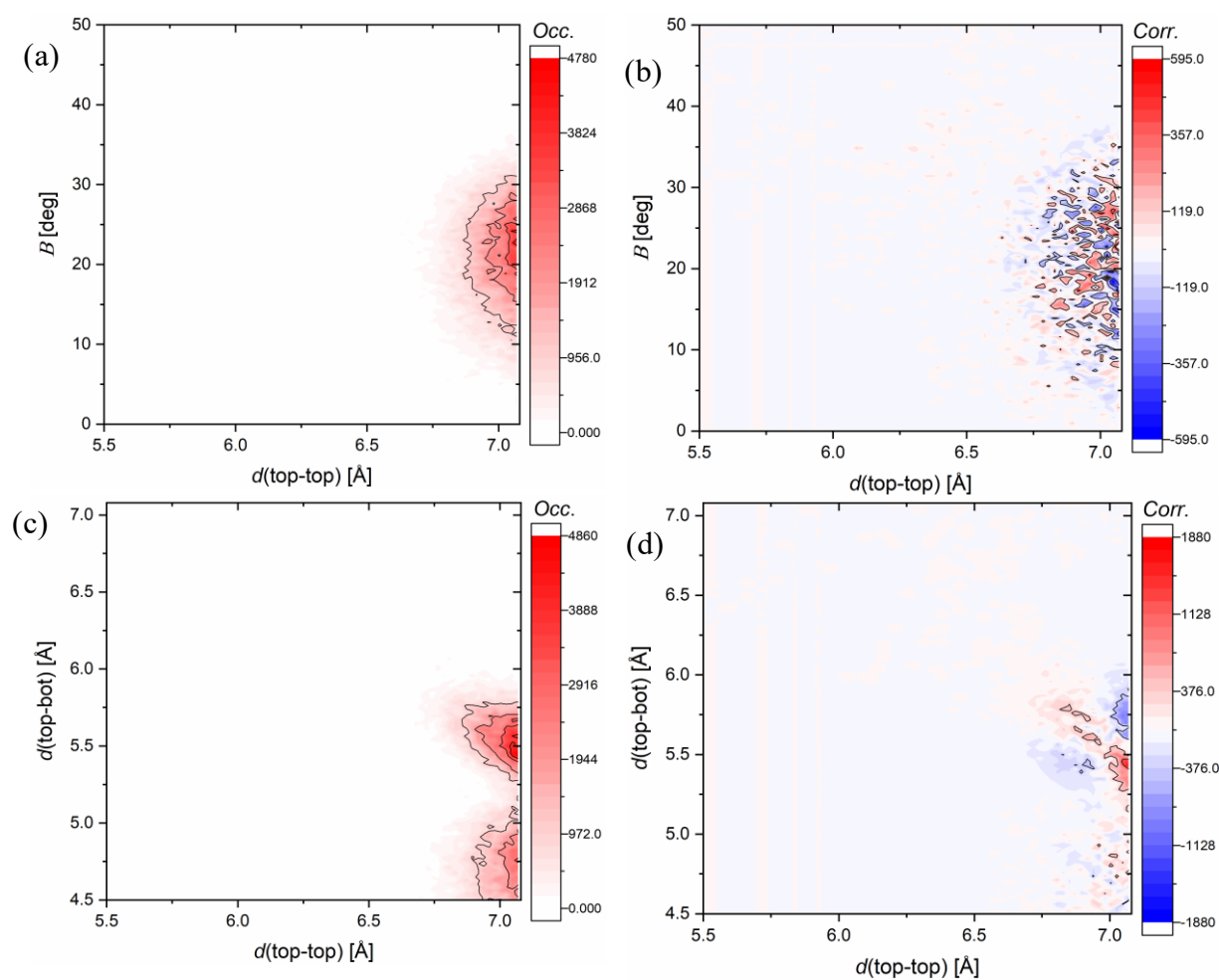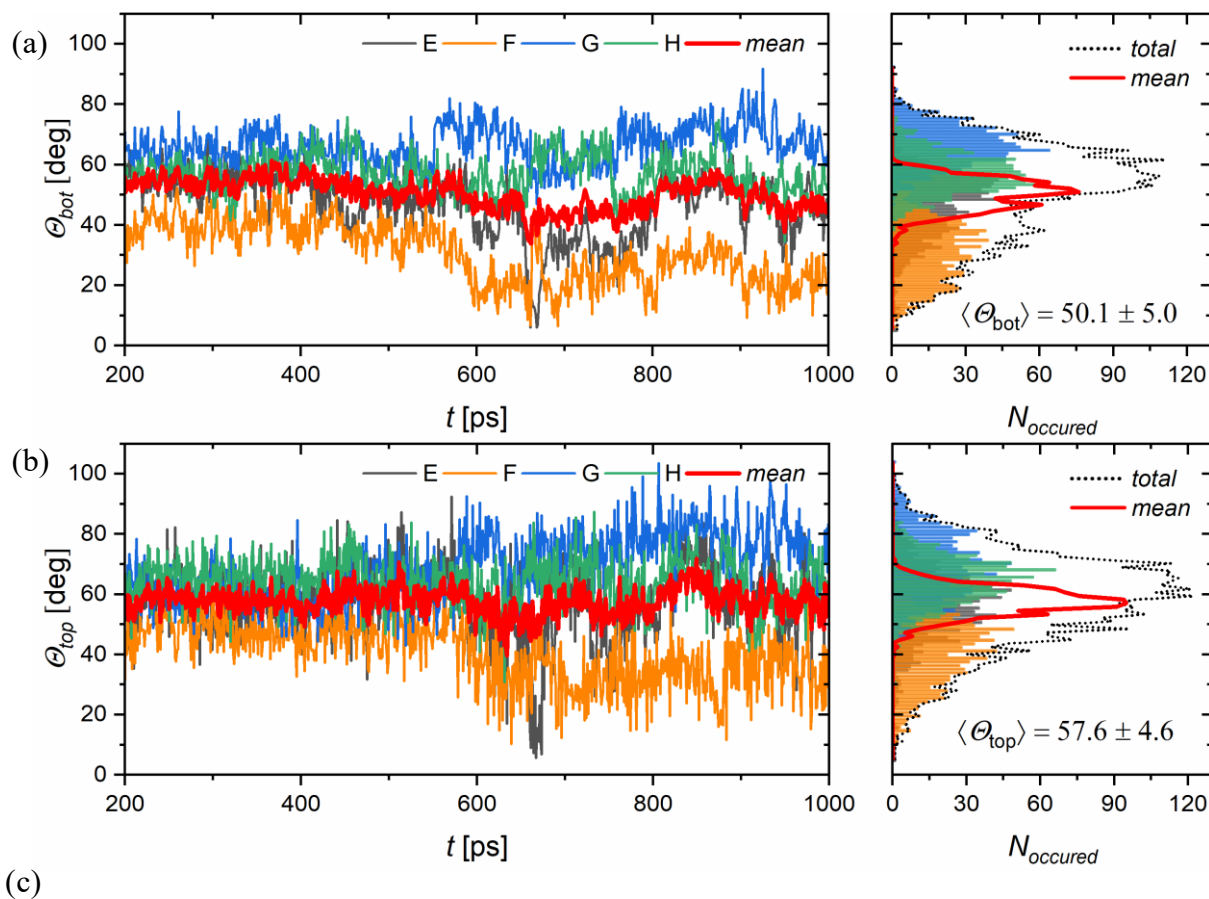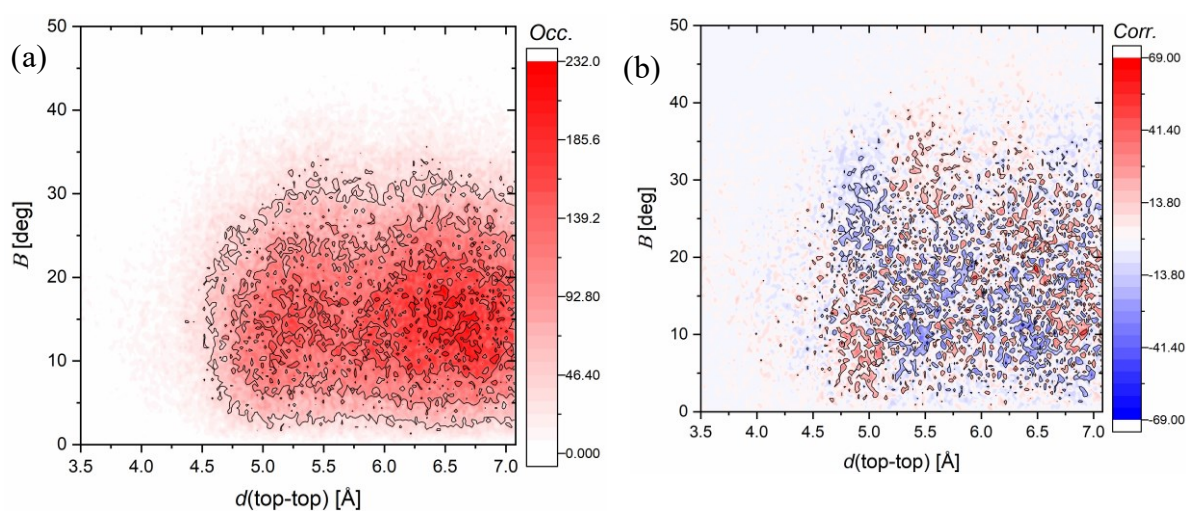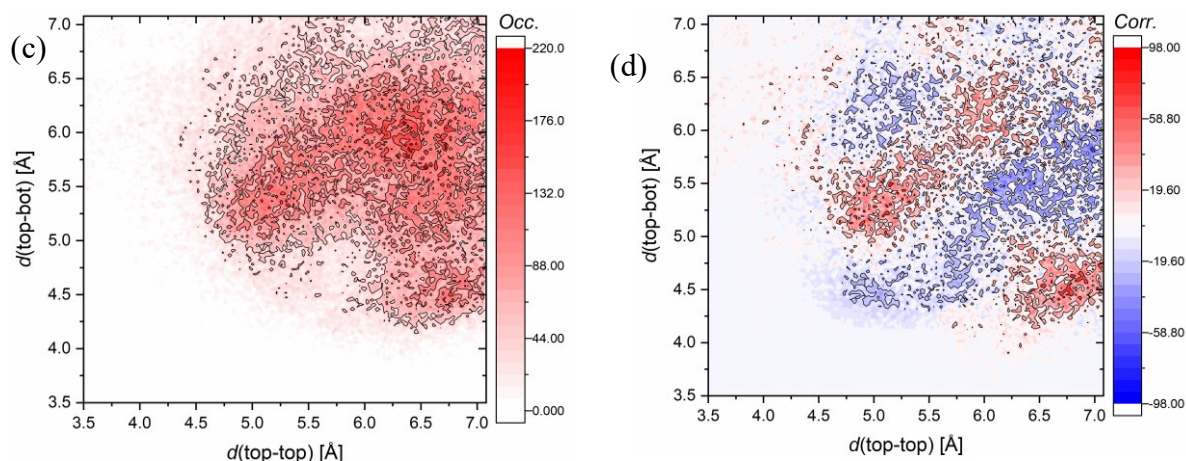
Figure S10. **Phe1** system: Combined distribution functions depicting: (a) distance between functional heads versus functional head bending angle and (b) its respective correlation plot. (c) distance between functional heads vs distance between functional heads and triptycene stands and (d) its correlation plot. Red depicts positive correlation, white depicts uncorrelated events, while blue depicts configurations less probable than in the uncorrelated case (see Methods, Trajectory Analysis section).

Table S4. Phe1 Kendall correlation matrices for tilting and bending of individual machines

| | Θ(bot) [deg] | | | | Θ(top) [deg] | | | | B [deg] | | | |
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	0.34	0.08	0.02	1.00	0.02	0.16	-0.02	1.00	-0.04	0.07	-0.01
F	0.34	1.00	-0.06	0.11	0.02	1.00	0.03	-0.08	-0.04	1.00	0.00	0.02
G	0.08	-0.06	1.00	-0.14	0.16	0.03	1.00	-0.08	0.07	0.00	1.00	0.00
H	0.02	0.11	-0.14	1.00	-0.02	-0.08	-0.08	1.00	-0.01	0.02	0.00	1.00

S6. PBC Molecular Dynamics: analysis and benchmarking details

Thermodynamics and stability
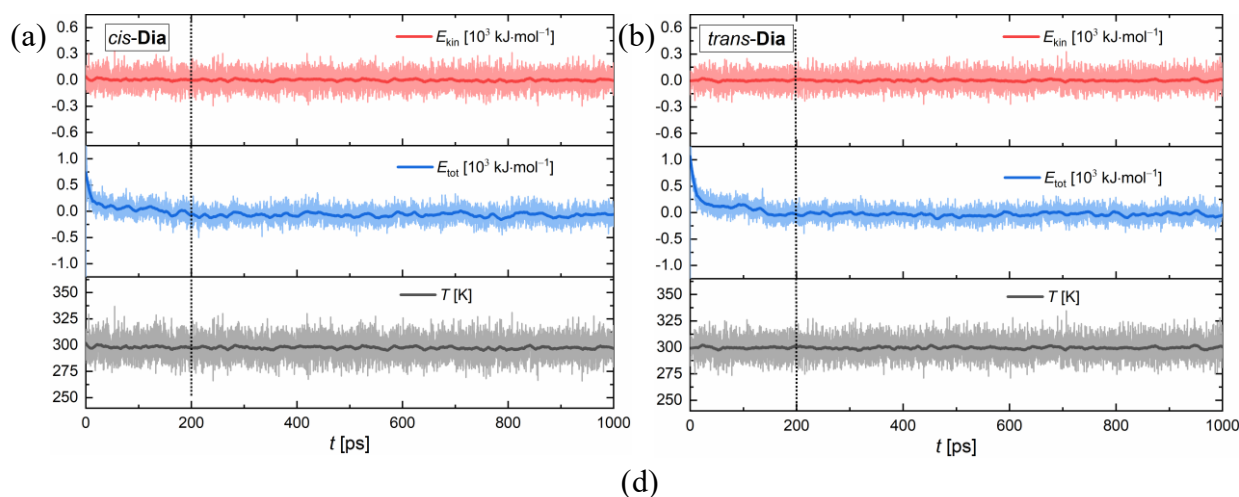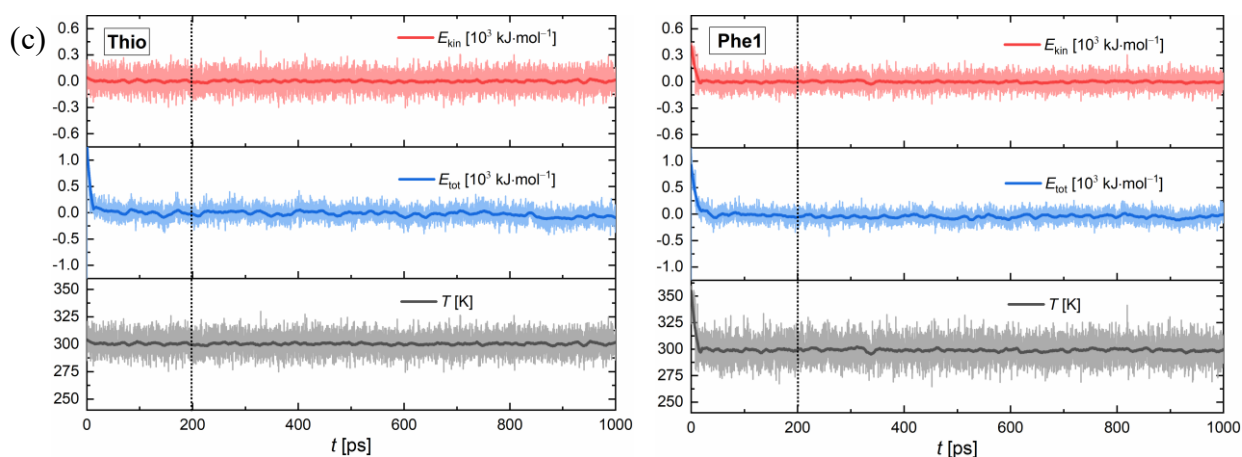
Figure S11. Kinetic energy blue, total energy red and temperature grey for (a) cis- Dia(b) trans-Dia(c) Thio and (d) Phe1 systems. The chosen equilibration cut-off is denoted by a dotted vertical line. Energy mean is shifted to zero.
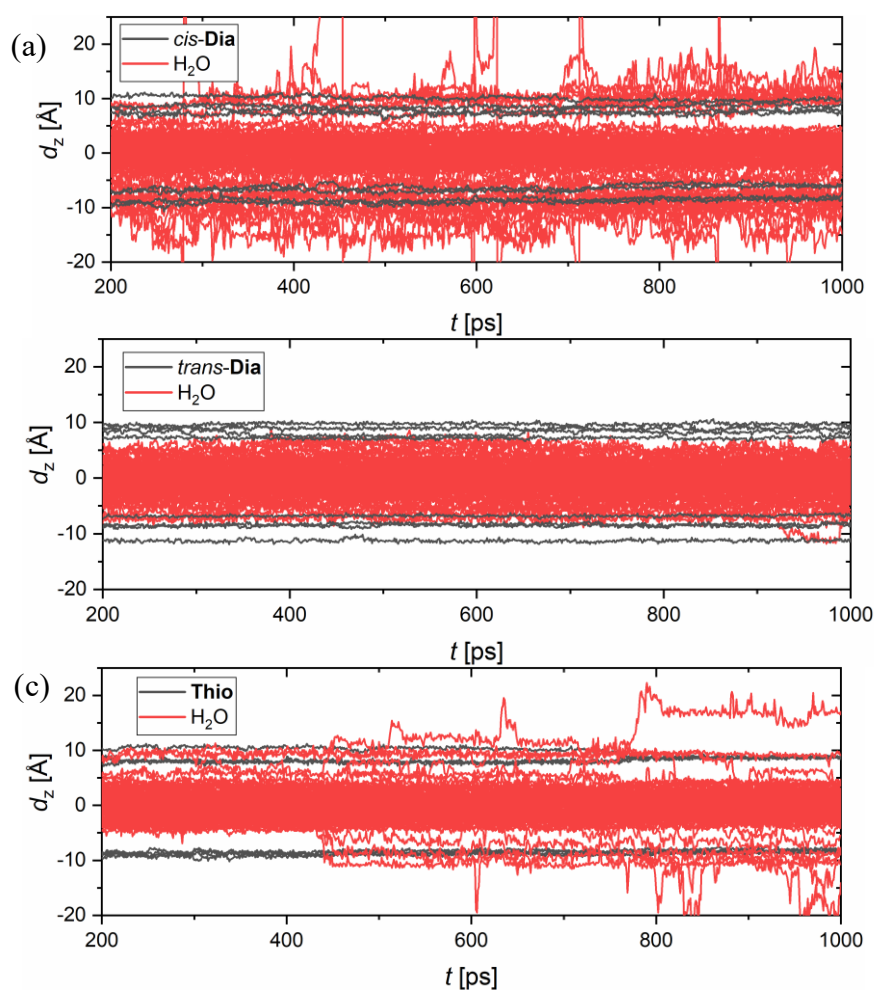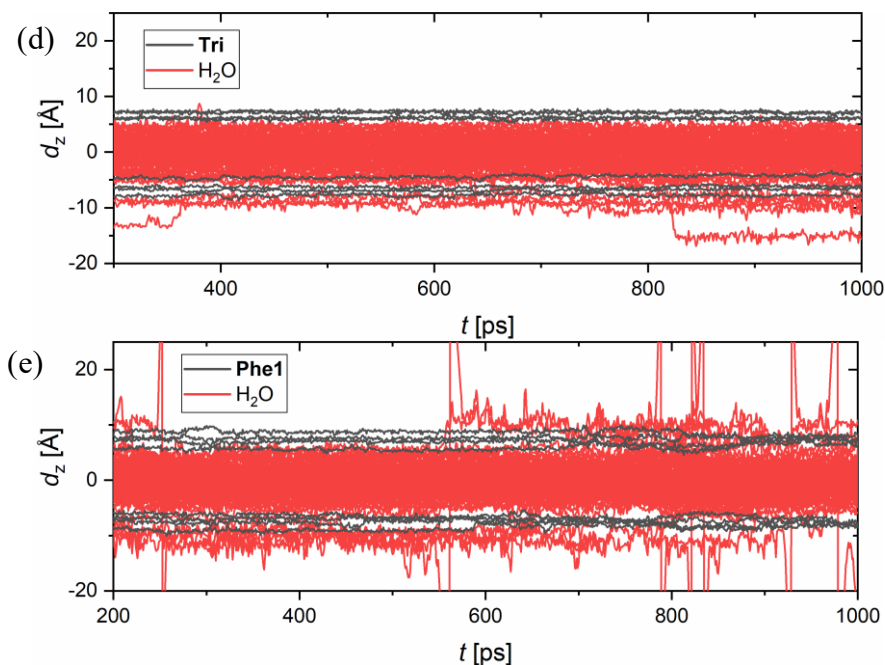
Figure S12. Temporal evolution of water (red) and triptycene stand (grey) densities in the z-direction (perpendicular to the film surface) for ((a) *cis-* **Dia**(b) *trans-***Dia**(c) **Thio**, (d) **Tri** and (e) **Phe1** systems. z-coordinates of water oxygen and upper bridgehead carbon atoms represent the density distribution of water molecules and individual triptycene stands, respectively. See Figure 14 in Section 4.2 for definition and details.

Table S5. Mean values of total energy and temperature of investigated systems

Sys.	E_{tot} [10^6 kJ·mol^{-1}]	T [K]	Sys.	E_{tot} [10^6 kJ·mol^{-1}]	T [K]
*trans-***Azo**	-2.9213	300.0	**Thio**	-3.7810	301.0
*cis-***Azo**	-2.9241	300.1	**Phe0**	-1.9693	299.9
*cis-***Dia**	-3.0124	302.6	**Phe1**	-2.1458	299.4
*trans-***Dia**	-3.0122	299.6	**Phe**	-2.4962	300.1
Fer	-3.3696	300.3	**Tri**	-3.3890	299.9

Figure S13 shows the projection of B and distances onto the x direction of the unit cell. The asymmetric collective tilting of the monolayer can clearly be observed.
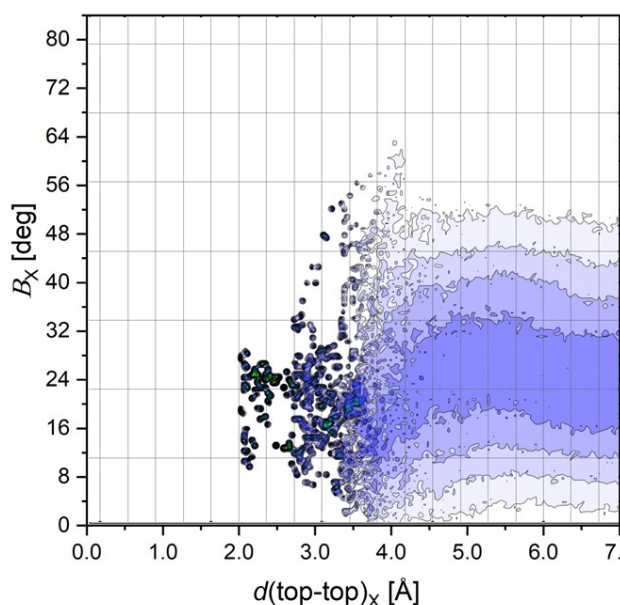
Figure S13. Average neighbouring functional head distance and bending angle projected onto *x* unit cell vector. Machines in the monolayer collectively tilt and their functional heads point approximately in direction direction of *x* base vector.

Kendall correlation matrices for the Phe0 and Phe systems

Table S6. Phe0 Kendall correlation matrices for tilting and bending of individual machines

	Θ(bot) [deg]				Θ(top) [deg]				B [deg]			
	A	**B**	**C**	**D**	**A**	**B**	**C**	**D**	**A**	**B**	**C**	**D**
A	1.00	-0.09	0.08	0.08	1.00	0.08	-0.06	0.04	1.00	-0.01	-0.02	-0.01
B	-0.09	1.00	-0.14	0.01	0.08	1.00	-0.05	0.07	-0.01	1.00	0.06	-0.07
C	0.08	-0.14	1.00	-0.14	-0.06	-0.05	1.00	-0.07	-0.02	0.06	1.00	0.02
D	0.08	0.01	-0.14	1.00	0.04	0.07	-0.07	1.00	-0.01	-0.07	0.02	1.00

Table S7. Phe Kendall correlation matrices for tilting and bending of individual machines

	Θ(bot) [deg]				Θ(top) [deg]				B [deg]			
	E	**F**	**G**	**H**	**E**	**F**	**G**	**H**	**E**	**F**	**G**	**H**
E	1.00	0.02	-0.01	0.06	1.00	-0.02	0.07	-0.14	1.00	-0.09	0.04	-0.05
F	0.02	1.00	0.03	0.03	-0.02	1.00	-0.21	0.02	-0.09	1.00	-0.02	0.11
G	-0.01	0.03	1.00	-0.02	0.07	-0.21	1.00	-0.10	0.04	-0.02	1.00	-0.04
H	0.06	0.03	-0.02	1.00	-0.14	0.02	-0.10	1.00	-0.05	0.11	-0.04	1.00

Cross-correlation between the upper and lower layer in the simulation cell.

Kendall correlation matrices show that, although correlation between the machines in upper layer is significant, it does not transfer to the lower, more stable layer. Since there is no

communication between the layers, lower and upper layer can be analysed indebendently. **Table S8-Table S10** show this on the example of *trans-***Azo** system, which was the *riskiest* to analyze, since the top layer gets destroyed resulting in energy drop of the system.

Table S8. Kendall correlation matrix for triptycene stand tilting of whole *trans-***Azo** system, including top and bottom layer of simulation cell.

| | Θ(bot) [deg] | | | | | | | |
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A	1.00	0.42	0.14	0.20	-0.02	-0.11	-0.06	-0.04
B	0.42	1.00	0.44	0.13	-0.03	-0.06	-0.01	0.01
C	0.14	0.44	1.00	0.02	0.00	-0.01	0.00	0.00
D	0.20	0.13	0.02	1.00	-0.04	-0.07	0.03	-0.01
E	-0.02	-0.03	0.00	-0.04	1.00	-0.01	-0.01	-0.13
F	-0.11	-0.06	-0.01	-0.07	-0.01	1.00	0.14	-0.03
G	-0.06	-0.01	0.00	0.03	-0.01	0.14	1.00	0.05
H	-0.04	0.01	0.00	-0.01	-0.13	-0.03	0.05	1.00

Table S9. Kendall correlation matrix for functional head tilting of whole *trans-***Azo** system, including top and bottom layer of simulation cell.

| | Θ(top) [deg] | | | | | | | |
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A	1.00	0.17	-0.22	0.06	0.07	-0.03	-0.01	0.02
B	0.17	1.00	-0.29	0.14	0.03	-0.01	0.01	0.03
C	-0.22	-0.29	1.00	-0.09	-0.05	0.01	0.00	-0.02
D	0.06	0.14	-0.09	1.00	0.01	-0.02	0.01	0.04
E	0.07	0.03	-0.05	0.01	1.00	0.03	-0.01	-0.09
F	-0.03	-0.01	0.01	-0.02	0.03	1.00	-0.07	-0.05
G	-0.01	0.01	0.00	0.01	-0.01	-0.07	1.00	0.07
H	0.02	0.03	-0.02	0.04	-0.09	-0.05	0.07	1.00

Table S10. Kendall correlation matrix for bending of whole *trans-***Azo** system, including top and bottom layer of simulation cell.

| | B [deg] | | | | | | | |
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**
A	1.00	-0.05	0.14	0.15	0.01	-0.05	-0.01	-0.01
B	-0.05	1.00	0.39	0.02	-0.05	-0.06	-0.05	0.04
C	0.14	0.39	1.00	0.11	-0.05	-0.08	-0.01	0.02
D	0.15	0.02	0.11	1.00	-0.04	-0.02	0.00	-0.03
E	0.01	-0.05	-0.05	-0.04	1.00	0.04	0.05	-0.11

F	-0.05	-0.06	-0.08	-0.02	0.04	1.00	0.03	-0.03
G	-0.01	-0.05	-0.01	0.00	0.05	0.03	1.00	0.05
H	-0.01	0.04	0.02	-0.03	-0.11	-0.03	0.05	1.00

Exemplary results of a meta-stable system (trans-Azo)

Taking *trans*-**Azo** system for an example once more, the presented graphs show the analysis of monolayer properties of trans-Azo metastable layer. Results for the rest of the metastable systems can be found on Github repository.[63]

(a)

(b)

(c)
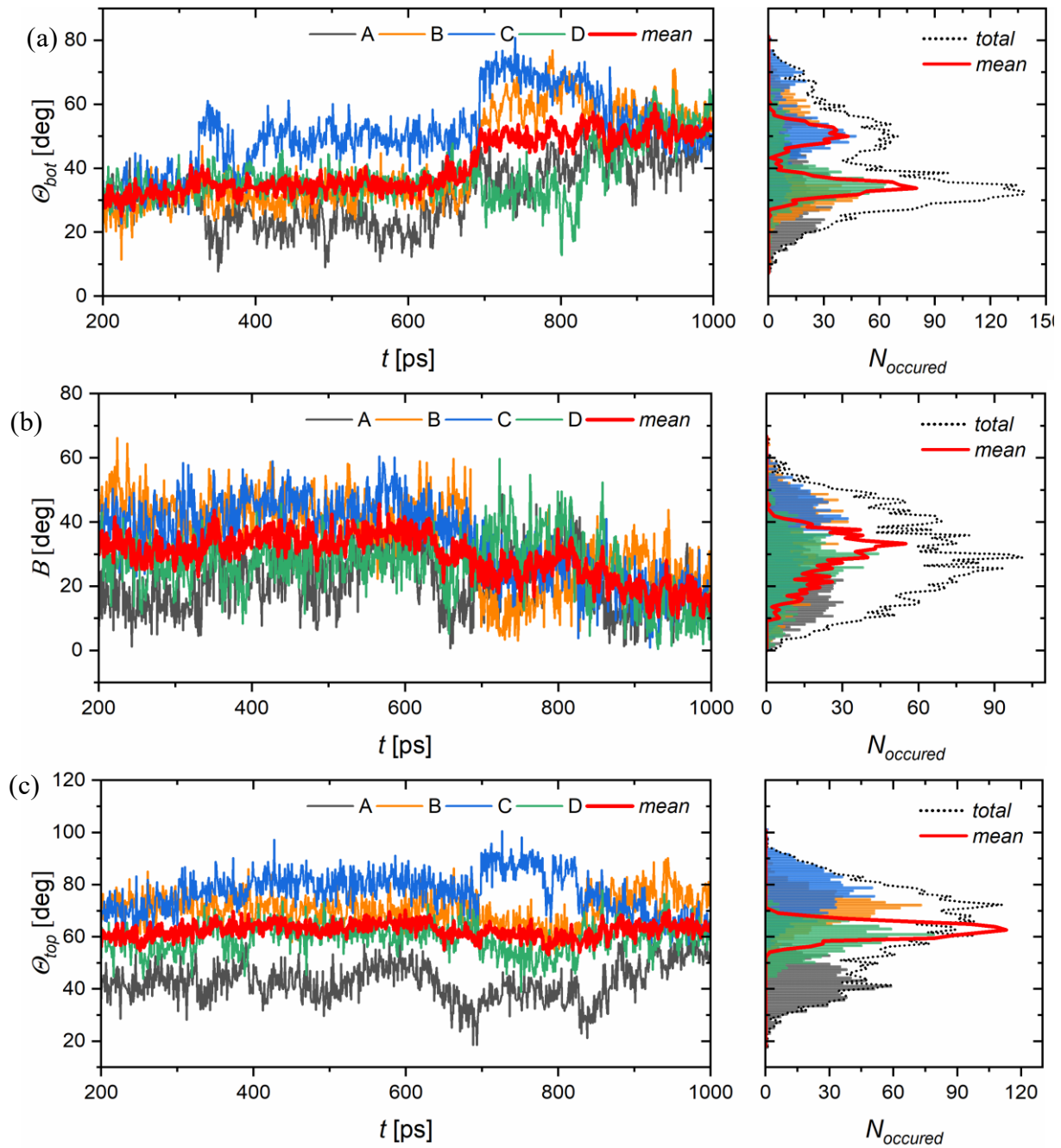
(d)
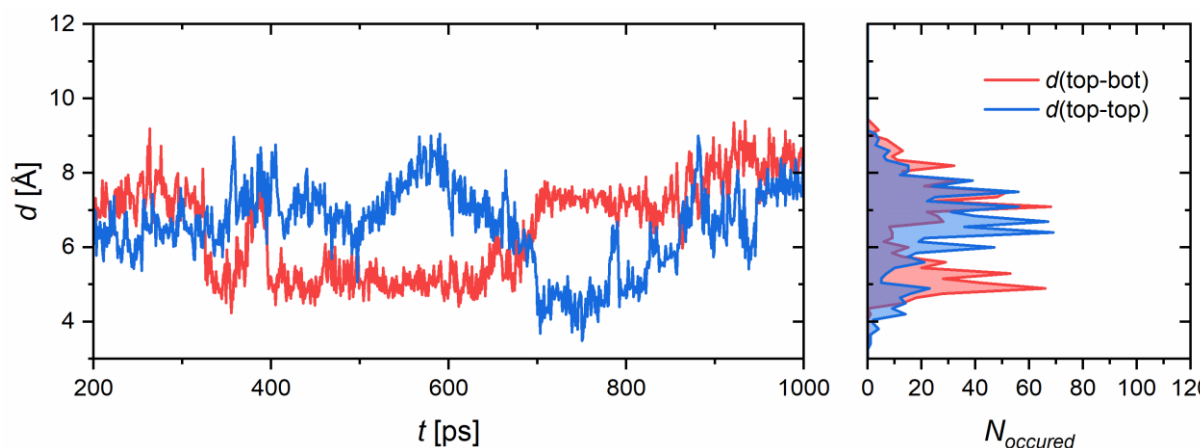
Evolution of metastable layer of *trans*-**Azo** film structural properties: (a) triptycene stand and (b) functional head tilt angles, (c) bending angle and (d) minimal observed distance between functional heads and either (*red*) triptycene stands or (*blue*) other functional heads, as defined in Table 2 in Methods, Section 4.2.

Table S11. Layer stabilities of different systems. More stable layers are presented in the main text. When evaluating the stability, the layer with better converged observables is denoted as the stable one and the one with phase transitions as the meta-stable. The top layer of the *trans*-**Azo** system is denoted as unstable, as the protrusion on the water involves changes in total energy.

Sys.	*Top layer*	*Bot layer*	Sys.	*Top layer*	*Bot layer*
trans-Azo	unstable	stable	Thio		
cis-Azo	meta-stable	stable	Phe0	stable	meta-stable
cis-Dia	meta-stable	stable	Phe1	meta-stable	stable
trans-Dia	meta-stable	stable	Phe	stable	stable
Fer	meta-stable	stable	Tri	stable	meta-stable

Thermostat benchmarking

The coupling strength: value of $3200\ \mathrm{cm}^{-1}$ was chosen as the least amount of coupling necessary to acquire the stable trajectory with the thermodynamic temperature oscillating around the value of 300 K.
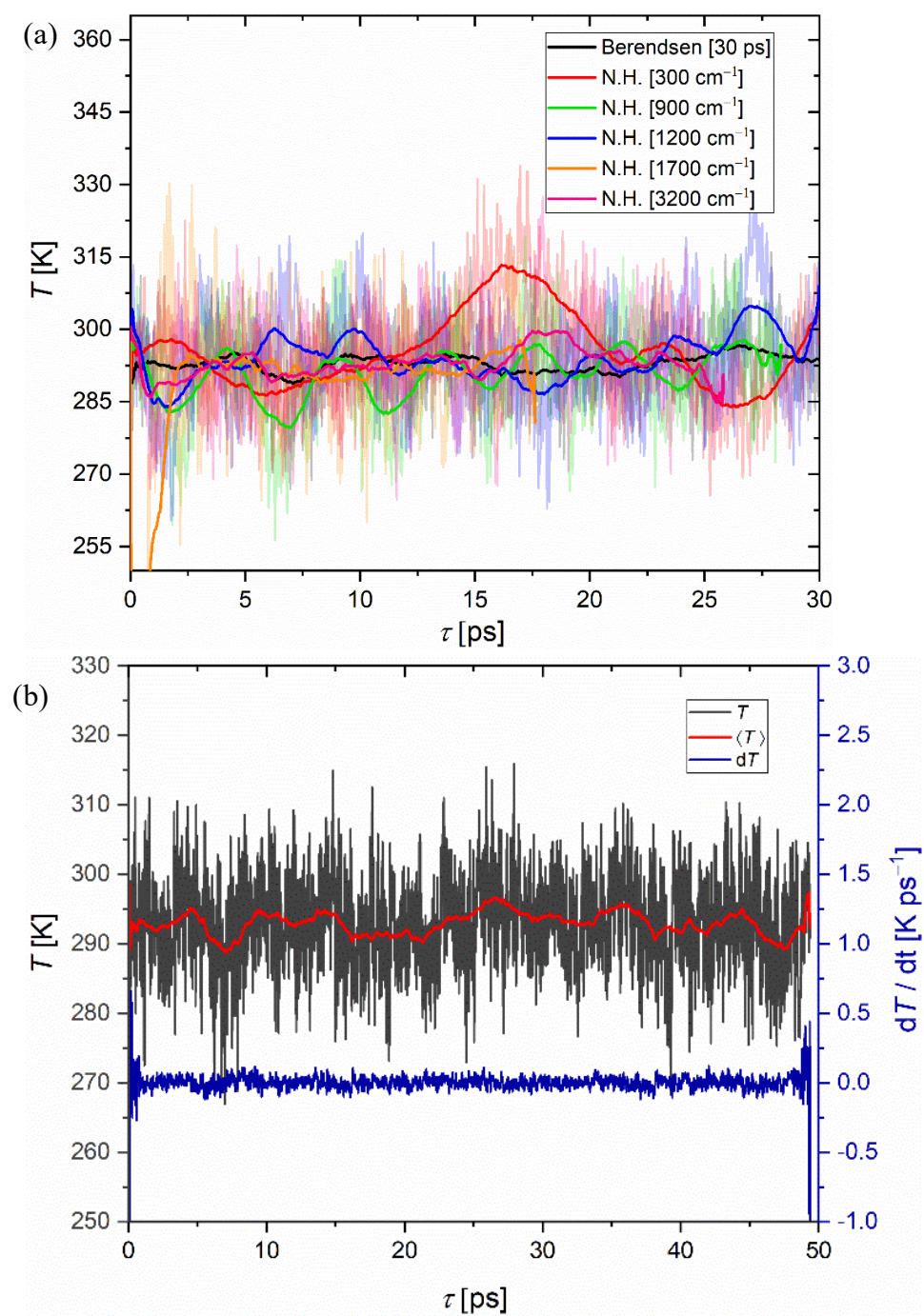
Figure S14. (a) Oscillations of temperature with Berendsen (black) and Nose-Hoover (coloured) thermostats at different coupling strengths.

S7. IR and Raman Spectroscopy
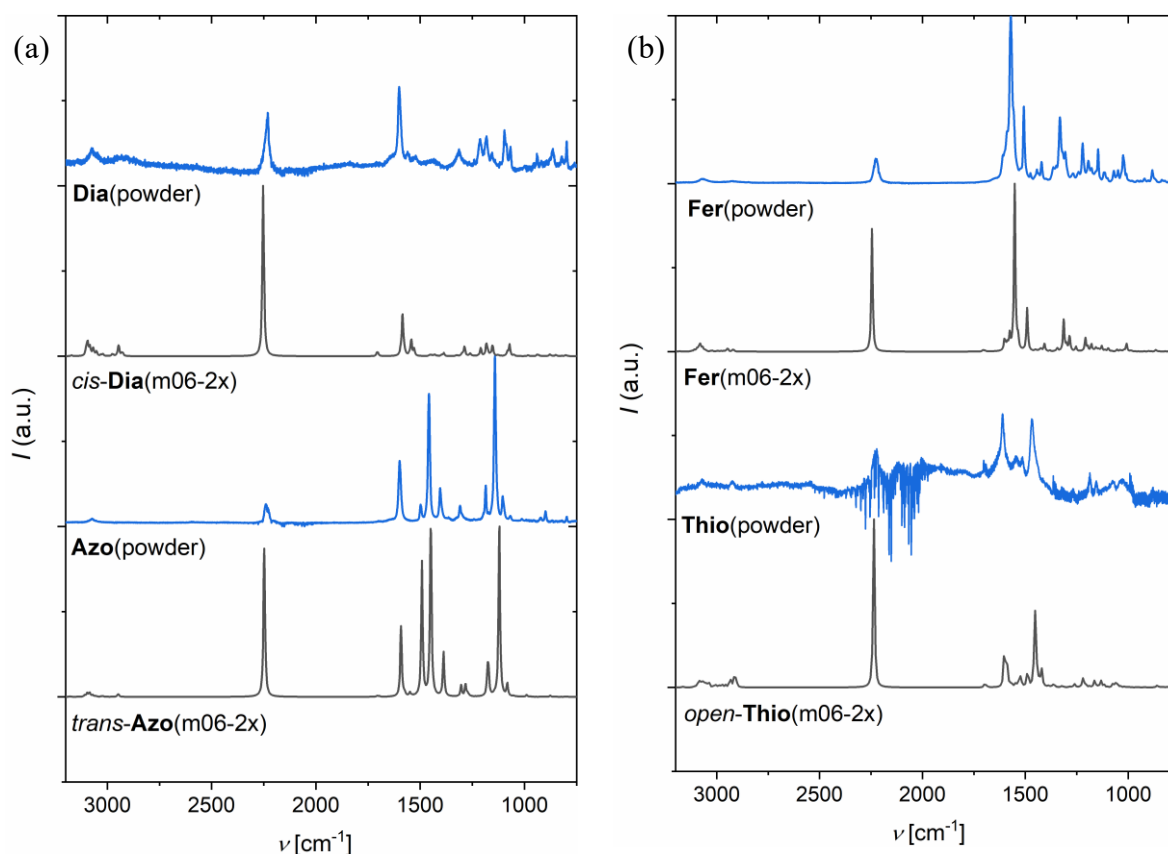
Figure S15. Experimental powder Raman spectra of all investigated systems vs their theoretical analogues calculated in vacuum. Limited quality of Thio spectrum monitored in powder is probably a result of the cooling system failure during the measurement.

Gaussian fitting procedure example

An example of deconvolution procedure of PM-IRRAS spectra by Gaussian fitting, performed in Origin Lab.

Figure S16. Deconvolution of **Dia** system PM-IRRAS peak at 1584 cm^{-1} by Gaussian fitting

S8. Scanning of local minima

Table 17. Film thickness, tilting and interaction energies of different motors and switches

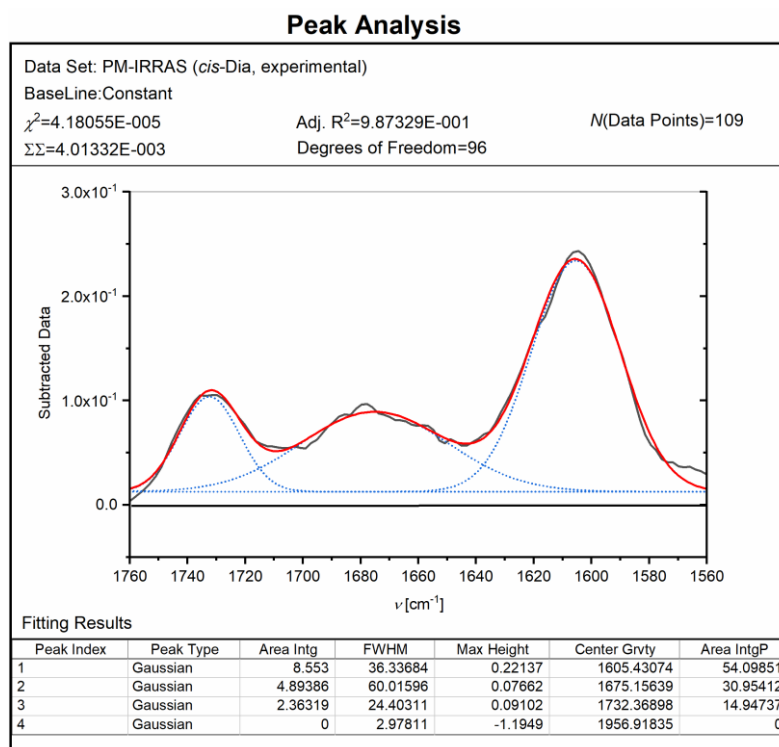| *trans*-**Azo**				*cis*-**Azo**			
d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]	d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]
14.59	69.45	37.90	-54.85	15.44	53.49	60.07	-45.33
17.75	70.36	53.74	-40.96	16.09	65.55	63.40	-44.69
18.48	70.55	60.01	-39.73	16.10	67.16	64.38	-45.33
20.14	72.55	64.58	-35.10	16.06	64.70	62.66	-44.38
21.59	78.60	71.40	-32.63	16.01	65.98	63.14	-44.34
22.30	78.09	75.91	-31.12	16.36	80.98	76.64	-44.12
22.46	78.04	78.71	-32.18	16.62	82.47	84.13	-43.03
22.68	80.32	84.04	-32.18	16.07	85.26	82.20	-45.62
22.44	92.73	93.89	-32.42	16.06	89.00	87.00	-45.69
cis-**Dia**				*trans*-**Dia**			
d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]	d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]
16.51	63.58	46.62	-49.88	13.49	60.56	46.68	-51.45
16.67	72.36	59.67	-48.46	13.97	62.72	47.06	-51.87
16.64	69.36	57.75	-48.23	16.25	69.61	57.16	-22.83
17.00	69.56	56.98	-45.94	14.78	66.68	47.88	-53.60

16.43	72.16	64.16	-49.28	15.43	67.76	47.02	-54.22
17.13	72.21	65.65	-46.92	19.93	78.52	75.57	-35.86
17.35	77.24	71.23	-45.61	20.65	83.72	81.53	-36.01
17.62	80.99	76.26	-44.70	18.52			-48.71
17.04	85.46	80.06	-44.62	21.35	89.86	87.54	-35.58
			-486.19				

Fer				Thio			
d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]	d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]
17.19			365.67	20.26	73.36	61.29	-63.01
17.41	70.82	59.72	-59.84	20.42	72.48	61.38	-63.08
17.65	69.69	59.42	-59.96	20.67	71.06	61.53	-62.78
18.14	70.13	60.88	-60.34	20.86	71.44	62.51	-62.59
18.65	70.25	62.56	-59.79	20.43	71.66	61.48	-55.60
19.02	71.41	64.30	-59.34	21.45	82.08	74.68	-56.86
20.20	78.73	73.87	-59.30	21.57	87.28	79.67	-56.98
20.77	80.91	77.43	-57.07	21.58	90.52	82.60	-56.93
18.00	69.56	59.65	-57.01	20.84	86.99	76.78	-52.27

Phe1				Phe			
d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]	d [Å]	Θ_{Bot} [°]	Θ_{Up} [°]	E_{film} [kcal·mol^{-1}]
11.38	49.82	50.58	-23.19	12.81	67.15	59.06	-33.96
10.68	44.34	51.19	-28.04	12.73	67.81	58.18	-39.49
10.67	44.11	54.00	-28.06	13.23	69.58	58.96	-39.52
14.13	83.31	76.67	-22.19	13.34	71.69	61.87	-39.33
14.10	83.04	76.12	-22.04	14.05	72.32	67.66	-38.06
14.07	82.25	72.82	-22.12	14.43	75.54	73.12	-37.87
14.19	83.11	80.47	-21.96	14.75	77.77	78.94	-38.09
14.27	84.04	83.85	-21.80	14.86	81.33	84.39	-37.71
14.44	83.69	82.79	-22.16	14.81	87.23	87.35	-37.57

Tri				
d [Å]	Θ_{Bot} [°]	Θ_{Mid} [°]	Θ_{Top} [°]	E_{film} [kcal·mol^{-1}]
18.29	62.70	32.24	49.94	-60.09
18.17	61.24	32.25	49.47	-59.43
20.24	62.34	63.79	54.30	-46.79
20.73	64.61	67.21	57.96	-46.74
21.43	70.68	78.44	66.48	-48.54
21.82	76.99	82.22	75.23	-49.41
21.94	79.76	83.83	79.35	-49.38
21.97	81.68	84.91	81.53	-49.37
21.91	85.94	87.93	85.81	-49.26

S9. GFN-FF method

Table S12. Choices of custom potential used during the optimisation and simulation steps in GFN-FF method

APM [Å^2]	H [Å]	R [Å]	$R_{\log}$ [Bohr]	R_{pol} [Bohr]	$N(H_2O)$
100.0	16.00	19.54	38.82	42.60	640
87.5	18.29	18.28	36.44	40.22	640
75.0	21.33	16.93	33.87	37.65	640
62.5	25.60	15.45	31.09	34.87	640
50.0	32.00	13.82	28.01	31.78	640
65.0	24.62	15.76	31.67	35.45	640
55.0	29.09	14.49	29.28	33.06	640
45.0	35.56	13.11	26.67	30.44	640
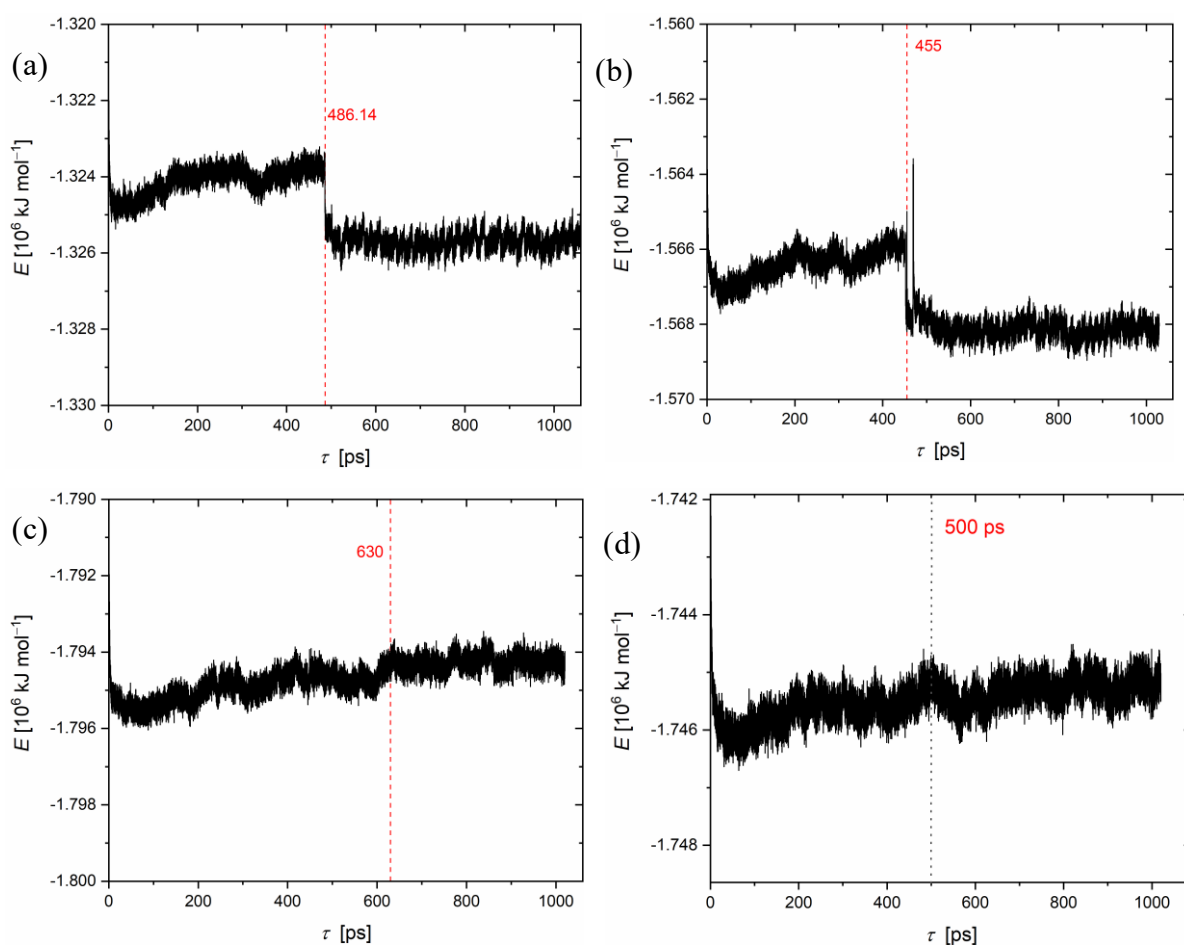
Table S13. Examples of energy profiles of other four simulated systems: (a) **Phe**, (b) **Dia**, (c) **Fer**, (d) **Thio**. Presented were simulated at their experimental APM. At other APMs, the monolayer destruction occurs even faster.

§ 11. CURRICULUM VITAE

Personal Information

Name and surname: Lujo Matasovic

Date of birth: 14. July 1996.

Place of birth: Zagreb

Education

2003–2011	OŠ Izidor Kršnjavi, Zagreb
2011–2015	Klasična gimnazija, Zagreb
2015–2019	Undergraduate studies in chemistry, Department of Chemistry, Faculty of Science, University of Zagreb
2021	International Student Exchange [ERASMUS+ programme, *Computational design of supramolecular building blocks*, Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, Prague, Czech Republic]

Honours and Awards

2021	Rector's award, *Tuning optoelectronic properties of di- and trinitrosoarene derivatives*, University of Zagreb, Croatia
2021	Medal of the Department of Chemistry, Faculty of Science, University of Zagreb, Croatia
2021	ERASMUS+ Scholarship
2017–2018	EU scholarship for excellent students in STEM fields

Activities in Popularization of Science

2015–2019	Open Chemistry Day, *active participation*
2015–2019	Simposium of Chemistry Students, *organiser*
2018	Computational Chemistry Day, *organiser*

Participation at Scientific Meetings and Conferences

1. **L. Matasović**, H. Vančik, I. Biljan, I. Rončević, *Tuning the optoelectronic properties of dinitrosoarene polymers*, Math/Chem/Comp 2021, Dubrovnik, 2021, page 40.

Scientific Publications

1. **L. Matasović**, B. Panić, M. Bubaš, H. Vančik, I. Biljan, I. Rončević, *J. Mat. Chem. C.*, **10** (2022) 5433−5446.