Accelerated Design of Near-Infrared-II Molecular Fluorophores via First-Principles Understanding and Machine Learning

Shidang Xu[+], Pengfei Cai[+], Jiali Li, Xianhe Zhang, Xianglong Liu, Xiaonan Wang[*] and Bin Liu[*]

Dr. S. Xu, Mr. J. Li, Mr. X. Zhang, Mr. X. Liu, Dr. X. Wang, Prof. B. Liu
Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore
E-mail: chewxia@nus.edu.sg
E-mail: cheliub@nus.edu.sg

Mr. P. Cai
Department of Materials Science and Engineering, National University of Singapore, 9 Engineering Drive 1, 117575, Singapore

Prof. B. Liu
Joint School of National University of Singapore and Tianjin University, International Campus of Tianjin University, Binhai New City, Fuzhou 350207, China

Organic molecular fluorophores in the second near-infrared window (NIR-II) have attracted much attention in the recent decade due to their great potentials in both fundamental research and practical applications. This is especially true for biomedical research, owing to their deep light penetration depth and low bioluminescence background at the long wavelength. However, the fluorescence quantum yields (QY) of most NIR-II materials are very low, which are not ideal for practical applications. Although there is a growing need to discover new NIR-II fluorophores, most of them were designed based on experience, and the structures were limited to few molecular motifs. Herein, we report the design of high QY NIR-II fluorophores in solutions based on enhancing the rigidity of the conjugated backbones, which could be quantified by the Seminario method. A deep neural network was trained to predict the HOMO-

LUMO energy gaps for a chemical library of NIR-II backbone structures. Hundreds of new NIR-II cores with low E_g were discovered, and eight of them across different acceptor cores are found to have relatively rigid conjugated backbones. With further molecular processing or formulation, the proposed new fluorophores should boost the development of NIR-II materials for applications in a wide range of fields.

1. Introduction

Organic molecular fluorophores in the second near-infrared window (NIR-II, 1000−1700 nm) have attracted much attention in fundamental research and biomedical applications.[1-4] Because of their low molecular mass, superior biocompatibility, and flexible molecular design, NIR-II organic dyes are more promising for biological applications than inorganic contrast agents.[5-8] In both research and clinical applications, NIR-II fluorescence imaging provides a spectacular imaging modality for disease diagnosis and image-guided surgery.[2,9-11] Fluorescence imaging does not require or emit hazardous radiations, unlike clinical modalities such as positron emission tomography, computed tomography, and single-photon emission computed tomography. However, due to the high absorption, scattering, and autofluorescence of biological tissues, the primary bottleneck in the development of fluorescent imaging lies in the shallow tissue penetration depth and poor signal-to-background ratio.[12-17] In the recent decade, fluorescence imaging in the NIR-II region has shown great potentials owing to its higher resolution at larger penetration depth (ca. 5−20 mm) in biological tissues as compared to that of light in the visible (400−700 nm) and first near-infrared (NIR-I, 700−900 nm) regions.[15, 18-23]

Fluorophores with NIR-II emission are under intense research interest because they are imperative for NIR-II imaging. Recent review papers have summarized the rational design,[24-26] HOMO/LUMO engineering, and strategies to modify NIR-II molecules with high brightness in different environments,[27-29] such as aqueous media and aggregate state. These molecules are generally limited to the benzo[1,2-c:4,5-c′]bis[1,2,5]thiadiazole (BBTD) core,[9, 24, 30-32] polymethine skeleton[27, 33-37] and few examples have been developed based on BODIPY,[38, 39]

squaraine,[40] as well as rhodamine.[41] It is difficult to design high-performance NIR-II fluorophores due to high vibronic coupling caused nonradiative decay between S_1 and S_0 states. Moreover, multiple factors, such as resonant energy transfer, π-π interactions, conjugated chain length, chain length alternations, have been reported to affect the performance of NIR-II fluorophores.[5, 36, 42] These factors further increase the difficulty in designing high-performance NIR-II molecules. It is critical if one could prioritize multiple factors in the design of NIR-II fluorophores and more precisely quantify them, and in turn, offer a clear design strategy towards many new NIR-II fluorophores. Detailed literature review reveals that shielding of excited center and reduction of π-π interaction could help boost NIR emission. [26, 36, 42]

To facilitate the design of NIR-II fluorophores, we summarized the current strategies and proposed rigidity of conjugated backbones as one important parameter for yielding high-performance NIR-II fluorophores. Owing to the strict HOMO-LUMO energy gap (E_g) requirement, NIR-II molecular fluorophores are mainly focused on two types of architectures, namely, donor–π-acceptor-π–donor (DπAπD) fluorophores and D$^+$$\pi$D (heterocyclic D$^+$ with positive charge serving as the acceptor). The design of new fluorophores is therefore reliant on D/A/π engineering. We chose to focus on rigidity engineering and further propose to quantify the molecular rigidity by using Seminario and Compliance methods. All the QYs (Φf) of selected molecules in the monomolecular state (solvent) show good agreement with the rigidity calculations.

To design new NIR-II fluorophores with high rigidity, we also trained a deep neural network capable of predicting E_g using a molecular graph convolutional neural network. We

performed predictions on a chemical library made up of NIR-II backbone structures in DA and DAD arrangements. As a result of accurate prediction and accelerated screening, hundreds of new NIR-II cores with low E_g are discovered. Of which, eight of them are quantified to have relatively rigid conjugated backbones. To ensure the performance of the newly designed NIR-II fluorophores in practical applications, strategies to prevent the two primary quenching effects, excited center-polar media (e.g. water) interaction and π-π interaction are also summarized. With further molecular processing or formulation, all the proposed new fluorophores should boost the development of NIR-II materials in a wide range of fields.
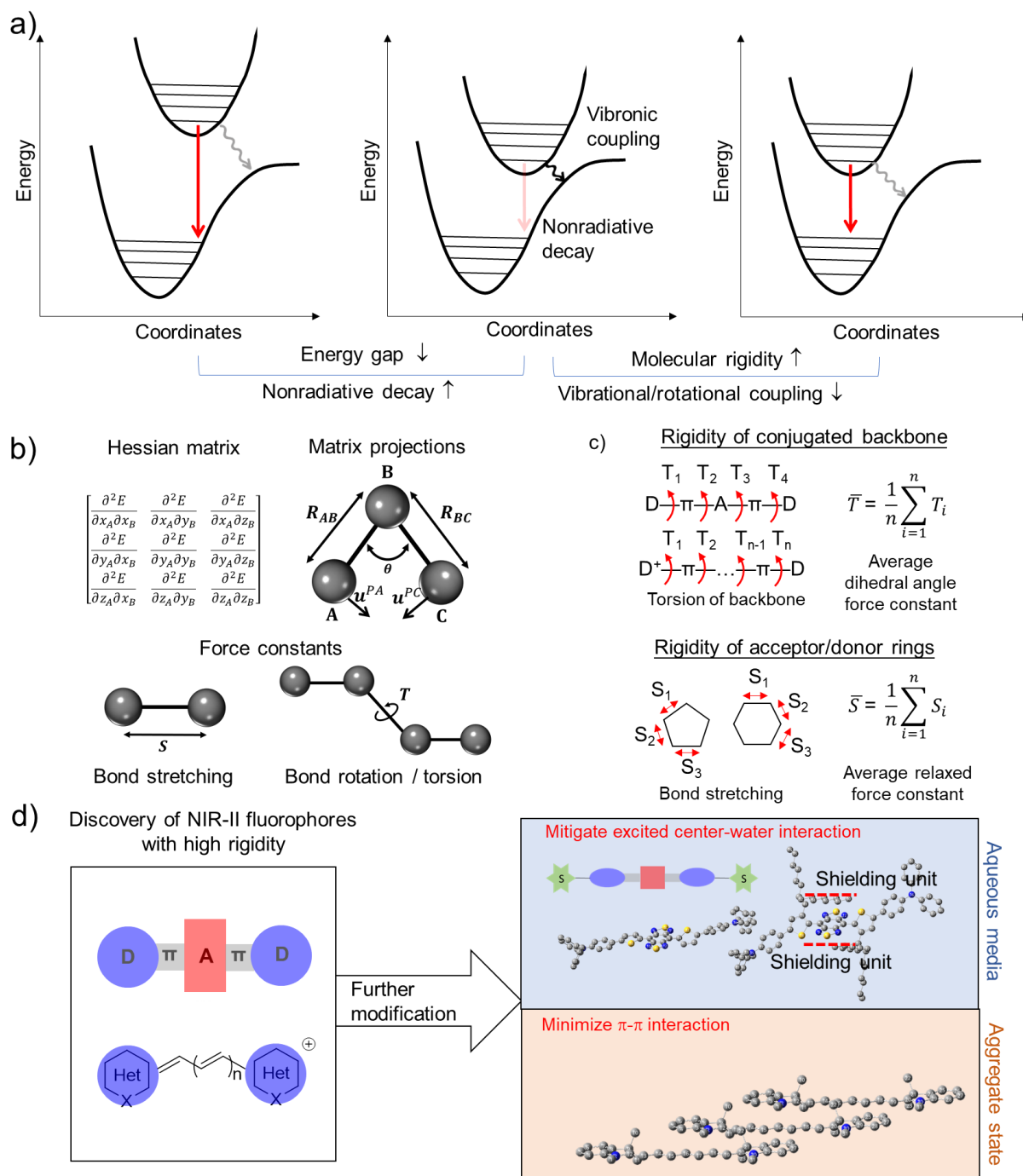
2. Results and Discussion

2.1. Molecular Design of NIR-II Fluorophore

The difficulty in designing NIR-II molecules with high Φf lies in their very high non-radiative decay which can be attributed to the strong vibronic coupling between S_0 and S_1 states, although other factors may also play a role. With a small energy gap between S_0 and S_1, the low vibronic state of S_1 can decay to the high vibronic state of S_0 efficiently via vibrational relaxation. The longer the absorption wavelength for a molecule, the higher the chance for the molecule to show low Φf. One solution to resolve the conflict between the low S_0-S_1 energy gap and high Φf is to restrict the vibrational relaxation from S_1 to S_0, which can be realized through the engineering of molecular rigidity. As the vibrational energy relaxation of the NIR-II molecule in its excited state is dependent on a vibronic coupling process, increasing molecular rigidity, especially the rigidity of the core structure, could effectively deviate the intramolecular motion of S_1 and thereby increase the Φf of the molecule (Figure 1a). In addition,

to achieve efficient S_1-S_0 transition, the configuration of S_1 should preferably be π-π* instead of n-π* as the latter is spatially forbidden.

The next question is how to quantify the molecular rigidity of NIR-II fluorophores. Since most NIR-II fluorophores are composed of conjugated backbones formed by donor, acceptor, and π-bridge, the rigidity of any molecule can be evaluated by calculating the rigidity of both rings involved and the conjugated backbone structures. The rigidity of a ring can be represented by the amount of bond stretching of all the bonds involved and this can be quantified with average relaxed force constants ($\bar{S}$, Figure 1c). Herein, the Compliance method [46, 47] is used to derive individual relaxed force constants (S) for bond stretching between 2 adjacent atoms. Compliance constants are elements of an inverted Hessian matrix and the relaxed force constants can be obtained by taking the reciprocals of the compliance constants [47, 48]. On the other hand, the rigidity of a conjugated backbone can be represented by the average dihedral angle force constant ($\bar{T}$) to quantify the amount of bond torsion or rotation (Figure 1c). The Seminario method allows the derivation of force constants through diagonalization of the partial Hessian matrix [49], and herein, only the individual dihedral angle force constants (T) are of focus. For both Seminario and Compliance methods to work, a structure's Hessian matrix (Cartesian 2nd order partial derivatives of a molecule's energy function with respect to its geometric coordinates [46]) is imperative. With density functional theory (DFT), the Hessian matrix can be computed with vibrational frequency calculations on a geometry optimized structure. Both geometry optimization and vibrational frequency calculations were completed with the same method (B3LYP hybrid functional and 6-31g(d) basis set) in this work. It is

6

important to note that each group of the examples selected in this study has similar backbone

structures so that the contributions of other factors that influence Φf could be normalized to

highlight the importance of molecular rigidity in designing high-performance NIR-II

fluorophores.

a)

Energy

Coordinates

Energy

Vibronic coupling

Nonradiative decay

Coordinates

Energy

Coordinates

Energy gap ↓

Nonradiative decay ↑

Molecular rigidity ↑

Vibrational/rotational coupling ↓

b)

Hessian matrix

$$\begin{bmatrix} \dfrac{\partial^2 E}{\partial x_A \partial x_B} & \dfrac{\partial^2 E}{\partial x_A \partial y_B} & \dfrac{\partial^2 E}{\partial x_A \partial z_B} \\ \dfrac{\partial^2 E}{\partial y_A \partial x_B} & \dfrac{\partial^2 E}{\partial y_A \partial y_B} & \dfrac{\partial^2 E}{\partial y_A \partial z_B} \\ \dfrac{\partial^2 E}{\partial z_A \partial x_B} & \dfrac{\partial^2 E}{\partial z_A \partial y_B} & \dfrac{\partial^2 E}{\partial z_A \partial z_B} \end{bmatrix}$$

Matrix projections

B

R_{AB} R_{BC}

θ

u^{PA} u^{PC}

A   C

Force constants

s

Bond stretching

T

Bond rotation / torsion

c)

Rigidity of conjugated backbone

T_1 T_2 T_3 T_4

D—π—A—π—D

T_1 T_2 T_{n-1} T_n

D^+—π—…—π—D

Torsion of backbone

$$\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$$

Average dihedral angle force constant

Rigidity of acceptor/donor rings

S_1

S_2

S_3

S_1

S_2

S_3

Bond stretching

$$\bar{S} = \frac{1}{n}\sum_{i=1}^{n} S_i$$

Average relaxed force constant

d)

Discovery of NIR-II fluorophores with high rigidity

D π A π D

Het X   Het X ⊕
       n

Further modification

Mitigate excited center-water interaction

S   S

Shielding unit

Shielding unit

Aqueous media

Minimize π-π interaction

Aggregate state

7

Figure 1. Molecular rigidity for the design of high-performance NIR-II fluorophores. a) Schematic illustration of the importance of molecular rigidity on the design of high-performance NIR-II fluorophores. b) Schematic illustration on how relaxed force constant (S) for bond stretching and dihedral angle force constant (T) for bond torsion are related to the individual atoms and are derivable from the Hessian matrix; u refers to bond vector; R_{AB} and R_{CB} are the two bond lengths; θ is equilibrium bond angle. c) Rigidity quantification for conjugated backbone and acceptor/donor rings with average dihedral angle force constants ($\bar{T}$) and average relaxed force constants ($\bar{S}$), respectively. d) Molecular engineering strategies for the modification of NIR-II fluorophores in a high-polarity environment and aggregate state.

Several factors could cause the quenching of excited fluorophores such as π-π interaction in aggregate state and excited state center-water interaction in an aqueous medium[42]. For instance, owing to the high-energy vibrational band of -OH in the NIR region, a resonant energy transfer process with water molecules will dominate the excited-state decay of NIR-II fluorophores. This explains why many NIR-II fluorophores show severe quenching effects in aqueous media. To alleviate the excited center-water interaction, a shielding unit, particularly an alkyl or alkoxy chain that surrounds the backbone, could isolate the excited state NIR-II backbone from surrounding water molecules and thus favours fluorescence emission. The effectiveness of the shielding unit is affected by its position, size, and orientation (Figure 1d).

When NIR-II molecules are formulated into nanoparticles, they tend to form face-to-face π-π stacking and result in intermolecular interaction to cause exciton quenching.[26] To address this issue, one way is to design the backbone structure with certain distortion, which can reduce the face-to-face π-π interaction. The other way to prevent closed π stacking is to introduce bulky groups or aggregation-induced emission (AIE) molecules around the emission center, which could yield high steric hindrance around the backbone to effectively prevent close molecular packing. Nonetheless, fine-tuning the molecular structure for optimized

performance requires the fundamental design of backbones that fall within the NIR-II region with good potentials to be a high-performance fluorophore.

2.2. Molecular Strategies towards Rigid NIR-II Conjugated Backbone
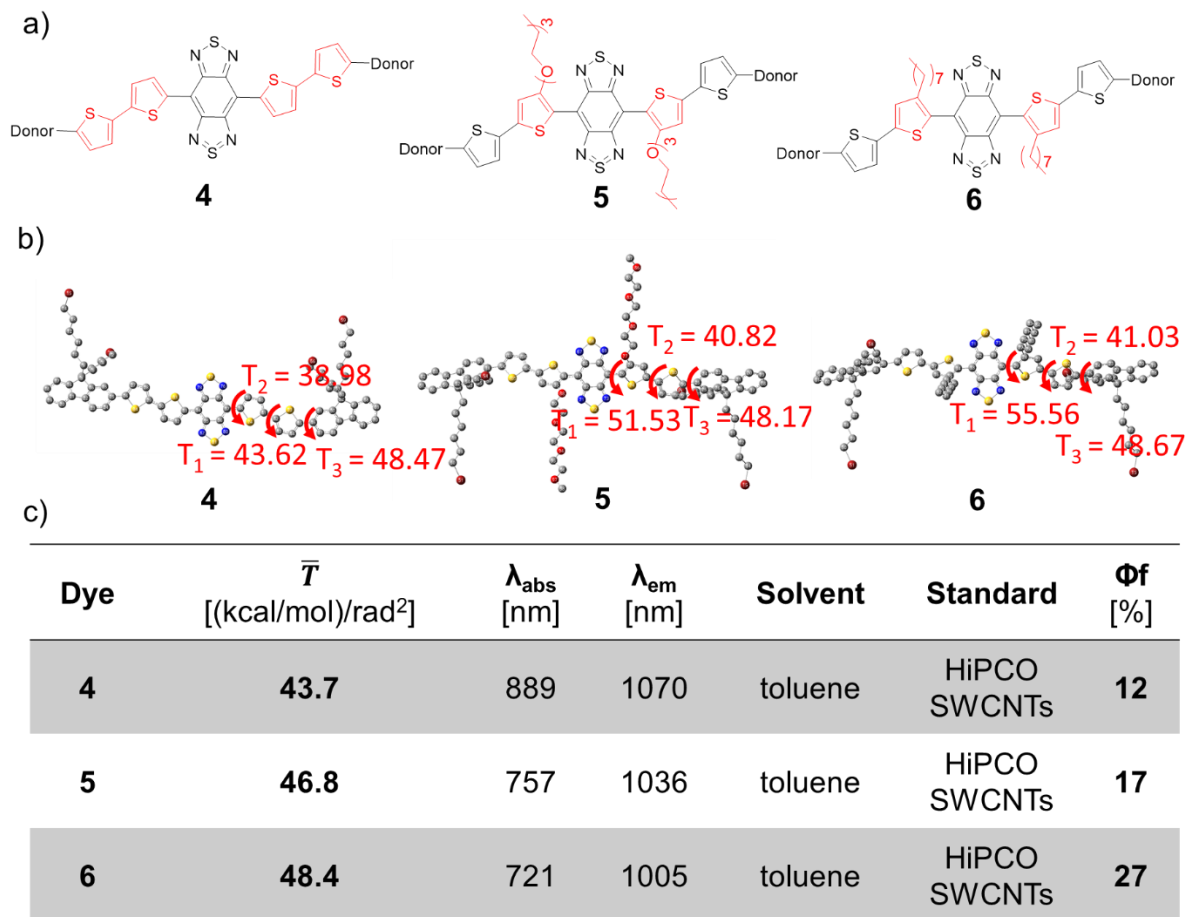
2.2.1. π -Bridge Engineering on DπAπD NIR-II Fluorophore

Compounds **1**, **2**, and **3** represent DπAπD NIR-II fluorophores with different π bridges (Figure 2). According to dihedral force constant calculation on the two rotatable single bonds along the conjugated backbone, compounds **2** and **3** with 3,4-propylenedioxy thiophene (PDOT) ring show relatively higher torsional forces, leading to more rigid conjugated backbones as compared to **1**. In compound **3**, the additional alkyl chains cause the increase of dihedral force constants along all rotatable single bonds compared. As a result, **3** has the highest Φf of 39% while **1** has the lowest Φf (19%).[43, 44] These results reveal the importance of the dihedral force constant ($\bar{T}$) and rigidity of conjugated backbone in the design of high-performance DπAπD NIR-II molecules.
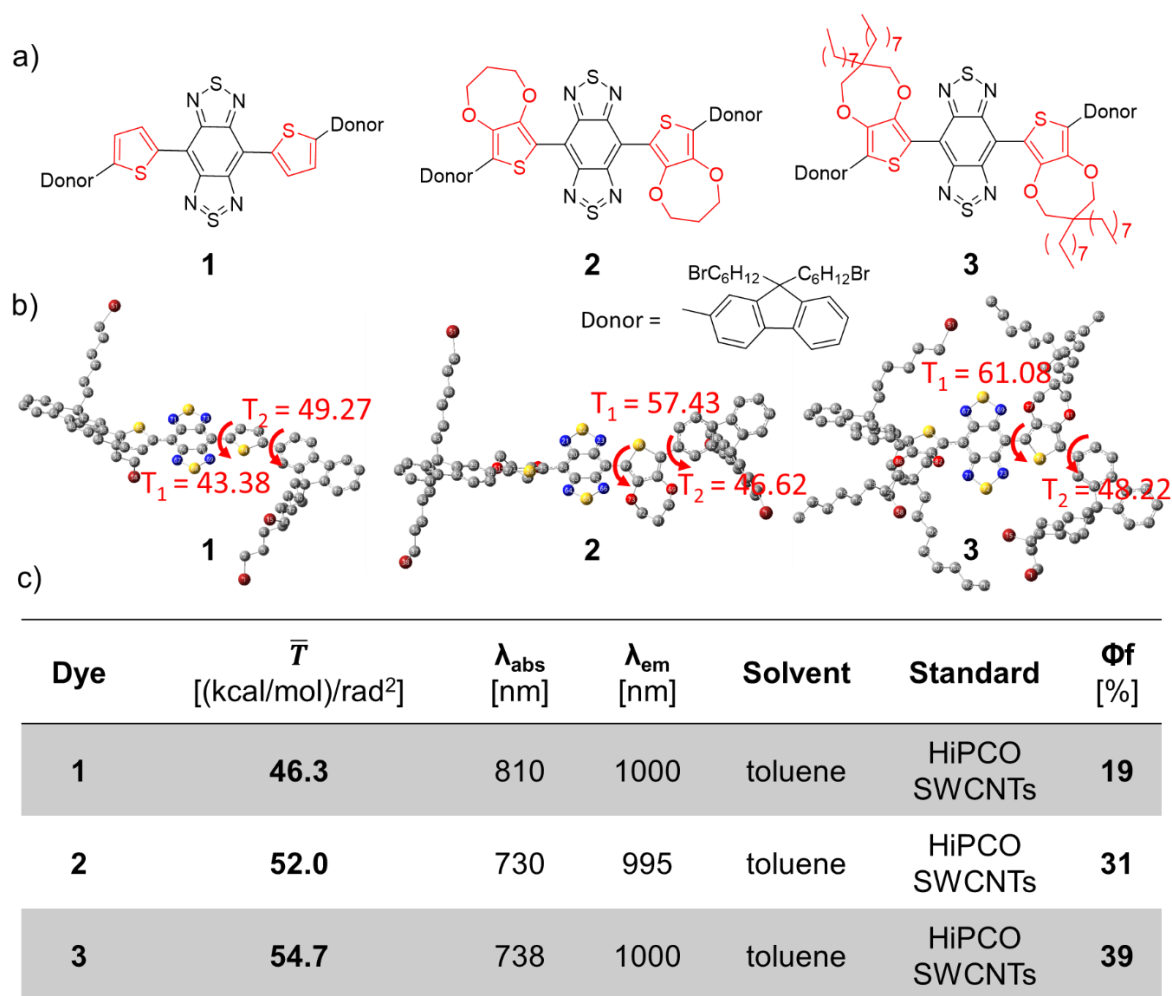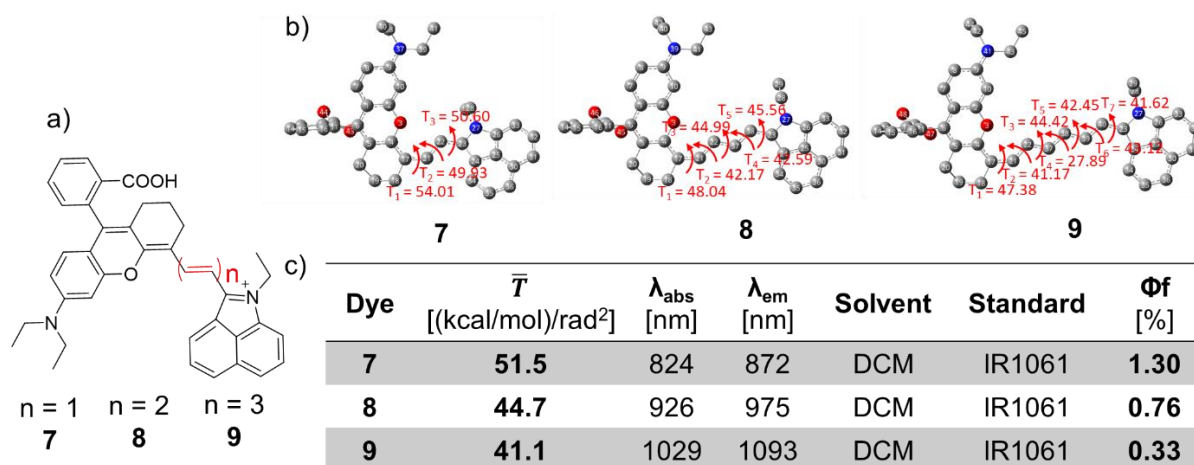
Figure 2. π-Bridge engineering on DπAπD NIR-II fluorophores. a) Molecular structures and b) optimized structures of compounds **1**-**3**. c) Force constants and photophysical properties[43, 44] of **1**-**3**. HiPCO SWCNTs: pressure carbon monoxide conversion single-walled carbon nanotubes.

Dye	$\overline{T}$ [(kcal/mol)/rad^2]	λ_{abs} [nm]	λ_{em} [nm]	Solvent	Standard	Φf [%]
1	46.3	810	1000	toluene	HiPCO SWCNTs	19
2	52.0	730	995	toluene	HiPCO SWCNTs	31
3	54.7	738	1000	toluene	HiPCO SWCNTs	39

With an additional thiophene on the conjugated backbone, compound **4** (Figure 3) shows a weaker dihedral force constant as compared with **1**. Compared with **5** and **6**, compound **4** has the lowest value of $\overline{T}$, and it also has the lowest Φf of 12%.[43, 44] Compounds **5** and **6** show relatively higher dihedral force constant by introducing alkoxyl chain and alkyl chain to the thiophene rings. As a result, the dihedral angle force constants between acceptor and π bridge reach 51.53 and 55.56 kcal·mol^{-1}·rad^{-2} for **5** and **6**, respectively, which are much higher than

that of compound **4** (43.62 kcal·mol^{-1}·rad^{-2}). Such a result reveals the effectiveness of π bridge engineering on increasing molecular rigidity and improving fluorescence efficiency.

Figure 3. π-Bridge engineering on DπAπD NIR-II fluorophores. a) Molecular structures and b) optimized structures of compounds **4-6**. c) Force constant and photophysical properties[43, 44] of **4-6**.

Dye	$\overline{T}$ [(kcal/mol)/rad^2]	λ_{abs} [nm]	λ_{em} [nm]	Solvent	Standard	Φf [%]
4	43.7	889	1070	toluene	HiPCO SWCNTs	12
5	46.8	757	1036	toluene	HiPCO SWCNTs	17
6	48.4	721	1005	toluene	HiPCO SWCNTs	27

2.2.2 π-Bridge Engineering on D π A NIR-II Fluorophore

Conjugated double bonds are the most widely used π bridge in the design of DπA NIR fluorophores. For compounds **7–9**, they have the same electron donor and acceptor on both sides of double bonds. The average dihedral angle force constants of the bonds, $\overline{T}$, decrease (Figure 4) with the increase in the number of conjugated double bonds. When n = 3, $\overline{T}$ for the central double bond (C37-C38) is as small as 27.89 kcal·mol^{-1}·rad^{-2}. $\overline{T}$ values for compounds

7-9 are 51.5, 44.7, and 41.1 kcal·mol^{-1}·rad^{-2}, respectively. Clearly, the decrease of Φf from 1.3% to 0.33% agrees well with the decrease in backbone rigidity for **7-9**.[45] In general, the extension of conjugation by increasing the number of conjugated double bonds decreases both the HOMO-LUMO energy gap and molecular rigidity.

Figure 4. π-Bridge engineering on DπA NIR-II fluorophores. a) Optimized structures and b) molecular structures of compounds **7–9**. c) Force constant and photophysical properties[45] of **7–9**. DCM: dichloromethane.

Dye	$\overline{T}$ [(kcal/mol)/rad^2]	λ_{abs} [nm]	λ_{em} [nm]	Solvent	Standard	Φf [%]
7	51.5	824	872	DCM	IR1061	1.30
8	44.7	926	975	DCM	IR1061	0.76
9	41.1	1029	1093	DCM	IR1061	0.33

2.2.3 Donor Engineering on D$^+$πD NIR-II Fluorophore

Compounds **10-11** belong to the D$^+$πD type of NIR-II molecules with different D$^+$ (Figure 5). The primary difference between compounds **10** and **11** is the size of the ring conjugated with π bridges. These rings are constructed with the double π bridge to form the conjugated backbone. As such, we analyzed the rigidity of the conjugated backbone by calculating the dihedral angle force constant between D$^+$ and D. Compound **10** shows relatively higher rigidity of backbone as compared with **11**, indicating that cyclopentane forms a more rigid conjugated double bond structure than cyclohexane. Since cyclopentane and cyclohexane rings are the main component of acceptor for compounds **10** and **11**, the ring rigidity was further analyzed

by calculating the relaxed force constants of the involved bonds. The average relaxed force constants ($\bar{S}$) of compounds **10** and **11** are 4.52 and 4.51 mdyn·Å^{-1}, respectively (Figure 5b). Considering that sulfur atom could lead to heavy-atom caused fluorescence quenching, the much higher Φf [36] of **10** as compared to **11** can be attributed to its higher backbone rigidity.
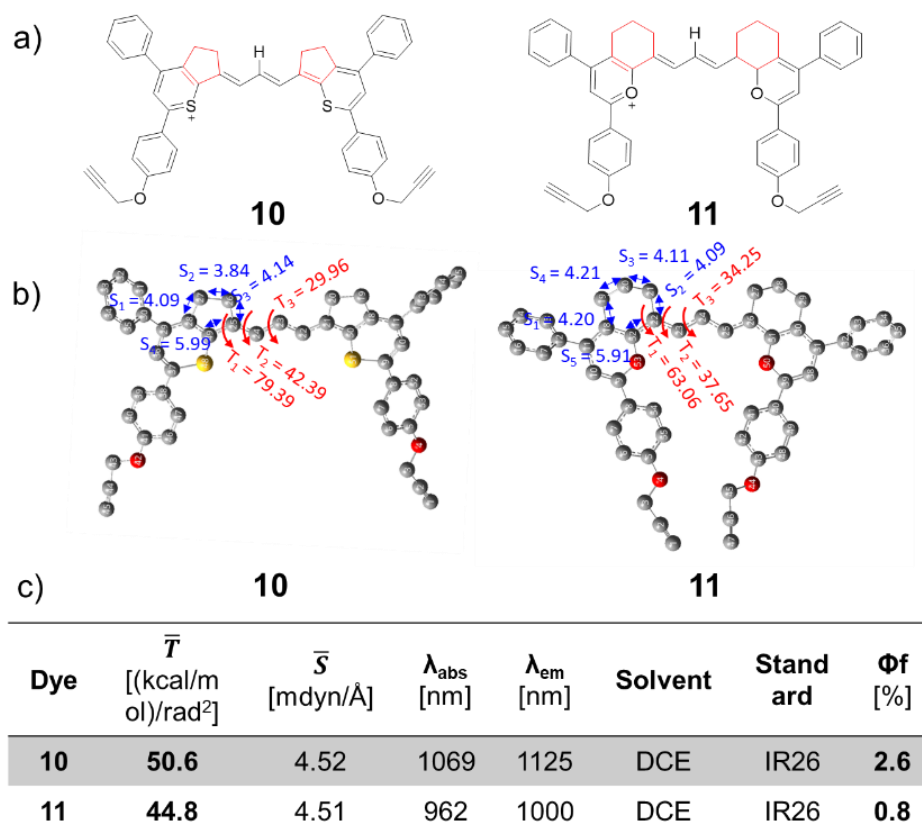
Figure 5. Donor engineering on D$^+$πD NIR-II fluorophores. a) Molecular structures and b) optimized structures of compounds **10-11**. c) Force constant and photophysical properties [36] of **10-11**. Average bond stretching force constant ($\bar{S}$). DCE: 1,2-dichloroethane.

Dye	$\bar{T}$ [(kcal/mol)/rad^2]	$\bar{S}$ [mdyn/Å]	λ_{abs} [nm]	λ_{em} [nm]	Solvent	Standard	Φf [%]
10	50.6	4.52	1069	1125	DCE	IR26	2.6
11	44.8	4.51	962	1000	DCE	IR26	0.8

2.3. Design of NIR-II Conjugated Backbone by Virtual Screening

Based on the understanding of the structure-property relationship between NIR-II conjugated backbone and fluorescence efficiency, virtual screening of NIR-II cores with DA and DAD structures based on HOMO-LUMO energy gap (E_g) was conducted using a deep

neural network based active learning strategy [57] to expedite the search of low energy gap structures.

Firstly, based on a molecular space generation approach [57], the full chemical library of 11,745 structures (6,210 DA and 5,535 DAD) was generated through iterative combinations of a list of starting backbone components. A total of 9 donors, 46 acceptors, and 15 bridges (14 π and 1 single bond) were selected either from literature or proposed as new components for NIR-II molecular backbones (the full list is available in Figures S1-S3). A DA (or DπA) structure was formed by combining a donor with an acceptor molecule, with either a single bond or a π bridge in between, at designated respective atomic positions. A DAD (or DπAπD) structure was formed similarly but with two identical donor and bridge components bonded to two atomic positions of a central acceptor. The combination algorithm adopted the RDKIT [54] python package for conversions from simplified molecular-input line-entry systems (SMILES) to molecular structures and bond connections between component structures.

Next, to learn the structure-property relationship between the NIR-II molecular backbone and its HOMO-LUMO energy gap (E_g), a deep neural network was trained to predict E_g. Herein, a molecular graph convolutional neural network [51] was used for the prediction of E_g with a molecular structure as an input. In our dataset, SMILES was used to represent the input molecular structures and the E_g label values for all structures were computed with structural optimizations via quantum calculations based on DFT. A preliminary model was first trained on an initial dataset of 673 structures that were randomly selected from the full chemical library. To accelerate the search of promising NIR-II cores with low E_g from the whole molecular search space, a Bayesian optimization-based active learning [52] strategy adapted from our

14

previous work [57] was used herein. In each active learning cycle, a small subset of 20 new structures with the highest expected improvement score, based on a tradeoff between exploitation and exploration of the search space, was suggested for labeling by DFT before they were added to the current training set for the next cycle (Figure 6). This strategy allowed the model to improve its accuracy in the target region of the entire search space, specifically structures with low E_g, while requiring fewer structures to be labeled [52, 57].
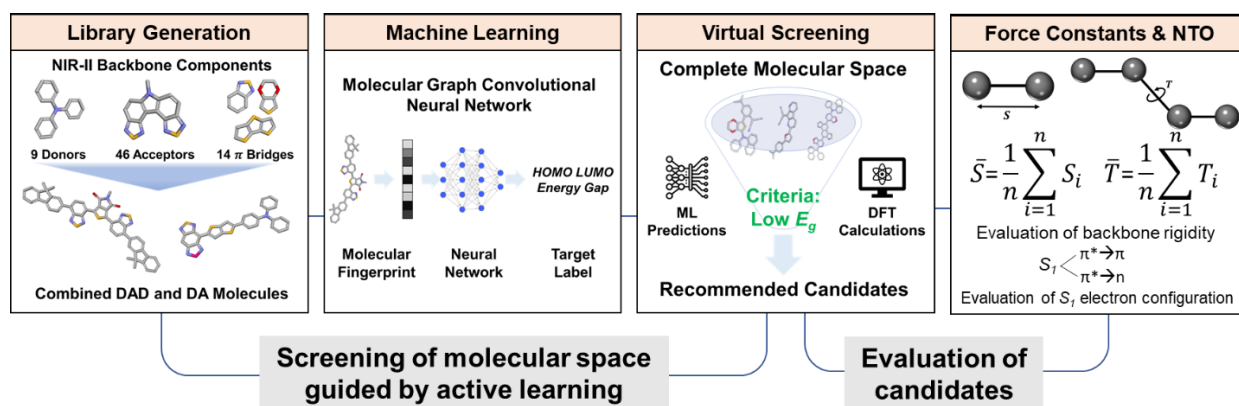
Figure 6. Schematic overview of virtual screening of molecular space guided by active learning with a molecular graph convolutional neural network used for prediction of energy gaps. The evaluation of recommended candidates is based on force constants ($\bar{T}$ and $\bar{S}$) and natural transition orbitals (NTO) computations.

After the completion of 5 active learning cycles, a total of 773 NIR-II fluorophore cores were labeled with their E_g values. A fixed test set of 135 structures (making up 20% of the initial random dataset) was used to evaluate the final model prediction accuracy with mean absolute error (MAE) by fitting the model on the training dataset (638 structures) over 200 epochs (Figure 7a-b). With 5 repetitions of training and evaluation, an average MAE of 0.1315 eV for the final model was obtained. This accuracy was comparable with many DFT methods that have MAE of approximately 0.1 eV for the same type of energy gap predictions for organic

molecules. With the use of the trained deep learning model, the predictions could be accelerated to seconds compared to up to a few hours for a DFT calculation of the same molecule, thus allowing a high-throughput screening of a large molecular space.

After 5 cycles of active learning have been completed, the final model trained on all labeled datasets (i.e. 773 molecules) was used to predict the E_g values for the remaining unlabeled molecular space. A total of 403 structures were predicted to meet the low energy gap criteria of < 1.25 eV. To suggest NIR-II fluorophores with a variety of acceptors, 24 structures (8 sets of 3 structures, with each set consisting of structures with the same acceptor) were selected for comparisons, and their $\bar{T}$ and E_g values were calculated with DFT. In each set, the structure with the highest calculated rigidity (quantified by $\bar{T}$) of the conjugated backbone was finally recommended. A range of structures is well-represented across different absorption wavelengths and rigidity (Figure 7c) and the 8 new cores with high rigidity across different acceptors are presented (Figure 7d, details for each set in Figures S4-S11). With a high rigidity of conjugated backbone, these 8 new cores can be potentially used to form NIR-II molecules with high Φf. Among the 8 of them, 4 cores, namely NDAD_4731, NDA_3155, NDAD_2873, and NDAD_4204, are already in the NIR-II absorption region (Figure 7c-d).
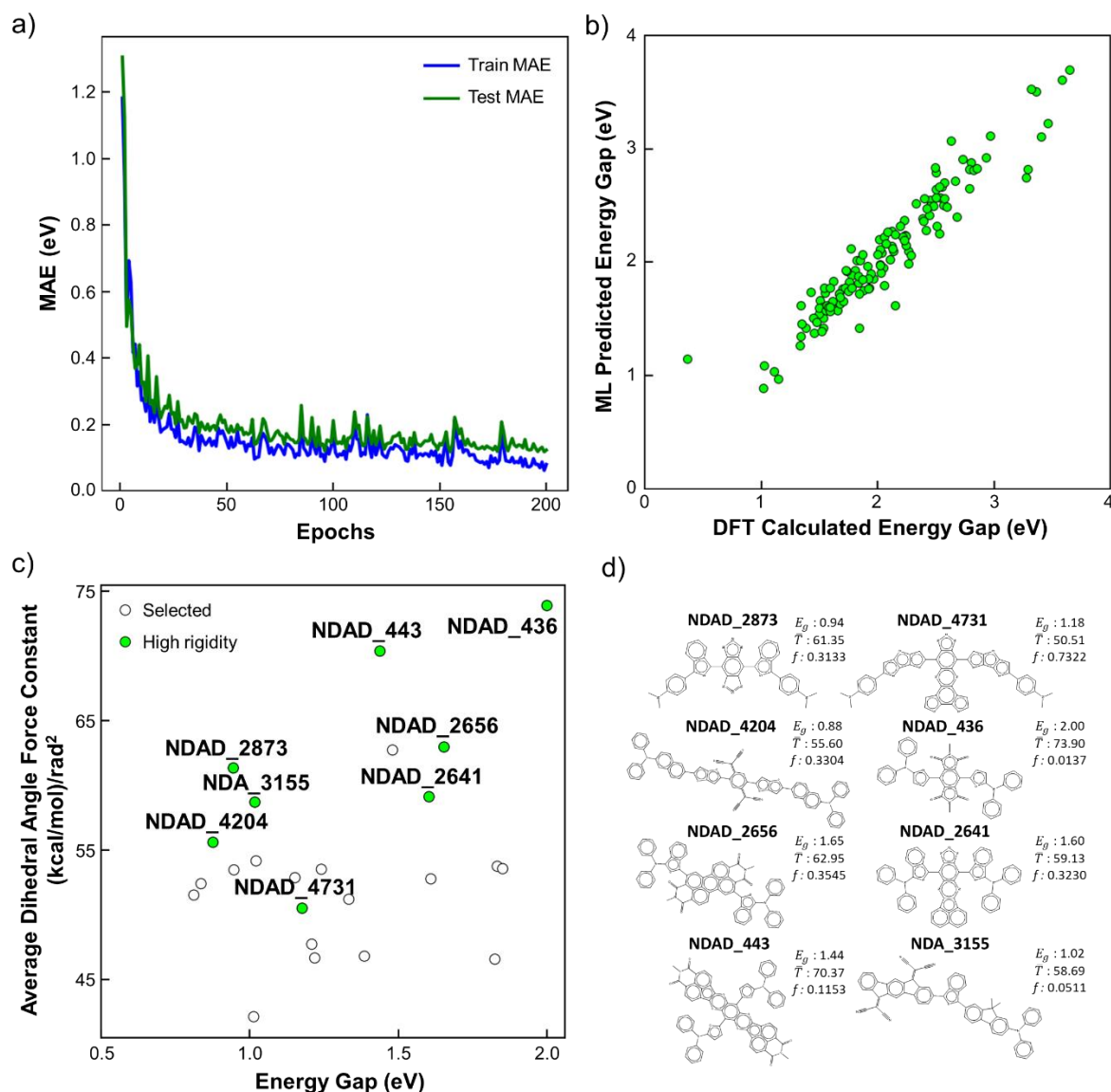
Figure 7. Virtual screening to discover new NIR-II fluorophores with both low E_g and rigid conjugated backbones with the help of a deep neural network and active learning. a) Performance of a final E_g prediction model that was fitted on a training set (638 structures) for 200 epochs and b) E_g predictions on the fixed test set (135 structures) made by the same model compared against DFT calculated values. c) Distribution of average $\bar{T}$ and E_g for the selected 24 NIR-II fluorophore cores predicted by the final model ($\bar{T}$ and E_g values here were calculated with DFT). d) NIR-II fluorophore cores identified with low predicted E_g (in eV), high oscillator strength (f) for high extinction coefficient, and high rigidity of conjugated backbone quantified by $\bar{T}$ (in (kcal/mol)/rad^2).

17

2.4. NIR-II Fluorophore Modification

2.4.1. Mitigate Water-Caused Quenching

For biomedical applications, either the fluorophores are encapsulated into polymer matrices to form nanoparticles, or they are modified to be dispersible in aqueous environment. In both cases, finetuning the structures of NIR-II fluorophores is necessary for optimized optical performance. To effectively reduce the excited NIR-II fluorophore-water interaction, three strategies have been used to improve the hydrophobicity around the excited center. Firstly, the hydrophobicity around the center can be enhanced significantly by π bridge engineering in 3D size/space. Compounds **12-15** were designed with increasing size of π bridge (Table 1). The radial distribution function (RDF) of oxygen atoms in water molecules and the counted number of water molecules versus the radius showed that **15** exhibited the smallest RDF value in three compounds; therefore, there were the least amount of water molecules around the BBTD center of **15**.[43, 44] The increasing hydrophobicity revealed by RDF from **13** to **15** can be attributed to the enhanced protection against water molecules from the increasing 3D size of π bridge and the alkyl chains.
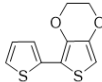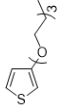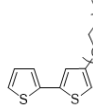
Table 1. a) Molecular skeleton for compounds **12-20**. Properties of b) compounds **12-15** [43, 44] and c) **13**, **16-20** under different conditions.

b)

π bridge	12	13	14	15
Φf_water [%]	0.02	2	4.8	6
Φf_toluene [%]	19	31	31	39
RDF*@8Å	-	**0.55**	**0.36**	**0.25**
n(H₂O)**	-	23.5	12.5	7.5

*RDF: Radical distribution function of oxygen atoms in water at radius of 8 Å around the BBTD acceptor; **number of water molecules around the BBTD center within radius of 8 Å.

c)

π bridge	13	16	17	18	19	20
Φf_water [%]	2	1	-	5.3	1.9	1.4
Φf_toluene [%]	31	15		27	22	17
RDF@8Å	0.55	**0.65**	0.35	**0.63**	-	**0.6**
n(H₂O)	23.5	**32**	12.5	**30**	-	**27.5**
Torsion (°)*	56	**66**	75	**87**	66	**77**

*Torsion between A (acceptor) and S (shieling unit) represented by the sum of dihedral angles between π bridges and A/S of the compound.

Secondly, the hydrophobicity around the center will be affected by the orientation of the shielding chain. As compared with **13**, **17,** and **19**, **16**, **18,** and **20** possess additional thiophene rings within the π bridge, which could increase the torsion between the acceptor and the shielding units on the donor. As shown in Table 1c, the additional thiophene ring increases the sum of dihedral angles by 11-12 degrees for compounds **16**, **18,** and **20**. As a result, **16**, **18,** and **20** show higher hydrophobicity around the center with RDF values at 8 angstroms of 0.65, 0.63, and 0.60, respectively, as compared with **13, 17,** and **19**. A reasonable explanation is that the

alkoxy chain on the donor provides a better shielding effect when there is less torsion between donor and acceptor.

Thirdly, the shielding effect will also be affected by the length/size of shielding units. Compounds **21-23** were designed with shielding units of different chain lengths (Table S2). From the simulation, the RDF values for oxygen atoms in water at 8 angstroms are 0.46, 0.38, and 0.29 for **21**, **22,** and **23**, respectively (Table S2). The increasing hydrophobicity around the center from **21** to **23** can be attributed to the enhanced protection toward water molecules from the increasing size of π bridge around the center. Overall, hydrophobicity around the center plays a very important role in determining the fluorescence Φf of compounds **22-23** in aqueous media, [58] suggesting that the described strategies could be useful in further modification of screened rigid NIR-II fluorophores.

2.4.2. Prevention of π-π interaction-caused quenching

Nanoparticle formation, on the other hand, could cause NIR-II fluorophores to face fluorescence quenching due to π-π interaction in aggregate state. Backbone distortion could alleviate π-π interaction caused quenching in solids. As shown in Figure S14, the dihedral angles between the π bridge and benzo[1,2-c:4,5-c']bis[1,2,5]thiadiazole (BBTD) unit/π bridge of **24** at S_0 state were calculated to be 45/26°, smaller than 56/27° of **25**, indicating that compound **24** exhibits smaller backbone distortion than **25**. As a result, compound **25** shows higher Φf as compared with **24** in aggregate state. Dihedral angles between the π bridge and acceptor unit/donor unit of **26** are 55/25°, smaller than that of **25**, but larger than that of **24**,

resulting in a mediocre Φf of 9.1%.[59] These results suggest that certain backbone distortion favors the fluorescence efficiency of NIR-II fluorophores in nanoparticles.

3. Conclusion

We demonstrate that a high-rigidity conjugated backbone plays an important role in the development of new NIR-II fluorophores with high Φf. Both Seminario and Compliance methods have been used to successfully quantify the rigidity by calculating the average dihedral angle force constants and average relaxed force constants, respectively. Molecular strategies towards the design of rigid conjugated backbone are demonstrated with concrete examples. Based on the good understanding of the structure-property relationship between NIR-II conjugated backbone and fluorescence efficiency, a virtual screening system based on a deep neural network was designed and applied in the design of new NIR-II conjugated backbones. Hundreds of new NIR-II cores with low E_g are discovered, and eight of them have relatively rigid conjugated backbones with the highest $\bar{T}$ among the same series. The proposed strategies toward prevention of the two primary quenching effects, excited center-water interaction and π-π interaction in a different environment, will further raise the usefulness of the screened new fluorophores across a range of applications.

4. Methods

Quantum Mechanics Calculations: Density functional theory (DFT) calculations were performed with B3LYP hybrid functional and a 6-31g(d) basis set. Geometry optimizations and vibrational frequency calculations were performed for all structures with the same method. To build up the training dataset for the energy gap prediction task, geometry optimizations are done with B3LYP hybrid functional and a 6-31g(d) basis set. For the calculation of oscillator strength, excited-state characteristics were calculated by time-dependent density functional theory (TD-DFT) using the optimized ground state geometries based on the B3LYP functional and 6-31g(d) basis set. All calculations are completed with Gaussian16 (Revision A.03) software [53].

Force Constants Calculation: The Hessian matrix is central to both the Seminario and Compliance methods and it can be read from Gaussian output files (specifically fchk and log files) after geometry optimization and vibrational frequency calculations are completed for a given structure at the same level of theory (in this case, B3LYP/6-31g(d)). The Seminario method was used to calculate dihedral angle force constants. For a given dihedral angle along 4 atoms, CartHess2FC.py code from AmberTools17 [50] was used to calculate the dihedral angle force constant by reading the Hessian matrix from a Gaussian formatted checkpoint file (.fchk file). The Compliance method is used to compute each relaxed force constant for the bond stretching between two atoms. Compliance 3.0.2 software [46, 47] is a graphical interface used to compute the Compliance constants for every bond of interest, and in this case, the bonds in an acceptor ring. With the software, the Compliance constants are the diagonal values of the

22

Compliance matrix. Each relaxed force constant was calculated by taking a reciprocal of the respective Compliance constant before an average value is obtained for comparison across molecules.

Machine Learning: To construct the chemical library of DA and DAD structures, a molecular space generation algorithm [57] was used. The python script adopted the RDKIT python package [54] for SMILES conversion and molecular bonding operations. OpenBabel 3.1.1 [55] software was used for the conversion of molecular structures from SMILES to Gaussian input files (in .gjf). The deep learning model made use of a graph convolutional neural network implemented using the DeepChem [56] python package. The model's hyperparameters (Table S1) were used consistently for the initial model, all models throughout the active learning cycles, and the final prediction model. For the evaluation of model performance (by mean absolute error) on the prediction of the energy gap, a train-test dataset split (80% and 20%, respectively) was done on the initial dataset (673 structures) and a fixed test set of 135 structures was used for evaluation. The active learning framework used herein was adapted based on our previous work [57]. In each active learning cycle, the model was trained for 200 epochs on all the labeled dataset (molecular structures and their energy gap values calculated by DFT) so far (i.e. the initial model was trained on 673 structures while the subsequent model after cycle 1 was trained on 693 structures with the increment of 20 new structures recommended by the first cycle of active learning). In each cycle of active learning, the currently trained model predicted structures in the remaining unlabeled molecular space. An expected improvement (EI) value was calculated based on the mean values of the predicted

energy gap and the uncertainty of the prediction for each structure. The uncertainty prediction was conducted using the DeepChem package [56]. In each cycle, the top 20 structures with the highest EI values were selected, calculated with DFT, and added to the training set for the next active learning cycle. This process was repeated 5 times and a final model was obtained. The final model was then trained on a total labeled dataset of 773 structures before being used to predict the energy gaps of the remaining unlabeled molecular space of 10,972 structures. The final trained model files are available at https://github.com/cpfpengfei/NIRII-ML-Design.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

References

[1] K. Welsher, Z. Liu, S. P. Sherlock, J. T. Robinson, Z. Chen, D. Daranciang, H. Dai, *Nat. Nanotech.* **2009**, 4, 773.

[2] A. L. Antaris, H. Chen, K. Cheng, Y. Sun, G. Hong, C. Qu, S. Diao, Z. Deng, X. Hu, B. Zhang, X. Zhang, O. K. Yaghi, Z. R. Alamparambil, X. Hong, Z. Cheng, H. Dai, *Nat. Mater.* **2016**, 15, 235.

[3] S. Liu, H. Ou, Y. Li, H. Zhang, J. Liu, X. Lu, R. T. K. Kwok, J. W. Y. Lam, D. Ding, B. Z. Tang, *J. Am. Chem. Soc.* **2020**, 142, 15146.

[4] W. Xu, Z. Zhang, M. Kang, H. Guo, Y. Li, H. Wen, M. M. S. Lee, Z. Wang, R. T. K. Kwok, J. W. Y. Lam, K. Li, L. Xi, S. Chen, D. Wang, B. Z. Tang, *ACS Mater. Lett.* **2020**, 2, 1033.

[5] A. L. Antaris, H. Chen, K. Cheng, Y. Sun, G. Hong, C. Qu, S. Diao, Z. Deng, X. Hu, B. Zhang, X. Zhang, O. K. Yaghi, Z. R. Alamparambil, X. Hong, Z. Cheng, H. Dai, *Nat. Mater.* **2016**, 15, 235.

[6] N. G. Horton, K. Wang, D. Kobat, C. G. Clark, F. W. Wise, C. B. Schaffer, C. Xu, *Nat. Photon.* **2013**, 7, 205.

[7] Y. Sun, C. Qu, H. Chen, M. He, C. Tang, K. Shou, S. Hong, M. Yang, Y. Jiang, B. Ding, Y. Xiao, L. Xing, X. Hong, Z. Cheng, *Chem. Sci.* **2016**, 7, 6203.

[8] L. Sun, R. Li, W. Zhan, Y. Yuan, X. Wang, X. Han, Y. Zhao, *Nat. Commun.* **2019**, 10.

[9] Y. Sun, C. Qu, H. Chen, M. He, C. Tang, K. Shou, S. Hong, M. Yang, Y. Jiang, B. Ding, Y. Xiao, L. Xing, X. Hong, Z. Cheng, *Chem. Sci.* **2016**, 7, 6203.

[10] H. Liu, G. Hong, Z. Luo, J. Chen, J. Chang, M. Gong, H. He, J. Yang, X. Yuan, L. Li, X. Mu, J. Wang, W. Mi, J. Luo, J. Xie, X.-D. Zhang, *Adv. Mater.* **2019**, 31.

[11] Y. Fan, P. Wang, Y. Lu, R. Wang, L. Zhou, X. Zheng, X. Li, J. A. Piper, F. Zhang, *Nat. Nanotech.* **2018**, 13, 941.

[12] S. Liu, C. Chen, Y. Li, H. Zhang, J. Liu, R. Wang, S. T. H. Wong, J. W. Y. Lam, D. Ding, B. Z. Tang, *Adv. Funct. Mater.* **2020**, 30, 1908125.

[13] M. Zhao, J. Wang, Z. Lei, L. Lu, S. Wang, H. Zhang, B. Li, F. Zhang, *Angew. Chem. Int. Ed.* **2021**, 60, 5091.

[14] W. Zhu, M. Kang, Q. Wu, Z. Zhang, Y. Wu, C. Li, K. Li, L. Wang, D. Wang, B. Z. Tang, *Adv. Funct. Mater.* **2021**, 31, 2007026.

[15] F. Zhang, Y. Chen, P. Pei, Z. Lei, X. Zhang, D. Yin, *Angew. Chem. Int. Ed.* https://doi.org/10.1002/anie.202103071

[16] Q. Wang, Y. Dai, J. Xu, J. Cai, X. Niu, L. Zhang, R. Chen, Q. Shen, W. Huang, Q. Fan, *Adv. Funct. Mater.* **2019**, 29, 1901480.

[17] X. Ni, X. Zhang, X. Duan, H.-L. Zheng, X.-S. Xue, D. Ding, *Nano Lett.* **2019**, 19, 318.

[18] O. T. Bruns, T. S. Bischof, D. K. Harris, D. Franke, Y. Shi, L. Riedemann, A. Bartelt, F. B. Jaworski, J. A. Carr, C. J. Rowlands, M. W. B. Wilson, O. Chen, H. Wei, G. W. Hwang, D. M. Montana, I. Coropceanu, O. B. Achorn, J. Kloepper, J. Heeren, P. T. C. So, D. Fukumura, K. F. Jensen, R. K. Jain, M. G. Bawendi, *Nat. Biomed. Engin.* **2017**, 1.

[19] K. Shou, C. Qu, Y. Sun, H. Chen, S. Chen, L. Zhang, H. Xu, X. Hong, A. Yu, Z. Cheng, *Adv. Funct. Mater.* **2017**, 27.

[20] Y. Sun, X. Zeng, Y. Xiao, C. Liu, H. Zhu, H. Zhou, Z. Chen, F. Xu, J. Wang, M. Zhu, J. Wu, M. Tian, H. Zhang, Z. Deng, Z. Cheng, X. Hong, *Chem. Sci.* **2018**, 9, 2092.

[21] M. Zhao, B. Li, P. Wang, L. Lu, Z. Zhang, L. Liu, S. Wang, D. Li, R. Wang, F. Zhang, *Adv. Mater.* **2018**, 30.

[22] S. Zhu, B. C. Yung, S. Chandra, G. Niu, A. L. Antaris, X. Chen, *Theranostics* **2018**, 8, 4141.

[23] F. Ding, Y. Fan, Y. Sun, F. Zhang, *Adv. Health. Mater.* **2019**, 8.

[24] Q. Yang, Z. Ma, H. Wang, B. Zhou, S. Zhu, Y. Zhong, J. Wang, H. Wan, A. Antaris, R. Ma, X. Zhang, J. Yang, X. Zhang, H. Sun, W. Liu, Y. Liang, H. Dai, *Adv. Mater.* **2017**, 29.

[25] B. Li, M. Zhao, F. Zhang, *ACS Mater. Lett.* **2020**, 2, 905.

[26] Q. Yang, H. Ma, Y. Liang, H. Dai, *Acc. Mater. Res.* **2021**, 2, 170.

[27] S. Wang, Y. Fan, D. Li, C. Sun, Z. Lei, L. Lu, T. Wang, F. Zhang, *Nat. Commun.* **2019**, 10.

[28] K. Li, X. Duan, Z. Jiang, D. Ding, Y. Chen, G. Q. Zhang, Z. Liu, *Nat. Commun.* **2021**, 12, 2376.

[29] W. Xu, D. Wang, B. Z. Tang, *Angew. Chem. Int. Ed.* **2021**, 60, 7476.

[30] H. Wan, J. Yue, S. Zhu, T. Uno, X. Zhang, Q. Yang, K. Yu, G. Hong, J. Wang, L. Li, Z. Ma, H. Gao, Y. Zhong, J. Su, A. L. Antaris, Y. Xia, J. Luo, Y. Liang, H. Dai, *Nat. Commun.* **2018**, 9.

[31] Y. Fang, J. Shang, D. Liu, W. Shi, X. Li, H. Ma, *J. Am. Chem. Soc.* **2020**, 142, 15271.

[32] S. Liu, C. Chen, Y. Li, H. Zhang, J. Liu, R. Wang, S. T. H. Wong, J. W. Y. Lam, D. Ding, B. Z. Tang, *Adv. Funct. Mater.* **2020**, 30, 1908125.

[33] P. Prosposito, M. Casalboni, F. De Matteis, M. Glasbeek, A. Quatela, E. van Veldhoven, H. Zhang, *J. Lumin.* **2001**, 94, 641.

[34] E. D. Cosco, J. R. Caram, O. T. Bruns, D. Franke, R. A. Day, E. P. Farr, M. G. Bawendi, E. M. Sletten, *Angew. Chem. Int. Ed.* **2017**, 56, 13126.

[35] B. Li, L. Lu, M. Zhao, Z. Lei, F. Zhang, *Angew. Chem. Int. Ed.* **2018**, 57, 7483.

[36] B. Ding, Y. Xiao, H. Zhou, X. Zhang, C. Qu, F. Xu, Z. Deng, Z. Cheng, X. Hong, *J. Med. Chem.* **2019**, 62, 2049.

[37] Z. Lei, C. Sun, P. Pei, S. Wang, D. Li, X. Zhang, F. Zhang, *Angew. Chem. Int. Ed.* **2019**, 58, 8166.

[38] G. Xu, Q. Yan, X. Lv, Y. Zhu, K. Xin, B. Shi, R. Wang, J. Chen, W. Gao, P. Shi, C. Fan, C. Zhao, H. Tian, *Angew. Chem. Int. Ed.* **2018**, 57, 3626.

[39] Q. Wu, Z. Kang, Q. Gong, X. Guo, H. Wang, D. Wang, L. Jiao, E. Hao, *Org. Lett.* **2020**, 22, 7513.

[40] P. Sun, Q. Wu, X. Sun, H. Miao, W. Deng, W. Zhang, Q. Fan, W. Huang, *Chem. Commun.* **2018**, 54, 13395.

[41] C. S. L. Rathnamalala, J. N. Gayton, A. L. Dorris, S. A. Autry, W. Meador, N. I. Hammer, J. H. Delcamp, C. N. Scott, *J. Org. Chem.* **2019**, 84, 13186.

[42] J. Maillard, K. Klehs, C. Rumble, E. Vauthey, M. Heilemann, A. Fürstenberg, *Chem. Sci.* **2021**, 12, 1352.

[43] Q. Yang, Z. Hu, S. Zhu, R. Ma, H. Ma, Z. Ma, H. Wan, T. Zhu, Z. Jiang, W. Liu, L. Jiao, H. Sun, Y. Liang, H. Dai, *J. Am. Chem. Soc.* **2018**, 140, 1715.

[44] H. Ma, C. Liu, Z. Hu, P. Yu, X. Zhu, R. Ma, Z. Sun, C.-H. Zhang, H. Sun, S. Zhu, Y. Liang, *Chem. Mater.* **2020**, 32, 2061.

[45] Y. Shi, W. Yuan, Q. Liu, M. Kong, Z. Li, W. Feng, K. Hu, F. Li, *ACS Mater. Lett.* **2019**, 1, 418.

[46] K. Brandhorst, J. Grunenberg, *J. Chem. Phys.* **2010**, 132, 9160.

[47] K. Brandhorst, J. Grunenberg, *Chem. Soc. Rev*. **2008**, 37, 1558.

[48] J. Grunenberg, *Chem. Sci.* **2015**, 6, 4086.

[49] J. M. Seminario, *Int. J. Quantum Chem.* **1996**, 60, 1271.

[50] D.A. Case, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, D. Greene, N. Homeyer, S. Izadi, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2017), AMBER 2017, University of California, San Francisco.

[51] D. Duvenaud, D. Maclaurin, J. Aguilera-Iparraguirre, R. Gómez-Bombarelli, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, *Convolutional Networks on Graphs for Learning Molecular Fingerprints,* **2015**.

[52] T. Lookman, P. V. Balachandran, D. Xue, R. Yuan, *NPJ Comput. Mater.* **2019**, 5, DOI 10.1038/s41524-019-0153-8.

[53] 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, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, et al., *Gaussian 16, Revision A.03*, **2016**, http://gaussian.com/citation_a03/, Gaussian Inc. Wallingford CT.

[54] G. Landrum, RDKit: Open-source cheminformatics, http://www.rdkit.org, **2021**.

[55] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, *T. Vandermeersch, G. R. Hutchison, J. Cheminform.* **2011**, 3, 1.

[56] B. Ramsundar, P. Eastman, P. Walters, V. Pande, Deep learning for the life sciences: applying deep learning to genomics, microscopy, drug discovery, and more, O'Reilly Media, Inc., **2019**.

[57] S. Xu, J. Li, P. Cai, X. Liu, B. Liu, X. Wang, *ChemRxiv*. Preprint. https://doi.org/10.26434/chemrxiv.14757975.v1, **2021**.

[58] H. Wan, H. Ma, S. Zhu, F. Wang, Y. Tian, R. Ma, Q. Yang, Z. Hu, T. Zhu, W. Wang, Z. Ma, M. Zhang, Y. Zhong, H. Sun, Y. Liang, H. Dai, *Adv. Funct. Mater.* **2018**, 28, 1804956.

[59] Y. Li, Z. Cai, S. Liu, H. Zhang, S. T. H. Wong, J. W. Y. Lam, R. T. K. Kwok, J. Qian, B. Z. Tang, *Nat. Commun.* **2020**, 11, 1255.

Supporting Information

Accelerated Design of Near-Infrared-II Molecular Fluorophores via First-Principles Understanding and Machine Learning

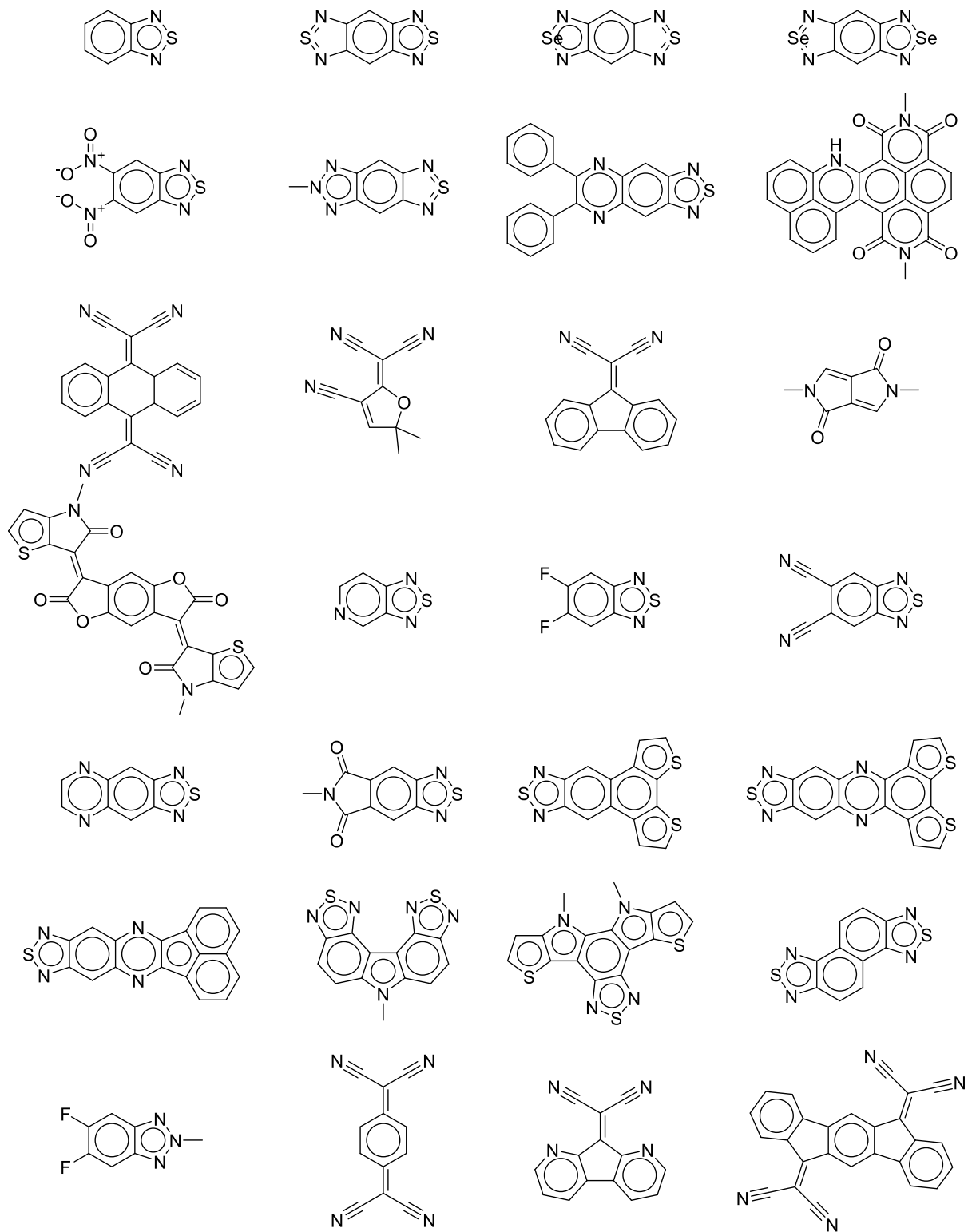*Shidang Xu[+], Pengfei Cai[+], Jiali Li, Xianhe Zhang, Xianglong Liu, Xiaonan Wang[*] and Bin Liu[*]*
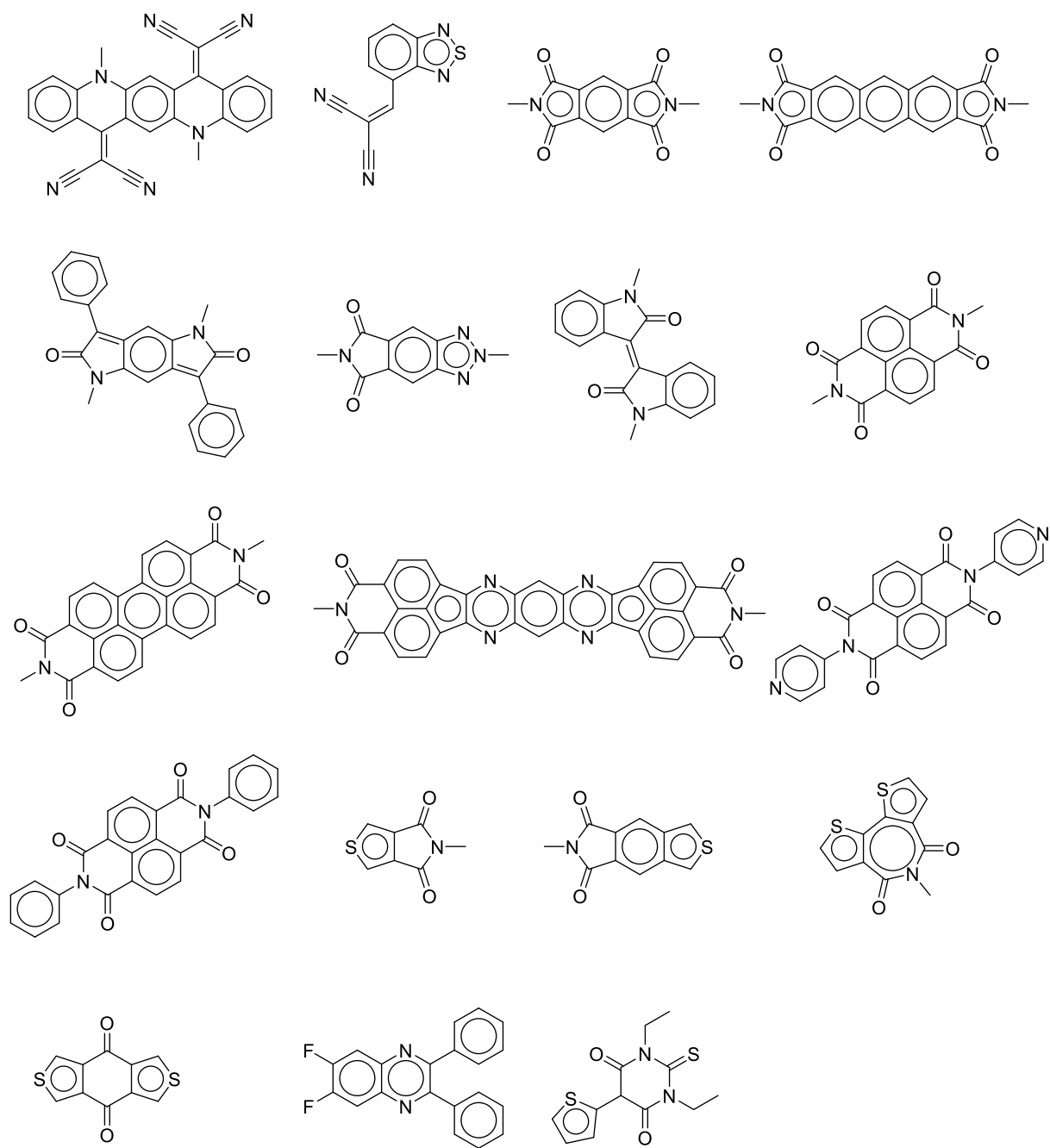
Dr. S. Xu, Mr. J. Li, Mr. X. Zhang, Mr. X. Liu, Dr. X. Wang, Prof. B. Liu
Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, 117585, Singapore
E-mail: chewxia@nus.edu.sg
E-mail: cheliub@nus.edu.sg

Mr. P. Cai
Department of Materials Science and Engineering, National University of Singapore, 9 Engineering Drive 1, 117575, Singapore

Prof. B. Liu

Joint School of National University of Singapore and Tianjin University, International

Campus of Tianjin University, Binhai New City, Fuzhou 350207, China

**Figure S1.** List of acceptors used in molecular space generation.
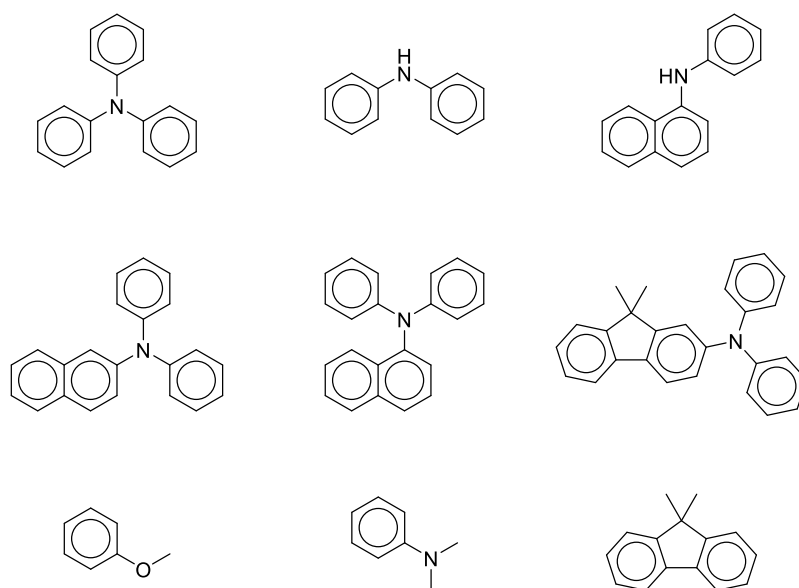
Figure S2. List of donors used in molecular space generation.
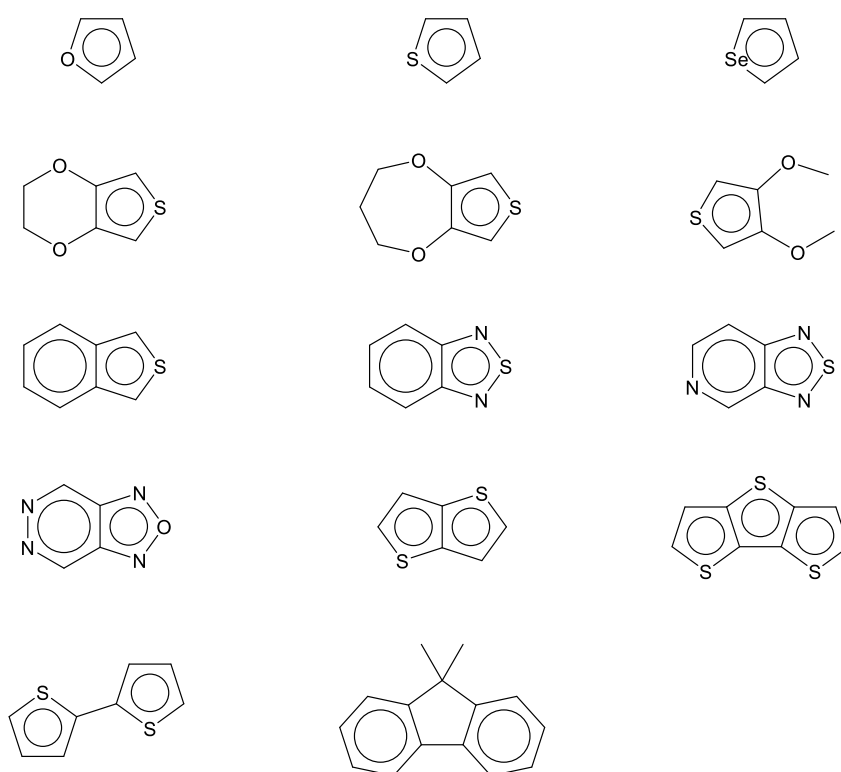
**Figure S3.** List of π bridges used in molecular space generation.

Table S1. Model hyperparameters used for Graph Convolutional Neural Network

Graph convolution layers	[512, 512, 512, 512]
Dense layers	[128, 128, 128]
Learning rate	0.001
Dropout	0.01
Batch size	10
Epochs	200

NDAD_4185
E_g : 0.8125 eV

NDAD_659
E_g : 0.9464

NDAD_2873
E_g : 0.9445
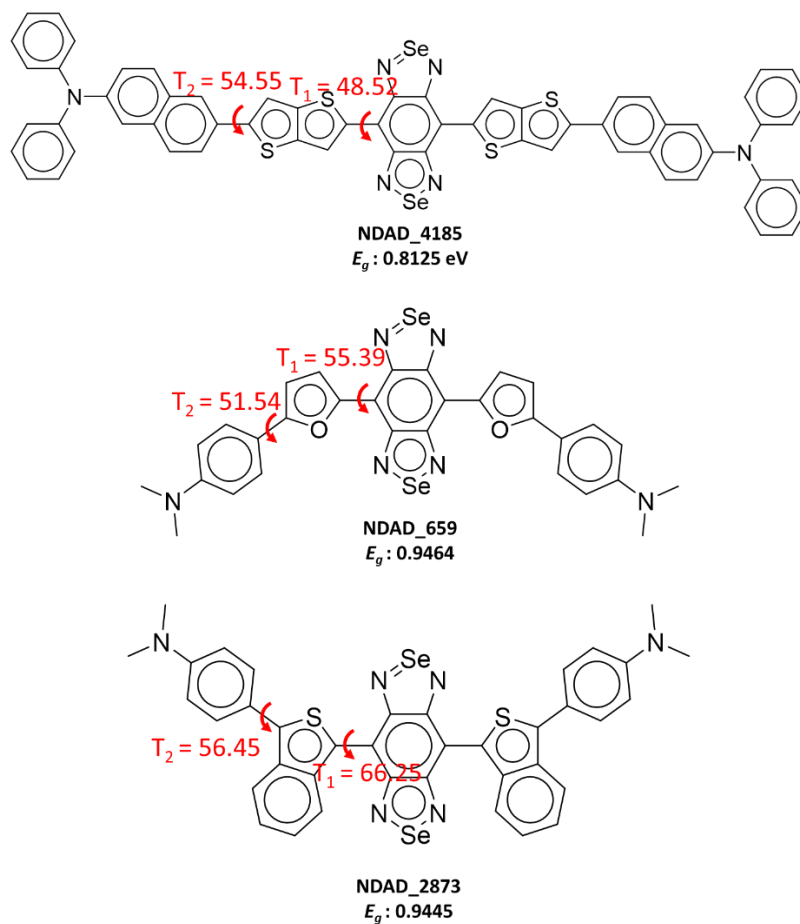
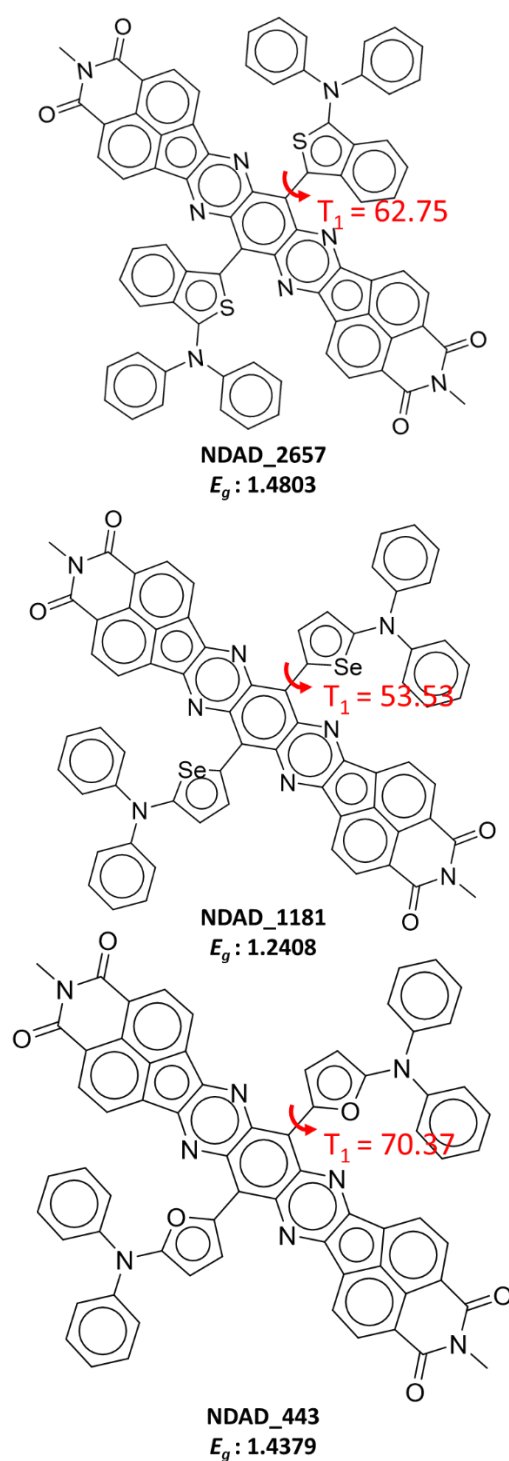**Figure S4.** Selected structures set 1 and their corresponding dihedral angle force constants and energy gaps.

NDAD_4942
E_g : 0.8362

NDAD_4204
E_g : 0.8765

NDAD_2523
E_g : 1.1532
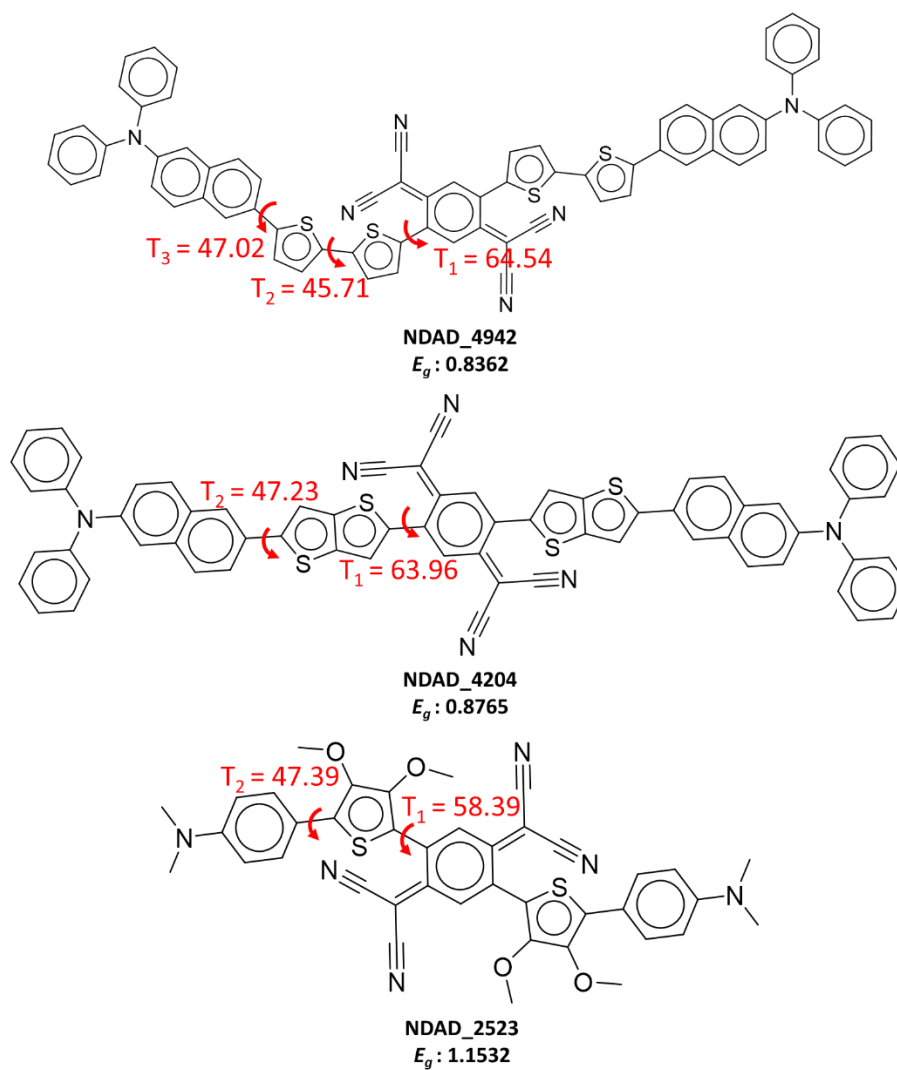
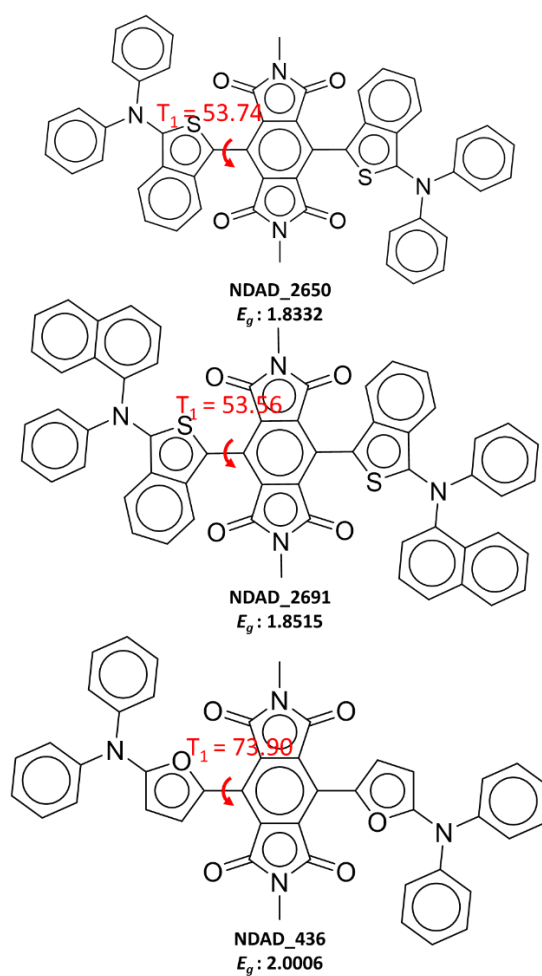**Figure S5.** Selected structures set 2 and their corresponding dihedral angle force constants and energy gaps.

NDAD_2656
E_g : 1.6539

NDAD_442
E_g : 1.8251

NDAD_2082
E_g : 1.6087
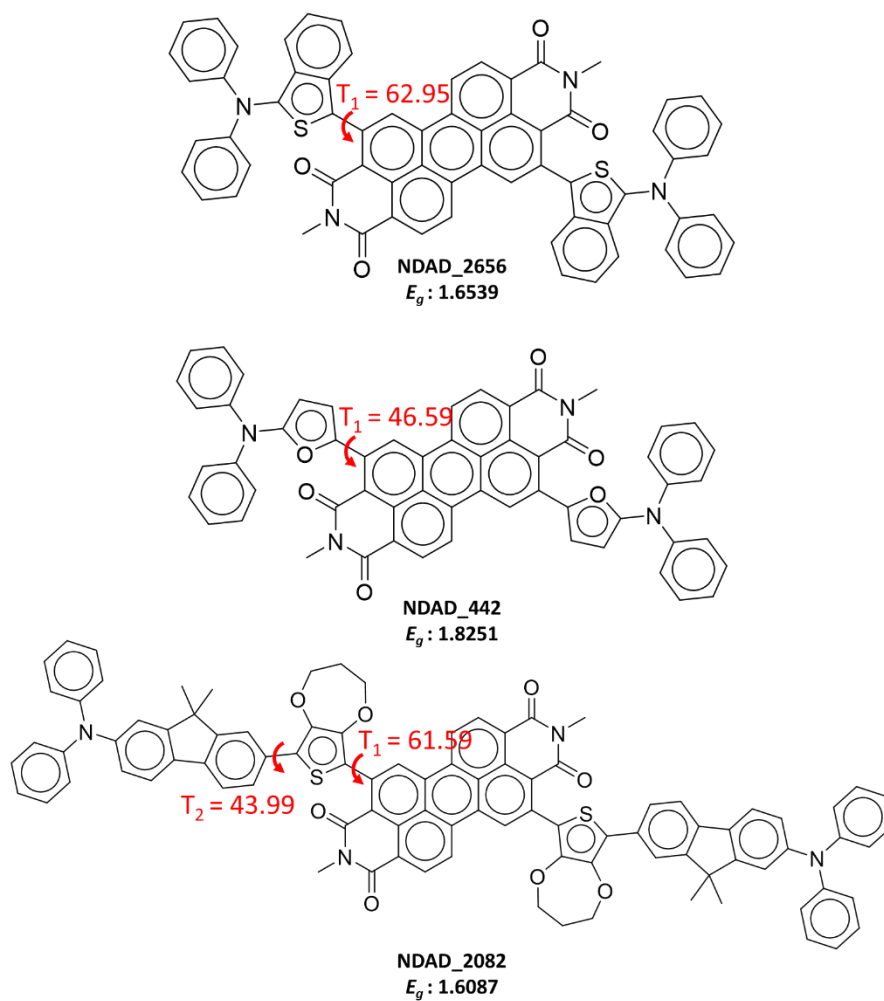
Figure S6. Selected structures set 3 and their corresponding dihedral angle force constants and energy gaps.

NDAD_2657
E_g : 1.4803

**NDAD_1181**
E_g : 1.2408

NDAD_443
E_g : 1.4379

Figure S7. Selected structures set 4 and their corresponding dihedral angle force constants and energy gaps.

NDAD_4813
E_g: 1.2082

NDAD_4731
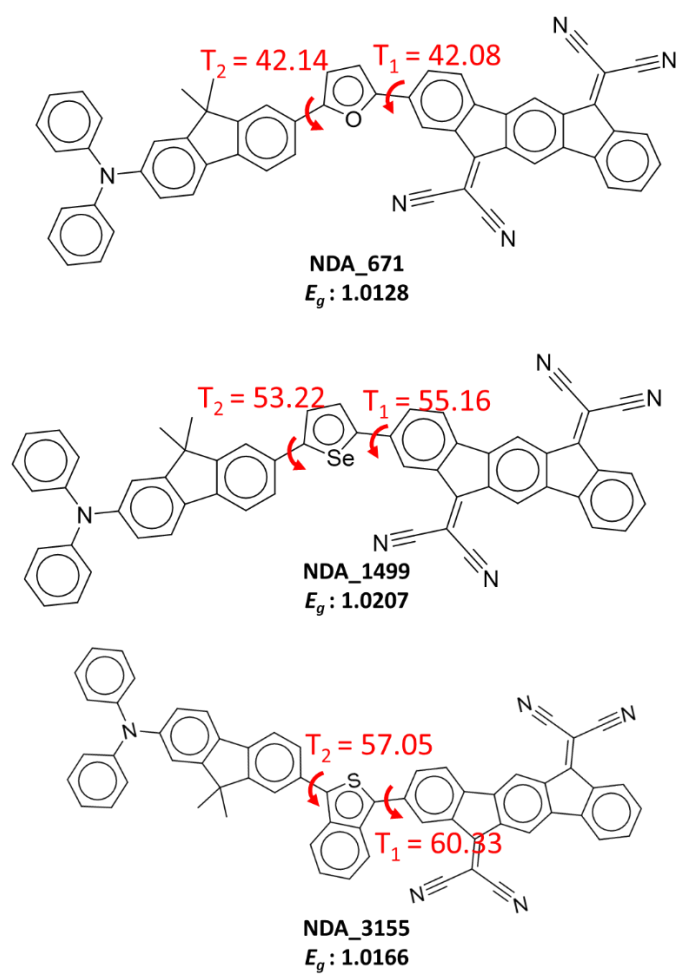E_g: 1.1761

NDAD_4936
E_g: 1.2193
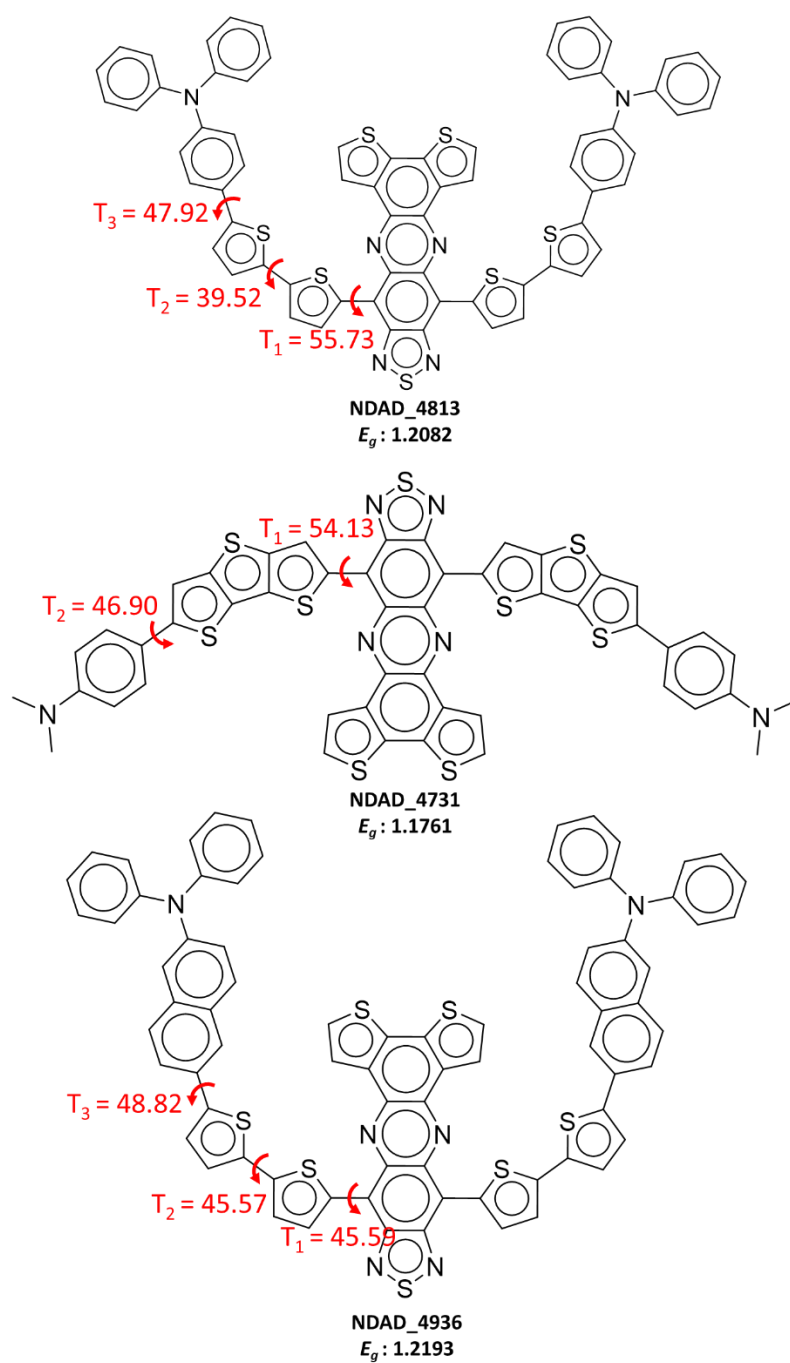
Figure S8. Selected structures set 5 and their corresponding dihedral angle force constants and energy gaps.

NDAD_2650
E_g: 1.8332

**NDAD_2691**
E_g: 1.8515

NDAD_436
E_g: 2.0006

Figure S9. Selected structures set 6 and their corresponding dihedral angle force constants and energy gaps.

NDAD_4609
E_g: 1.3859

NDAD_2641
E_g: 1.6028

NDAD_2723
E_g: 1.3334
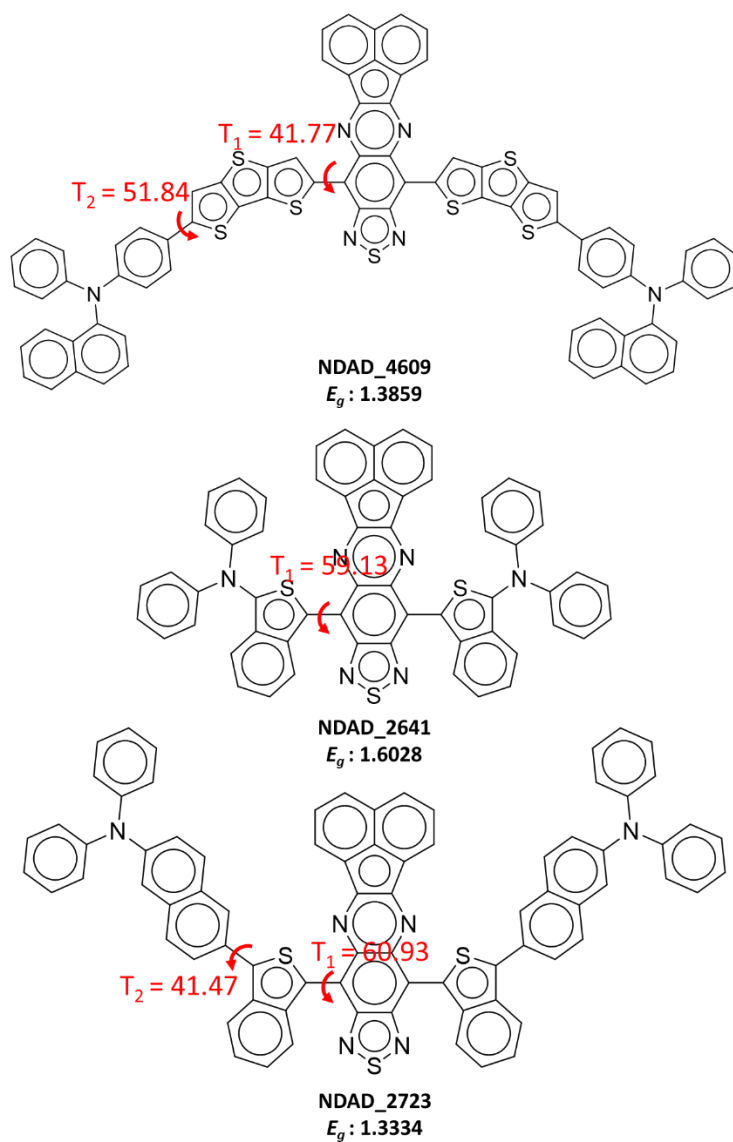
Figure S10. Selected structures set 7 and their corresponding dihedral angle force constants and energy gaps.

NDA_671
E_g : 1.0128

$T_2 = 42.14$ $T_1 = 42.08$

$T_2 = 53.22$ $T_1 = 55.16$

NDA_1499
E_g : 1.0207

$T_2 = 57.05$

$T_1 = 60.33$
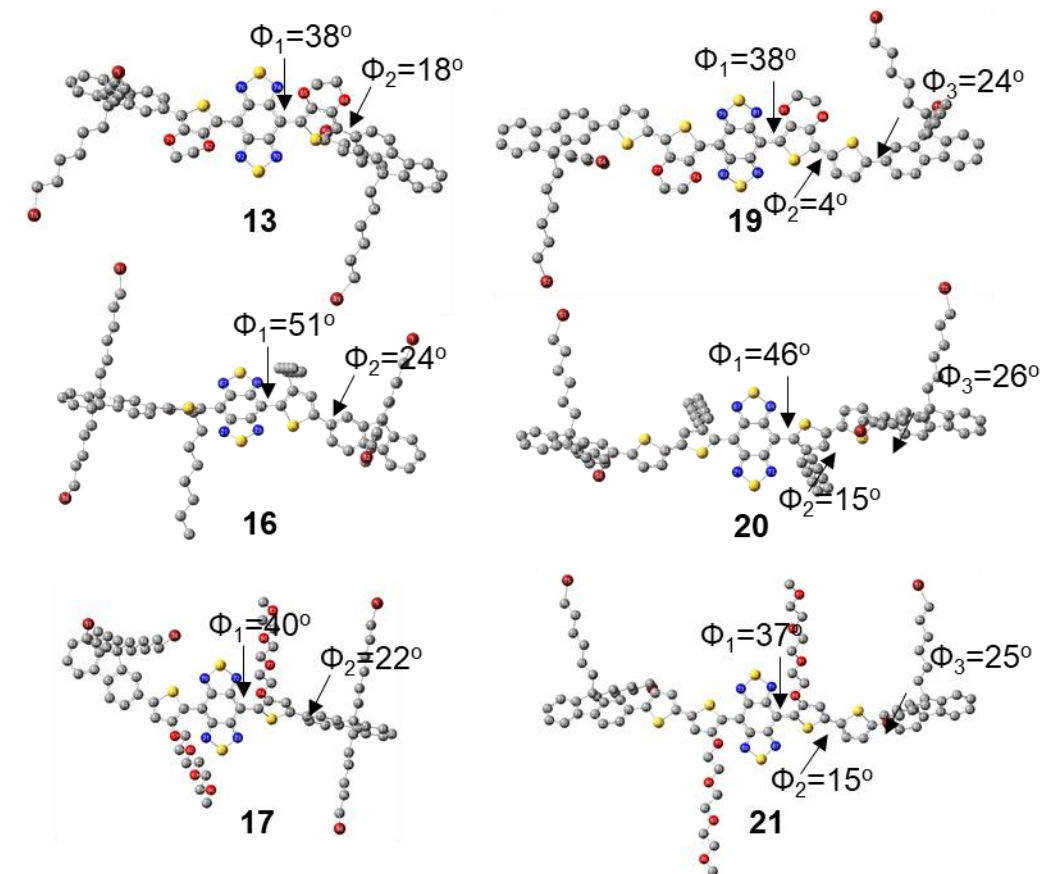
NDA_3155
E_g : 1.0166

Figure S11. Selected structures set 8 and their corresponding dihedral angle force constants and energy gaps.
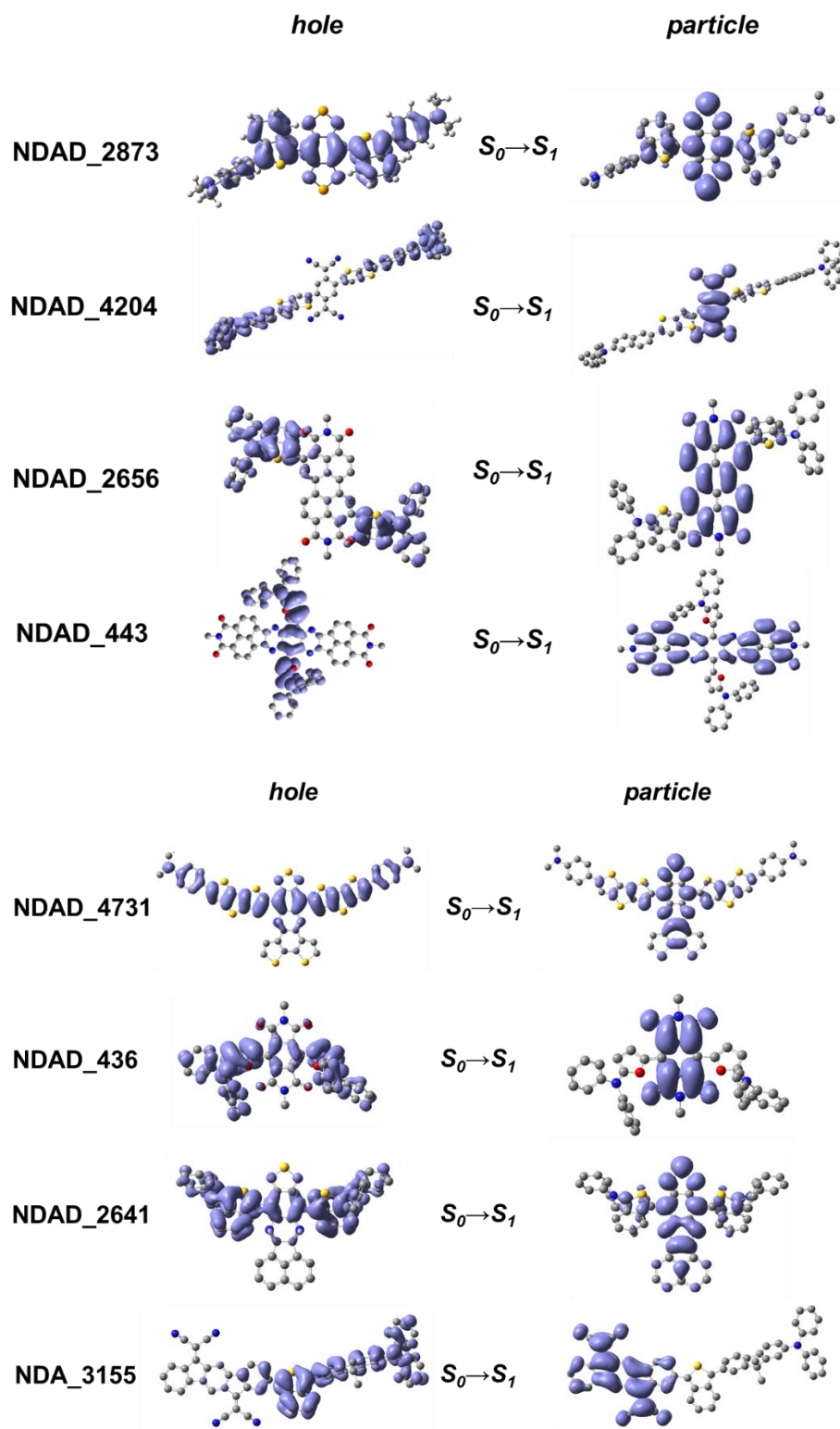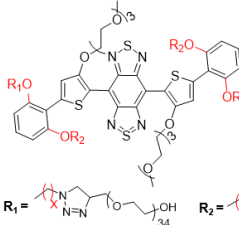
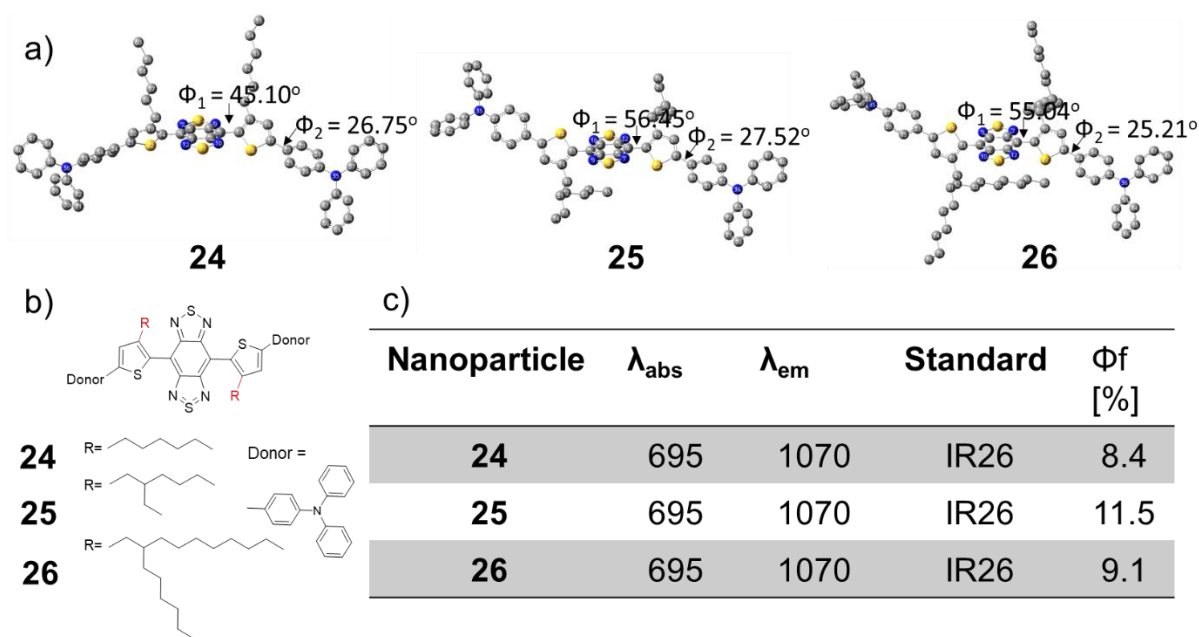**Figure S12.** NTO distribution of S_1 of recommended structures.

Figure S13. Optimized structures of compounds **13**, **16**, **17**, **19**, **20**, and **21**.

Table S2. Hydrophobicity of compounds **21-23**.[S1]

	21 X = 3	**22** X = 6	**23** X = 9
Φf_{water} [%]	0.9	1.5	3.0
RDF@8Å	0.46	0.38	0.29

Figure S14. Increasing backbone distortion in aggregate state. a) Molecular structures and b) optimized structures of compounds **24–26**. c) Photophysical properties of **24–26** [S2].

Supplementary References

[S1] H. Wan, H. Ma, S. Zhu, F. Wang, Y. Tian, R. Ma, Q. Yang, Z. Hu, T. Zhu, W. Wang, Z. Ma, M. Zhang, Y. Zhong, H. Sun, Y. Liang, H. Dai, *Adv. Funct. Mater.* **2018**, 28, 1804956.

[S2] Y. Li, Z. Cai, S. Liu, H. Zhang, S. T. H. Wong, J. W. Y. Lam, R. T. K. Kwok, J. Qian, B. Z. Tang, Nat. Commun. 2020, 11, 1255.