Supporting Information

When Does Cytosine Dimerize? A Protonation Window for the Charge- and Resonance-Assisted [CytH$^+$···Cyt] Synthon

*Olga Książkiewicz [a,b], Zdzisław Kinart [a] , Marcin Palusiak [a], Kinga Wzgarda-Raj *a*

[a] Department of Physical Chemistry, University of Lodz, Pomorska 163/165, 91-236 Lodz, Poland

[b] Bio-Med-Chem Doctoral School of University of Lodz and Lodz Institutes of the Polish Academy of Sciences, Matejki 21/23, 90-237 Lodz, Poland

S1. Experimental Section

S2. Supplementary Results

S3. References

S1. Experimental Section:

S1. 1 Synthesis and crystallization (I)–(IV)

Preparation of salt (**I**). A 1:1 ratio of 3,5 dinitrosalicylic acid (0.0114 g) and cytosine (0.0056 g) was mixed and transferred to a 25 ml crystallizing dish. The mixture was then dissolved in a solution of ethanol (4.5 ml) and water (4 ml). Because of the limited solubility of the two reactants, the mixture was heated on a magnetic stirrer with a heating plate. Once the substances were completely dissolved, the crystallizing dish was sealed with parafilm and the product allowed to crystallize for three weeks.

Preparation of salt (**II**). In the crystallizations of 3,5-dinitrobenzoic acid with cytosine in the ratio 1:1, 1:2 and 1:4 the same cocrystals were obtained. For the X-ray studies described in the manuscript, a crystal from the synthesis of 3,5-dinitrobenzoic acid (0.0114 g) and cytosine (0.0056 g) in the ratio 1:1 was used. The weighted substances were transferred to a 25 ml crystallizing dish. The mixture was then dissolved in a solution of ethanol (4 ml) and water (5 ml). Because of the limited solubility of the two reactants, the mixture was heated on a magnetic stirrer with a heating plate. Once the substances were completely dissolved, the crystallizing dish was sealed with parafilm and the product allowed to crystallize for three weeks.

Preparation of salt (**III**). A 1:4 ratio of 4-nitrobenzoic acid (0.0084g) and cytosine (0.0224 g) was mixed and transferred to a 25 ml crystallizing dish. The mixture was then dissolved in a solution of ethanol (4 ml) and water (6 ml). Because of the limited solubility of the two reactants, the mixture was heated on a magnetic stirrer with a heating plate. Once the substances were completely dissolved, the crystallizing dish was sealed with parafilm and the product allowed to crystallize for three weeks.

Preparation of salt (**IV**). A 1:1 ratio of 4-nitrobenzoic acid (0.0084g) and cytosine (0.0056 g) was mixed and transferred to a 25 ml crystallizing dish. The mixture was then dissolved in a solution of ethanol (4 ml) and water (4 ml). Because of the limited solubility of the two reactants, the mixture was heated on a magnetic stirrer with a heating plate. Once the substances were completely dissolved, the crystallizing dish was sealed with parafilm and the product allowed to crystallize for three weeks.

S1.2 Refinement

Crystal data, data collection and structure refinement details are summarized in **Table S2.1**. The positions of H atoms of NH, NH_2 groups, -COOH group and water molecules were found in the Fourier difference map and refined isotropically without any restrictions. The remaining H atoms were introduced in calculated positions with an idealized geometry and constrained using a rigid-body model.

S1. 3 Cambridge Structural Database (CSD) search

A search through Crystal Structure Database was performed, using ver. 5.45 (with update from March 2024 included) of CSD [1]. As a main criterion of the search we used the structural molecular motif. We tightened the search to high quality crystal structure only. As a high quality structures we understand those crystal structure CSD [1] depositions which fulfil the following criteria: only non-disordered structures; with no errors and polymeric structures; only single crystal organic structures. In the discussion we refer to CSD search exclusively limited to such defined high quality data. For the CSD data selection we were using ConQuest tool [1]. Firstly we performed search for all high quality (see definition of high quality data in "CSD Search Details" section) CSD [1] depositions containing the unmodified (see **Figure S2.4**) cytosine molecule.

While searching the CSD [1] we identified in such a way 88 crystal structures. Note that we included in that search those crystal structures which contain N-protonated cytosine. It should be noted that from this structure selection we excluded all multiple precipitates, doing it manually if necessary and always selecting only the one with the smallest R value. For example, in the case of the hydrated cytosine structure - the natural crystal state of cytosine [2,3] - we selected only one of all 28 precipitates (refcode: CYTOSM21), high-resolution refinement, R=1.57% [4]. Of these 89 structures, in 61 cases the cytosine molecule was protonated, which indicates the presence of a counterion in the given crystal. The most interesting situation from the point of view of our research is the case when cytosine formed a dimeric structure, as shown in **Figure S2.5**.

S1. 4 Computational details

The protonated dimer of cytosine at geometry from the X-ray experiment was chosen for further theoretical calculations. Calculations were done with use of Gaussian 09 set of codes [5]. The full optimization of the geometry was performed at DFT-M062X [6] functional with dispersion correction, in conjunction with Def2TZVP [7] basis set. After the optimization procedure the interaction energy was estimated by means of super molecular method. At this stage the correction for basis set superposition error was implemented, using counterpose procedure in Gaussian 09 [5].

The complex of $[Cyt_2]^+$ can formally be classified as a complex stabilized by charge assisted hydrogen bonds (CAHB+) [8-14]. This is due to the fact that one of cytosine molecules is bearing formal positive charge due to protonation. On the other side, due to the specific location of formally single and double bonds linking individual proton-donating and proton-accepting centers in the complex of $[Cyt_2]^+$ can, the hydrogen bridges in that complex can also be classified as resonance-assisted ones [15-18]. Since both from X-ray experiment and from the results of optimization procedure in computational experiment the proton giving the formal charge of the complex is covalently bound to one of cytosine molecules, there was no problem of identification of fragments within the complex, which is crucial for the purpose

of super molecular approach. The interaction energy estimated is high, being of 49,73 kcal/mol. For comparison, the strength of H-bond in water dimer is of about 3-4 kcal/mol [19], while resonance assisted H-bond in formic acid dimer is close to 15 kcal/mol [20]. Therefore, The cytosine cationic dimer shall be considered stable molecular entity and cytosine itself will tend to form such dimer if only the environmental conditions will allow that.

S1. 5 ITC measurements

For ITC measurements, a solution of cytosine (for the working chamber) and a solution of hydrochloric acid (for the syringe) were prepared for each titration by dissolving weighed samples in distilled water and mixing overnight to obtain: 10 ml of a 0.5 mM cytosine solution (0.00074 g) and 10 ml of a 5 mM hydrochloric acid solution (0.1817 g). The thermodynamics of cytosine complexation with hydrochloric acid was studied using ITC titrations (Malvern MicroCal PEAQ-ITC, TA Instruments Nano ITC) in aqueous solutions at a temperature of 298.15 K. A 1.423 ml sample of a 0.5 mM aqueous cytosine solution (in the cuvette) was titrated with a 5 mM aqueous hydrochloric acid (from the syringe). Each titration included 50 injections performed every 500 s. For each individual injection, a volume of 5 μl was injected for 10 seconds from the syringe (containing the hydrochloric acid solution) into the working cell (containing the cytosine solution) at a constant stirring rate of 357 rpm. To correct for background, a titration of the pure solvent (water in a cuvette) with the hydrochloric acid solution (from the syringe) of identical acid concentration was performed and subtracted from the obtained enthalpograms. Thermodynamic parameters were calculated using a nonlinear multiparameter regression method with a single set of binding site models (in the dedicated MicroCal Origin 7.0 software).

S2. Supplementary Results

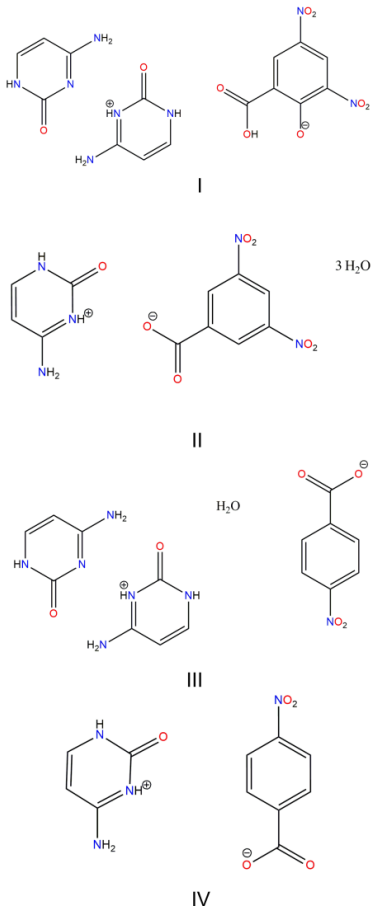

Figure S2. 1 The molecular formulas of (**I**) – (**IV**) multicomponent crystals (**I**) – cytosine dimer with 3,5-dinitrosalicylic acid; (**II**) –cytosine with 3,5-dinitrobenzoic acid trihydrate; (**III**) – cytosine dimer with 4-nitrobenzoic acid monohydrate; (**IV**) – cytosine with 4-nitrobenzoic acid.

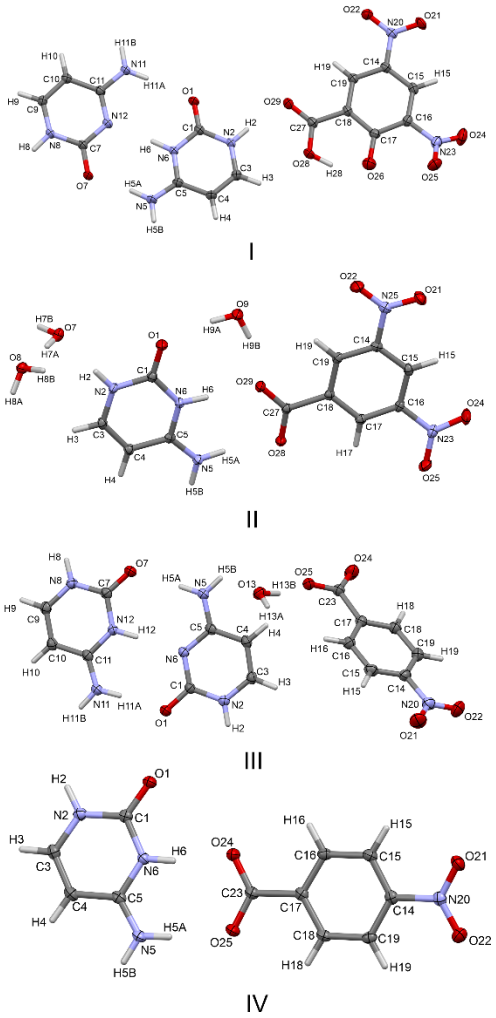

Figure S2. 2 The atom-labelling schemes for the molecular components of cocrystal (**I**)-(**IV**). Displacement ellipsoids for the non-H atoms are drawn at the 50% probability level.

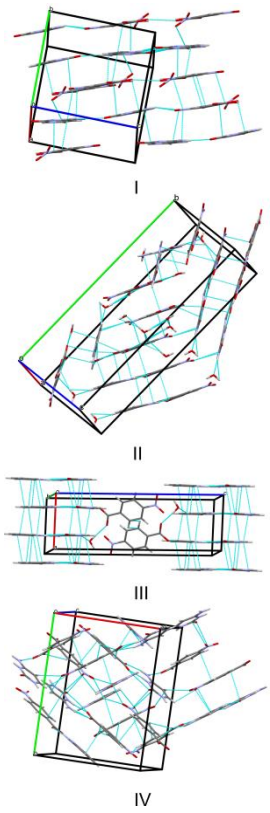

Figure S2. 3 Scheme of the selected π-π stacking interactions of the packing in the unit cells in structures (**I**)-(**IV**).

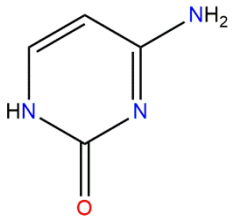

Figure S2. 4 The diagram shows an unmodified cytosine molecule.

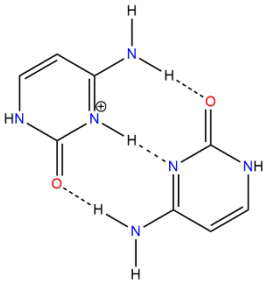

Figure S2. 5 Structural diagram of cytosine - the main criterion for the CSD search discussed.

Table S2. 1 Experimental details.

Crystal data	**Structure I**	**Structure II**	**Structure III**	**Structure IV**
Chemical formula	$C_4H_5N_3O_1{\cdot}C_4H_6N_3O_1{\cdot}\ C_7H_2N_2O_7$	$C_4H_6N_3O_1{\cdot}\ C_7H_3N_2O_6{\cdot}(H_2O)_3$	$C_4H_5N_3O_1{\cdot}C_4H_6N_3O_1{\cdot}\ C_7H_4N_1O_4{\cdot}\ H_2O$	$C_4H_6N_3O_1{\cdot}\ C_7H_4N_1O_4$
Formula weight	450.34	377.28	407.35	278.23
Temperature/ K	100(2)	100(2)	100(2)	100(2)
Radiation	Cu Kα (λ = 1.54184)	Cu Kα (λ = 1.54184)	Cu Kα (λ = 1.54184)	Cu Kα (λ = 1.54184)
Crystal system, space group	Triclinic, $P\bar{1}$	Monoclinic, $P2_1/c$	Triclinic, $P\bar{1}$	Orthorhombic, P*ccn*
a, b, c (Å)	8.3555(2) 9.9675(2) 10.8443(2)	4.93124(3) 24.69141(15) 12.91178(7)	6.6543(2) 7.3960(3) 19.1733(6)	11.90240(10) 15.66930(10) 12.40120(10)
α, β, γ (º)	93.2367(17) 95.2184(19) 99.8519(19)	90 91.8731(5) 90	86.316(3) 87.762(2) 66.189(3)	90 90 90
V ($Å^3$)	883.75(3)	1571.289(16)	861.43(5)	2312.85(3)
Z	2	4	2	8
μ (mm^{-1})	1.240	1.248	1.091	2.109
Crystal size (mm)	0.110 x 0.186 x 0.311	0.081 x 0.143 x 0.598	0.079 x 0.082 x 0.123	0.167 x 0.379 x 0.917
Reflections collected	22892	34572	28167	24293
Idependent reflections	3517[R_{int}=0.0340, R_{sigma}=0.0185]	3104[R_{int}=0.0504, R_{sigma}=0.0189]	3349[R_{int}=0.0469 R_{sigma}=0.0213]	2359[R_{int}=0.0493, R_{sigma}=0.0177]
Data/restraints/ parameters	3517/0/323	3104/0/276	3349/0/298	2259/0/198
Goodness-of-fit on F^2	1.042	1.042	1.021	1.083
Final R indexes [I>=2σ(I)]	R_1=0.0492, wR_2=0.1403	R_1=0.0303, wR_2=0.0845	R_1=0.0394, wR_2=0.1103	R_1=0.0350, wR_2=0.0990
Final R indexes [all data]	R_1=0.0515, wR_2=0.1431	R_1=0.0313, wR_2=0.0855	R_1=0.0431, wR_2=0.1138	R_1=0.0360, wR_2=0.1000
$\Delta\rho_{max}$, $\Delta\rho_{min}$ (e $Å^{-3}$)	1.38/-0.47	0.29/-0.29	0.21/-0.32	0.38/-0.27
CCDC number	2533644	2533643	2533641	2533642

Table S2. 2 Parameters of selected hydrogen bonds of structure (**I**)-(**IV**) (Å,º).

	D − H ⋯ A	**D − H**	**H ⋯ A**	**D ⋯ A**	**D − H ⋯ A**	***Symmetry codes***
(**I**)	N2 − H2 ⋯ O29	0.85(3)	2.01(3)	2.858(1)	172(3)	*x, y, z*
	N5 − H5A ⋯ O7	0.92(2)	1.82(2)	2.737(2)	176(2)	*x, y, z*
	N5 − H5B ⋯ O22	0.87(3)	2.23(3)	2.956(1)	141(3)	*-1+x, y, 1+z*
	N6 − H6 ⋯ N12	0.91(2)	1.95(2)	2.855(2)	174(2)	*x, y, z*
	N8 − H8 ⋯ O7	0.92(2)	1.89(2)	2.797(1)	171(1)	*-1-x, -y, 2-z*
	N11 − H11A ⋯ O1	0.87(2)	2.13(2)	2.999(2)	177(2)	*x, y, z*
	N11 − H11B ⋯ O1	0.89(2)	2.04(2)	2.874(1)	157(2)	*-x, -y, 1-z*
	O28 − H28 ⋯ O26	0.96(5)	1.54(5)	2.424(2)	151(4)	*x, y, z*
Intra	C3 − H3 ⋯ O25	0.95	2.50	3.258(2)	137	*2-x, 1-y, 2-z*
	C3 − H3 ⋯ O28	0.95	2.42	3.088(2)	127	*x, y, z*
	C4 − H4 ⋯ O24	0.95	2.59	3.388(2)	142	*2-x, 1-y, 2-z*

	C9 − H9 ⋯ O22	0.95	2.42	3.358(2)	169	*-x, -y, 1-z*
	C10 − H10 ⋯ O29	0.95	2.55	3.489(2)	171	*-x, -y, 1-z*
(**II**)	N2 − H2 ⋯ O7	0.86(1)	1.91(1)	2.768(1)	178(1)	*x, y, z*
	N5 − H5A ⋯ O28	0.87(1)	2.00(1)	2.869(1)	176(1)	*x, y, z*
	N5 − H5B ⋯ O9	0.88(1)	1.95(1)	2.811(1)	162(1)	*-1+x, ½ -y, ½ +z*
	N6 − H6 ⋯ O29	0.90(1)	1.87(1)	2.776(1)	179(1)	*x, y, z*
	O7 − H7A ⋯ O8	0.87(1)	2.03(2)	2.875(1)	165(1)	*-1-x, -y, 1-z*
	O7 − H7B ⋯ O8	0.85(1)	1.95(1)	2.795(1)	177(1)	*x, y, z*
	O8 − H8B ⋯ O7	0.87(2)	2.05(2)	2.889(1)	160(1)	*-1+x, y, z*
	O8 − H8A ⋯ O28	0.91(2)	1.87(2)	2.770(1)	172(1)	*-1+x, ½ -y, -½ +z*
	O9 − H9A ⋯ O1	0.85(2)	1.91(2)	2.744(1)	168(1)	*x, y, z*
	O9 − H9B ⋯ O29	0.87(2)	1.91(2)	2.761(1)	167(2)	*x, y, z*
	C4 − H4 ⋯ O9	0.95	2.45	3.158(1)	131	*-1+x, ½ -y, ½ +z*
	C4 − H4 ⋯ O22	0.95	2.52	3.263(1)	135	*-2+x, ½ -y, ½ +z*
	C17 − H17 ⋯ O8	0.95	2.59	3.511(1)	162	*1+x, ½ -y, ½ +z*
(**III**)	N2 − H2 ⋯ O25	0.92(2)	1.84(2)	2.735(1)	163(1)	*x, -1+y, z*
	N5 − H5A ⋯ O7	0.88(1)	2.08(1)	2.951(1)	173(1)	*1-x, -y, -z*
	N5 − H5B ⋯ O1	0.89(2)	2.08(2)	2.863(1)	146(1)	*x, 1+y, z*
	N8 − H8 ⋯ O13	0.96(2)	1.74(2)	2.697(1)	172(1)	*1+x, -1+y, z*
	N11 − H11A ⋯ O1	0.89(1)	1.85(1)	2.740(1)	178(3)	*1-x, -y, -z*
	N11 − H11B ⋯ O7	0.90(2)	2.02(2)	2.869(1)	156(1)	*x, 1+y, z*
	N12 − H12 ⋯ N6	0.91(1)	1.94(1)	2.846(1)	177(1)	*1-x, -y, -z*
	O13 − H13A ⋯ O24	0.92(3)	1.81(3)	2.724(1)	175(1)	*-1+x, y, z*
	O13 − H13B ⋯ O25	0.90(3)	1.92(3)	2.797(1)	164(2)	*x, y, z*
	C4 − H4 ⋯ O25	0.95	2.29	3.233(1)	175	*x, y, z*
	C9 − H9 ⋯ O22	0.95	2.36	3.259(1)	157	*1-x, -y, 1-z*
	C10 − H10 ⋯ O13	0.95	2.55	3.475(1)	166	*1+x, y, z*
(**IV**)	N2 − H2 ⋯ O1	0.89(1)	1.93(1)	2.811(1)	169(1)	*2-x, 1-y, 1-z*
	N5 − H5A ⋯ O25	0.92(1)	1.85(1)	2.764(1)	171(1)	*x, y, z*
	N5 − H5B ⋯ O24	0.87(2)	2.17(2)	3.019(1)	165(1)	*$^{3}/_{2}$-x, y, ½+z*
	N6 − H6 ⋯ O24	0.93(2)	1.75(2)	2.668(1)	171(1)	*x, y, z*
	C19 − H19 ⋯ O21	0.95	2.54	3.243(1)	131	*½-x, y, ½ +z*

Table S2. 3 Parameters of the selected π-π stacking interactions (Å,º).

π-π stacking interaction	**Cg ⋯ Cg**	**PLN1, PLN2**	**d(Cg ⋯ PLN)**	**Slippage**	***Symmetry codes***
Structure I					
Cg(ACID)A ⋯ Cg(ACID)*A*	3.7292(9)	0.02(8)	3.2616(7), 3.2616(7)	1.808	*2-x, 1-y, 1-z*
Cg(ACID)A ⋯ Cg(CYT)	3.8599(9)	1.98(8)	3.3016(7), 3.2429(6)	2.093	*1+x, y, z*
Cg(ACID)A ⋯ Cg(CYT)	5.4957(9)	8.17(8)	3.2748(7), 3.0330(6)	4.583	*1+x, y, z*
Cg(CYT) ⋯ Cg(ACID)A	3.8599(9)	1.98(8)	3.2429(6), 3.3016(7)	1.999	*-1+x, y, z*
Cg(CYT) ⋯ Cg(CYT)	4.7855(9)	9.76(8)	3.2369(6), 2.9457(6)	3.771	*1+x, y, z*
Cg(CYT) ⋯ Cg(CYT	4.4405(9)	9.76(8)	3.0504(6), 3.5426(6)	2.677	*-x, -y, 2-z*
Cg(CYT) ⋯ Cg(ACID)A	5.4956(9)	8.17(8)	3.0330(6), 3.2749(7)	4.413	*-1+x, y, z*
Cg(CYT) ⋯ Cg(CYT)	4.7856(9)	9.76(8)	2.9457(6), 3.2369(6)	3.525	*-1+x, y, z*
Cg(CYT) ⋯ Cg(CYT)	4.4406(9)	9.76(8)	3.5427(6), 3.0505(6)	3.227	*-x, -y, 2-z*
Cg(CYT) ⋯ Cg(CYT)	5.6847(9)	0.00(8)	2.8788(6), 2.8788(6)	4.902	*-1-x, -y, 1-z*
Structure II					
Cg(ACID)B ⋯ Cg(ACID)*B*	4.9312(6)	0.03(5)	3.2456(4), 3.2455(4)	3.713	*-1+x, y, z*
Cg(ACID)B ⋯ Cg(ACID)B	4.9313(6)	0.03(5)	3.2456(4), 3.2456(4)	3.713	*1+x, y, z*
Cg(ACID)B ⋯ Cg(ACID)*B*	5.5890(6)	0.03(5)	2.8924(4), 2.8925(4)	4.782	*2-x, -y, -z*
Cg(ACID)B ⋯ Cg(CYT)	5.7454(6)	4.07(5)	3.2704(4), 2.9269(4)	4.944	*1+x, y, z*
Cg(CYT) ⋯ Cg(ACID)*B*	5.7453(6)	4.07(5)	2.9269(4), 3.2703(4)	4.724	*-1+x, y, z*

Cg(CYT) ··· Cg(CYT)	4.9313(6)	0.00(5)	3.0924(4), 3.0924(4)	3.841	*-1+x, y, z*
Cg(CYT) ··· Cg(CYT)	4.9312(6)	0.00(5)	3.0923(4), 3.0923(4)	3.841	*1+x, y, z*
Structure III					
Cg(ACID)C ··· Cg(ACID)C	3.8831(8)	0.00(6)	3.5382(6), 3.5382(6)	1.600	*-x, 1-y, 1-z*
Cg(ACID)C ··· Cg(ACID)C	4.1641(8)	0.00(6)	3.4326(6), 3.4326(6)	2.357	*1-x, -y, 1-z*
Cg(CYT) ··· Cg(CYT)	3.4538(8)	1.84(7)	3.1392(6), 3.1073(6)	1.508	*-1+x, y, z*
Cg(CYT) ··· Cg(CYT)	3.6091(8)	1.84(7)	3.0760(6), 3.1202(6)	1.814	*x, y, z*
Cg(CYT) ··· Cg(CYT)	3.6091(8)	1.84(7)	3.1202(6), 3.0760(6)	1.888	*x, y, z*
Cg(CYT) ··· Cg(CYT)	3.4537(8)	1.84(7)	3.1073(6), 3.1392(6)	1.440	*1+x, y, z*
Cg(CYT) ··· Cg(CYT)	5.7290(8)	0.00(7)	3.0639(6), 3.0639(6)	4.841	*1-x, -y, -z*
Cg(CYT) ··· Cg(CYT)	5.4207(8)	0.00(7)	3.1512(6), 3.1512(6)	4.411	*2-x, -y, -z*
Structure IV					
Cg(ACID)C ··· Cg(ACID)*C*	3.9929(7)	7.35(5)	3.4197(5), 3.4197(5)	2.061	*½-x, ½-y, z*
Cg(ACID)C ··· Cg(CYT)	3.8629(7)	3.09(5)	3.3336(5), 3.2339(5)	2.113	*$^3/_2$-x, ½-y, z*
Cg(ACID)C ··· Cg(CYT)	5.0645(7)	3.09(5)	2.9336(5), 3.0975(5)	4.007	*-½+x, -½+y, 1-z*
Cg(CYT) ··· Cg(ACID)*C*	3.8630(7)	3.09(5)	3.2340(5), 3.3336(5)	1.952	*$^3/_2$-x, ½-y, z*
Cg(CYT) ··· Cg(ACID)*C*	5.9550(7)	69.93(5)	2.1896(5), 3.9511(5)	-	*½+x, ½+y, 1-z*
Cg(CYT) ··· Cg(ACID)C	5.0645(7)	3.09(5)	3.0978(5), 2.9335(5)	4.128	*$^3/_2$-x, y, ½+z*

Notes: Cg is the centre of gravity of the ring; (ACID)A is 3,5 dinitrosalicylic acid; (ACID)B is 3,5-dinitrobenzoic acid; (ACID)C is 4-nitrobenzoic acid; (CYT) is cytosine; PLN corresponds to the ring plane calculated by the least-squares method based of the positions of all ring-forming atoms; PLN1,PLN2 is the dihedral angle between neighbouring planes; d(Cg···PLN) is yhe mean perpendicular distance between Cg and the neighbouring stacking plane; slippage is the distance between Cg and d(Cg···PLN); A denotes the second molecular stacking of the asymmetric unit.

Table S2. 4 Parameters of hydrogen bonds occurring in the dimer motif of structures (II), (IV) and ZEGDUM (Å,º).

Structure	**N/O − H ··· O**	**N − H ··· N/O**	**N − H ··· O**
II	O9 − H9A ··· O1	N6 − H6 ··· O29	N5 − H5A ··· O28
	2.7439(12)	2.7758(12)	2.8689(12)
ZEGDUM	2.762	2.782	2.862
IV	N5 − H5A ··· O25	N6 − H6 ··· O24	-
	2.7643(14)	2.6684(14)	-

Table S3. 1 The estimation of the basis set superposition error correction.

Data:

M062X/Def2TZVP sp counterpoise=2

```
---------
Symbolic Z-matrix:
Charge =  1 Multiplicity = 1 in supermolecule
Charge =  0 Multiplicity = 1 in fragment      1.
Charge =  1 Multiplicity = 1 in fragment      2.
```

```
 O(Fragment=1)         1.24319  -2.18623   0.00047
 N(Fragment=1)         3.29925  -1.27408   0.00004
 N(Fragment=1)         1.38759   0.08404   0.00031
 N(Fragment=1)         1.60009   2.35647   0.0002
 H(Fragment=1)         0.58796   2.4365    0.00036
 H(Fragment=1)         2.15874   3.19185  -0.00015
 C(Fragment=1)         1.92169  -1.15864   0.00027
 C(Fragment=1)         3.60403   1.04876  -0.00047
 H(Fragment=1)         4.2366    1.92205  -0.00077
 C(Fragment=1)         4.11916  -0.19505  -0.00042
 H(Fragment=1)         5.18153  -0.39596  -0.00068
 C(Fragment=1)         2.17578   1.16016  -0.00002
 H(Fragment=1)         3.66475  -2.21521   0.00001
 O(Fragment=2)        -1.34715   2.25115   0.00024
 N(Fragment=2)        -3.37458   1.22976  -0.00017
 N(Fragment=2)        -1.41982  -0.02047   0.00012
 N(Fragment=2)        -1.43612  -2.30632  -0.00021
 H(Fragment=2)        -0.38904  -2.30854   0.00008
 H(Fragment=2)        -1.93742  -3.18003  -0.00011
 C(Fragment=2)        -1.99475   1.23267   0.0001
 C(Fragment=2)        -2.10739  -1.18497  -0.00009
 C(Fragment=2)        -3.53515  -1.12479  -0.00021
 H(Fragment=2)        -4.12171  -2.02908  -0.00034
 C(Fragment=2)        -4.11442   0.09137  -0.0002
 H(Fragment=2)        -5.18782   0.22313  -0.0003
 H(Fragment=2)        -0.35922  -0.00812   0.00016
 H(Fragment=2)        -3.81118   2.14115   0.00025
---------
```

Results:

```
Counterpoise corrected energy =    -790.317608622677
sum of monomers =    -790.238363125507
BSSE energy =       0.000804467772 = 0.51 kcal/mol
```

Table S3. 2 Cytosine dimer optimisation results with use of solvent (water) continuum model.

```
# M062X/Def2TZVP opt(MaxCycles=100) SCF(MaxCycles=999)
SCRF(Solvent=Water)

         Item               Value     Threshold  Converged?
 Maximum Force            0.000081     0.000450     YES
 RMS     Force            0.000016     0.000300     YES
 Maximum Displacement     0.000793     0.001800     YES
 RMS     Displacement     0.000277     0.001200     YES
 Predicted change in Energy=-1.069145D-07
 Optimization completed.
    -- Stationary point found.
      1          8           0       -3.617523   -0.113255
10.451433
      2          7           0       -4.906116   -1.161436
8.928689
      3          7           0       -2.751692   -0.430266
8.371264
      4          7           0       -1.937307   -0.776703
6.268106
      5          1           0       -1.073848   -0.319104
6.544558
      6          1           0       -2.031177   -1.125325
5.329788
      7          6           0       -3.729570   -0.540754
9.296852
      8          6           0       -4.146998   -1.556897
6.747996
      9          1           0       -4.283115   -1.942960
5.750646
     10          6           0       -5.109746   -1.657393
7.684526
     11          1           0       -6.067837   -2.123267
7.504955
     12          6           0       -2.930603   -0.911783
7.139019
     13          1           0       -5.622474   -1.232604
9.636555
     14          8           0        0.479773    0.577283
7.264764
     15          7           0        1.724423    1.627165
8.848891
     16          7           0       -0.409668    0.880693
9.332794
     17          7           0       -1.361884    1.155103
11.393872
     18          1           0       -2.221613    0.681290
11.063345
     19          1           0       -1.314615    1.487923
12.343805
     20          6           0        0.592606    0.998489
8.398324
     21          6           0       -0.332994    1.332468
10.606081
```

```
     22          6           0        0.867680    1.978368
11.019972
     23          1           0        0.968795    2.352248
12.025229
     24          6           0        1.857956    2.100574
10.114560
     25          1           0        2.801150    2.575579
10.341453
     26          1           0       -1.279264    0.399270
8.991714
     27          1           0        2.475858    1.725898
8.180064


HF=-790.3925398
```

S3. References

[1] C. R. Groom, I. J. Bruno, M. P. Lightfoot, S. C. Ward, *Acta Crystallogr., Sect. B.* 2016, **72**, 171–179

[2] M. E. Fleming, T. A. Watts, K. A. McKenna, E. K. Miehls and J. A. Swift, *Cryst. Growth Des*. 2022, **22**(12), 7443–7451

[3] G. A. Jeffrey and Y. Kinoshita, *Acta Crystallogr.* 1963, **16**, 20-28

[4] K. K. Jha, B. Gruza, P. Kumar, M. L. Chodkiewicza and P. M. Dominiak, *Cryst. Growth Des.* 2022, **22**(12), 7443–7451

[5] M. J. Frisch, W. H. Trucks, B. Schlegel, G.E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Li. H. Nakatsuji, 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, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K Throssell, J. A. Jr. Montgomery, Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman and D. J. Fox, 2019 Gaussian 16, Revision C.01, Gaussian, Inc., Wallingford CT.C

[6] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc*. 2008, **120**, 215–241

[7] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* 2005, **7**, 3297–3305

[8] C. B. Aakeröy, A. M. Beatty and K. R. Lorimer, K. R. *Struct. Chem.* 1999, **10**, 229–242

[9] D. Braga and F. Grepioni, *Acc. Chem. Res.* 2000, **33**, 601–608

[10] D. Braga, L. Maini, F. Grepioni, A. De Cian, O. Félix, J. Fischer and M. W. Hosseini, *New J. Chem.* 2000, **24**, 547–553

[11] F. Grepioni, G. Cojazzi,S. M. Draper, N. Scully and D. Braga, *Organometallics* 1998, **17**, 296–307

[12] D. Braga, S. M. Draper, E. Champeil and F. Grepioni, *J. Organomet. Chem.* 1999, **573**, 73–77

[13] G. Gilli and P. Gilli, *J. Mol. Struct.* 2000, **552**(1-3), 1-15

[14] B. Bankiewicz ana M. Palusiak, *Computational and Theoretical Chemistry* 2011, **966**, 113-119

[15] G. Gilli, F. Bellucci, V. Ferretti and V Bertolasi, *J. Am. Chem. Soc.* 1989, **111**, 1023–1028

[16] G. Gilli and P. Gilli, Oxford University Press: Oxford, 2009.

[17] V. Bertolasi, L. Pretto, G. Gilli and P. Gilli, *Acta Crystallogr. B* 2006, **62**, 850–863

[18] M. Domagała, M. Jabłoński, A. T. Dubis, M. Zabel, A. Pfitzner and M. Palusiak, *Int. J. Mol. Sci.* 2022, **23**, 14719

[19] M. W. Feyereisen, D. Feller and D. A. Dixon, *J. Phys. Chem.* 1996, **100** (8), 2993-2997

[20] R. Kalescky, E. Kraka and D. Cremer, *J. Chem. Phys.* 2014, **140**, 084315