Supporting Information

From Handbooks to High-Throughput: Rule-Based Prediction of Electronic Absorption Maxima from SMILES with ChromoPredict

Connor Forster[1], Carolin Müller[1,*]

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Computer-Chemistry-Center, Nägelsbachstraße 25, Erlangen, 91052, Germany

[*] correspondence: carolin.cpc.mueller@fau.de

Contents

1 ChromoPredict **2**
1.1 Example Predictions . . . 3
1.1.1 Enones . . . 3
1.1.2 Dienes . . . 4
1.1.3 Polyenes . . . 4

2 Reference Datasets and Predictive Frameworks **5**
2.1 Enones . . . 5
2.2 Coumarines . . . 12

1 ChromoPredict

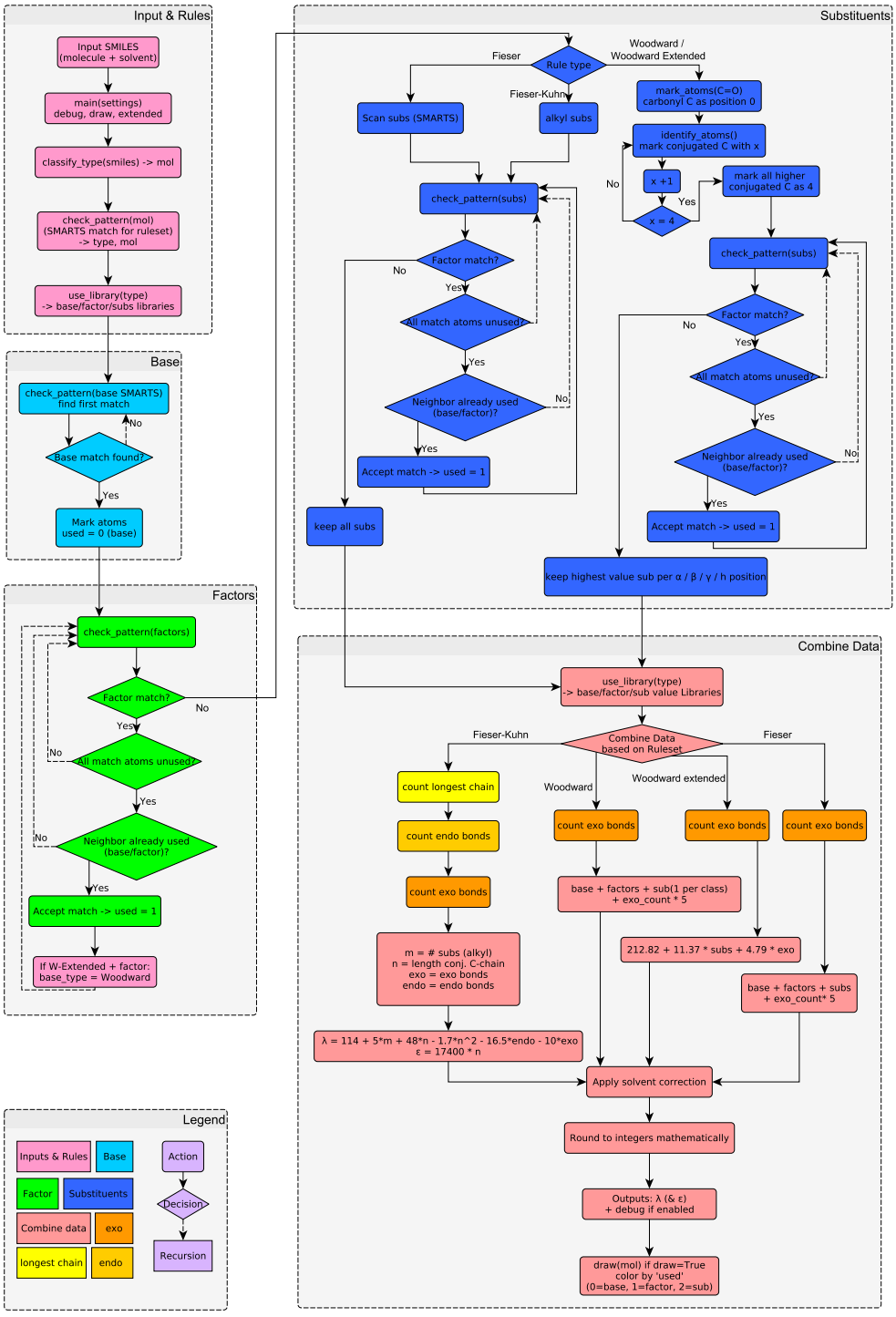


Figure S1: Flowchart of ChromoPredict [1]. The `predict` function receives the SMILES of the target molecule and the solvent as input (pink). The resulting molecular graph is analyzed to identify the base chromophore (light blue), followed by determination of relevant factors (green), such as exocyclic or additional conjugated double bonds. Substituents are then matched (dark blue), and all contributions are combined in the final estimation step according to the calculation schemes of Woodward-Fieser [2, 3], Woodward extended [4], Fieser-Kuhn [5], or Fieser [5].

Table S1: Summary of solvent contributions to the estimation of λ_{max} using ChromoPredict [1].

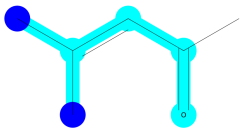
Solvent	SMILES	c_{solv} / nm
Water	`O`	-8
Acetone	`CO`	-5
Ethanol	`CCO`	-1
Chloroform	`C(Cl)(Cl)Cl`	± 0
Dioxane	`O1CCOCC1`	$+5$
Diethyl ether	`CCOCC`	$+7$
Hexane	`CCCCCC`	$+11$
Cyclohexane	`C1CCCCC1`	$+11$

1.1 Example Predictions

1.1.1 Enones

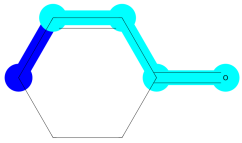
Mesityl Oxide

SMILES	`O=C(\C=C(/C)C)C`
Solvent	`CCO`
Exp. λ_{max}	227.5 nm [6]
Base	215
Factor	0
α	0
β	12
γ	0
Higher	0
Exo	0
Solvent	-1
Pred. λ_{max}	226



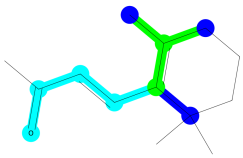
2-Cyclohexen-1-one

SMILES	`O=C1CCCC=C1`
Solvent	`C(Cl)Cl`
Exp. λ_{max}	224 nm [7]
Base	215
Factor	0
α	0
β	12
γ	0
Higher	0
Exo	0
Solvent	0 (no DCM value)
Pred. λ_{max}	227



β-Ionone

SMILES	`CC(=O)\C=C\C1=C(C)CCCC1(C)C`
Solvent	`CCCCCC`
Exp. λ_{max}	293 nm [8]
Base	215
Factor	30
α	0
β	0
γ	18
Higher	18
Exo	0
Solvent	11
Pred. λ_{max}	292



Retinal

SMILES	`CC1=C(C(CCC1)(C)C)/C=C/C(=C/C=C/C(=C/C=O)/C)/C`
Solvent	non-polar
Exp. λ_{max}	$\approx$ 370 nm [9]
Base	210
Factor	120
α	0
β	12
γ	0
Higher	18
Exo	0
Solvent	11
Pred. λ_{max}	371

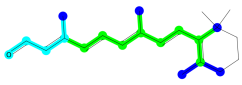
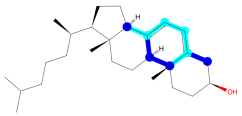


Figure S2: Example breakdown of the calculation of λ_{max} for four representative enones using the Woodward-Fieser rules.

1.1.2 Dienes

7-Dehydrocholesterol

SMILES	`C[C@H](CCCC(C)C)[C@H]1CC[C@@H][C@@]3(CC[C@@H](C4)O)C)C`
Solvent	`CCO`
Exp. λ_{max}	271, 282, 294 nm [10]
Fieser calc.	
Base	253
Subs	+20
Exo	+10
Solvent	−1
Pred. λ_{max}	282



Trans-2,4-Hexadiene

SMILES	`C/C=C/C=C/C`
Solvent	`CCO`
Exp. λ_{max}	226 nm [11]
Fieser calc.	
Base	217
Subs	+10
Exo	+0
Solvent	−1
Pred. λ_{max}	226

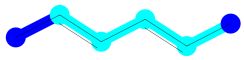


Figure S3: Example breakdown of the calculation of λ_{max} for two representative dienes using the Fieser rules.

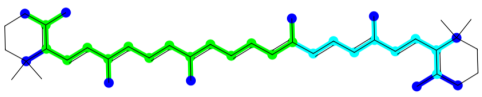
1.1.3 Polyenes

β-Carotene

Solvent	`CCCCCC`
Exp. λ_{max}	453 nm [12]
Fieser-Kuhn calc.	
M	10
N	11
Endo	2
Exo	0
Solvent	11
Pred. λ_{max}	464

Lycopene

Solvent	`CCCCCC`
Exp. λ_{max}	≈ 470 nm [13]
Fieser-Kuhn calc.	
M	8
N	11
Endo	0
Exo	0
Solvent	11
Pred. λ_{max}	487



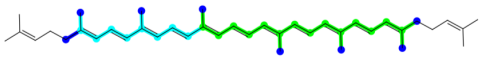
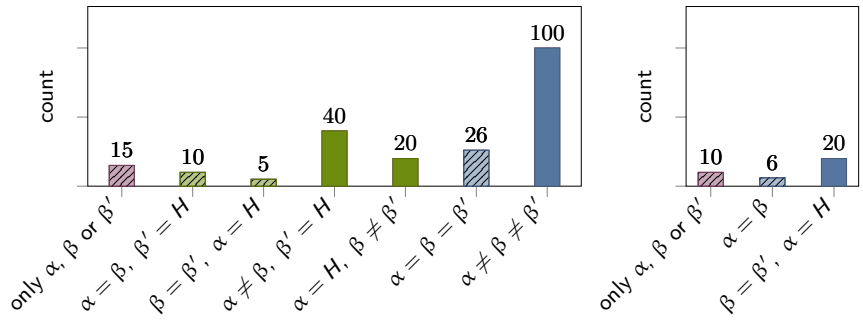


Figure S4: Example breakdown of the calculation of λ_{max} for two representative polyenes using the Fieser-Kuhn rules.

2 Reference Datasets and Predictive Frameworks

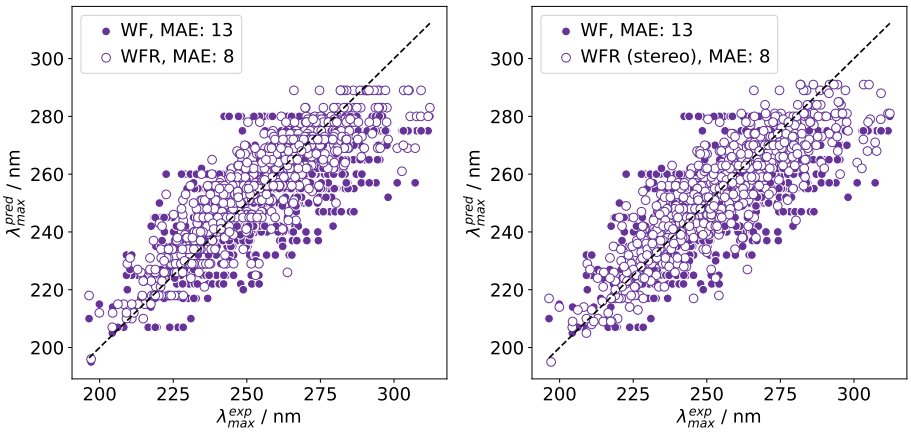
2.1 Enones

For training/testing of machine learning models and refinement of the Woodward-Fieser rules, we computed the absorption properties for in total of 720 enones: 216 per acyclic compound class (ketone, aldehyde, and carboxylic acid), and 36 per cyclic enone (cyclopentenone and cyclohexenone). For the acyclic compounds, all possible permutations of mono- (α or β), di- (α, β), and tri-substitution (α, β, β') were systematically constructed (see Figure S5). For the cyclic enones, due to the presence of only one accessible β-position, mono- and di-substitution patterns were generated, covering α-, β-, and α, β-substitution.



(a) acyclic **(b)** cyclic

Figure S5: Distribution of data points per class for acyclic (a) and cyclic (b) molecules. For the acyclic aldehydes, ketones, and acids, three substitution sites (α, β, β') were considered, whereas for cyclic molecules (cyclopentenon and cyclohexenon) only the α and β positions were analyzed, as the β' position belongs to the ring. The substituents included `Cl`, `Br`, `O`, `OC`, and `H`. Bars are colored according to the number of substituents (1: purple, 2: green, 3: blue) and patterned to indicate substitution type (mono-type substitution: lined bars; mixed substitution: solid bars).



(a) **(b)**

Figure S6: Comparison of predicted λ_{max} values for 720 α, β-unsaturated carbonyl compounds using the original Woodward–Fieser rules (WF, filled symbols) and the refined Woodward–Fieser rules (WFR, unfilled symbols). Panel (a) shows predictions without explicit consideration of stereochemistry, while panel (b) includes stereochemistry, all compared against TD-B3LYP/def2-TZVP/D4 reference calculations.

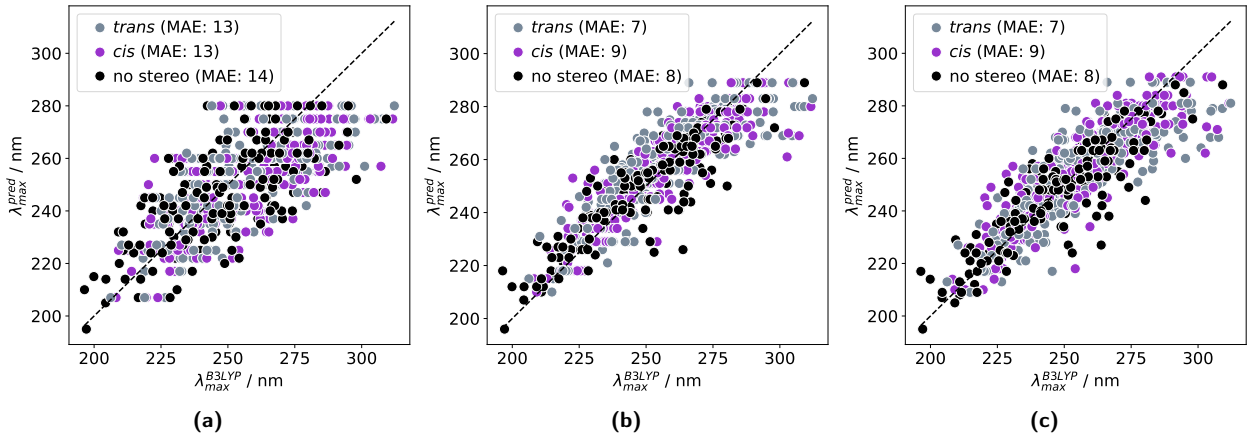


Figure S7: Comparison of predicted λ_{max} values for 720 α,β-unsaturated carbonyl compounds using the original Woodward-Fieser rules (a) and the refined Woodward–Fieser rules without explicit consideration of stereochemistry (b) and including stereochemistry (c). All values are compared against TD-B3LYP/def2-TZVP/D4 reference values. The symbols are color coded by the stereochemistry of the reference molecules.

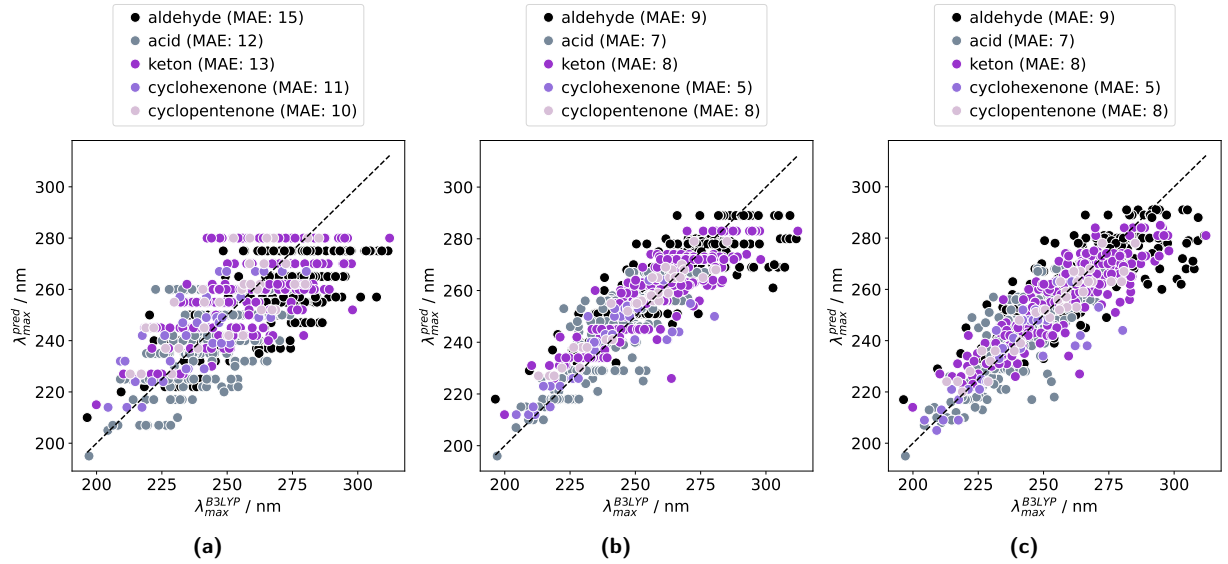


Figure S8: Comparison of predicted λ_{max} values for 720 α,β-unsaturated carbonyl compounds using the original Woodward-Fieser rules (a) and the refined Woodward–Fieser rules without explicit consideration of stereochemistry (b) and including stereochemistry (c). All values are compared against TD-B3LYP/def2-TZVP/D4 reference values. The symbols are color coded by molecular class (aldehydes, ketones, acids, cyclopentenones, and cyclohexenones).

```
base_acid_cis                   197
base_acid_no_stereo             197
base_acid_trans                 198
base_aldehyde_cis               219
base_aldehyde_no_stereo         219
base_aldehyde_trans             221
base_cyclohexenone_no_stereo    208
base_cyclopentenone_no_stereo   193
base_ketone_cis                 211
base_ketone_no_stereo           215
base_ketone_trans               214
alpha_H                           0
alpha_alkoxy                     28
alpha_alkyl                      10
alpha_bromo                      38
alpha_chloro                     27
alpha_hydroxy                    38
beta_H                            0
beta_alkoxy                      21
beta_alkyl                       17
beta_bromo                       32
beta_chloro                      21
beta_hydroxy                     13
dtype: int64
```

Figure S9: Optimized stereochemistry-dependent base chromophore values and substituent increments at the α- and β-positions, obtained by dual annealing within the Woodward–Fieser framework, based on 720 TD-DFT absorption maxima of α,β-unsaturated carbonyls (aldehydes, acids, ketones, cyclopentenones, and cyclohexenones).

Table S2: Experimental, Woodward-Fieser estimated and simulated $\pi\pi^*$ absorption maxima (λ_{max}) at B3LYP/TD-B3LYP/def2-TZVP/D4 [14–18] and xTB/TD-B3LYP/def2-TZVP/D4 [14–19] level of theory of enone derivatives given in SMILES. The rule-based predictions employed the original (WF) and refined (WFR) Woodward-Fieser schemes. In the TD-DFT simulations the 10 lowest singlet states were computed in vacuum upon geometry optimization at B3LYP/def2-TZVP/D4 (DFT) [14–18] or xTB [19] level, respectively. All quantum chemical calculations were performed using VeloxChem [20]. Experimental values are taken from references [4, 21–23].

ID	SMILES	base	α	β	γ	higher	λ_{max} / nm				
							Exp.	DFT	xTB	WF	WFR
E01	C=CC(=O)C	keton	H	H	H	H	214	200	199	215	212
E02	C=CC(=O)CCCCCCCCCC	keton	H	H	H	H	212	211	210	215	212
E03	C=CC(=O)CC1CCCCC1	keton	H	H	H	H	213	221	223	215	212
E04	C=CC(=O)CC1=CC=CC=C1	keton	H	H	H	H	210	272	269	215	212
E05	C=CC(=O)C1CCCCC1	keton	H	H	H	H	212	218	213	215	212
E06	C=CC(=O)C1CCC2(CC1)OCCO2	keton	H	H	H	H	214	263	258	215	212
E07	C=CC(=O)C1CCCCCC1	keton	H	H	H	H	213	218	216	215	212
E08	C=C1CCCC1=O	cyclopenton	alkyl	H	H	H	231	226	227	230	228
E09	C=C1CCCCC1=O	cyclohexon	alkyl	H	H	H	230	224	224	230	228
E10	C=C1CCCCCC1=O	cyclohepton	alkyl	H	H	H	230	300	301	230	228
E11	C=C1C[C@H]2[C@@H]3CC[C@@H]4C[C@@H](CC[C@]4(C)[C@H]3C(=O)C[C@]2(C)C1=O)OC(=O)C	cyclopenton	alkyl	H	H	H	228	241	245	230	228
E12	C=C1C[C@H]2[C@@H]3CC[C@@H]4C[C@@H](CC[C@]4(C)[C@H]3[C@H](C[C@]2(C)C1=O)O)O	cyclopenton	alkyl	H	H	H	227	245	245	230	228
E13	C1CCC(=C2CCCC2=O)C1	cyclopenton	alkyl	alkyl	H	H	259	256	256	242	247
E14	C1CCC(=C2CCCCC2=O)CC1	cyclohexon	alkyl	alkyl	H	H	257	249	252	242	243
E15	CCCCCCCC[C@H]1CC(=O)C2=C3CC[C@H]4C[C@H](CC[C@]4(C)[C@H]3CC[C@]12C)OC(=O)C	cyclopenton	alkyl	alkyl	H	H	257	249	252	242	247
E16	CCCCCCCC[C@H]1C[C@H](C2=C3[C@H](CC[C@]12C)[C@@]4(C)CC[C@@H](C[C@@H]4CC3=O)OC(=O)C)OC(=O)C	cyclohexon	alkyl	acyl	H	H	256	260	261	260	250
E17	CCCCCCCC[C@H]1C[C@@H](C2=C3[C@H](CC[C@]12C)[C@@]4(C)CC[C@@H]C[C@@H]4CC3=O)OC(=O)C)OC(=O)C	cyclohexon	alkyl	acyl	H	H	257	266	266	260	250
E18	CC(=O)C1=C(C)CCCC1	cyclohexenon	alkyl	alkyl	H	H	247	238	243	237	238
E19	CC(C)=C1CCC(C)CC1=O	cyclohexon	alkyl	alkyl	H	H	252	247	242	242	247
E20	CC1=CC(=O)CC(C)(C)C1	cyclohexenon	H	alkyl	H	H	235	219	219	227	227
E21	CC1=C2CCCCC2CCC1=O	cyclohexenon	alkyl	alkyl	H	H	239	241	242	242	243
E22	CC(=C1CCCCC1)C(=O)C	keton	alkyl	alkyl	H	H	242	244	243	242	247
E23	C[C@@H]1CCC(=C(C)C)C(=O)C1	cyclohexon	alkyl	alkyl	H	H	252	247	246	242	247
E24	C[C@H](CCCC(C)C)[C@H]1CC[C@@H]2[C@@]1(CC[C@H]3[C@H]2CCC4=C(C(=O)CC[C@]34C)Br)C	cyclohexenon	alkyl	alkyl	H	H	256	273	274	257	270
E25	CC12CCC(=O)C(Br)=C1CC[C@@H]1[C@@H]2CC[C@]2(C)C(=O)CC[C@@H]12	cyclohexenon	alkyl	alkyl	H	H	254	294	292	257	270
E26	CC1=C(C(=O)C(CC1)C(C)C)O	cyclohexenon	hydroxy	alkyl	H	H	270	263	264	262	265
E27	C[C@@]12CCC[C@H]1[C@H]3[C@H](CC2)[C@]4(CCC=CC4=CC3=O)C	cyclohexenon	H	alkyl	H	alkyl	278	–	–	280	280
E28	CC1=CCCC2(C)CCC(=O)C=C12	cyclohexenon	H	alkyl	alkyl	alkyl	287	–	–	298	298
E29	CC(=O)C=CC1=C(C)CCCC1(C)C	keton	H	H	alkyl	alkyl	283	–	–	281	278
E30	CC(=O)OC1(C)C=CC=CC1=O	cyclohexenon	H	H	H	acyloxy	300	–	–	276	238

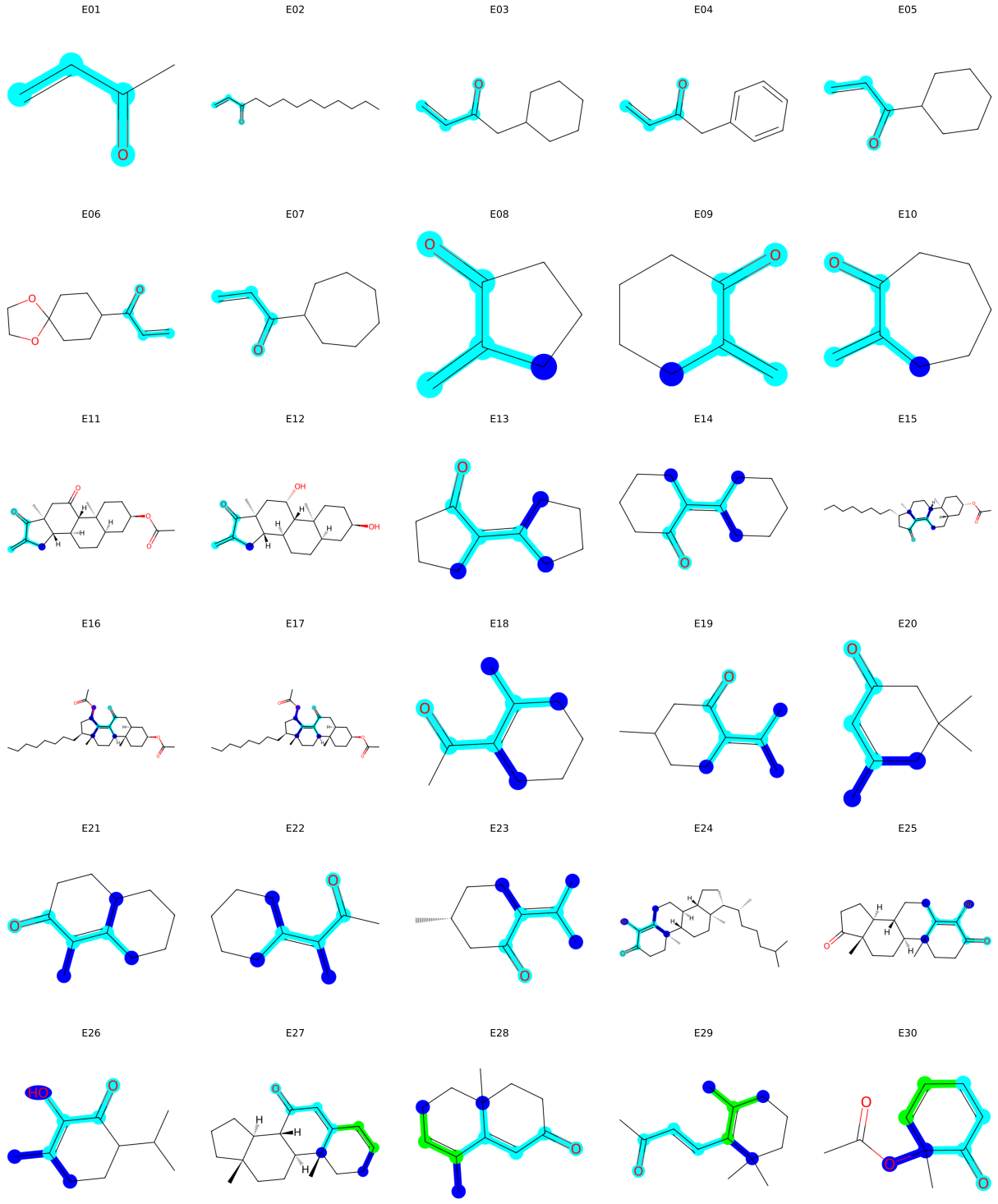


Figure S 10: Structural formulas of the enones **E01–E30** (see Table S2), with the base chromophore (light blue), substituents (dark blue), and additional conjugated double bonds (green) highlighted as identified by `ChromoPredict`.

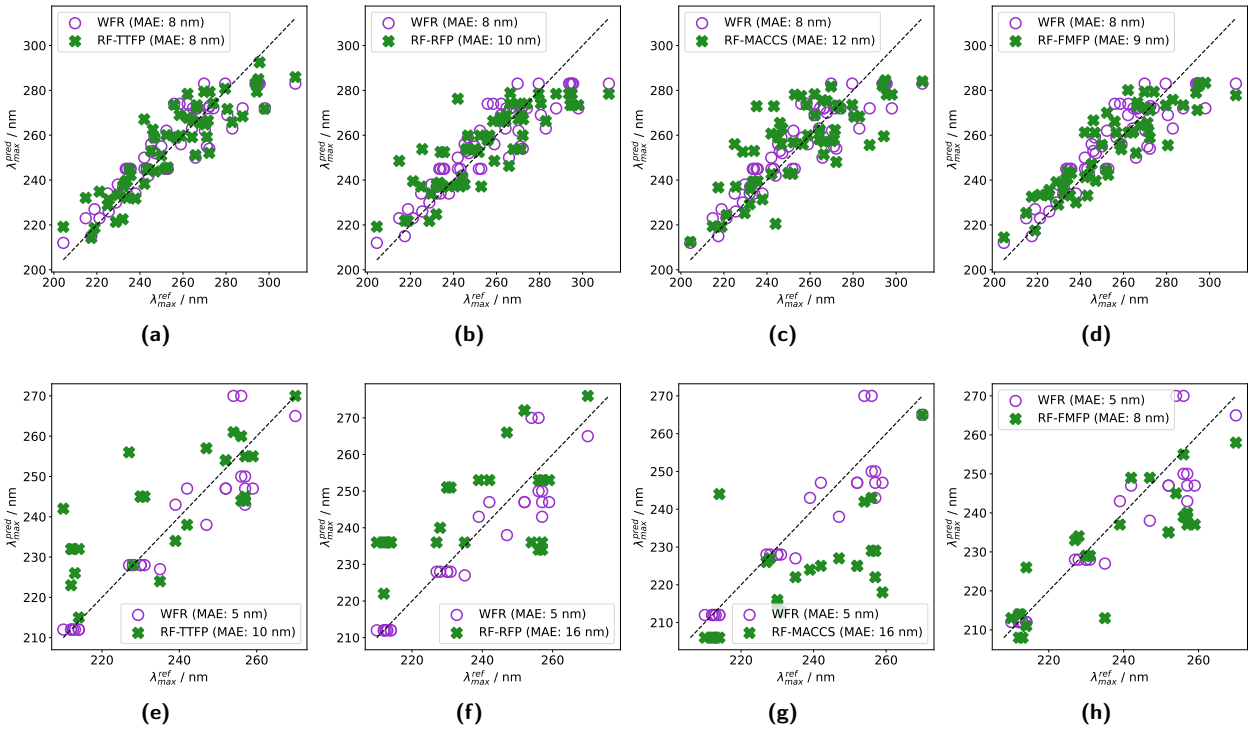


Figure S11: Estimations of λ_{max} with the refined Woodward-Fieser (unfilled, purple symbols) and random forest models (green symbols) employing topological torsion fingerprints (a, e), rooted topological torsions fingerprints (b, f), MACCS keys (c, g), and feature Morgan fingerprints (d, h) as structural descriptors. Predictions are plotted against TD-B3LYP simulated absorption maxima for 20 % (#57) of enones from the computational dataset (#288 acyclic ketones, cyclohexenone and cyclopentenon derivatives, see Figure S5) in panels a–d and against experimental spectra of **E01–E27** (see Table 2) in panels e–h. Enones **E28–E30** are omitted because they contain an additional conjugated double bond relative to the core chromophore, representing out-of-distribution systems not included in model training.

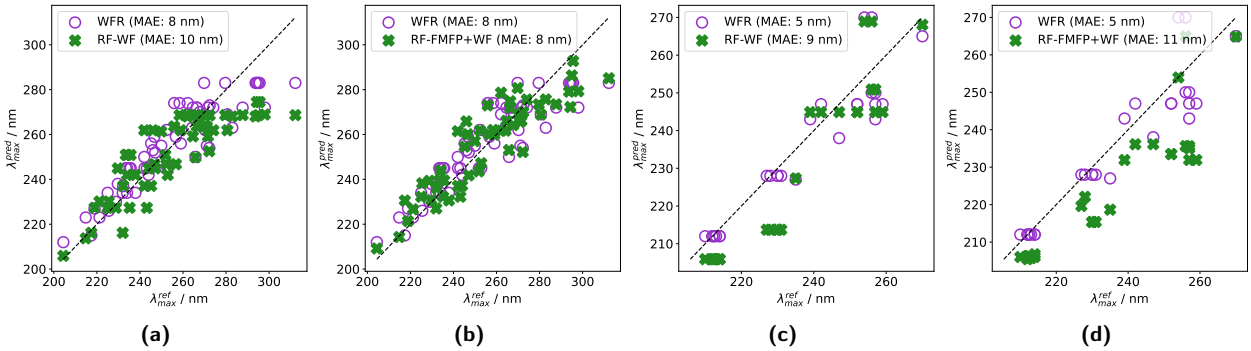
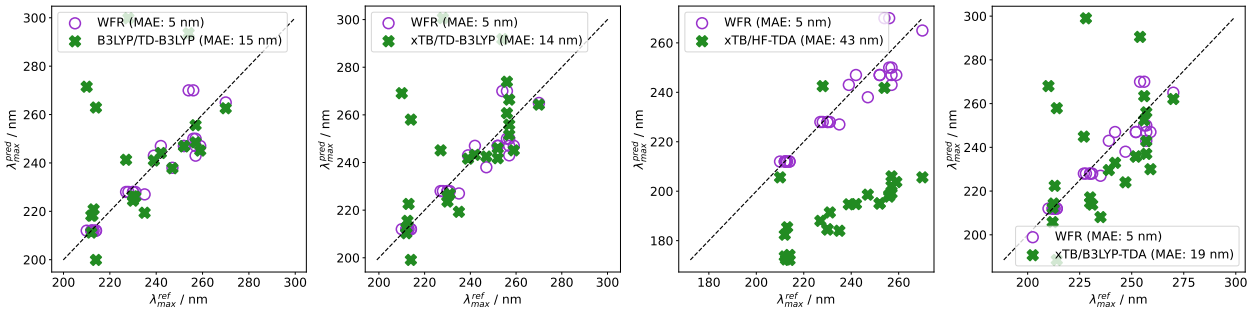


Figure S12: Predicted λ_{max} values from Woodward–Fieser (unfilled purple symbols, panels a–d) and random forest models (green symbols) informed by Woodward–Fieser increment values. Models use (i) increments for exocyclic double bonds, α- and β-positions, and the total number of substituents (a, c), or (ii) the same features combined with Morgan fingerprints (b, d). Predictions are plotted against TD-B3LYP simulated absorption maxima for 20 % (#57) of enones from the computational dataset (a, b) and against experimental spectra of **E01–E27** (c, d).



(a) B3LYP/TD-B3LYP **(b)** xTB/TD-B3LYP **(c)** xTB/HF-TDA **(d)** xTB/B3LYP-TDA

Figure S13: Predicted absorption maxima (λ_{max}) obtained from empirical Woodward-Fieser rules (open purple symbols, panels a–d) and *ab initio* electronic structure calculations (green symbols). Panels show results computed at the (a) B3LYP/TD-B3LYP [14, 15], (b) xTB/TD-B3LYP [14, 15, 19], (c) xTB/HF-TDA [19], and (d) xTB/B3LYP-TDA [14, 15, 19] levels of theory. Linear-response TDDFT (RPA) was used for panels (a) and (b), while panels (c) and (d) employed the Tamm-Dancoff Approximation (TDA) based on Hartree–Fock and B3LYP reference wave functions, respectively. All simulations included the ten lowest singlet excited states and utilized the def2-TZVP basis set [18] and D4 dispersion correction [16, 17]. Calculations were performed with VeloxChem [20]. Predicted λ_{max} values are compared with experimental data for compounds **E01–E26**.

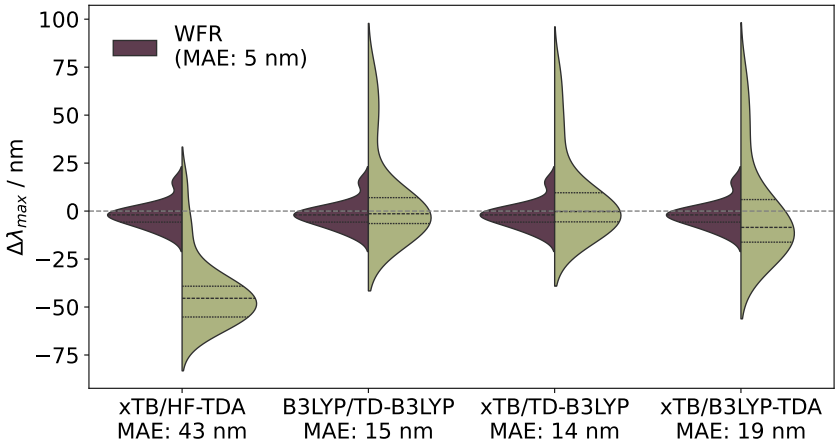


Figure S14: Violin plots showing the prediction errors between the calculated absorption maxima (λ_{max}) from empirical Woodward-Fieser rules (purple) and *ab initio* methods (green) relative to the experimental λ_{max} values of enones **E01–E26**. Simulations were performed at the (a) B3LYP/TD-B3LYP [14, 15] and (b) xTB/TD-B3LYP [14, 15, 19] levels of theory using linear-response TDDFT (RPA), and at the (c) xTB/HF-TDA [19]; and (d) xTB/B3LYP-TDA [14, 15, 19] levels employing the Tamm-Dancoff Approximation (TDA) based on Hartree–Fock and B3LYP reference wave functions, respectively. All simulations included the ten lowest singlet excited states and utilized the def2-TZVP basis set [18] and D4 dispersion correction [16, 17]. Calculations were performed with VeloxChem [20]. The respective dataset comprising the experimental and calculated λ_{max} values is available as csv file on GitHub.

2.2 Coumarines

Table S3: Experimental and Woodward-Fieser estimated $\pi\pi^*$ absorption maxima (λ_{max}) of coumarin derivatives given in SMILES representation in the respective solvent envrionemnt given as SMILES. The rule-based predictions employed the Woodward-Fieser scheme refined for the indicated coumarin dyes by means of dual annealing optimization of the base value and summands contributing to λ_{max}. The respective values are listed in Table 5. The database is available as csv file on GitHub.

ID	SMILES	Solvent	α	β	higher	λ_{max} / nm		
						Exp.	WF	B3LYP
C01	C1=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	H	310 [24]	311	299
C02	C1=CC=C2C(=C1)C(C)=CC(=O)O2	CCO	H	alkyl	H	310 [24]	310	296
C03	C1=CC=C2C(=C1)C=C(C)C(=O)O2	CCO	alkyl	H	H	310 [24]	310	296
C04	C1(C)=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	alkyl	322 [24]	326	313
C05	C1=CC=C2C(=C1)C(C)=C(C)C(=O)O2	CCO	alkyl	alkyl	H	309 [24]	309	298
C06	C1=CC=C2C(=C1)C(Cl)=CC(=O)O2	CCO	H	chloro	H	316 [24]	308	305
C07	C1=CC=C2C(=C1)C=C(Cl)C(=O)O2	CCO	chloro	H	H	315 [24]	315	308
C08	C1(Cl)=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	chloro	321 [24]	321	314
C09	C1=CC=C2C(=C1)C(Cl)=C(Cl)C(=O)O2	CCO	chloro	chloro	H	310 [24]	312	311
C10	C1=CC=C2C(=C1)C(Br)=CC(=O)O2	CCO	H	bromo	H	316 [25]	316	306
C11	C1=CC=C2C(=C1)C=C(Br)C(=O)O2	CCO	bromo	H	H	315 [25]	315	311
C12	C1=CC=C2C(=C1)C(O)=CC(=O)O2	CCO	H	hydroxy	H	317 [26]	314	287
C13	C1=CC=C2C(=C1)C=C(O)C(=O)O2	CCO	hydroxy	H	H	310 [24]	310	303
C14	C1(O)=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	hydroxy	349 [24]	338	341
C15	C1=CC=C2C(=C1)C(O)=C(O)C(=O)O2	CCO	hydroxy	hydroxy	H	313 [26]	313	303
C16	C1(O)=CC=C2C(=C1)C(O)=CC(=O)O2	CCO	H	hydroxy	hydroxy	320 [24]	341	323
C17	C1=CC=C2C(=C1)C(OC)=CC(=O)O2	CCO	H	alkoxy	H	303 [27]	303	284
C18	C1=CC=C2C(=C1)C=C(OC)C(=O)O2	CCO	alkoxy	H	H	304 [28]	304	303
C19	C1(OC)=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	alkoxy	344 [24]	341	345
C20	C1(O)=CC=C2C(=C1)C(C)=CC(=O)O2	CCO	H	alkyl	hydroxy	342 [24]	337	335
C21	C1(C)=CC=C2C(=C1)C(O)=CC(=O)O2	CCO	H	hydroxy	alkyl	334 [24]	329	297
C22	C1(OC)=CC=C2C(=C1)C(OC)=CC(=O)O2	CCO	H	alkoxy	alkoxy	342 [24]	333	321
C23	C1(OC)=CC=C2C(=C1)C(C)=CC(=O)O2	CCO	H	alkyl	alkoxy	340 [24]	340	338
C24	C1(O)=CC=C2C(=C1)C(OC)=CC(=O)O2	CCO	H	alkoxy	hydroxy	320 [24]	330	318
C25	C1(OCC)=CC=C2C(=C1)C(C)=CC(=O)O2	CCO	H	alkyl	alkoxy	330 [29]	340	342
C26	C1(OC)=CC=C2C(=C1)C(O)=C(CC)C(=O)O2	CCO	alkyl	hydroxy	alkoxy	322 [30]	343	324

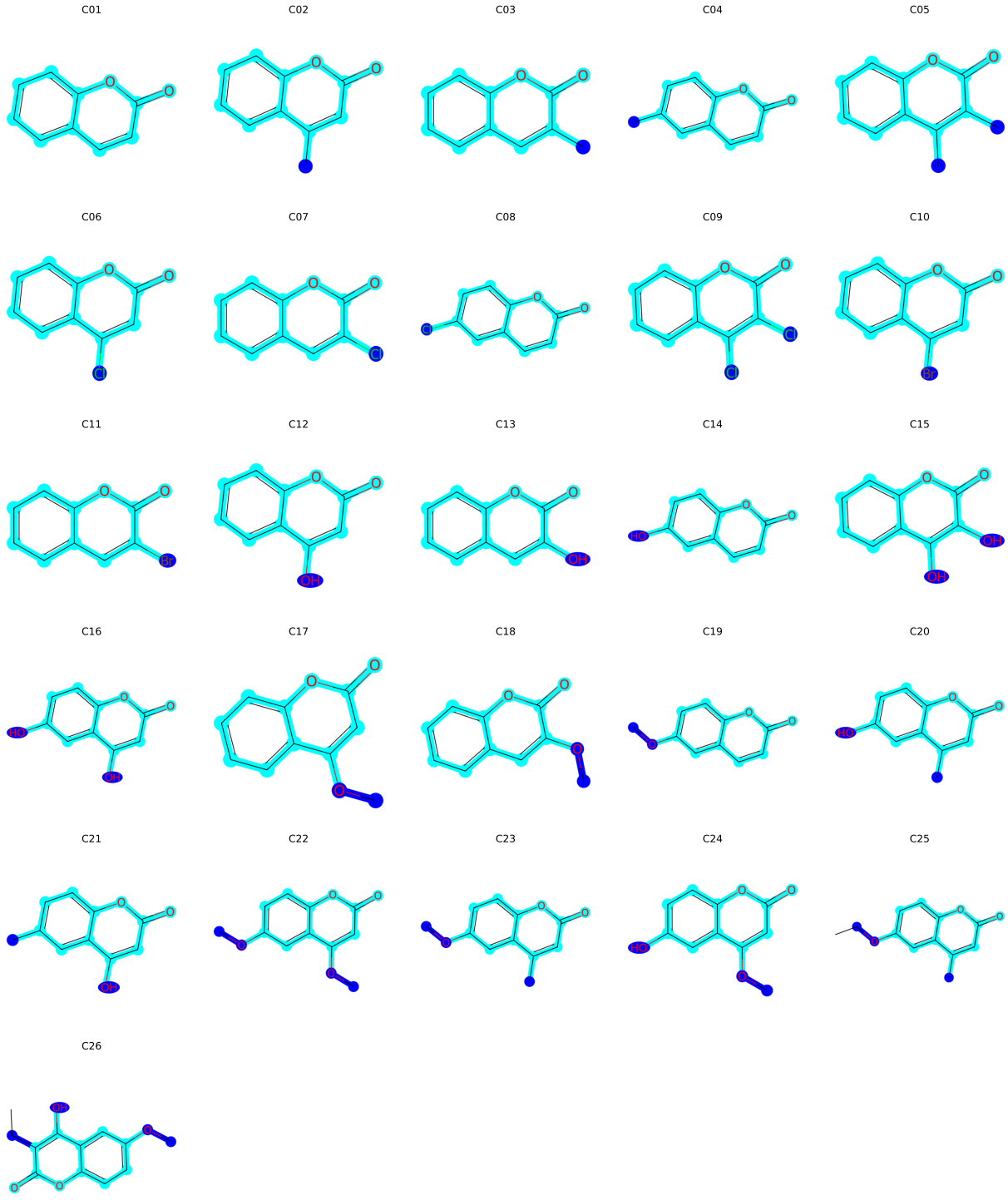


Figure S15: Structural formulas of the coumarines **C01–C26** (see Table S3), with the base chromophore and substituents identified by ChromoPredict are highlighted in light blue and dark blue, respectively.

Table S4: Experimental and Woodward-Fieser estimated $\pi\pi^*$ absorption maxima (λ_{max}) of coumarin derivatives given in SMILES representation in the respective solvent environment given as SMILES. The rule-based predictions employed the Woodward-Fieser scheme refined on the basis of the coumarin dyes as listed in Table S3. The respective values are listed in Table 5. The database is available as csv file on GitHub.

ID	SMILES	Solvent	3- (α)	4- (β)	6- (higher)	λ_{max} / nm		
				position		Exp.	WF	B3LYP
C27	C1(C)=CC=C2C(=C1)C(C)=C(C)C(=O)O2	CO	alkyl	alkyl	alkyl	318 [24]	320	308
C28	C1(O)=CC=C2C(=C1)C(O)=C(C(c1ccccc1)CC(=O)C)C(=O)O2	CCO	alkyl	hydroxy	hydroxy	330 [31]	340	339
C29	C1(OC)=CC=C2C(=C1)C=C(C)C(=O)O2	CCO	alkyl	H	alkoxy	337 [24]	340	364
C30	C1(OC)=CC=C2C(=C1)C(CC1)=CC(=O)O2	CCO	H	alkyl	alkoxy	345 [32]	340	364
C31	C1(OCC(=O)O)=CC=C2C(=C1)C(C)=CC(=O)O2	CO	H	alkyl	hydroxy	352 [33]	333	360
C32	C(C)C=1C(OC2=CC=C(C=C2C1O)OC)=O	CCO	H	alkyl	alkoxy	335 [34]	340	331
C33	C1(OC)=CC=C2C(=C1)C(CCO)=CC(=O)O2	CO	H	alkyl	alkoxy	337 [35]	336	342
C34	C1(OC(C)C#C)=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	alkoxy	337 [36]	341	341
C35	C1(OC(C)(C)C#C)=CC=C2C(=C1)C=CC(=O)O2	CCO	H	H	alkoxy	333 [36]	341	341
C36	C1=CC=C2C(=C1)C=C(CCCC1=CCCCC1)C(=O)O2	CN	alkyl	H	H	311 [37]	311	324

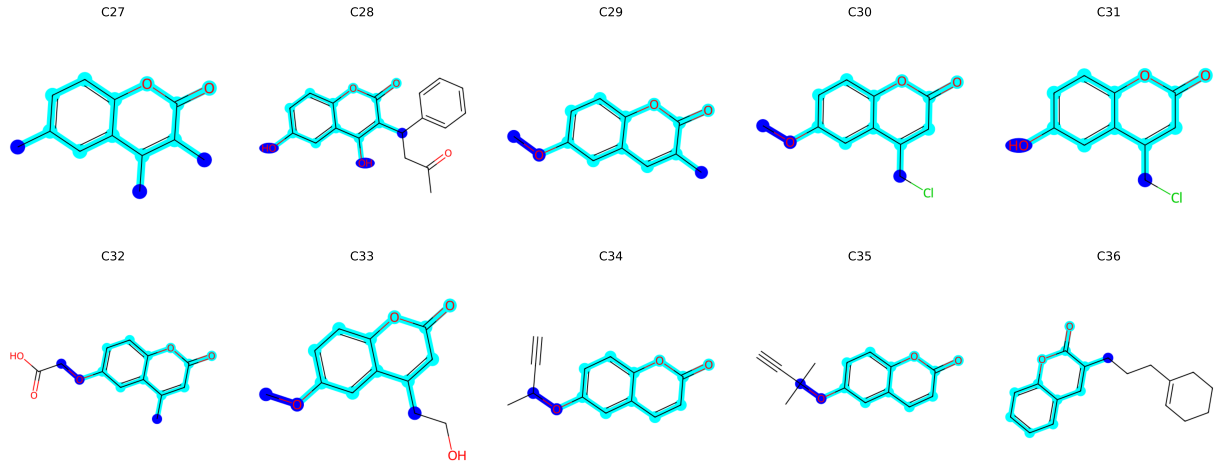
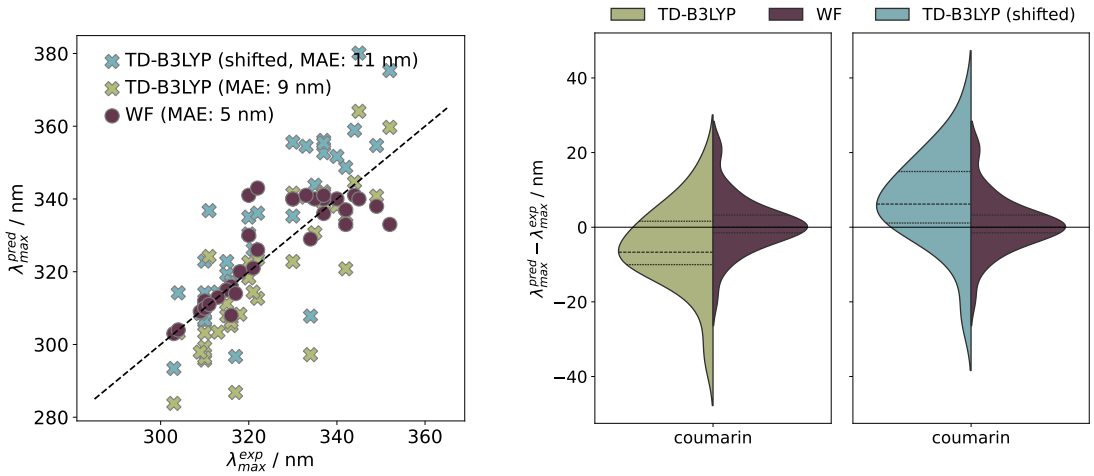


Figure S16: Structural formulas of the coumarines **C27–C36** (see Table S4), with the base chromophore and substituents identified by ChromoPredict are highlighted in light blue and dark blue, respectively.

Table S 5: Summary of Woodward-Fieser-type rules for predicting λ_{max} of mono-, di-, and tri-substituted coumarins, focusing on substitution at the 3-, 4-, and 6-positions. The table includes the base value, increment values for extended conjugation (not fitted in this work), and position-specific substituent corrections.

Coumarin (`[#6]1=[#6]-[#6]=[#6]2-[#6](=[#6]-1)-[#6]=[#6]-[#6](-[#8]-2)=[#8]`							
compound category	**base value**	**conjugated features**	**increment**	**substituent**	3- (α)	4- (β)	6- ($>\gamma$)
coumarin	312	–		–alkyl	-1	-1	$+15$
				–Cl	$+4$	-3	$+10$
				–Br	$+4$	$+5$	$+10$
				–OH	-1	$+3$	$+27$
				–O-alkyl	-7	-8	$+30$



(a) **(b)**

Figure S 17: Pair plots (a) show the predicted λ_{max} values for coumarins obtained using our refined Woodward–Fieser rule increments (purple symbols; parameters listed in Table S5) alongside TD-B3LYP/def2-TZVP/D4/CPCM(Ethanol) calculations. Violin plots (b) illustrate the prediction errors relative to the experimental absorption maxima, presented both without spectral alignment (green) and after applying a uniform shift of -0.143 eV to match the experimental maximum of **C01**.

References

[1] ChromoPredict, **2025**, https://github.com/CompPhotoChem/ChromoPredict.

[2] R. B. Woodward, *J. Am. Chem. Soc.* **1942**, *64*, 72–75, DOI 10.1021/ja01253a018.

[3] R. B. Woodward, *J. Am. Chem. Soc.* **1942**, *64*, 76–77, DOI 10.1021/ja01253a019.

[4] G. Kang, Y. Kang, F.-A. Kang, *Journal of Physical Organic Chemistry* **2021**, *34*, e4186, DOI 10.1002/poc.4186, https://onlinelibrary.wiley.com/doi/abs/10.1002/poc.4186.

[5] L. F. Fieser, M. Fieser, S. Rajagopalan, *The Journal of organic chemistry* **1948**, *13 6*, PMID: 18106021, 800–806, DOI 10.1021/jo01164a003, https://doi.org/10.1021/jo01164a003.

[6] L. Láng in *Absorption Spectra in the Ultraviolet and Visible Region*, Academic Press, New York, **1961**, pp. 92–115.

[7] M. T. Peschel, P. Kabaciński, D. P. Schwinger, E. Thyrhaug, G. Cerullo, T. Bach, J. Hauer, R. de Vivie-Riedle, *Angewandte Chemie International Edition* **2021**, *60*, 10155–10163, DOI 10.1002/anie.202016653.

[8] W. Kimel, J. D. Surmatis, J. Weber, G. O. Chase, N. W. Sax, A. Ofner, *Journal of Organic Chemistry* **1957**, *22*, 1611–1618, DOI 10.1021/jo01363a021.

[9] C. Chen, E. Tsina, M. C. Cornwall, R. K. Crouch, S. Vijayaraghavan, Y. Koutalos, *Biophysical Journal* **2005**, *88*, 2278–2287, DOI 10.1529/biophysj.104.054254.

[10] A. Honda, A. K. Batta, G. Salen, G. S. Tint, T. S. Chen, S. Shefer, *American Journal of Medical Genetics* **1997**, *68*, 288–293, DOI https://doi.org/10.1002/(SICI)1096-8628(19970131)68:3<288::AID-AJMG8>3.0.CO;2-I, https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-8628%2819970131%2968%3A3%3C288%3A%3AAID-AJMG8%3E3.0.CO%3B2-I.

[11] W. F. Forbes, R. Shilton, A. Balasubramanian, *Journal of Organic Chemistry* **1964**, *29*, 3527–3531, DOI 10.1021/jo01035a020.

[12] M. Hagos, M. Redi-Abshiro, B. S. Chandravanshi, E. E. Yaya, *International Journal of Analytical Chemistry* **2022**, *2022*, 9363692, DOI 10.1155/2022/9363692.

[13] M. Cámara, M. Mata, V. Fernández-Ruiz, R. Camara, S. Manzoor, J. Caceres in *Vol. 40*, **2013**, pp. 383–426, DOI 10.1016/B978-0-444-59603-1.00011-4.

[14] A. D. Becke, *The Journal of Chemical Physics* **1993**, *98*, 5648–5652, DOI 10.1063/1.464913, https://doi.org/10.1063/1.464913.

[15] C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789, DOI 10.1103/PhysRevB.37.785, https://link.aps.org/doi/10.1103/PhysRevB.37.785.

[16] E. Caldeweyher, C. Bannwarth, S. Grimme, *The Journal of Chemical Physics* **2017**, *147*, 034112, DOI 10.1063/1.4993215, https://doi.org/10.1063/1.4993215.

[17] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, S. Grimme, *The Journal of Chemical Physics* **2019**, *150*, 154122, DOI 10.1063/1.5090222, https://doi.org/10.1063/1.5090222.

[18] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297, DOI 10.1039/b508541a.

[19] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, S. Grimme, *WIREs Computational Molecular Science* **2021**, *11*, e1493, DOI https://doi.org/10.1002/wcms.1493, https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1493.

[20] Z. Rinkevicius, X. Li, O. Vahtras, K. Ahmadzadeh, M. Brand, M. Ringholm, N. H. List, M. Scheurer, M. Scott, A. Dreuw, P. Norman, *WIREs Computational Molecular Science* **2020**, *10*, e1457, DOI https://doi.org/10.1002/wcms.1457, https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1457.

[21] V. Wathelet, J. Preat, M. Bouhy, M. Fontaine, E. A. Perpète, J.-M. André, D. Jacquemin, *International Journal of Quantum Chemistry* **2006**, *106*, 1853–1859, DOI https://doi.org/10.1002/qua.20982, https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.20982.

[22] R. C. Cambie, P. A. Craw, R. J. Hughes, P. S. Rutledge, P. D. Woodgate, *Australian Journal of Chemistry* **1982**, *35*, 2111–2130, DOI 10.1071/CH9822111.

[23] J. A. Marshall, D. J. Schaeffer, *The Journal of Organic Chemistry* **1965**, *30*, 3642–3646, DOI 10.1021/jo01022a013, https://doi.org/10.1021/jo01022a013.

[24] A. Mangini, R. Passerini, *Gazzetta Chimica Italiana* **1957**, *87*, 243–266.

[25] T. S. Reddy, H. Moon, M.-S. Choi, *New Journal of Chemistry* **2020**, *44*, 4992–5000, DOI 10.1039/D0NJ00037J, https://doi.org/10.1039/D0NJ00037J.

[26] W. W. Mantulin, P.-S. Song, *Journal of the American Chemical Society* **1973**, *95*, 5122–5129, DOI 10.1021/ja00797a004, https://doi.org/10.1021/ja00797a004.

[27] J. Kolodziejczyk-Czepas, S. Kozachok, Łukasz Pecio, S. Marchyshyn, W. Oleszek, *Phytochemistry* **2021**, *190*, 112861, DOI 10.1016/j.phytochem.2021.112861, https://doi.org/10.1016/j.phytochem.2021.112861.

[28] J. Avó, S. Martins, A. Parola, J. Lima, J. Ramalho, P. Branco, A. Pereira, **2013**, DOI 10.1002/cplu.201300118, https://doi.org/10.1002/cplu.201300118.

[29] F. Celikezen, C Orek, A. Parlak, K Sarac, H. Turkez, Ö. Ö. Tozlu, *Journal of Molecular Structure* **2020**, *1205*, 127577, DOI 10.1016/j.molstruc.2019.127577, http://doi.org/10.1016/j.molstruc.2019.127577.

[30] W. Chen, G. Wang, K. Mei, J. Zhu, *Rec. Nat. Prod* **2021**, *15*, 356–362, DOI 10.25135/rnp.225.21.01.1935, http://doi.org/10.25135/rnp.225.21.01.1935.

[31] A. O. Obaseki, J. E. Steffen, W. R. Porter, *Journal of heterocyclic chemistry* **1985**, *22*, 529–533, DOI 10.1002/jhet.5570220264, https://doi.org/10.1002/jhet.5570220264.

[32] G. Wang, S. B. Adhikari, P. Aryal, I. Okafor, A. Chen, A. Duffney, *Carbohydrate Research* **2025**, 109526, DOI 10.1016/j.carres.2025.109526, https://doi.org/10.1016/j.carres.2025.109526.

[33] C. Loarueng, B. Boekfa, S. Jarussophon, P. Pongwan, N. Kaewchangwat, K. Suttisintong, N. Jarussophon, *Arkivoc* **2019**, *6*, 116–127, DOI 10.24820/ark.5550190.p010.982, https://doi.org/10.24820/ark.5550190.p010.982.

[34] M. Mujahid, N. Trendafilova, A. F. Arfa-Kia, G. Rosair, K. Kavanagh, M. Devereux, M. Walsh, S. McClean, B. S. Creaven, I. Georgieva, *Journal of Inorganic Biochemistry* **2016**, *163*, 53–67, DOI 10.1016/j.jinorgbio.2016.07.010, https://doi.org/10.1016/j.jinorgbio.2016.07.010.

[35] T. Eckardt, V. Hagen, B. Schade, R. Schmidt, C. Schweitzer, J. Bendig, *The Journal of organic chemistry* **2002**, *67*, 703–710, DOI 10.1021/jo010692p, https://doi.org/10.1021/jo010692p.

[36] K. C. Majumdar, P. Chatterjee, *ChemInform* **1997**, *28*, DOI 10.1016/j.tetlet.2013.07.163, https://doi.org/10.1016/j.tetlet.2013.07.163.

[37] D. P. Kranz, A. G. Griesbeck, R. Alle, R. Pérez Ruiz, J. M. Neudörfl, K. Meerholz, H.-G. Schmalz, *Angewandte Chemie International Edition* **2012**, *51*, 6000–6004, DOI 10.1002/anie.201201222, https://doi.org/10.1002/anie.201201222.