

# Organic & Biomolecular Chemistry

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.



Journal Name

ARTICLE

# Synthesis and structural characterization of monomeric and dimeric Peptide Nucleic Acid prepared by microwave-promoted multicomponent reaction†

Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Reuben Ovadia<sup>a</sup>, Aurélien Lebrun<sup>b</sup>, Ivan Barvik<sup>d</sup>, Jean-Jacques Vasseur<sup>c</sup>, Carine Baraguey<sup>c</sup> and Karine Alvarez<sup>a</sup>

A solution phase synthesis of Peptide Nucleic Acid monomers and dimers was developed by using microwave-promoted Ugi multicomponent reactions. The mixture of a functionalized amine, a carboxymethyl nucleobase, paraformaldehyde and an isocyanide as building blocks generates PNA monomers which are then partially deprotected and used in a second Ugi 4CC reaction, leading to PNA dimers. Conformational rotamers were identified by NMR and MD simulations.

## Introduction

In our search for efficient inhibitors of viral polymerases to control infections originating from human viruses, we have recently described as HCV polymerase inhibitors several 5'-GC-3' phosphoramidate dinucleosides<sup>1, 2</sup>. These active compounds were rationally designed to hence their inhibition properties by blocking the active site of the HCV polymerase.

We took advantage of this original « dimeric » approach to conceive new compound family, (1) with higher binding to the active site of the polymerase and to the RNA template, (2) with higher stability in biological media and (3) easier to synthesize to afford molecular diversity to target a panel of viral polymerases.

Peptide Nucleic Acids (PNAs) appeared as good candidates to reach our goal. PNAs, described first in 1991 by Nielsen and his colleagues<sup>3</sup>, are structural homologs of DNA and RNA with a pseudo-peptide backbone of 2-amino-ethylglycine units to which purine and pyrimidine nucleobases are linked via a methylene carbonyl linkage. Oligomeric PNAs afford several advantages over

natural oligonucleotides as they show a high binding affinity to complementary DNA and RNA sequences and display stability towards enzymes which degrade peptides or nucleic acids. Many practical applications of PNAs and PNA analogues as gene therapy drugs or molecular probes for diagnostics have attracted much interest for over 20 years<sup>4, 5</sup>.

PNA oligomers are in most cases synthesized by solid phase peptide assembly from N-protected monomeric building blocks, via automated synthesis<sup>6</sup>. For short PNA sequences, liquid phase synthesis is preferred and a variety of N-protected unit syntheses<sup>7, 8</sup> has been described with different strategies of protection for the amino function of the N-terminus and for the amino function of the nucleobases. Dömling and colleagues have described<sup>9</sup>, for the first time in 1998, the original concept of combinatorial PNA synthesis<sup>10</sup> to obtain PNA monomers and multimers by consecutive Ugi-4CC reactions. Several reports demonstrated the usefulness of Ugi-4CC in the synthesis of PNA monomers<sup>11, 12</sup> in one-pot reaction, by mixing a functionalized amine, a carboxymethyl nucleobase, an oxo-component and an isocyanide. Despite the originality of this strategy to generate PNA multimers with a highly diverse substitution pattern, to date, only one publication<sup>13</sup> describes the synthesis of protected TT/CT sequences PNA dimers in solution.

We describe here an attractive synthetic strategy of PNA monomers (T, C, A, G, U) prepared by one-pot Ugi-4CC reaction and dimers (TT, AC, GC, GU, AG) prepared by two consecutive Ugi-4CC reactions (Scheme 1). In order to obtain target compounds with high yields and shorter reaction times, we turned our attention to microwave irradiation (MWI). Complete NMR assignment of PNA monomers and dimers was performed and allowed us to characterize conformational rotamers.

<sup>a</sup> Laboratoire d'Architecture et Fonction des Macromolécules Biologiques, Université Aix-Marseille UMR CNRS 7257, Equipe "Antiviral Medicinal Chemistry", Parc scientifique de Luminy, 163 av. de Luminy, 13288 Marseille Cedex 9, France.

<sup>b</sup> Institut des Biomolécules Max Mousseron (IBMM), UMR 5247 CNRS - Université Montpellier-ENSCM, Laboratoire de Mesures Physiques - Plateau Technique, Bâtiment 17, Campus Triolet, Place Eugène Bataillon, 34095 Montpellier Cedex 5, France.

<sup>c</sup> Institut des Biomolécules Max Mousseron (IBMM), UMR 5247 CNRS - Université Montpellier-ENSCM, Equipe « Oligonucléotides Modifiés », Bâtiment 17, Campus Triolet, Place Eugène Bataillon, CC1704, 34095 Montpellier Cedex 5, France.

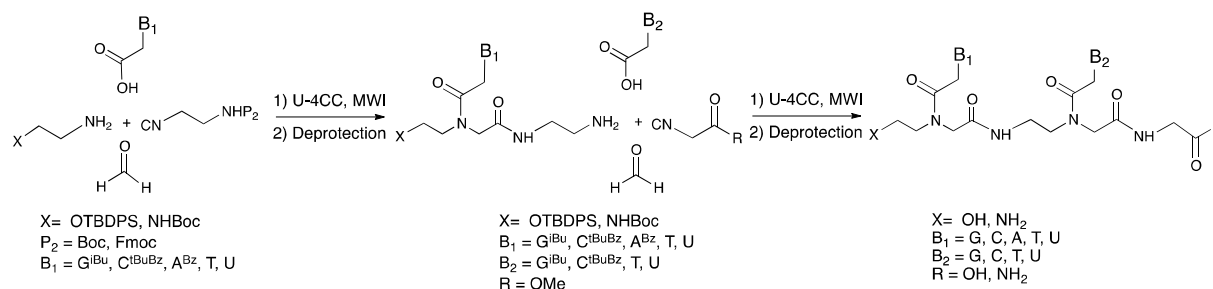
<sup>d</sup> Institute of Physics, Faculty of Mathematics and Physics, Charles University, Ke Karlovu 5, Prague 2, 121 16, Czech Republic.

† Electronic Supplementary Information (ESI) available: Experimental details and spectroscopic characterization. See DOI: 10.1039/x0xx00000x



Journal Name

ARTICLE

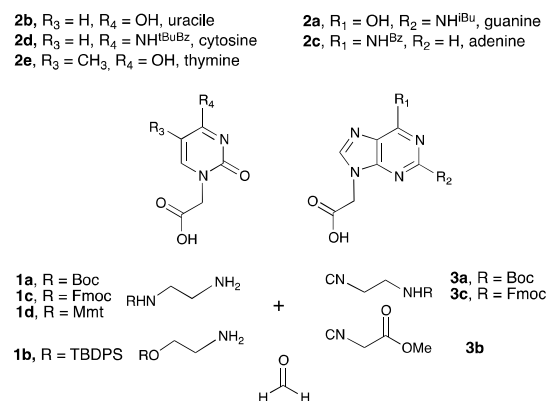


Scheme 1 Synthetic scheme of PNA dimers via two consecutive Ugi-4CC reactions

## Results and discussion

Largely used in combinatorial chemistry, the Ugi-4CC plays a significant role on the development of atom-economic and synthetically effective syntheses. Ugi-4CC is also known to be one of the most versatile tools for access to peptide like derivatives<sup>14-16</sup> by one-pot condensation of a functionalized amine, a carboxymethyl nucleobase, an oxo-component and an isocyanide.

To conceive PNA monomers and dimers, we first designed the different building blocks and the strategy of protection of the *N*-protected units (Scheme 1 and Figure 1) to modulate *N*-terminus and *C*-terminus patterns and offer different conditions of deprotection. Our choices were also driven by the ease of access to the designed building blocks. For the *N*-terminus, we chose hydroxylamine protected by a *tert*-butyl-diphenylsilyl group **1b** to generate hydroxyl end or ethylene diamine protected by the acid labile Boc group **1a** and Mmt group **1d** or by the Fmoc group **1c** to generate amino end. The carboxymethyl nucleobases **2a**, **2c** and **2d** were protected by base labile acyl groups. We designed also two isocyanides able to perform oligomerization **3a** and **3c** and one unable **3b**.

Fig. 1 Structure of the building blocks : **1** – functionalized amine, **2**- carboxymethyl nucleobase, **3**- isocyanide

The carboxymethyl nucleobases **2a**, **2b**, **2c** and **2e** were prepared in three steps according to literature procedures<sup>17, 18</sup> with 22%, 28% and 31% global yield, respectively. Hydroxylamine **1b** and ethylene diamine **1a** were synthesized according to the literature procedures<sup>19, 20</sup> with 76% and quantitative yield, respectively. Isocyanides **3a** and **3c** were prepared in 2 steps. Compound **1a** and **1c** were dissolved in ethyl formate and heated at reflux for 18 h to furnish the corresponding formamides. Dehydration of formamides with  $\text{POCl}_3$  in THF afforded isocyanides **3a** and **3c** with 80% and 38% global yield, respectively.

### Monomer synthesis

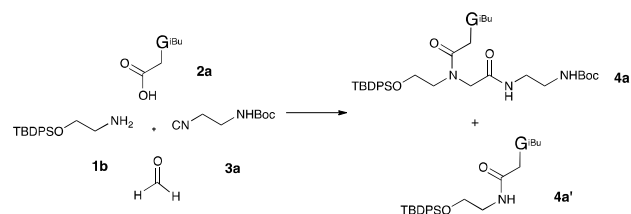
Commonly, the preparation in liquid phase of *N*-protected monomers requires two steps: (1) the synthesis of the backbone unit *via* reductive amination of a glycine ester with an *N*-protected

aminoacetaldehyde, or *via* an alkylation for mono-protected ethylenediamine with a bromoacetic acid ester, and (2) the linkage of the protected nucleobases *via* an amide bond using a coupling agent. For oligomer synthesis, the major drawback of this method is that prior preparation of each PNA monomer is needed. Likewise, any modification of the backbone requires the multi-steps monomer synthesis and a complete resynthesis of the oligomer.

With consecutive Ugi-4CC reaction method, the monomer is built up along the elongation of the chain and the oligomer formation is performed by cycle of coupling of nucleobase unit and then deprotection phase. This strategy can be used advantageously to introduce a diversity of building blocks during each cycle of elongation.

To probe the strategy to synthesize PNA monomers by Ugi-4CC previously proposed by P. Xu<sup>13</sup>, the Ugi product **4a** (guanine PNA monomer model) served our exploration to study the effect of concentration of building blocks, solvent, reaction time, temperature, MW irradiation, and pre-incubation of protected hydroxylamine **1b** and paraformaldehyde. We mixed protected hydroxylamine **1b**, carboxymethyl nucleobase **2a**, paraformaldehyde and isocyanide **3a** and modulated experimental conditions. Conditions and yields are gathered in Table 1. The aim of the study was to obtain the favorable ratio between compound **4a** and the byproduct **4a'**, which was identified by HPLC/MS as the amide resulting from the direct coupling between protected hydroxylamine **1b** and carboxymethyl nucleobase **2a**.

**Table 1** Optimization of reaction conditions for the synthesis of PNA monomer model **4a**.



Entry	Solvent	T (°C)	Time	Pre-incubation	MWI	Yield <b>4a</b>	Yield <b>4a'</b>
1	MeOH	rt	24h	NO	NO	26%	6%
2	Isopropanol	rt	24h	NO	NO	5%	2%
3	Isopropanol	rt	48h	NO	NO	12%	4%
4	Isopropanol	60	10h	NO	NO	23%	36%
5	MeOH	rt	24h	YES	NO	19%	2%
6	Isopropanol	rt	24h	YES	NO	15%	Trace
7	Isopropanol	60	10h	YES	NO	35%	12%
8	Isopropanol	60	3h	YES	YES	45%	4%
9	Isopropanol	70	1h30	YES	YES	40%	14%
10	Isopropanol	80	45min	YES	YES	36%	15%
11	Isopropanol	100	15min	YES	YES	33%	20%
12	Isopropanol	130	1min	YES	YES	30%	25%

Reaction conditions: **1b** (0.3 mmol), **2a** (0.3 mmol), **3a** (0.3 mmol), paraformaldehyde (0.3 mmol). Yields are expressed in percentage of byproduct **4a'** formation followed by HPLC and percentage of recovery of expected compound **4a**.

Ugi-4CC is generally performed at high concentration (0.5 M to 2 M) in protic solvent as methanol, ethanol or isopropanol at room temperature, during 24 to 48 h. Using classical conditions, with 0.5 M concentration of starting material in equimolar mixture at rt during 24 h in MeOH, the compound **4a** was obtained with 26% yield and the byproduct **4a'** formation was around 6% (entry 1, Table 1). The monomer precipitates in the medium when it is formed to facilitate its recovery as a white powder by simple filtration and washing. At higher concentration of starting material, until 2M, the yield was not improved due to the unfavorable precipitation of some building blocks. The addition of 0.05 or 0.1 equivalent of compound **1b** and paraformaldehyde *versus* compounds **2a** and **3a**, did not either improve the yield. By substituting the MeOH by isopropanol, the yield decreased (entry 2, Table 1) and was not widely improved neither even if the reaction was continued longer (48h) (entry 3, Table 1).

The effect of the reaction temperature was then evaluated. At 60 °C, the compound **4a** was formed faster and reaction time was reduced by half. However, these conditions were in favor of the formation of the byproduct **4a'** (36% yield) and unfavorable for the formation of the expected compound **4a** (23% yield) (entry 4, Table 1).

In order to reduce the formation of the byproduct **4a'**, the effect of the pre-incubation of the amine **1b** and paraformaldehyde leading to the first mechanism intermediate imine was evaluated. Thus, the protected hydroxylamine **1b** and paraformaldehyde were stirred during 1 h at room temperature in MeOH or isopropanol, the carboxymethyl nucleobase **2a** and isocyanide **3a** were then added and the mixture was stirred at rt for 24 h (entry 5 and 6, Table 1). In these conditions, the formation of the byproduct **4a'** was slightly reduced.

However, when the reaction was stirred at 60 °C for 10 h (entry 7, Table 1), the compound **4a** formation was improved from 23 to 35% yield and the formation of the byproduct **4a'** was drastically reduced from 36 to 12% yield.

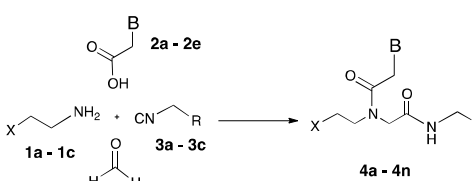
Because the use of energy-efficient microwave heating to facilitate chemical reactions often results in a dramatic acceleration of reactions, resulting in cleaner outcomes and increased yields, the effect of microwave irradiation (MWI) was evaluated at the same temperature of 60 °C. Thus, the protected hydroxylamine **1b** and paraformaldehyde were pre-incubated during 1 h, the carboxymethyl nucleobase **2a** and isocyanide **3a** were then added and the mixture was irradiated at 60 °C for 3 h. The reaction was faster and no longer changed after 3 h. The compound **4a** was obtained with 45% yield and the byproduct **4a'** formation was negligible (entry 8, Table 1). Under MWI, the rate of the reaction is governed by the law of Arrhenius. The reaction time is halved each time the temperature is raised of 10 °C, from 3 h reaction at 60 °C to 1 min at 130 °C. To evaluate the thermal effect of MWI, reaction was carried out at 70 °C for 1h30, 80 °C for 45 min, 100 °C for 15 min and 130 °C for 1 min (entry 9 to 12, Table 1). In all conditions, the compound **4a** was obtained with modest yields (30 to 40%) but a substantial proportion (14 to 25%) of byproduct **4a'** was also observed (14 to 25%). Therefore, if MWI has shown its ability to reduce the reaction time, up to 1000 times when the reaction was carried out at 130 °C and its ability to modestly improve the yield,

however the criterion of minimizing the formation of the byproduct **4a'** was not fulfilled when temperatures above 60 °C were used.

Optimal conditions should be a compromise between reaction time, yield and purity of the compound **4a**. Thus, MWI at 60 °C for 3 h in isopropanol preceded by a pre-incubation of the hydroxylamine **1b** and paraformaldehyde during 1 h seems to satisfy (entry 8, Table 1).

To explore the generality and the substrate scope of the developed condensation, several protected amines (**1a-d**), carboxymethyl nucleobases (**2a-e**) and isocyanides (**3a-c**) were used. The corresponding Ugi products (**4a-4n**) were obtained in good yields (Table 2) and purity higher than 95%, after purification by silica gel column chromatography. Compounds (**4a-4n**) were characterized by <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy and HRMS.

**Table 2** Synthesis of PNA monomers **4a** to **4n**.



Entry	R	P	X	R	B	Yield
1	<b>1b, 2a, 3a</b>	<b>4a</b>	OTBDPS	CH <sub>2</sub> NHBoc	G <sup>Bu</sup>	45%
2	<b>1b, 2b, 3a</b>	<b>4b</b>	OTBDPS	CH <sub>2</sub> NHBoc	C <sup>tBuBz</sup>	40%
3	<b>1b, 2c, 3a</b>	<b>4c</b>	OTBDPS	CH <sub>2</sub> NHBoc	A <sup>Bz</sup>	41%
4	<b>1b, 2d, 3a</b>	<b>4d</b>	OTBDPS	CH <sub>2</sub> NHBoc	T	65%
5	<b>1b, 2e, 3a</b>	<b>4e</b>	OTBDPS	CH <sub>2</sub> NHBoc	U	73%
6	<b>1b, 2b, 3b</b>	<b>4f</b>	OTBDPS	COOMe	C <sup>tBuBz</sup>	51%
7	<b>1b, 2d, 3b</b>	<b>4g</b>	OTBDPS	COOMe	T	62%
8	<b>1b, 2e, 3b</b>	<b>4h</b>	OTBDPS	COOMe	U	54%
9	<b>1a, 2b, 3b</b>	<b>4i</b>	NHBoc	COOMe	C <sup>tBuBz</sup>	37%
10	<b>1a, 2c, 3b</b>	<b>4j</b>	NHBoc	COOMe	A <sup>Bz</sup>	40%
11	<b>1a, 2d, 3b</b>	<b>4k</b>	NHBoc	COOMe	T	45%
12	<b>1a, 2e, 3b</b>	<b>4l</b>	NHBoc	COOMe	U	41%
13	<b>1d, 2b, 3b</b>	<b>4m</b>	NHMT	COOMe	C <sup>tBuBz</sup>	41%
14	<b>1b, 2a, 3c</b>	<b>4n</b>	OTBDPS	CH <sub>2</sub> NHfmoc	G <sup>Bu</sup>	43%

Reaction conditions: Yields are expressed in percentage of recovery of expected compound after purification by column chromatography. Conditions: 1 equiv of each reagent, pre-incubation, MWI, 60 °C, 3 h, isopropanol. R = reagent, P = product.

The PNA monomers **4a-e** and **4n** were designed to be used in the synthesis of PNA dimers where they will play the role of the amine by their C-terminus in the second Ugi-4CC reaction. The monomers **4f-m** were designed to be used in the synthesis of chimeric RNA-PNA dimers.

As indicated in Table 2 (entries 4, 5, 7, 8), monomers **4d**, **4e**, **4g** and **4h**, bearing pyrimidine nucleobases (thymine and uracil) were obtained in excellent yields (54 to 73%), comparable or rather higher to those described by P. Xu on similar PNA templates<sup>13</sup>. Moreover, for all products, the reaction conditions are improved (less equivalent) and reaction time is much shorter (3h *versus* 48h). A slightly decrease in the efficiency of the reaction was observed with protected carboxymethyl nucleobases (entries 1, 2, 3, 6, 9, 10,

13, 14, Table 2), however, the yields remain similar or better than those described by W. Maison<sup>11,12</sup>.

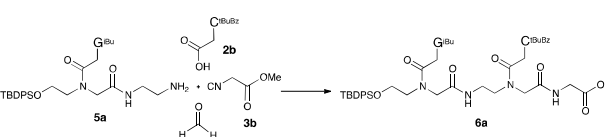
For **4f-h** series (entries 6, 7, 8, Table 2) only monomers bearing pyrimidine nucleobases were synthesized and good yields were observed (50-60%). For the **4i-l** series (entries 9, 10, 11, 12, Table 2), slightly lower yields were obtained, due to the slight decomposition of protected amine **1a** in the reaction conditions.

To be used in the second Ugi-4CC reaction as building blocks, the monomers **4a**, **4c**, **4d** and **4e** were first deprotected at their C-terminus. To remove Boc protection, monomers were treated with standard 50% TFA/DCM solution at 0 °C during 1h. After treatment scavenger resin was added to quench the trifluoroacetate ions generated during Boc removal. After filtration and evaporation, deprotected PNA monomers **5a**, **5c**, **5d** and **5e** were obtained as white powders, respectively, with 84, 81, 85 and 69% yield.

### Dimer synthesis

To probe the strategy to synthesize PNA dimers, the Ugi product **6a** (GC PNA dimer model) served our exploration to study the effect of solvent, reaction time and temperature under MW irradiation. We mixed deprotected PNA monomer **5a**, carboxymethyl nucleobase **2b**, paraformaldehyde and isocyanide **3b** and modulated experimental conditions. Results are gathered in Table 3.

**Table 3** Optimization of reaction conditions for the synthesis of PNA dimer model **6a**.



Entry	Solvent	T (°C)	Time	Pre-incubation	MWI	Yield <b>6a</b>
1	MeOH	rt	48h	YES	NO	7%
2	Isopropanol	rt	48h	YES	NO	17%
3	DMF	rt	48h	YES	NO	19%
4	Isopropanol	60	3h	YES	YES	30%
5	Isopropanol	70	1h30	YES	YES	39%
6	DMF	100	45 min	YES	YES	40%
7	Isopropanol	100	45 min	YES	YES	43%

Reaction conditions: **5a** (0.1 mmol), **2b** (0.1 mmol), **3b** (0.1 mmol), paraformaldehyde (0.1 mmol). Yields are expressed in percentage of recovery of expected compound **6a** after purification by column chromatography.

Using the standard Ugi-4CC conditions reactions, with 0.5 M concentration of building blocks, in equimolar mixture, in MeOH, isopropanol or DMF, at rt during 48 h, compound **6a** was obtained with 7%, 17% and 19% yield, respectively (entries 1, 2, 3, Table 3). In all cases, a pre-incubation between deprotected PNA monomer **5a** and paraformaldehyde during 1 h was performed as recommended for the synthesis of PNA monomers. The very low yield obtained in MeOH was due to the unfavorable precipitation of some building blocks, nevertheless the efficiency of the Ugi-4CC in



these standard conditions is weak. As expected, the substitution of MeOH by isopropanol or DMF affected positively the yield of the reaction (entries 2 and 3, Table 3).

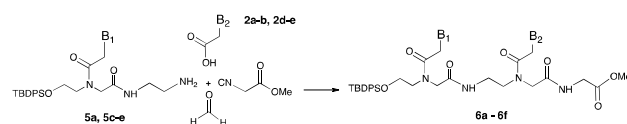
The effect of the MWI was then evaluated. Thus, the deprotected PNA monomer **5a** and paraformaldehyde were stirred during 1 h at room temperature in isopropanol, the carboxymethyl nucleobase **2b** and isonitrile **3b** were then added and the mixture was irradiated at 60 °C for 3 h. The compound **6a** was formed faster, the reaction time was reduced 16 times and the yield was improved from 19 to 30% (entry 4, Table 3).

To evaluate the thermal effect of MWI, reaction was carried out at 70 °C for 1h30 and 100 °C for 45 min (entries 4, 5, 7, Table 3). In all conditions, the yields are modest but serviceable (39 to 43%).

In conclusion, the synthesis of the dimer **6a** by Ugi-4CC is inefficient at rt, whatever the solvent used, due to the weak solubility of PNA monomer **5a** and carboxymethyl nucleobase **2b**. The use of MWI proved to be particularly beneficial, on the reaction time (60 times reduced) and on the efficiency of the reaction (6 times improved). Thus, irradiation at 100 °C for 45 min in isopropanol preceded by a pre-incubation of the deprotected PNA monomer **5a** and paraformaldehyde seems to satisfy (entry 7, Table 3).

To explore the generality and the substrate scope of the developed condensation, several PNA monomers (**5a**, **5c**, **5d** and **5e**), carboxymethyl nucleobases (**2a**, **2b**, **2d** and **2e**), isonitrile (**3b**) and paraformaldehyde were used. The corresponding Ugi products (**6a** – **6f**) were obtained in moderate to good yields (Table 4) and purity higher than 95%, after purification by silica gel column chromatography. Compounds (**6a** – **6f**) were characterized by <sup>1</sup>H and <sup>13</sup>C NMR spectroscopy and HRMS.

**Table 4** Synthesis of PNA dimers dimers **6a** to **6f**.



Entry	R	P	B <sub>1</sub>	B <sub>2</sub>	Yield
1	<b>5a, 2b</b>	<b>6a</b>	G <sup>Bu</sup>	C <sup>tBuBz</sup>	43%
2	<b>5a, 2e</b>	<b>6b</b>	G <sup>Bu</sup>	U	35%
3	<b>5c, 2b</b>	<b>6c</b>	A <sup>Bz</sup>	C <sup>tBuBz</sup>	30%
4	<b>5c, 2a</b>	<b>6d</b>	A <sup>Bz</sup>	G <sup>Bu</sup>	25%
5	<b>5d, 2d</b>	<b>6e</b>	T	T	40%
6	<b>5e, 2b</b>	<b>6f</b>	U	C <sup>tBuBz</sup>	30%

Reaction conditions: Yields are expressed in percentage of recovery of expected compound after purification by column chromatography. Conditions: 1 equivalent of each reagent, pre-incubation, MWI, 100 °C, 45 min, isopropanol. R = reagent, P = product.

The PNA dimers **6a-f** were designed to represent dimer sequences bearing all purine and pyrimidine nucleobases. Yields vary between 25 and 43%. In comparison with the PNA monomer synthesis, a decrease in the efficiency of the Ugi-4CC is observed and could be explained by the lower solubility of PNA monomers **5a-e** in the

reaction conditions. To bypass this solubility bottleneck, one solution should be to add an ionic liquid to the reaction mixture to improve the phase transfer and strengthen the MWI effect.

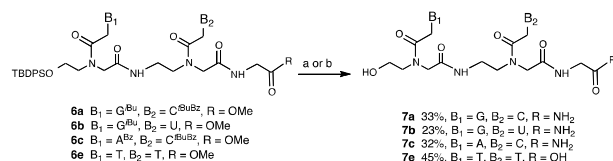
#### Deprotection of dimers **6a-c** and **6e**

The PNA dimers **6a**, **6b** and **6c** possess three types of protecting group. Among the different strategies of protection – deprotection described in literature for the synthesis of PNA oligomers<sup>21</sup>, the protections were chosen to be compatible with the synthesis of chimeric RNA-PNA or PNA-RNA<sup>22, 23</sup> and to avoid final deprotection in strong acidic conditions. The nucleobases are protected by acyl groups and the C-terminus as a methyl ester function cleavable in alkaline conditions. The N-terminus is protected by a silyl group cleavable with fluoride anions<sup>24</sup>.

Our first deprotection strategy was to apply a unique treatment to cleave all protections in one step. A solution of NaOH (0.2 N or 5 N) in MeOH or EtOH has been described<sup>25, 26</sup> to perform acyl and silyl group deprotection and ester hydrolysis. By using a solution of NaOH (0.2 N) in MeOH, we observed the rapid (< 1 h) saponification of the methyl ester function, the partial deprotection of the acyl groups and the cleavage of the silyl group. An extended treatment of 16 h was required to observe complete deprotection of acyl groups on purine nucleobases. However, this reaction time was responsible of an important degradation of the PNA dimer by cleavage of the amide bond between the nucleobase and the acetamide backbone. Therefore, a strong alkaline treatment for an unique deprotection, though attractive, is not possible.

Because aqueous ammonia solution is currently used for milder nucleobases deprotection and solid support cleavage in conventional oligonucleotide<sup>27</sup> or PNA oligomers synthesis<sup>28, 29</sup>, we used alternatively NH<sub>4</sub>OH solution (28%) in THF/MeOH at 55 °C (Scheme 2). After 6 h, complete deprotection of nucleobases and ammonolysis of the ester function were observed. After evaporation, TBAF (1M) solution in THF was added. Complete deprotection of the silyl group was observed after 1 h. If the treatment with TBAF solution was realized first, the methyl ester function was partially hydrolyzed (40%) along with the silyl group deprotection, leading to the formation of a mixture of one dimer with acidic C-terminus and one with amide C-terminus after ammoniacal treatment. The lack of orthogonality between protecting groups imposes the order in which deprotections should be performed. Thus, the dimers **6a**, **6b** and **6c** were solubilized in MeOH / THF (1:3, v:v), a 28% NH<sub>4</sub>OH solution was added and the mixture was stirred 6 h at 55 °C in a sealed tube. After evaporation, the residue was dissolved in THF and treated 1 h at rt with a solution of 1M TBAF in THF to yield the PNA dimers **7a-c** with 33, 23 and 32% yield respectively after reverse phase chromatography.

In order to generate a dimer with acidic C-terminus and because thymine does not require nucleobase protection, dimer **6e** was deprotected using alternative procedure (Scheme 2). Compound **6e** was first treated with a solution of 1M TBAF in THF during 1 h at rt to remove silyl group and then treated with NaOH 0.2 N in MeOH, during 1 h at rt to hydrolyze the ester function, leading to compound **7e** with 45% yield after reverse phase chromatography.



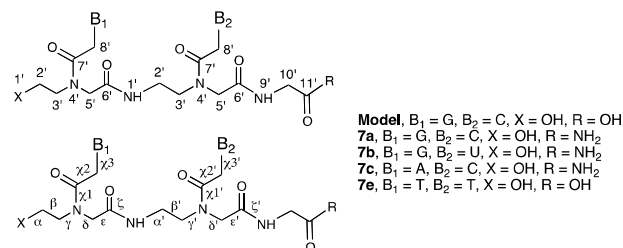
**Scheme 2** Synthesis of deprotected dimers **7a-c** and **7e**. Yields are expressed in percentage of recovery of expected compound after purification by reverse phase column chromatography. Reagents and Conditions: (a) 1) MeOH / THF (1:3, v:v), 28%  $NH_4OH$ , 55 °C, 6 h; 2) 1M TBAF in THF, rt, 1 h; (b) 1) 1M TBAF in THF, rt, 1 h; 2) NaOH 0.2N in MeOH, rt, 1 h.

### NMR characterization of PNA dimers **7a-c** and **7e** in aqueous solution

Nuclear Magnetic Resonance is among the analytical tools systematically used to validate the structure of synthetic compounds in a complementary way to other techniques such as mass spectroscopy. However in the case of small PNA oligomers, its use is limited by the complexity of the spectra. Indeed, for each monomer of a PNA oligomer, the restricted rotation around the amide bond linking the nucleobase to the acetamide backbone leads to the existence of two stable conformers<sup>30, 31</sup>. The energetic barrier between them is high and the interconversion rate is often slow enough (in the order of  $1 \text{ sec}^{-1}$ )<sup>32</sup> to observe a separate data set for each conformer. The coexistence of several species in solution thus explains why rather few NMR data exists to describe short PNAs oligomers<sup>30, 33</sup>. Indeed, lists of the  $^1H$  and  $^{13}C$  chemical shifts are usually given without assignment to the corresponding nuclei, or with a reliable but partial assignment. A possibility to simplify the analysis consists in accelerating the isomerization rate in order to approach the fast exchange regime and thus get a single averaged data set. This strategy was applied by Plöger *et al.* to study fluor-containing short PNA oligomers in DMSO<sup>34</sup>. If working at elevated temperature effectively led to simplified spectra, improved but nevertheless still incomplete assignments were obtained. Moreover this method is limited by the possible alteration of the compounds. We thus investigated NMR study in the conditions of the slow exchange regime occurring in  $D_2O$ .

We first used variation of temperature experiments to emphasize the existence of the rotamers. The  $^1H$  NMR spectrum of compound **7e** at 298 K (S29, Supporting Information) presented four sets of signals attributed to the putative existence of the 4 conformers. Several  $^1H$  spectra acquired on a temperature range of 298 to 363 K (S30, Supporting Information) showed a broadening of the resonances line shapes and coalescence was reached around 360 K. Resulting in the acceleration of the exchange rate between the different conformations, this behavior confirmed the existence of several conformers. In  $D_2O$ , the four PNA dimers **7a-c** and **7e** exhibited similar behaviors. The  $^1H$  spectra revealed several data sets of rather thin peaks, characteristic of slow exchanges at the NMR time scale. In each case one of the conformers appeared to be significantly predominant over the others (at more than 50 %) and it was thus possible to completely assign both the  $^1H$  and  $^{13}C$  nuclei of these major conformers.

We describe below the complete NMR characterization of 4 PNA dimers **7a-c** and **7e** in aqueous solution (For 1D  $^1H$  and  $^{13}C$  NMR spectra, see Experimental Section and Supporting Information S26-S29) based on the major rotamers. To describe the dimers, we used a nomenclature inspired from Brown *et al.*<sup>35</sup> and commonly used by others<sup>36</sup>. The numbering inside the nucleobases  $B_1$  and  $B_2$  follows the IUPAC recommendations defined for the nucleic acids (Figure 2).

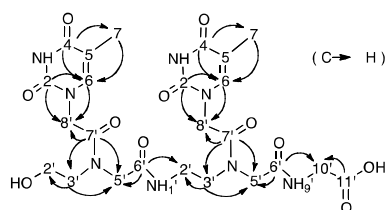


**Fig. 2** Atom numbering and torsion angles convention in the PNA dimers **7a-c**, **7e** and dimer model.

The  $^1H$  and  $^{13}C$  assignments were obtained by a combined analysis of  $J$ -coupling correlation spectroscopy (DQF-COSY<sup>37</sup>), heteronuclear multiple bond (HMBC<sup>38, 39</sup>) and single bond (HSQC<sup>40</sup>) correlation experiments. In addition to the conformational equilibrium and its effects on the spectra, the structure of the PNA itself with its abundance in heteroatoms and quaternary carbons makes the assignment more difficult than for natural oligonucleotides. Consequently and in order to discard any error, the identification of the nuclei was essentially based on a careful examination of the heteronuclear couplings. Through-space interactions were used to confirm the assignments and get conformational data and ROESY experiment<sup>41</sup> allowed to distinguish the dipolar interactions from the conformational exchange between the conformers since the latter had opposite sign.

In a general way, the assignment process was done using as starting point the aromatic protons of the nucleobases ( $H_8$  for the purines and  $H_6$  for the pyrimidines). For example for the dimer **7e**, the resonances of the nuclei composing each thymidine were identified from the correlations to the aromatic  $H_6$  protons (at 7.19 and 7.48 ppm respectively), the  $C_4$  quaternary carbons (169.3 and 169.4 ppm) being distinguished from  $C_2$  (154.8 ppm) according to their correlation to the methyl groups. Each methylcarbonyl linker was then connected to the corresponding nucleobase by the correlations implying the methylene  $H_8'$  protons with the  $C_2$  and  $C_6$  (146.4 ppm) carbons, a third correlation allowing to assign the carbonyl  $C_7'$ . On the pseudo-peptidic backbone, the two  $H_2$ - $H_3'$  spin systems of the aminoethyl moieties located on the COSY spectrum were differentiated from the higher chemical shift of the  $2'$  protons (at 3.81 ppm) of the first monomeric unit deshielded by the terminal hydroxyl group. The  $C_3$ - $H_5'$  and  $C_6$ - $H_5'$  correlations were used to complete the peptidic chain assignment of each monomer. From the correlations between the carbonyls  $C_7'$  and the methylenes  $H_3'$  and  $H_5'$  in the backbone, each fragment constituted of the base with its carboxymethyl linker were distinctly linked to the peptidic sequence. Finally the acid terminal end was connected to the rest of the molecule as shown by the correlations of the

methylene H<sub>10'</sub> with the two carbonyls C<sub>6'</sub> and C<sub>11'</sub>. The main correlations observed for the dimer **7e** are represented on Scheme 3.



**Scheme 3** Selected heteronuclear long-range correlations in the dimer **7e**. For a better clarity, most of the protons were not reported on the structure and some correlations were omitted.

The <sup>1</sup>H spectrum at 323 K of dimer **7a** revealed several data sets, one of the conformers prevailing significantly at 62%. The resonances of all the non exchanging protons were assigned. For the <sup>13</sup>C nuclei, only the C<sub>2</sub> and C<sub>6</sub> of the guanine were not determined. The <sup>1</sup>H spectrum of dimer **7b** revealed also a major conformer (50 %), for which all the protons resonances were assigned. The quaternary carbons of the bases were observed with difficulty, and exhibited particularly broad signals while C<sub>8</sub> was completely missing on the 1D <sup>13</sup>C NMR spectrum. Its chemical shift was nevertheless deduced from its correlation to H<sub>8'</sub> visible on the HMBC spectrum. No long range correlation was detectable for the carbons C<sub>2</sub> and C<sub>6</sub> of the guanine and the two corresponding resonances at 161.3 and 156.6 ppm were thus not individually assigned. The dimer **7c** at 298 K was found to exist as a mixture of 4 slowly interconverting conformers in a ratio of 67%, 17%, 9% and 8% as determined from the integration of the isolated H<sub>6</sub> protons of the cytidine. All the resonances were individually assigned. The <sup>13</sup>C resonances were determined from the collected HMBC experiment data. For dimer **7e**, the respective population of the 4 conformers was estimated as 58%, 24%, 12% and 6% according to the integration of the H<sub>6</sub> signals (S30, Supporting Information). The complete assignment of the major conformer was achieved, and the <sup>1</sup>H resonances were also determined for the second predominant one.

The whole of the chemical shifts (<sup>1</sup>H and <sup>13</sup>C) for the four dimers are reported in the Experimental Section. These NMR data taken together with the mass spectrometry analysis, confirmed the structure of the dimers **7a-c** and **7e**.

#### Conformational preference of the tertiary amide bond in GC PNA dimer model by MD simulations

The PNA monomers were extensively described in the literature and many examples report that among the 2 main conformers resulting of the *cis-trans* equilibrium around the methylcarbonyl linker amide bond, the *cis* form (that is to say, the one corresponding to the C<sub>7</sub> carbonyl pointing towards the glycine moiety) significantly (70:30) predominates<sup>31, 32, 42</sup>. The *cis* conformational preference becomes even more pronounced in PNA duplexes, where it's now well established from crystallography and

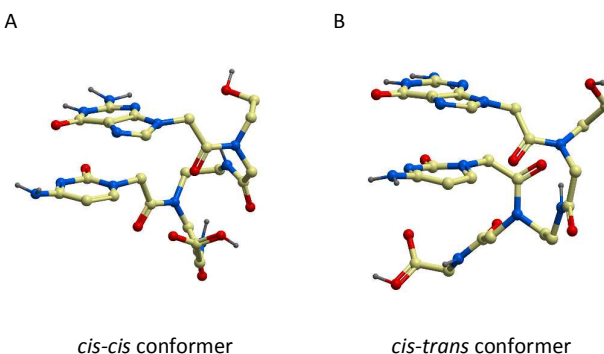
NMR studies that the carbonyls of the base linkers are oriented toward the C-terminal extremity of the PNA strand<sup>35, 36</sup>.

The rate of exchange was calculated as 0.5-2 s<sup>-1</sup> at 37°C<sup>34</sup>. The activation energy for the interconversion between *cis*- and *trans*-conformers has been determined by variable-temperature 1H-NMR spectroscopy on four PNA monomers as 19+2 kcal/mol<sup>34</sup>. This energy barrier is apparently too high, that could be overcome on a time scale of about one microsecond. Therefore, our MD simulations were started from various initial conformers (see S62, Supporting Information). While the secondary amide bond (i.e. backbone torsion angle ζ) was always set to *trans*, for tertiary amide bonds in linkages connecting bases to backbone (i.e. χ<sub>1</sub> and χ<sub>1'</sub> torsion angles), all possible combinations of conformers (i.e. *cis-cis*, *cis-trans*, *trans-cis* and *trans-trans*) were used leading to starting structures I-IV (see S62, Supporting Information). In fact, two four-membered sets of initial structures were constructed. In crystal structures, backbone torsion angles β'/δ prefer synchronously either *+gauche* or *-gauche* conformers. The *+gauche* conformers are usually found in crystal structures of right-handed homo- or heteroduplexes that contained at least one PNA strand<sup>35, 43</sup>. In contrast, the ideal value for β'/δ in left-handed PNA duplexes is -60°<sup>36</sup>. Therefore, both conformers (i.e. +/-*gauche*) were initially set for β'/δ in our simulated systems leading to two four-membered sets of initial structures (i.e. I-IVa vs. I-IVb - see S62, Supporting Information). Microsecond MD simulations were produced using programmable GPUs, which is a big step forward considering that previous calculations on PNA were mere minimizations of either molecular-mechanical energy of helical structures<sup>44</sup> or quantum-mechanical energy of dinucleotides<sup>45</sup> or at best short 100 ps MD simulations of CT-dimer<sup>32</sup> or 1.5 ns MD simulations of PNA single strands<sup>46</sup>.

Many inter-conversions of torsion angles β'/δ between *+gauche*/*-gauche* conformers were observed in our MD trajectories (see S61, Supporting Information). Many other torsion angles in PNA backbone (i.e. δ, ε, α', β', γ') or in backbone-base linkages (i.e. χ<sub>2</sub>, χ<sub>3</sub>, χ<sub>2'</sub>, χ<sub>3'</sub>) were also conformationally fluidic, which is usually considered as an asset with respect to binding of PNA to complementary DNA/RNA strand. Here, it caused even a complete de-stacking of dimers that was more or less reversible (see time evolutions of either N<sub>3</sub>/N<sub>1</sub> or C<sub>8</sub>/C<sub>6</sub> inter-atomic distances in S61, Supporting Informations). In MD simulations Ia, Ib, IIa, IIb (i.e. for PNA dimers with either *cis-cis* or *cis-trans* conformers in terms of the χ<sub>1</sub>/χ<sub>1'</sub> torsion angles), the base-stacked ordered structure appeared over long periods throughout MD trajectories lasting for 900-1.200 ns. In MD simulations IIIa, IIIb (i.e. for the *trans-cis* conformers in terms of the χ<sub>1</sub>/χ<sub>1'</sub> torsion angles), destacked conformers prevailed. Finally, in MD simulations IVa, IVb (i.e. for the *trans-trans* conformers of the χ<sub>1</sub>/χ<sub>1'</sub> torsion angles) stacking was observed only rarely. Taken together, relative viability of base stacking could at least partially explain uneven abundance of PNA conformers observed in our NMR experiments. Representative conformers of *cis-cis* and *cis-trans* PNA structures with the most viable base-stacking in MD simulations Ia, IIa are depicted in Figure 3. The *cis-cis* PNA structure (see Figure 3A) is one that is usually found in crystal structures of either duplexes or triplexes including



at least one PNA strand. The *cis-cis* conformation puts the PNA C<sub>7</sub> carbonyl groups in positions isosteric to an RNA C<sub>2</sub>, facilitating maximal solvent exposure of the two carbonyl oxygens. This arrangement appears to be optimal if one considers polarizations of individual atoms, as it puts into a close spatial proximity the oxygen atom from the carbonyl group of the residue with guanine base and the hydrogen atom on the secondary nitrogen along the backbone. In references<sup>44</sup>, the “inter-residue” hydrogen bonding stabilizing PNA-DNA and PNA-RNA hybrids was proposed on the base of molecular mechanics calculations produced in the gas phase. Nevertheless, it was not found neither in crystal structures of helical complexes nor in short (100 ps) MD simulations produced for a CT-dimer in a cube of water<sup>32</sup> and this hydrogen bonding was not stabilized even in the course of our microsecond MD runs. In MD simulation Ia (i.e. starting from the left-hand *cis-cis* conformer) we observed fluctuations for  $\beta'$ , however,  $\delta$  was stable. This means that we did not observe a complete transition from the left- to right-handed arrangement. Nevertheless, transition in the opposite direction was found in the MD simulations IIa. It could indicate, that PNA tends to be pre-organized in the left-handed arrangement (which is well-known from PNA:PNA duplexes) and the right-handed arrangement is imposed on PNA by DNA/RNA counterparts. According to our best knowledge, the *cis-trans* PNA structure (see Figure 3B) has not yet been studied. This conformer is potentially stabilized by the intra-molecular hydrogen bond established between the secondary backbone amidic proton and the C<sub>7</sub> carbonyl group from the monomer with cytosine base (see Figure 3B, and S61, Supporting Information). Earlier energy-minimized structures of Watson-Crick base-paired decameric duplexes of PNA with A-, B- and Z-DNA and A-RNA indicated that such “intra-residue” hydrogen bonding does not contribute to their stability<sup>44</sup>. However, these oligonucleotides were rather in the “all-trans” arrangement, that doesn’t lead to viable base stacking in our MD simulations IVab (see Figure S61, Supporting Information). On the other hand, the *cis-trans* PNA could be destabilized by unfavorable arrangement of oxygens from both carbonyl groups which are directed each against the other. It must lead to a substantial repulsion. Maybe it opens up space for the stabilization of this conformer by replacing the carbonyl group in the monomer with the guanine base. Compatibility of *cis-cis* and *cis-trans* PNA structures with the active site of RdRp will be subject to further examination.



**Fig. 3** Representative conformers of GC PNA dimer model with the most viable base-stacking in MD simulations. a) MD run 1a (t = 800 ns): torsion angle  $\beta'$  in -gauche,  $\chi_{\text{B}}$  in *cis*,  $\chi_{\text{B}}'$  in *cis*. The *cis-cis* conformation puts the PNA C<sub>7</sub> carbonyl group in a position isosteric to an RNA C<sub>2</sub>, facilitating maximal solvent exposure of the two carbonyl oxygens. b) MD run IIa (t = 600 ns): torsion angle  $\beta'$  in -gauche,  $\chi_{\text{B}}$  in *cis*,  $\chi_{\text{B}}'$  in *trans*.

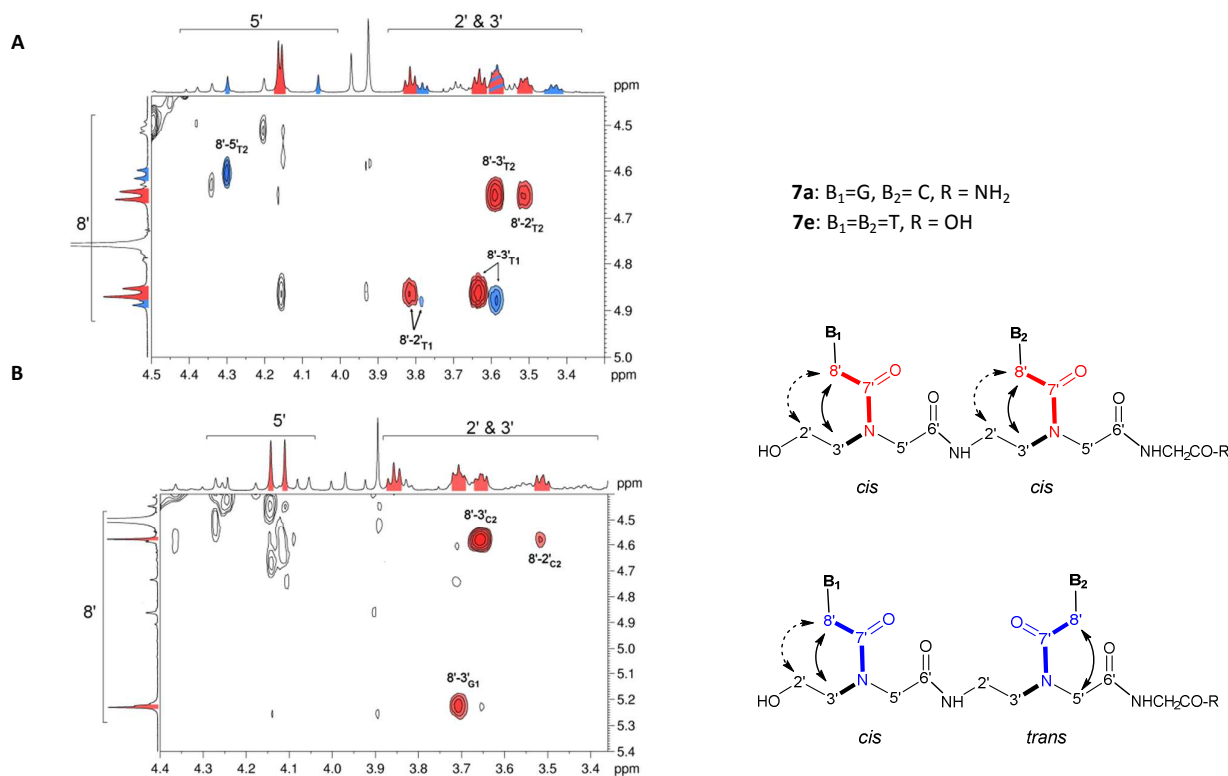
#### Validation of the conformational preference of the tertiary amide bond in dimers 7a and 7e by NMR studies

The conformational preference around the N<sub>4</sub>-C<sub>7</sub> amide bonds can be established from the ROESY spectrum, using the correlations of the carboxymethyl linker protons H<sub>8</sub>' with the methylenes of the backbone<sup>32, 47</sup>. ROESY spectra were thus acquired for **7a** and **7e** (Figure 4A and 4B respectively). For both compounds, the two H<sub>8</sub>' groups were found to have strongest cross correlations with the protons of the ethylenediamino moiety H<sub>3</sub>, suggesting the C<sub>7</sub> carbonyl groups to be turned towards the C-terminal extremity (*cis-cis* conformation). Among studied dimers, the dimer **7e** is particularly interesting because the splitting of the resonances and the respective population of the 4 conformers made possible to assign the <sup>1</sup>H resonances of the second majoritary conformer as well. In this second conformer, the strong H<sub>8</sub>'-H<sub>3</sub> correlation was found for the N-terminal residue (T<sub>1</sub>) while the conformation around the amide bond of the C-terminal residue was determined as *trans*, with the H<sub>8</sub>' methylene oriented to the glycine, as deduced from the H<sub>8</sub>'-H<sub>5</sub>' correlation (Figure 4B). The populations of the two minor conformers were too weak to get reliable NOE data. The result observed at the dimeric stage is in agreement with the monomeric behavior.



Journal Name

ARTICLE



**Fig. 4** (Left) Selected region of the ROESY spectra of **7a** (A) and **7e** (B). (Right) Favoured orientations of the N<sub>4</sub>-C<sub>7'</sub> bonds determined from the through-space correlations observed between H8' and the backbone's methylenes for the major conformer (in red) and second major conformer (for **7e** only).

Finally, from the chemical shifts collected for all the dimers (**7a-e** as well as the protected ones **6a** and **6c**) it appeared that among the two H8' methylenes of the nucleobase linkers, the one at the C terminal extremity was systematically found at a lower value. The effect was found to be more important when the N terminal extremity was bearing a purine as nucleobase (-0.4 to -1.0 ppm) rather than a pyrimidine (-0.2 ppm). This shield can be explained by the presence of this C-terminal H8' in the vicinity of the magnetic anisotropy cone of the N-terminal nucleobase and is consistent with the structures obtained for the GC model (*vide supra*) and for the TT dimer<sup>45</sup>.

## Conclusions

Peptide Nucleic Acids have received great attention as tools in molecular biology and biotechnology. Many analogues of PNA were designed to improve their physico-chemical and biological properties. Faster and more efficient methodologies for their synthesis need to be explored to circumvent the limitations of the classical peptide chemistry. This work describes an efficient, rapid and attractive solution phase strategy to obtain Peptide Nucleic Acid monomers and dimers by using microwave-promoted Ugi multicomponent reactions. Combining these two versatile tools is a particularly profitable method in current organic synthesis. Hence, we generated efficiently 14 PNA monomers (**4a** to **4n**, 37 to 73% isolated yield) by using one Ugi-4CC reaction and 6 PNA dimers (**6a** to **6f**, 25 to 43% isolated yield) by using two consecutive Ugi-4CC

reactions under microwave irradiation. This method is particularly attractive in terms of molecular diversity, simplicity and atom economy along with using readily available building blocks.

Although the PNA dimers exist in solution as 4 conformational rotamers, we have shown that it was possible to assign the resonances of the major species and thus to get a complete data sets for both  $^1\text{H}$  and  $^{13}\text{C}$  nuclei. MD simulations were run independently on a GC dimer model and established the supremacy of the *cis-cis* conformer, as result that was experimentally confirmed by NMR.

## Experimental

### General Information

TLC was performed on Merck silica coated plates 60F<sub>254</sub>. Compounds were revealed on UV light (254 nm). Column chromatography was performed on silica gel Merck G60 230-240 mesh. Flash chromatography was performed on a Biotage SP4 using silica gel column with an elution profile adapted to the crude. All moist-sensitive reactions were carried out under anhydrous conditions using dry glassware, anhydrous solvents and argon atmosphere. All anhydrous solvents were purchased in Aldrich. Microwave-promoted syntheses were performed on a Biotage<sup>®</sup> initiator using sealed vessels and the reaction temperature was monitored using an external IR sensor.

The NMR experiments were acquired on Bruker spectrometers equipped of 14.1 and 9.4 T magnets (AVANCE III 600 and AVANCE III HD 400 NMR systems). The samples were prepared by directly dissolving the products into deuterated solvents purchased from Eurisotop (Saint Aubin, France). The concentrations of the dimers were ranging from 15 to 27 mM for analysis in 5 mm diameter tubes. NMR microtubes with System Match (Bruker) were used for the products available in smaller quantities (dimers **7b** and **7c**) in order to reach concentrations of 70 and 8.3 mM respectively. The NMR data of **7c** and **7d** were collected at 298 K and the less soluble **7a** and **7b** were studied at 323 K.

In organic solution, the spectra were internally referenced according to the solvent residual peak (in DMSO:  $\delta^{13}\text{C}$  39.5 ppm;  $\delta^1\text{H}$  2.50 ppm; in  $\text{CDCl}_3$ :  $\delta^{13}\text{C}$  77.2 ppm;  $\delta^1\text{H}$  7.26 ppm; in  $\text{CD}_3\text{OD}$ :  $\delta^{13}\text{C}$  49.0 ppm;  $\delta^1\text{H}$  3.31 ppm). In aqueous medium, the DSS resonance was used to calibrate the  $^1\text{H}$  spectra at 0 ppm while the  $^{13}\text{C}$  spectra were indirectly referenced with  $\Xi_{\text{C}}=25.144953^{48}$ . The 2D spectra were obtained from standard pulse sequences with 1K to 2K data points in F2 and 256 to 512 data points in F1. Phase-sensitive ROESY spectra were carried out with a mixing time of 300 ms. Multiplicity-edited HSQC experiments using echo-antiecho were acquired with an evolution delay optimized for  $^1J_{\text{CH}}=145$  Hz. Magnitude mode  $^1\text{H}$ - $^{13}\text{C}$  HMBC spectra were acquired with an evolution period optimized for an average long-range coupling of 8.5 Hz.

Chemical shifts are given in parts per million (ppm) and coupling constants are given in hertz. Signal splitting patterns are described as singlet (s), doublet (d), triplet (t), quartet (q), broad signal (bs), doublet of doublet (dd), doublet of triplet (dt), doublet of quartet (dq), and multiplet (m). Low-resolution mass spectra (LRMS) were

recorded with a Q-TRAP mass analyzer under electrospray ionization (ESI) in positive or negative ionization mode detection. High-resolution mass spectra (HRMS) were recorded with a Q-Tof mass analyzer under electrospray ionization (ESI) in positive or negative ionization mode detection. LC/MS spectra were recorded on an Accela 600 thermofisher (HPLC) with an Hypersil-Gold C18 (2.5  $\mu\text{m}$ , 2.1x50 mm) column and Finnigan survey MSQ (MS) under electrospray ionization (ESI) in positive and negative ionization mode detection. Compounds were eluted with a gradient of a solution of 2 to 100% of 0.01% formic acid in ACN solution with a 550  $\mu\text{L}/\text{min}$  flow rate. High-pressure liquid chromatography (HPLC) spectra were recorded on a Waters apparatus using a Waters 600 pump and a Waters 916 photodiode array detector with an Xbridge C18 (10  $\mu\text{m}$ , 4.6x150 mm) column provided with a pre-column Xbridge (3.5  $\mu\text{m}$ , 100  $\text{\AA}$ , 4.6x20 mm). Compounds were eluted using a linear gradient of 10 to 100% of ACN or using a gradient of 0 to 100% of 0.05M triethylammonium bicarbonate buffer (pH 7.5) in 50% ACN with a 1mL/min flow rate.

**(2-isocyano-ethyl)-carbamic Acid tert-butyl Ester (3a).** *tert*-butyl N-(2-aminoethyl)carbamate **1a** (3.8 g, 23.70 mmol) was dissolved in ethyl formate (20 mL) and the solution was heated at reflux for 18 h. Upon cooling to room temperature, the reaction mixture was concentrated under reduced pressure to furnish the corresponding formamide **1'a** (4.4 g, quantitative yield).  $R_f$  0.55 (DCM/MeOH, 9/1);  $^1\text{H}$  NMR (250 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 8.12 (s, 1H, COH), 6.36 (s, 1H, NH), 4.89 (s, 1H, NH), 3.40-3.28 (m, 2H,  $\text{CH}_2$ ), 3.26-3.13 (m, 2H,  $\text{CH}_2$ ), 1.37 (s, 9H,  $\text{CH}_3$ ).  $^{13}\text{C}$  NMR (63 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 161.7 (CO), 156.9 (CO), 79.9 (C-tBu, C), 40.3 ( $\text{CH}_2$ ), 39.4 ( $\text{CH}_2$ ), 28.3 ( $\text{CH}_3$ ). TOF MS  $\text{ESI}^+$  189.1  $[\text{M}+\text{H}]^+$ .

A solution of formamide **1'a** (4.4 g, 23.30 mmol, 1.0 equiv.) and triethylamine (15.7 mL, 116.50 mmol, 5.0 equiv.) in anhydrous THF (72 mL) was cooled to 0°C, and phosphorous oxychloride (2.4 mL, 25.60 mmol, 1.1 equiv.) was added dropwise. The reaction mixture was stirred at 0 °C for 2 h and at room temperature for 1 h. The reaction mixture was then poured into cold  $\text{H}_2\text{O}$  (2 x reaction volume), the organic layer was separated, and the aqueous layer was extracted with  $\text{Et}_2\text{O}$  (3 x 100 mL). The combined organic fractions were washed with brine, dried over  $\text{Na}_2\text{SO}_4$  filtered and concentrated under reduced pressure. The crude was purified over silica gel column chromatography (1:1 EtOAc – hexane) to give (2-isocyano-ethyl)-carbamic acid *tert*-butyl ester **3a** as a white powder (2.83 g, 71% yield).  $^1\text{H}$  NMR (250 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 4.91 (s, 1H, NH), 3.46 (t, 2H,  $J = 5.3$  Hz,  $\text{CH}_2$ ), 3.37-3.25 (m, 2H,  $\text{CH}_2$ ), 1.39 (s, 9H,  $\text{CH}_3$ ).  $^{13}\text{C}$  NMR (63 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 157.6 (CN), 155.7 (CO), 80.3 (C-tBu, C), 42.1 ( $\text{CH}_2$ ), 39.9 ( $\text{CH}_2$ ), 28.4 ( $\text{CH}_3$ ). LC/MS  $\text{ESI}^+$  171.15  $[\text{M}+\text{H}]^+$ ; 340.89  $[\text{2M}+\text{H}]^+$ .

**(2-isocyano-ethyl)-carbamic acid 9H-fluoren-9-yl-methyl ester (3c).** N-(2-aminoethyl)carbamic acid 9H-Fluoren-9-ylmethyl ester hydrochloride **1c** (501 mg, 1.57 mmol, 1.0 equiv.) was dissolved in ethyl formate (9 mL) and the solution was heated at reflux and triethylamine was added (255  $\mu\text{L}$ , 1.89 mmol, 1.2 equiv.). The solution was stirred at reflux for 18 h. Upon cooling to room temperature, the reaction mixture was concentrated under reduced

pressure and the crude was purified over silica gel column chromatography (DCM - MeOH) to furnish the corresponding formamide **1c** (519 mg, 94% yield).  $^1\text{H}$  NMR (250 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 7.94 (s, 1H, COH), 7.82 (d, 2H,  $J = 7.2$  Hz, H arom), 7.61 (d, 2H,  $J = 7.2$  Hz, H arom), 7.30 (m, 4H, H arom), 4.24 (d, 2H,  $J = 6.6$  Hz,  $\text{CH}_2$ ), 4.14 (t, 1H,  $J = 6.6$  Hz, CH), 3.15-2.90 (m, 4H,  $2\text{CH}_2$ ).  $^{13}\text{C}$  NMR (63 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 161.2 (CO), 156.1 (CO), 143.9 (C), 140.7 (C), 127.6 (CH arom), 127.0 (CH arom), 125.1 (CH arom), 120.0 (CH arom), 65.3 ( $\text{CH}_2$ ), 46.7 (CH), 40.4 ( $\text{CH}_2$ ), 39.9 ( $\text{CH}_2$ ). LC/MS  $\text{ESI}^+$  310.91  $[\text{M}+\text{H}]^+$ ; 621.12  $[2\text{M}+\text{H}]^+$ .

A solution of formamide **1c** (500 mg, 1.61 mmol, 1.0 equiv.) and triethylamine (1.0 mL, 8.06 mmol, 5.0 equiv.) in anhydrous DCM (4 mL) was cooled to  $0^\circ\text{C}$ , and phosphorous oxychloride (165  $\mu\text{L}$ , 1.77 mmol, 1.1 equiv.) was added dropwise. The reaction mixture was stirred at  $0^\circ\text{C}$  for 2 h and at room temperature for 1 h. The reaction mixture was then poured into cold  $\text{H}_2\text{O}$  (2 x reaction volume), the organic layer was separated, and the aqueous layer was extracted 3 times with DCM. The combined organic fractions were washed with brine, dried over  $\text{Na}_2\text{SO}_4$  filtered and concentrated under reduced pressure. The crude isonitrile was purified over silica gel column chromatography (1:1 EtOAc - hexane) to give (2-isocyano-ethyl)-carbamic acid 9H-fluoren-9-yl-methyl ester **3c** as a white powder (194 mg, 41% yield).  $^1\text{H}$  NMR (250 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 7.70 (d, 2H,  $J = 7.2$  Hz, H arom), 7.51 (d, 2H,  $J = 7.3$  Hz, H arom), 7.39-7.12 (m, 4H, H arom), 4.37 (d, 2H,  $J = 6.6$  Hz,  $\text{CH}_2$ ), 4.15 (t, 1H,  $J = 6.6$  Hz, NH), 3.41 (m, 4H,  $\text{CH}_2$ ).  $^{13}\text{C}$  NMR (63 MHz,  $\text{CDCl}_3$ )  $\delta$  (ppm) 155.2 (CO), 142.7 (C), 140.4 (C), 126.8 (C arom), 126.1 (C arom), 123.9 (C arom), 119.0 (C arom), 66.0 ( $\text{CH}_2$ ), 46.2 (CH), 40.9 ( $\text{CH}_2$ ), 39.3 ( $\text{CH}_2$ ). LC/MS  $\text{ESI}^+$  292.93  $[\text{M}+\text{H}]^+$ .

#### Preparation of compounds 4a-n.

##### 2-[[[ $(\text{N}^2$ -(isobutanoyl)guanin-9-yl)acetyl]-(2-(tert-butyl(diphenylsilyl)oxy)ethyl)]]amino-N-[2-(tert-butyloxycarbonyl-amino)ethyl]acetamide (4a).

A mixture of 2-[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (500 mg, 1.67 mmol, 1.0 equiv.) and paraformaldehyde (50.2 mg, 1.67 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 3.3 mL) was placed in a sealed vessel (0.5 - 2 mL) and stirred at rt for 1 h. (2-Isocyano-ethyl)-carbamic acid tert-butyl ester **3a** (284.3 mg, 1.67 mmol, 1.0 equiv.) and  $\text{N}^2$ -(isobutanoyl)-9-(carboxymethyl) guanine **2a** (466.5 mg, 1.67 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W ( $60^\circ\text{C}$ ) for 3 h. The precipitate was filtered, washed with cold MeOH to give **4a** as a white powder (563 mg, 45% yield). Tr-HPLC: 23.6 min (> 95%, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ )  $\delta$  (ppm) 12.07 and 12.06 (bs, 1H,  $\text{H}_1$ ), 11.69 and 11.65 (bs, 1H, NH-iBu), 8.26 (t, 1H, rotamer 1,  $J = 5.4$  Hz,  $\text{H}_9$ ), 7.94 (t, 1H, rotamer 2,  $J = 5.2$  Hz,  $\text{H}_9$ ), 7.77 (s, 1H, rotamer 1,  $\text{H}_8$ ), 7.47 (s, 1H, rotamer 2,  $\text{H}_8$ ), 7.64 (d, 2H,  $\text{H}_{\text{ortho-TBDPS}}$ , rotamer 2), 7.59 (d, 2H,  $\text{H}_{\text{ortho-TBDPS}}$ , rotamer 1), 7.46-7.42 (m, 6H,  $\text{H}_{\text{meta/para-TBDPS}}$ ), 6.85 (t, 1H, rotamer 1,  $J = 5.4$  Hz,  $\text{H}_{12'}$ ), 6.74 (t, 1H, rotamer 2,  $J = 5.4$  Hz,  $\text{H}_{12'}$ ), 5.02 (s, 2H, rotamer 2,  $\text{H}_8$ ), 4.97 (s, 2H, rotamer 1,  $\text{H}_8$ ), 4.17 (s, 2H, rotamer 1,  $\text{H}_5$ ), 3.94 (s, 2H, rotamer 2,  $\text{H}_5$ ), 3.88 (t, 2H, rotamer 2,  $J = 5.4$  Hz,  $\text{H}_2$ ), 3.68 (t, 2H, rotamer 1,

$J = 6.1$  Hz,  $\text{H}_2$ ), 3.62 (t, 2H, rotamer 2,  $J = 5.4$  Hz,  $\text{H}_3$ ), 3.47 (t, 2H, rotamer 1,  $J = 6.1$  Hz,  $\text{H}_3$ ), 3.13 (m, 2H, rotamer 1,  $\text{H}_{10'}$ ), 3.04 (m, 2H, rotamer 2,  $\text{H}_{10'}$ ), 3.04 (m, 2H, rotamer 1,  $\text{H}_{11'}$ ), 2.94 (m, 2H, rotamer 2,  $\text{H}_{11'}$ ), 2.75 (sept, 1H,  $J = 6.9$  Hz, iBu-CH), 1.36 (s, 9H, tBu-Boc), 1.10 (d, 6H,  $J = 6.9$  Hz, iBu- $\text{CH}_3$ ), 1.03 (s, 9H, tBu-TBDPS).  $^{13}\text{C}$  NMR (100 MHz,  $\text{DMSO}-d_6$ )  $\delta$  (ppm) 180.1 (C-iBu, CO), 167.9 ( $\text{C}_6$ , CO), 167.1 ( $\text{C}_7$ , CO, rotamer 1), 166.4 ( $\text{C}_7$ , CO, rotamer 2), 155.7 (C-Boc, CO, rotamer 1), 155.6 (C-Boc, CO, rotamer 2), 154.8 ( $\text{C}_6$ , CO), 149.3 ( $\text{C}_4$ , C, rotamer 1), 149.2 ( $\text{C}_4$ , C, rotamer 2), 147.8 ( $\text{C}_2$ , C), 140.5 ( $\text{C}_8$ , CH, rotamer 1), 140.0 ( $\text{C}_8$ , CH, rotamer 2), 135.0 ( $\text{C}_{\text{ortho-TBDPS}}$ , CH), 132.8 ( $\text{C}_{\text{ph-TBDPS}}$ , C, rotamer 1), 132.6 ( $\text{C}_{\text{ph-TBDPS}}$ , C, rotamer 2), 130.0 ( $\text{C}_{\text{para-TBDPS}}$ , CH, rotamer 2), 129.9 ( $\text{C}_{\text{para-TBDPS}}$ , CH, rotamer 1), 128.0 ( $\text{C}_{\text{meta-TBDPS}}$ , CH, rotamer 2), 127.9 ( $\text{C}_{\text{meta-TBDPS}}$ , CH, rotamer 1), 119.5 ( $\text{C}_5$ , C), 77.7 (CtBu-Boc, C), 61.7 ( $\text{C}_2$ ,  $\text{CH}_2$ , rotamer 2) 60.8 ( $\text{C}_2$ ,  $\text{CH}_2$ , rotamer1), 50.6 ( $\text{C}_5$ ,  $\text{CH}_2$ , rotamer1) 48.4 ( $\text{C}_5$ ,  $\text{CH}_2$ , rotamer2), 49.3 ( $\text{C}_3$ ,  $\text{CH}_2$ , rotamer 1) 49.2 ( $\text{C}_3$ ,  $\text{CH}_2$ , rotamer2), 44.2 ( $\text{C}_8$ ,  $\text{CH}_2$ , rotamer 2), 44.1 ( $\text{C}_8$ ,  $\text{CH}_2$ , rotamer1), 39.4 ( $\text{C}_{11'}$ ,  $\text{CH}_2$ ), 38.7 ( $\text{C}_{10'}$ ,  $\text{CH}_2$ ), 34.6 (C-iBu, CH), 28.2 (C-tBu-Boc,  $3\text{CH}_3$ ), 26.7 (C-tBu-TBDPS,  $3\text{CH}_3$ , rotamer 2), 26.6 (C-tBu-TBDPS,  $3\text{CH}_3$ , rotamer 1), 18.8 (C-iBu,  $2\text{CH}_3$ ). HRMS  $\text{ESI}^+$  calcd for  $\text{C}_{38}\text{H}_{53}\text{N}_8\text{O}_7\text{Si}^+$  ( $\text{M}+\text{H}$ ) $^+$  761.3806, found 761.3799.

##### 2-[[[ $(\text{N}^4$ -(4-tert-butylbenzoyl)cytosin-1-yl)acetyl]-(2-(tert-butyl(diphenylsilyl)oxy)ethyl)]]amino-N-[2-(tert-butyloxycarbonyl-amino)ethyl]acetamide (4b).

A mixture of 2-[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (200 mg, 0.67 mmol, 1.0 equiv.) and paraformaldehyde (20 mg, 0.67 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.3 mL) was placed in a sealed vessel (0.5 - 2 mL) and stirred at rt for 1 h. (2-Isocyano-ethyl)-carbamic Acid tert-butyl Ester **3a** (113.7 mg, 0.67 mmol, 1.0 equiv.) and  $\text{N}^4$ -(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (220.2 mg, 0.67 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W ( $60^\circ\text{C}$ ) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4b** as a white powder (273.8 mg, 50% yield). Tr-HPLC: 26.9 min (> 99%, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz,  $\text{DMSO}-d_6$ )  $\delta$  (ppm) 11.1 (s, 1H, NH-tBuBz), 8.20 (t, 1H, rotamer 1,  $J = 5.8$  Hz,  $\text{H}_9$ ), 7.83 (t, 1H, rotamer 2,  $J = 5.8$  Hz,  $\text{H}_9$ ), 7.98-7.97 (m, 2H,  $\text{H}_{\text{ortho-tBuBz}}$ ), 7.83-7.80 (m, 1H,  $\text{H}_6$ ), 7.65-7.61 (m, 4H,  $\text{H}_{\text{ortho-TBDPS}}$ ), 7.54-7.52 (m, 2H,  $\text{H}_{\text{meta-tBuBz}}$ ), 7.46-7.44 (m, 6H,  $\text{H}_{\text{meta/para-TBDPS}}$ ), 7.32-7.29 (m, 1H,  $\text{H}_5$ ), 6.85 (t, 1H, rotamer 1,  $J = 5.8$  Hz,  $\text{H}_{12'}$ ), 6.77 (t, 1H, rotamer 2,  $J = 5.8$  Hz,  $\text{H}_{12'}$ ), 4.68 (s, 2H,  $\text{H}_8$ ), 4.17 (s, 2H, rotamer 1,  $\text{H}_5$ ), 3.94 (s, 2H, rotamer 2,  $\text{H}_5$ ), 3.86 (t, 2H, rotamer 1,  $J = 5.8$  Hz,  $\text{H}_2$ ), 3.70 (t, 2H, rotamer 2,  $J = 5.8$  Hz,  $\text{H}_2$ ), 3.59 (t, 2H, rotamer 1,  $J = 5.8$  Hz,  $\text{H}_3$ ), 3.49 (t, 2H, rotamer 2,  $J = 5.8$  Hz,  $\text{H}_3$ ), 3.14 (m, 2H, rotamer 1,  $\text{H}_{10'}$ ), 3.05 (m, 2H, rotamer 2,  $\text{H}_{10'}$ ), 3.04 (m, 2H, rotamer 1,  $\text{H}_{11'}$ ), 2.97 (m, 2H, rotamer 2,  $\text{H}_{11'}$ ), 1.37 (s, 9H, tBu-Boc), 1.31 (s, 9H, tBuBz), 1.01 (d, 9H, tBu-TBDPS).  $^{13}\text{C}$  NMR (100 MHz,  $\text{DMSO}-d_6$ )  $\delta$  (ppm) 168.4 ( $\text{C}_7$ , CO), 168.1 ( $\text{C}_6$ , CO), 167.6 ( $\text{C}_2$ , CO), 163.8 ( $\text{C}_4$ , C), 156.1 (C-Boc, CO), 155.6 (C-tBuBz, CO), 155.6 ( $\text{C}_{\text{para-tBuBz}}$ , C), 151.6 and 151.3 ( $\text{C}_6$ , rotamers, CH), 135.5 ( $\text{C}_{\text{ortho-TBDPS}}$ , CH), 133.3 ( $\text{C}_{\text{ph-TBDPS}}$ , C), 130.9 ( $\text{C}_{\text{ph-tBuBz}}$ , C), 130.3 ( $\text{C}_{\text{para-TBDPS}}$ , CH), 128.8 ( $\text{C}_{\text{meta-tBuBz}}$ , CH), 128.4 ( $\text{C}_{\text{meta-TBDPS}}$ , CH), 125.7 ( $\text{C}_{\text{ortho-tBuBz}}$ , CH), 96.2 ( $\text{C}_5$ ), 78.1 (C-



tBu-Boc, C), 62.4 and 61.6 (C<sub>2'</sub>, rotamers, CH<sub>2</sub>), 51.5 and 49.9 (C<sub>5'</sub>, rotamers, CH<sub>2</sub>), 50.2 and 50.0 (C<sub>3'</sub>, rotamers, CH<sub>2</sub>), 39.5 and 29.8 (C<sub>11'</sub>, rotamers, CH<sub>2</sub>), 35.3 (C-tBu-tBuBz, C), 31.3 (C-tBu-tBuBz, CH<sub>3</sub>), 28.3 (C-tBu-Boc, CH<sub>3</sub>), 27.1 (C-tBu-TBDPS, CH<sub>3</sub>), 19.1 (C-tBu-TBDPS, C). HRMS ESI<sup>+</sup> calcd for C<sub>44</sub>H<sub>59</sub>N<sub>6</sub>O<sub>7</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 811.4214, found 811.4208.

**2-[[[[(N<sup>6</sup>-(benzoyl)adenin-9-yl)acetyl]-[2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-(tert-butyloxycarbonyl-amino)ethyl]acetamide (4c).**

A mixture of 2-[[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (50 mg, 0.17 mmol, 1.0 equiv.) and paraformaldehyde (5.1 mg, 0.17 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 334 µL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. (2-Isocyano-ethyl)-carbamic acid *tert*-butyl ester **3a** (28.8 mg, 0.17 mmol, 1.0 equiv.) and N<sup>6</sup>-(benzoyl)-9-(carboxymethyl)adenine **2c** (50.4 mg, 0.17 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4c** as a white powder (52.9 mg, 41% yield). Tr-HPLC: 24.1 min (>98%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>): δ (ppm) 11.04 (bs, 1H, NH-Bz), 8.73 (s, 1H, H<sub>2</sub>), 8.32 (s, 1H, H<sub>8</sub>), 8.26 (t, 1H, rotamer 1, J=5.4 Hz, H<sub>9</sub>), 7.94 (t, 1H, rotamer 2, J=6.0 Hz, H<sub>9</sub>), 8.08 (m, 2H, H<sub>para</sub>-Bz), 7.67-7.64 (m, 1H, H<sub>meta</sub>-Bz), 7.58-7.51 (m, 2H, H<sub>ortho</sub>-Bz), 7.64-7.59 (m, 4H, H<sub>ortho</sub>-TBDPS), 7.46-7.42 (m, 6H, H<sub>meta/para</sub>-TBDPS), 6.85 (t, 1H, rotamer 1, J=5.4 Hz, H<sub>12'</sub>), 6.72 (t, 1H, rotamer 2, J=5.4 Hz, H<sub>12'</sub>), 5.40 (s, 2H, rotamer 1, H<sub>8</sub>), 5.23 (s, 2H, rotamer 2, H<sub>8</sub>), 4.17 (s, 2H, rotamer 1, H<sub>5</sub>), 3.96 (s, 2H, rotamer 2, H<sub>5</sub>), 3.91 (t, 2H, rotamer 1, J= 5.4 Hz, H<sub>2</sub>), 3.70 (t, 2H, rotamer 2, J= 6.1 Hz, H<sub>2</sub>), 3.62 (t, 2H, rotamer 1, J= 5.4 Hz, H<sub>3</sub>), 3.48 (t, 2H, rotamer 2, J= 6.1 Hz, H<sub>3</sub>), 3.13 (m, 2H, rotamer 1, H<sub>10'</sub>), 3.04 (m, 2H, rotamer 2, H<sub>10'</sub>), 3.04 (m, 2H, rotamers 1, H<sub>11'</sub>), 2.92 (m, 2H, rotamer 2, H<sub>11'</sub>), 1.36 (s, 9H, tBu-Boc), 1.03 (s, 9H, tBu-TBDPS). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>): δ (ppm) 167.9 (C<sub>6</sub>, CO), 167.1 and 166.4 (C<sub>7</sub>, CO, rotamers), 166.3 (C-Bz, CO), 155.6 (C-Boc, CO), 153.3 (C<sub>4</sub>, C), 151.9 (C<sub>2</sub>, C), 150.6 (C<sub>6</sub>, C), 146.1 (C<sub>8</sub>, CH), 135.0 (C<sub>ortho</sub>-TBDPS, CH), 134.2 (C<sub>ph</sub>-Bz, C), 132.9 (C<sub>meta</sub>-Bz, CH), 132.7 (C<sub>ph</sub>-TBDPS), 129.8 (C<sub>para</sub>-TBDPS, CH), 129.0 (C<sub>ortho/para</sub>-Bz, CH), 128.0 (C<sub>meta</sub>-TBDPS, CH), 125.0 (C<sub>5</sub>, C), 77.6 (C-tBu-Boc, C), 61.2 and 60.9 (C<sub>2'</sub>, CH<sub>2</sub>, rotamers), 50.3 and 47.9 (C<sub>5'</sub>, CH<sub>2</sub>, rotamers), 49.3 and 49.2 (C<sub>3'</sub>, CH<sub>2</sub>, rotamers), 46.7 and 43.2 (C<sub>8'</sub>, CH<sub>2</sub>, rotamers), 40.0 (C<sub>11'</sub>, CH<sub>2</sub>), 39.4 (C<sub>10'</sub>, CH<sub>2</sub>), 28.3 (C-tBu-Boc, CH<sub>3</sub>), 26.6 (C-tBu-TBDPS, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>41</sub>H<sub>51</sub>N<sub>8</sub>O<sub>6</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 779.3701, found 779.3708.

**2-[[[[(Thymin-1-yl)acetyl]-[2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-(tert-butyloxycarbonyl-amino)ethyl]acetamide (4d).**

A mixture of 2-[[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (150 mg, 0.50 mmol, 1.0 equiv.) and paraformaldehyde (15.1 mg, 0.50 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.0 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. (2-Isocyano-ethyl)-carbamic Acid *tert*-butyl Ester **3a** (85.3 mg, 0.50 mmol, 1.0 equiv.) and thymin-1-acetic acid **2d** (92.5 mg, 0.50 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was

irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4d** as a white powder (217.6 mg, 65% yield). Tr-HPLC: 23.1 min (>98%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>): δ (ppm) 11.31 (1H, s, H<sub>3</sub>), 8.19 (t, 1H, rotamer 1, J= 5.8 Hz, H<sub>9</sub>), 7.88 (t, 1H, rotamer 2, J= 5.8 Hz, H<sub>9</sub>), 7.67-7.62 (m, 4H, H<sub>ortho</sub>-TBDPS), 7.53 – 7.44 (m, 6H, H<sub>meta/para</sub>-TBDPS), 7.27 (s, 1H, rotamer 1, H<sub>6</sub>), 7.08 (s, 1H, rotamer 2, H<sub>6</sub>), 6.87 (t, 1H, rotamer 1, J= 5.8 Hz, H<sub>12'</sub>), 6.81 (t, 1H, rotamer 2, J= 5.8 Hz, H<sub>12'</sub>), 4.58 (s, 2H, rotamer 1, H<sub>8</sub>), 4.51 (s, 2H, rotamer 2, H<sub>8</sub>), 4.12 (s, 2H, rotamer 1, H<sub>5</sub>), 3.93 (s, 2H, rotamer 2, H<sub>5</sub>), 3.83 (t, 2H, rotamer 1, J= 5.8 Hz, H<sub>2</sub>), 3.70 (t, 2H, rotamer 2, J= 5.8 Hz, H<sub>2</sub>), 3.57 (t, 2H, rotamer 1, J= 5.8 Hz, H<sub>3</sub>), 3.48 (t, 2H, rotamer 2, J= 5.8 Hz, H<sub>3</sub>), 3.13 (m, 2H, rotamer 1, H<sub>10'</sub>), 3.06 (m, 2H, rotamer 2, H<sub>10'</sub>), 3.06 (m, 2H, rotamer 1, H<sub>11'</sub>), 2.98 (m, 2H, rotamer 2, H<sub>11'</sub>), 1.75 (d, 3H, H<sub>7</sub>), 1.39 (s, 9H, tBu-Boc), 1.02 (d, 9H, tBu-TBDPS). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>): δ (ppm) 168.7 (C<sub>7</sub>, CO), 168.1 (C<sub>6</sub>, CO), 165.2 (C<sub>4</sub>, CO), 156.5 (C-Boc, CO), 151.8 (C<sub>2</sub>, CO), 142.8 (C<sub>6</sub>, CH), 135.9 (C<sub>ortho</sub>-TBDPS, CH), 133.6 (C<sub>ph</sub>-TBDPS, C), 130.8 (C<sub>para</sub>-TBDPS, CH), 128.8 (C<sub>meta</sub>-TBDPS, CH), 108.9 (C<sub>5</sub>, C), 78.5 (C-tBu-Boc, C), 62.6 and 61.9 (C<sub>2'</sub>, CH<sub>2</sub>, rotamers), 51.6 and 50.3 (C<sub>5'</sub>, CH<sub>2</sub>, rotamers), 49.8 and 48.7 (C<sub>3'</sub>, CH<sub>2</sub>, rotamers), 48.2 (C<sub>8'</sub>, CH<sub>2</sub>), 40.0 and 39.3 (C<sub>11'</sub>, rotamers, CH<sub>2</sub>), 39.5 and 39.3 (C<sub>10'</sub>, CH<sub>2</sub>, rotamers), 29.1 (C-tBu-Boc, CH<sub>3</sub>), 27.5 (C-tBu-TBDPS, CH<sub>3</sub>), 19.5 (C-tBu-TBDPS, C), 12.7 (C<sub>7</sub>, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>34</sub>H<sub>46</sub>N<sub>5</sub>O<sub>7</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 664.3167, found 664.3166.

**2-[[[[(Uracil-1-yl)acetyl]-[2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-(tert-butyloxycarbonyl-amino)ethyl]acetamide (4e).**

A mixture of 2-[[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (100 mg, 0.33 mmol, 1.0 equiv.) and paraformaldehyde (10 mg, 0.33 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 669 µL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. (2-Isocyano-ethyl)-carbamic acid *tert*-butyl ester **3a** (56.8 mg, 0.33 mmol, 1.0 equiv.) and uracil-1-acetic acid **2e** (56.9 mg, 0.33 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4e** as a white powder (159.4 mg, 73% yield). Tr-HPLC: 22.6 min (>98%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>): δ (ppm) 11.27 (1H, s, H<sub>3</sub>), 8.16 (t, 1H, rotamer 1, J= 5.8 Hz, H<sub>9</sub>), 7.84 (t, 1H, rotamer 2, J= 5.8 Hz, H<sub>9</sub>), 7.67 – 7.57 (m, 4H, H<sub>ortho</sub>-TBDPS), 7.51 – 7.41 (m, 6H, H<sub>meta/para</sub>-TBDPS), 7.40 (d, 1H, rotamer 1, H<sub>6</sub>), 7.25 (d, 1H, rotamer 2, H<sub>6</sub>), 6.83 (t, 1H, rotamer 1, J= 5.8 Hz, H<sub>12'</sub>), 6.77 (t, 1H, rotamer 2, J= 5.8 Hz, H<sub>12'</sub>), 5.55 (dd, 1H, H<sub>5</sub>), 4.59 (s, 2H, rotamer 1, H<sub>8</sub>), 4.53 (s, 2H, rotamer 2, H<sub>8</sub>), 4.10 (s, 2H, rotamer 1, H<sub>5</sub>), 3.92 (s, 2H, rotamer 2, H<sub>5</sub>), 3.81 (t, 2H, rotamer 1, J= 5.8 Hz, H<sub>2</sub>), 3.69 (t, 2H, rotamer 2, J= 5.8 Hz, H<sub>2</sub>), 3.54 (t, 2H, rotamer 1, J= 5.8 Hz, H<sub>3</sub>), 3.47 (t, 2H, rotamer 2, J= 5.8 Hz, H<sub>3</sub>), 3.11 (m, 2H, rotamer 1, H<sub>10'</sub>), 3.05 (m, 2H, rotamer 2, H<sub>10'</sub>), 3.03 (m, 2H, rotamer 1, H<sub>11'</sub>), 2.97 (m, 2H, rotamer 2, H<sub>11'</sub>), 1.37 (s, 9H, tBu-Boc), 0.99 (d, 9H, tBu-TBDPS). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>): δ (ppm) 168.2 (C<sub>6</sub>, CO), 167.6 (C<sub>7</sub>, CO), 164.2 (C<sub>4</sub>, CO), 156.1 (C-Boc, CO), 151.4 (C<sub>2</sub>, CO), 146.9 (C<sub>6</sub>, CH), 135.5 (C<sub>ortho</sub>-TBDPS, CH), 133.2 (C<sub>ph</sub>-TBDPS, C),



130.4 ( $C_{\text{para}}\text{-TBDPS}$ , CH), 128.4 ( $C_{\text{meta}}\text{-TBDPS}$ , CH), 101.1 ( $C_5$ , CH), 78.1 (CtBu-Boc, C), 62.6 and 61.2 ( $C_2$ ,  $\text{CH}_2$ , rotamer), 49.9 and 49.7 ( $C_3$ ,  $\text{CH}_2$ , rotamer), 49.4 ( $C_5$ ,  $\text{CH}_2$ ), 48.4 ( $C_8$ ,  $\text{CH}_2$ ), 39.7 ( $C_{11}$ ,  $\text{CH}_2$ ), 39.4 ( $C_{10}$ ,  $\text{CH}_2$ ), 28.7 (C-tBu-Boc,  $\text{CH}_3$ ), 27.1 (C-tBu-TBDPS,  $\text{CH}_3$ ), 19.1 (C-tBu-TBDPS, C). HRMS  $\text{ESI}^-$  calcd for  $\text{C}_{33}\text{H}_{44}\text{N}_5\text{O}_7\text{Si}^-$  (M-H) $^-$  650.3010, found 650.3015.

**2-((((N<sup>4</sup>-(4-tert-butylbenzoyl)cytosin-1-yl)acetyl)-(2-tert-butyl(diphenylsilyloxy)ethyl))amino-N-(methyleacetate)acetamide (4f).**

A mixture of 2-[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (200 mg, 0.67 mmol, 1.0 equiv.) and paraformaldehyde (20.1 mg, 0.67 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.3 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocynoacetate **3b** (60.8  $\mu\text{L}$ , 0.67 mmol, 10 equiv.) and N<sup>4</sup>-(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (220.2 mg, 0.67 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4f** as a white powder (251.6 mg, 51% yield). Tr-HPLC: 25.9 min (95 %, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 11.12 (1H, s, NH-tBuBz), 8.67 (d, 1H, rotamer 1,  $J$  = 5.9 Hz,  $H_9$ ), 8.32 (d, 1H, rotamer 2,  $J$  = 5.9 Hz,  $H_9$ ), 7.98 (d, 1H,  $J$  = 8.5 Hz,  $H_{\text{ortho-tBuBz}}$ ), 7.94 (d, 1H, rotamer 1,  $J$  = 7.4 Hz,  $H_6$ ), 7.83 (d, 1H, rotamer 2,  $J$  = 7.4 Hz,  $H_6$ ), 7.64-7.59 (m, 4H,  $H_{\text{ortho-TBDPS}}$ ), 7.54 (d, 1H,  $J$  = 8.5 Hz,  $H_{\text{meta-tBuBz}}$ ), 7.46-7.42 (m, 6H,  $H_{\text{meta/para-TBDPS}}$ ), 7.32 (d, 1H,  $J$  = 7.4 Hz,  $H_5$ ), 4.71 (s, 2H,  $H_8$ ), 4.29 (s, 2H, rotamer 1,  $H_5$ ), 4.07 (s, 2H, rotamer 2,  $H_5$ ), 3.92 (d, 2H, rotamer 1,  $J$  = 5.9 Hz,  $H_{10}$ ), 3.86 (d, 2H, rotamer 2,  $J$  = 5.9 Hz,  $H_{10}$ ), 3.90 (t, 2H, rotamer 1,  $J$  = 5.9 Hz,  $H_2$ ), 3.73 (t, 2H, rotamer 2,  $J$  = 5.9 Hz,  $H_2$ ), 3.64 (s, 3H,  $\text{OCH}_3$ ), 3.60 (t, 2H, rotamer 1,  $J$  = 5.9 Hz,  $H_3$ ), 3.51 (t, 2H, rotamer 2,  $J$  = 5.9 Hz,  $H_3$ ), 1.3 (s, 9H, tBuBz), 1.03 (s, 9H, tBu-TBDPS).  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 170.1 ( $C_{11}$ , CO), 168.6 ( $C_6$ , CO), 167.3 ( $C_7$ , CO), 167.0 (C-tBuBz, CO), 163.3 ( $C_4$ , C), 155.7 ( $C_{\text{para-tBuBz}}$ , C), 151.0 and 150.8 ( $C_6$ , CH, rotamers), 135.0 ( $C_{\text{ortho-TBDPS}}$ , CH), 132.7 ( $C_{\text{ph-TBDPS}}$ , C), 130.4 ( $C_{\text{ph-tBuBz}}$ , C), 129.9 ( $C_{\text{para-TBDPS}}$ , CH), 128.3 ( $C_{\text{ortho-tBuBz}}$ , CH), 128.0 ( $C_{\text{meta-TBDPS}}$ , CH), 125.2 ( $C_{\text{meta-tBuBz}}$ , CH), 95.7 ( $C_5$ , CH), 61.9 and 61.1 ( $C_2$ ,  $\text{CH}_2$ , rotamers), 51.8 ( $\text{OCH}_3$ ,  $\text{CH}_3$ ), 49.7 ( $C_8$ ,  $\text{CH}_2$ ), 49.4 ( $C_3$ ,  $\text{CH}_2$ ), 50.8 and 48.9 ( $C_5$ ,  $\text{CH}_2$ , rotamers), 40.7 and 40.4 ( $C_{10}$ ,  $\text{CH}_2$ , rotamers), 34.8 (C-tBu-tBuBz, C), 30.8 (C-tBu-tBuBz,  $\text{CH}_3$ ), 26.6 (C-tBu-TBDPS,  $\text{CH}_3$ ), 18.7 (C-tBu-TBDPS, C). HRMS  $\text{ESI}^+$  calcd for  $\text{C}_{40}\text{H}_{50}\text{N}_5\text{O}_7\text{Si}^+$  (M+H) $^+$  740.3480, found 740.3482.

**2-((((Thymin-1-yl)acetyl)-(2-tert-butyl(diphenylsilyloxy)ethyl))amino-N-(methyleacetate)acetamide (4g).**

A mixture of 2-[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (100 mg, 0.33 mmol, 1.0 equiv.) and paraformaldehyde (10 mg, 0.33 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 668  $\mu\text{L}$ ) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocynoacetate **3b** (30.4  $\mu\text{L}$ , 0.33 mmol, 1.0 equiv.) and thymin-1-acetic acid **2d** (61.7 mg, 0.33 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60

°C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4g** as a brown powder (123.5 mg, 62% yield). Tr-HPLC: 21.9 min (95%, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 11.28 (s, 1H, rotamer 1,  $H_3$ ), 11.26 (s, 1H, rotamer 2,  $H_3$ ), 8.62 (d, 1H, rotamer 1,  $J$  = 5.8 Hz,  $H_9$ ), 8.32 (d, 1H, rotamer 2,  $J$  = 5.8 Hz,  $H_9$ ), 7.64-7.60 (m, 4H,  $H_{\text{ortho-TBDPS}}$ ), 7.45-7.40 (m, 6H,  $H_{\text{meta/para-TBDPS}}$ ), 7.22 (s, 1H, rotamer 1,  $H_6$ ), 7.07 (s, 1H, rotamer 2,  $H_6$ ), 4.56 (s, 2H, rotamer 1,  $H_8$ ), 4.51 (s, 2H, rotamer 2,  $H_8$ ), 4.21 (s, 2H, rotamer 1,  $H_5$ ), 4.02 (s, 2H, rotamer 2,  $H_5$ ), 3.90 (d, 2H, rotamer 1,  $J$  = 5.8 Hz,  $H_{10}$ ), 3.85 (d, 2H, rotamer 2,  $J$  = 5.8 Hz,  $H_{10}$ ), 3.84 (t, 2H, rotamer 1,  $J$  = 5.4 Hz,  $H_2$ ), 3.71 (t, 2H, rotamer 2,  $J$  = 5.9 Hz,  $H_2$ ), 3.63 (s, 3H,  $\text{OCH}_3$ ), 3.54 (t, 2H, rotamer 1,  $J$  = 5.4 Hz,  $H_3$ ), 3.48 (t, 2H, rotamer 2,  $J$  = 5.9 Hz,  $H_3$ ), 1.74 (d, 3H, rotamer 1,  $J$  = 1.0 Hz,  $H_7$ ), 1.70 (d, 3H, rotamer 2,  $J$  = 1.0 Hz,  $H_7$ ), 1.01-0.98 (s, 9H, tBu-TBDPS).  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 170.2 ( $C_{11}$ , CO), 168.7 ( $C_6$ , CO), 167.7 ( $C_7$ , CO), 164.4 ( $C_4$ , CO), 151.0 ( $C_2$ , CO), 142.0 ( $C_6$ , CH), 135.1 ( $C_{\text{ortho-TBDPS}}$ , CH), 132.9 ( $C_{\text{ph-TBDPS}}$ , C), 129.9 ( $C_{\text{para-TBDPS}}$ , CH), 127.9 ( $C_{\text{meta-TBDPS}}$ , CH), 108.1 ( $C_5$ , C), 61.7 and 61.1 ( $C_2$ ,  $\text{CH}_2$ , rotamers), 51.8 ( $\text{OCH}_3$ ,  $\text{CH}_3$ ), 50.5 and 48.6 ( $C_5$ ,  $\text{CH}_2$ , rotamers), 49.4 and 49.2 ( $C_3$ ,  $\text{CH}_2$ , rotamers), 47.7 ( $C_8$ ,  $\text{CH}_2$ ), 40.7 and 40.4 ( $C_{10}$ ,  $\text{CH}_2$ , rotamers), 26.7 and 26.6 (C-tBu-TBDPS,  $\text{CH}_3$ ), 18.7 (C-tBu-TBDPS, C), 11.9 ( $C_7$ ,  $\text{CH}_3$ ). HRMS  $\text{ESI}^+$  calcd for  $\text{C}_{30}\text{H}_{39}\text{N}_4\text{O}_7\text{Si}^+$  (M+H) $^+$  595.2588, found 595.2593.

**2-((((Uracil-1-yl)acetyl)-(2-tert-butyl(diphenylsilyloxy)ethyl))amino-N-(methyleacetate)acetamide (4h).**

A mixture of 2-[[tert-butyl(diphenyl)silyl]oxy]ethanamine **1b** (100 mg, 0.33 mmol, 1.0 equiv.) and paraformaldehyde (10 mg, 0.33 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 668  $\mu\text{L}$ ) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocynoacetate **3b** (30.4  $\mu\text{L}$ , 0.33 mmol, 1.0 equiv.) and uracil-1-acetic acid **2e** (56.9 mg, 0.33 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4h** as a brown powder (105.5 mg, 54% yield). Tr-HPLC: 21.2 min (95%, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 11.02 (s, 1H,  $H_3$ ), 8.64 (t, 1H, rotamer 1,  $J$  = 5.9 Hz,  $H_9$ ), 8.33 (t, 1H, rotamer 2,  $J$  = 5.9 Hz,  $H_9$ ), 7.63-7.59 (m, 4H,  $H_{\text{ortho-TBDPS}}$ ), 7.50-7.38 (m, 6H,  $H_{\text{meta/para-TBDPS}}$ ), 7.39 (d, 1H, rotamer 1,  $J$  = 1.0 Hz,  $H_6$ ), 7.27 (d, 1H, rotamer 2,  $J$  = 1.0 Hz,  $H_6$ ), 5.57 (d, 1H, rotamer 1,  $J$  = 1.0 Hz,  $H_5$ ), 5.51 (d, 1H, rotamer 2,  $J$  = 1.0 Hz,  $H_5$ ), 4.57 (s, 2H, rotamer 1,  $H_8$ ), 4.54 (s, 2H, rotamer 2,  $H_8$ ), 4.20 (s, 2H, rotamer 1,  $H_5$ ), 4.01 (s, 2H, rotamer 2,  $H_5$ ), 3.89 (d, 2H, rotamer 1,  $J$  = 5.9 Hz,  $H_{10}$ ), 3.83 (d, 2H, rotamer 2,  $J$  = 5.9 Hz,  $H_{10}$ ), 3.82 (t, 2H, rotamer 1,  $J$  = 5.8 Hz,  $H_2$ ), 3.69 (t, 2H, rotamer 2,  $J$  = 5.8 Hz,  $H_2$ ), 3.62 (s, 3H,  $\text{OCH}_3$ ), 3.52 (t, 2H, rotamer 1,  $J$  = 5.8 Hz,  $H_3$ ), 3.47 (t, 2H, rotamer 2,  $J$  = 5.8 Hz,  $H_3$ ), 1.00 (d, 9H, tBu-TBDPS).  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 171.0 ( $C_{11}$ , CO), 169.4 ( $C_6$ , CO), 168.5 ( $C_7$ , CO), 164.6 ( $C_4$ , CO), 151.8 ( $C_2$ , CO), 147.1 ( $C_6$ , CH), 135.9 ( $C_{\text{ortho-TBDPS}}$ , CH), 133.6 ( $C_{\text{ph-TBDPS}}$ , C), 130.8 ( $C_{\text{para-TBDPS}}$ , CH), 128.8 ( $C_{\text{meta-TBDPS}}$ , CH), 101.5 ( $C_5$ , CH), 62.6 and 61.9

(C<sub>2</sub>, CH<sub>2</sub>, rotamers), 52.6 (OCH<sub>3</sub>, CH<sub>3</sub>), 51.4 and 50.2 (C<sub>5</sub>, CH<sub>2</sub>, rotamers), 49.5 and 48.7 (C<sub>3</sub>, CH<sub>2</sub>, rotamers), 48.7 and 48.6 (C<sub>8</sub>, CH<sub>2</sub>, rotamers) 41.6 and 41.3 (C<sub>10</sub>, CH<sub>2</sub>, rotamers), 27.5 (C-tBu-TBDPS, CH<sub>3</sub>), 19.5 (C-tBu-TBDPS, C). HRMS ESI<sup>+</sup> calcd for C<sub>29</sub>H<sub>37</sub>N<sub>4</sub>O<sub>7</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 581.2432, found 581.2438.

**2-((((N<sup>4</sup>-(4-tert-butylbenzoyl)cytosin-1-yl)acetyl)-(2-(tert-butylloxycarbonyl-amino)ethyl))amino-N-(methylacetate)acetamide (4i).**

A mixture of *tert*-butyl N-(2-aminoethyl)carbamate **1a** (50 mg, 0.31 mmol, 1.0 equiv.) and paraformaldehyde (9.4 mg, 0.31 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 625 µL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocynoacetate **3b** (28.4 µL, 0.31 mmol, 1.0 equiv.) and N<sup>4</sup>-(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (113.0 mg, 0.31 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4i** as a brown powder (69.3 mg, 37% yield). Tr-HPLC: 20.2 min (> 95%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>): δ (ppm) 11.1 (s, 1H, NH-tBuBz), 8.68 (d, 1H, rotamer 1, J = 5.8 Hz, H<sub>9</sub>), 8.35 (d, 1H, rotamer 2, J = 5.8 Hz, H<sub>9</sub>), 7.98 (d, 2H, J = 8.8 Hz, H<sub>ortho</sub>-tBuBz), 7.95 (m, 1H, H<sub>6</sub>), 7.53 (d, 2H, J = 8.8 Hz, H<sub>meta</sub>-tBuBz), 7.33 (m, 1H, H<sub>5</sub>), 6.96 (t, 1H, rotamer 1, J = 5.8 Hz, H<sub>1</sub>), 6.72 (t, 1H, rotamer 2, J = 5.8 Hz, H<sub>1</sub>), 4.85 (s, 2H, rotamer 1, H<sub>8</sub>), 4.68 (s, 2H, rotamer 2, H<sub>8</sub>), 4.19 (s, 2H, rotamer 1, H<sub>5</sub>), 4.00 (s, 2H, rotamer 2, H<sub>5</sub>), 3.94 (d, 1H, rotamer 1, J = 5.8 Hz, H<sub>10</sub>), 3.86 (d, 1H, rotamer 2, J = 5.8 Hz, H<sub>10</sub>), 3.65 (s, 3H, OCH<sub>3</sub>), 3.43 (t, 2H, rotamer 1, J = 6.8 Hz, H<sub>3</sub>), 3.34 (t, 2H, rotamer 2, J = 6.8 Hz, H<sub>3</sub>), 3.22 (dt, 2H, rotamer 1, J = 5.8-6.8 Hz, H<sub>2</sub>), 3.05 (dt, 2H, rotamer 2, J = 5.8-6.8 Hz, H<sub>2</sub>), 1.39 (s, 9H, tBu-Boc), 1.32 (s, 9H, tBuBz). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>): δ (ppm) 170.0 (C<sub>11</sub>, CO), 168.7 (C<sub>6</sub>, CO), 167.6 (C<sub>7</sub>, CO), 167.0 (C-tBuBz, CO), 163.3 (C<sub>4</sub>, C), 155.7 (C-Boc, CO), 155.7 (C<sub>para</sub>-tBuBz, C), 155.2 (C<sub>2</sub>, CO), 151.0 (C<sub>6</sub>, CH), 130.3 (C<sub>ph</sub>-tBuBz, C), 128.3 (C<sub>ortho</sub>-tBuBz, CH), 125.2 (C<sub>meta</sub>-tBuBz, CH), 95.7 (C<sub>5</sub>, CH), 77.8 (C-tBu-Boc, C), 50.2 and 48.6 (C<sub>5</sub>, CH<sub>2</sub>, rotamers), 51.6 (OCH<sub>3</sub>, CH<sub>3</sub>), 49.6 (C<sub>8</sub>, CH<sub>2</sub>), 47.2 (C<sub>3</sub>, CH<sub>2</sub>), 40.8 and 40.5 (C<sub>10</sub>, CH<sub>2</sub>, rotamers), 40.0 and 38.0 (C<sub>2</sub>, CH<sub>2</sub>, rotamers), 34.8 (C-tBu-tBuBz, C), 30.8 (C-tBu-tBuBz, CH<sub>3</sub>), 28.2 (C-tBu-Boc, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>29</sub>H<sub>41</sub>N<sub>6</sub>O<sub>8</sub><sup>+</sup> (M+H)<sup>+</sup> 601.2986, found 601.2987.

**2-((((N<sup>6</sup>-(benzoyl)adenin-9-yl)acetyl)-(2-(tert-butylloxycarbonyl-amino)ethyl))amino-N-(methylacetate)acetamide (4j).**

A mixture of *tert*-butyl N-(2-aminoethyl)carbamate **1a** (100 mg, 0.63 mmol, 1.0 equiv.) and paraformaldehyde (18.7 mg, 0.63 mmol, 1.0 equiv.), in anhydrous isopropanol (0.5 M, 1.25 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocynoacetate **3b** (56.8 µL, 0.63 mmol, 1.0 equiv.) and N<sup>6</sup>-(benzoyl)-9-(carboxymethyl)adenine **2c** (185.5 mg, 0.63 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4j** as a brown powder (141.9 mg, 40% yield). Tr-HPLC: 15.9 min (> 88%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>):

δ (ppm) 11.19 (s, 1H, NH-Bz), 8.72 (s, 1H, H<sub>2</sub>), 8.77 (t, 1H, rotamer 1, H<sub>9</sub>), 8.42 (t, 1H, rotamer 2, H<sub>9</sub>), 8.33 (s, 1H, H<sub>8</sub>), 8.08 (m, 2H, H<sub>para</sub>-Bz), 7.67-7.64 (m, 1H, H<sub>meta</sub>-Bz), 7.58-7.51 (m, 2H, H<sub>ortho</sub>-Bz), 6.84 (t, 1H, rotamer 1, H<sub>1</sub>), 6.75 (t, 1H, rotamer 2, H<sub>1</sub>), 5.37 (s, 2H, rotamer 1, H<sub>8</sub>), 5.22 (s, 2H, rotamer 2, H<sub>8</sub>), 4.29 (s, 2H, rotamer 1, H<sub>5</sub>), 4.02 (s, 2H, rotamer 2, H<sub>5</sub>), 3.98 (t, 2H, rotamer 1, J = 5.8 Hz, H<sub>10</sub>), 3.87 (t, 2H, rotamer 2, J = 5.8 Hz, H<sub>10</sub>), 3.64 (s, 3H, OCH<sub>3</sub>), 3.54 (m, 2H, rotamer 1, H<sub>3</sub>), 3.34 (m, 2H, rotamer 2, H<sub>3</sub>), 3.15 (m, 2H, rotamer 1, H<sub>2</sub>), 3.07 (m, 2H, rotamer 2, H<sub>2</sub>), 1.38 (s, 9H, tBu-Boc). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>): δ (ppm) 170.7 (C<sub>11</sub>, CO), 169.1 (C<sub>5</sub>, CO), 166.7 (C<sub>7</sub>, CO), 166.3 (C-Bz, CO), 156.2 (C-Boc, CO), 153.3 (C<sub>4</sub>, C), 151.9 (C<sub>2</sub>, CH), 150.6 (C<sub>6</sub>, C), 146.0 (C<sub>8</sub>, CH), 134.2 (C<sub>ph</sub>-Bz, C), 132.9 (C<sub>meta</sub>-Bz, CH), 129.0 (C<sub>para</sub>-Bz, CH), 129.0 (C<sub>ortho</sub>-Bz, CH), 125.7 (C<sub>5</sub>, C), 77.8 (C-tBu-Boc, C), 52.2 (OCH<sub>3</sub>, CH<sub>3</sub>), 50.5 and 48.9 (C<sub>5</sub>, CH<sub>2</sub>, rotamers), 47.7 (C<sub>3</sub>, CH<sub>2</sub>), 44.4 and 44.2 (C<sub>8</sub>, CH<sub>2</sub>, rotamers), 41.3 and 40.9 (C<sub>10</sub>, CH<sub>2</sub>, rotamers), 38.4 and 37.9 (C<sub>2</sub>, CH<sub>2</sub>, rotamers), 28.8 (C-tBu-Boc, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>26</sub>H<sub>31</sub>N<sub>8</sub>O<sub>7</sub><sup>+</sup> (M-H)<sup>+</sup> 567.2316, found 567.2316.

**2-((((Thymin-1-yl)acetyl)-(2-(tert-butylloxycarbonyl-amino)ethyl))amino-N-(methylacetate)acetamide (4k).**

A mixture of *tert*-butyl N-(2-aminoethyl)carbamate **1a** (100 mg, 0.63 mmol, 1.0 equiv.) and paraformaldehyde (18.7 mg, 0.63 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.25 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocynoacetate **3b** (56.8 µL, 0.63 mmol, 1.0 equiv.) and thymin-1-acetic acid **2d** (115.2 mg, 0.63 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4k** as a brown powder (128.9 mg, 45% yield). Tr-HPLC: 14.3 min (97%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>): δ (ppm) 11.28 (s, 1H, H<sub>3</sub>), 8.69 (t, 1H, rotamer 1, H<sub>9</sub>), 8.40 (t, 1H, rotamer 2, H<sub>9</sub>), 7.32 (s, 1H, H<sub>6</sub>), 7.00 (t, 1H, rotamer 1, H<sub>1</sub>), 6.77 (t, 1H, rotamer 2, H<sub>1</sub>), 4.66 (s, 2H, rotamer 1, H<sub>8</sub>), 4.50 (s, 2H, rotamer 2, H<sub>8</sub>), 4.14 (s, 2H, rotamer 1, H<sub>5</sub>), 3.99 (s, 2H, rotamer 2, H<sub>5</sub>), 3.95 (d, 2H, rotamer 1, J = 5.8 Hz, H<sub>10</sub>), 3.88 (d, 2H, rotamer 2, J = 5.8 Hz, H<sub>10</sub>), 3.66 (s, 3H, OCH<sub>3</sub>), 3.39 (m, 2H, rotamer 1, H<sub>3</sub>), 3.34 (m, 2H, rotamer 2, H<sub>3</sub>), 3.19 (m, 2H, rotamer 1, H<sub>2</sub>), 3.08 (m, 2H, rotamer 2, H<sub>2</sub>), 1.77 (s, 3H, H<sub>7</sub>), 1.40 (s, 9H, tBu-Boc). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>): δ (ppm) 170.6 (C<sub>11</sub>, CO), 169.2 (C<sub>8</sub>, CO), 168.2 (C<sub>7</sub>, CO), 164.8 (C<sub>4</sub>, CO), 156.1 (C-Boc, CO), 151.5 (C<sub>2</sub>, CO), 142.5 (C<sub>6</sub>, CH), 108.6 (C<sub>5</sub>, C), 78.4 (C-tBu-Boc, C), 52.2 (OCH<sub>3</sub>, CH<sub>3</sub>), 50.4 and 48.8 (C<sub>5</sub>, CH<sub>2</sub>, rotamers), 48.2 and 48.0 (C<sub>8</sub>, CH<sub>2</sub>, rotamers), 47.6 and 47.4 (C<sub>3</sub>, CH<sub>2</sub>, rotamers), 41.2 and 40.9 (C<sub>10</sub>, CH<sub>2</sub>, rotamers), 38.3 and 37.9 (C<sub>2</sub>, CH<sub>2</sub>, rotamers), 28.7 (C-tBu-Boc, CH<sub>3</sub>), 12.4 (C<sub>7</sub>, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>19</sub>H<sub>28</sub>N<sub>5</sub>O<sub>8</sub> (M-H)<sup>+</sup> 454.1938, found 454.1943.

**2-((((Uracil-1-yl)acetyl)-(2-(tert-butylloxycarbonyl-amino)ethyl))amino-N-(methylacetate)acetamide (4l).**

A mixture of *tert*-butyl N-(2-aminoethyl)carbamate **1a** (100 mg, 0.63 mmol, 1.0 equiv.) and paraformaldehyde (18.7 mg, 0.63 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.25 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl

isocyanoacetate **3b** (56.8  $\mu$ l, 0.63 mmol, 1.0 equiv.) and uracil-1-acetic acid **2e** (106.2 mg, 0.63 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4l** as a brown powder (111.3 mg, 41% yield). Tr-HPLC: 13.5 min (> 99%, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 11.30 (s, 1H, H<sub>3</sub>), 8.71 (t, 1H, rotamer 1, H<sub>9</sub>), 8.42 (t, 1H, rotamer 2, H<sub>9</sub>), 7.46 (s, 1H, rotamer 1, H<sub>6</sub>), 7.40 (s, 1H, rotamer 2, H<sub>6</sub>), 6.99 (t, 1H, rotamer 1, H<sub>1</sub>), 6.77 (t, 1H, rotamer 2, H<sub>1</sub>), 5.62-5.59 (m, 1H, H<sub>5</sub>), 4.71 (s, 2H, rotamer 1, H<sub>8</sub>), 4.54 (s, 2H, rotamer 2, H<sub>8</sub>), 4.15 (s, 2H, rotamer 1, H<sub>5</sub>), 3.99 (s, 2H, rotamer 2, H<sub>5</sub>), 3.95 (d, 2H, rotamer 1,  $J$  = 5.8 Hz, H<sub>10</sub>), 3.88 (d, 2H, rotamer 2,  $J$  = 5.8 Hz, H<sub>10</sub>), 3.66 (s, 3H, OCH<sub>3</sub>), 3.31 (m, 2H, rotamer 1, H<sub>3</sub>), 3.19 (m, 2H, rotamer 2, H<sub>3</sub>), 3.19 (m, 2H, rotamer 1, H<sub>2</sub>), 3.01 (m, 2H, rotamer 2, H<sub>2</sub>), 1.40 (s, 9H, tBu-Boc).  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 170.1 (C<sub>11</sub>, CO), 168.7 (C<sub>6</sub>, CO), 167.6 (C<sub>7</sub>, CO), 163.8 (C<sub>4</sub>, CO), 155.7 (C-Boc, CO), 151.0 (C<sub>2</sub>, CO), 146.3 (C<sub>6</sub>, CH), 100.6 (C<sub>5</sub>, CH), 78.0 (C-tBu-Boc, C), 51.7 (OCH<sub>3</sub>, CH<sub>3</sub>), 50.1 and 48.3 (C<sub>5</sub>, CH<sub>2</sub>, rotamers), 47.9 and 47.8 (C<sub>8</sub>, CH<sub>2</sub>, rotamers), 47.7 and 47.1 (C<sub>3</sub>, CH<sub>2</sub>, rotamers), 40.7 and 40.4 (C<sub>10</sub>, CH<sub>2</sub>, rotamers), 37.8 and 37.5 (C<sub>2</sub>, CH<sub>2</sub>, rotamers), 28.2 (C-tBu-Boc, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>18</sub>H<sub>26</sub>N<sub>5</sub>O<sub>8</sub><sup>+</sup> (M-H)<sup>+</sup> 440.1781, found 440.1787.

**2-[[[ $(\text{N}^4$ -(4-tert-butylbenzoyl)cytosin-1-yl)acetyl]-(2-(4-monomethoxytrityl)oxy)ethyl]]amino-N-(methylacetate)acetamide (4m).**

A mixture of 2-{[4-monomethoxytrityl]oxy}ethanamine **1d** (56.1 mg, 0.17 mmol, 1.0 equiv.) and paraformaldehyde (5.1 mg, 0.17 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 600  $\mu$ l) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocyanoacetate **3b** (15.3  $\mu$ l, 0.17 mmol, 1.0 equiv.) and N<sup>4</sup>-(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (55.6 mg, 0.17 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4m** as a brown powder (53.3 mg, 41% yield). Tr-HPLC: 24.9 min (99%, 10-100% ACN in 40 min).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 11.07 (s, 1H, NH-tBuBz), 8.65 (d, 1H, rotamer 1,  $J$  = 5.8 Hz, H<sub>9</sub>), 8.26 (d, 1H, rotamer 2,  $J$  = 5.8 Hz, H<sub>9</sub>), 7.94 (s, 1H, H<sub>6</sub>), 7.93 (d, 2H,  $J$  = 8.8 Hz, H<sub>ortho</sub>-tBuBz), 7.50 (d, 2H,  $J$  = 8.8 Hz, H<sub>meta</sub>-tBuBz), 7.33 (s, 1H, H<sub>5</sub>), 7.42-76.81 (m, 14H, H<sub>arom</sub>-Mmt), 5.71 (s, 1H, H<sub>1</sub>), 4.98 (s, 2H, rotamer 1, H<sub>8</sub>), 4.66 (s, 2H, rotamer 2, H<sub>8</sub>), 4.31 (s, 2H, rotamer 1, H<sub>5</sub>), 3.91 (s, 2H, rotamer 2, H<sub>5</sub>), 3.87 (d, 2H, rotamer 1,  $J$  = 5.9 Hz, H<sub>10</sub>), 3.90 (t, 2H, rotamer 2,  $J$  = 5.9 Hz, H<sub>10</sub>), 3.73 (t, 2H, rotamer 2,  $J$  = 5.9 Hz, H<sub>2</sub>), 3.69-3.67 (s, 3H, OCH<sub>3</sub>), 3.62-3.60 (s, 3H, OCH<sub>3</sub>), 3.48 (m, 2H, rotamer 1, H<sub>3</sub>), 3.29 (m, 2H, rotamer 2, H<sub>3</sub>), 2.11 (m, 2H, H<sub>2</sub>), 1.27 (s, 9H, tBuBz).  $^{13}\text{C}$  NMR (101 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 170.7 (C<sub>11</sub>, CO), 169.1 (C<sub>6</sub>, CO), 167.9 (C<sub>7</sub>, CO), 167.7 (C-tBuBz, CO), 157.9 (C<sub>2</sub>, CO), 156.3 (C<sub>1</sub>-Mmt, C), 151.6 (C<sub>para</sub>-tBuBz, C), 146.91 (C<sub>a</sub>-Mmt, C), 138.4 (C<sub>e</sub>-Mmt, C), 130.9 (C<sub>ph</sub>-tBuBz, C), 130.2 (C<sub>ortho</sub>-tBuBz, CH), 130.05 (C<sub>f</sub>-Mmt, CH), 128.9 (C<sub>c</sub>-Mmt, CH), 128.2 (C<sub>b</sub>-Mmt, CH), 126.5 (C<sub>d</sub>-Mmt, CH), 125.8 (C<sub>meta</sub>-tBuBz, CH), 113.5 (C<sub>g</sub>-Mmt, CH), 96.2 (C<sub>5</sub>, CH), 70.4 (C-Mmt, C), 55.4 (OCH<sub>3</sub>-Mmt, CH<sub>3</sub>),

52.2 (OCH<sub>3</sub>, CH<sub>3</sub>), 50.1 (C<sub>8</sub>, CH<sub>2</sub>), 48.9 (C<sub>5</sub>, CH<sub>2</sub>), 48.9 (C<sub>3</sub>, CH<sub>2</sub>), 40.8 and 41.2 (C<sub>10</sub>, CH<sub>2</sub>), 40.9 (C<sub>2</sub>, CH<sub>2</sub>), 35.4 (C-tBuBz, C), 31.4 (C-tBuBz, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>44</sub>H<sub>49</sub>N<sub>6</sub>O<sub>7</sub><sup>+</sup> (M+H)<sup>+</sup> 773.3663, found 773.3654.

**2-[[[ $(\text{N}^2$ -(isobutanoyl)guanine-9-yl)acetyl]-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))]amino-N-[2-(9H-fluoren-9-yl)-methoxy]carbonyl]-amino]ethyl]acetamide (4n).**

A mixture of 2-{[tert-butyl(diphenyl)silyl]oxy}ethanamine **1b** (45.8 mg, 0.15 mmol, 1.0 equiv.) and paraformaldehyde (4.6 mg, 0.15 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 500  $\mu$ l) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. (2-Isocyano-ethyl)-carbamic acid 9H-fluoren-9-yl-methyl ester **3c** (44.7 mg, 0.15 mmol, 1.0 equiv.) and N<sup>2</sup>-(isobutanoyl)-9-(carboxymethyl)guanine **2a** (42.7 mg, 0.15 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 50 W (60 °C) for 3 h. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 5% MeOH) to afford **4l** as a white powder (57.4 mg, 43% yield). HPLC/MS ESI<sup>+</sup>/ESI<sup>-</sup> 883.6 [M+H]<sup>+</sup>; 882.46 [M-H]<sup>-</sup>.

**General procedure for deprotection of PNA monomers 5**

PNA monomer **4** was dissolved in a 50% TFA/DCM solution and the mixture was stirred 1 h at 0 °C. After evaporation of the TFA, the residue was taken up in DCM and washed with saturated NaHCO<sub>3</sub> solution. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and evaporated. The residue was then taken up in DCM and scavenger resin benzyltriethylammonium chloride (3 equiv./mmol of **4**) was added. The mixture was stirred 3 h at rt. The resin was filtered, washed with DCM and the filtrate was evaporated to give the deprotected PNA monomer **5** as a white powder.

**2-[[[ $(\text{N}^2$ -(isobutanoyl)guanine-9-yl)acetyl]-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))]amino-N-[2-aminoethyl]acetamide (5a)**

Following the general procedure, compound **4a** (550 mg, 0.72 mmol) afforded **5a** as a white powder (400 mg, 84% yield). Tr-HPLC/MS: 7.7 min (> 99%, 2-100% ACN in 9 min).  $^1\text{H}$  NMR (400 MHz, DMSO- $d_6$ )  $\delta$  (ppm) 12.06 (bs, 1H, H<sub>1</sub>), 11.66 (bs, 1H, NH-iBu), 8.28 (t, 1H, rotamer 1,  $J$  = 5.3 Hz, H<sub>9</sub>), 7.91 (t, 1H, rotamer 2,  $J$  = 5.3 Hz, H<sub>9</sub>), 7.76 (s, 1H, rotamer 1, H<sub>8</sub>), 7.46 (s, 1H, rotamer 2, H<sub>8</sub>), 7.64-7.60 (m, 4H, H<sub>ortho</sub>-TBDPS), 7.47-7.42 (m, 6H, H<sub>meta/para</sub>-TBDPS), 5.04 (s, 2H, rotamer 1, H<sub>8</sub>), 4.96 (s, 2H, rotamer 2, H<sub>8</sub>), 4.20 (s, 2H, rotamer 1, H<sub>5</sub>), 3.93 (s, 2H, rotamer 2, H<sub>5</sub>), 3.89 (t, 2H, rotamer 1,  $J$  = 5.4 Hz, H<sub>2</sub>), 3.70 (t, 2H, rotamer 2,  $J$  = 6.1 Hz, H<sub>2</sub>), 3.77 (bs, 2H, H<sub>12</sub>), 3.64 (t, 2H, rotamer 1,  $J$  = 5.4 Hz, H<sub>3</sub>), 3.50 (t, 2H, rotamer 2,  $J$  = 6.1 Hz, H<sub>3</sub>), 3.17 (m, 2H, rotamer 1, H<sub>10</sub>), 3.07 (m, 2H, rotamer 2, H<sub>10</sub>), 2.68 (m, 2H, rotamer 1, H<sub>11</sub>), 2.55 (m, 2H, rotamer 2, H<sub>11</sub>), 2.71 (sept, 1H,  $J$  = 6.9 Hz, H-iBu), 1.10 (d, 6H,  $J$  = 6.9 Hz, iBu), 1.03 (s, 9H, tBu-TBDPS).  $^{13}\text{C}$  NMR (100 MHz, DMSO- $d_6$ ):  $\delta$  (ppm) 180.6 (C-iBu, CO), 168.0 (C<sub>6</sub>, CO), 167.1 and 166.5 (C<sub>7</sub>, CO, rotamers), 155.2 (C<sub>6</sub>, CO), 149.5 (C<sub>4</sub>, C), 148.7 (C<sub>2</sub>, C), 140.4 and 139.8 (C<sub>8</sub>, CH, rotamers), 135.0 (C<sub>ortho</sub>-TBDPS, CH), 132.7 (C<sub>ph</sub>-TBDPS, C), 130.0 (C<sub>para</sub>-TBDPS, CH), 128.0 (C<sub>meta</sub>-TBDPS, CH), 119.5 (C<sub>5</sub>, C), 61.9 and

60.9 (C<sub>2'</sub>, CH<sub>2</sub>, rotamers), 50.7 and 49.1 (C<sub>5'</sub>, CH<sub>2</sub>, rotamers), 49.5 and 49.2 (C<sub>3'</sub>, CH<sub>2</sub>, rotamers), 44.4 and 44.0 (C<sub>8'</sub>, CH<sub>2</sub>, rotamers), 42.1 and 41.7 (C<sub>10'</sub>, CH<sub>2</sub>, rotamers), 41.0 and 40.8 (C<sub>11'</sub>, CH<sub>2</sub>, rotamers), 35.1 (C-iBu, CH), 26.7 (C-tBu-TBDPS, 3CH<sub>3</sub>), 19.0 (C-iBu, 2CH<sub>3</sub>), 18.7 (C-tBu-TBDPS, C). HRMS ESI<sup>+</sup> calcd for C<sub>33</sub>H<sub>45</sub>N<sub>8</sub>O<sub>5</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 661.3282, found 661.3280. HPLC/MS ESI<sup>+</sup>/ESI<sup>-</sup> 661.10 [M+H]<sup>+</sup>; 659.46 [M-H]<sup>-</sup>.

**2-((((N<sup>6</sup>-(benzoyl)adenin-9-yl)acetyl)-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))amino-N-[2-aminoethyl]acetamide (5c).**

Following the general procedure, compound **4c** (300 mg, 0.39 mmol) afforded **5c** as a white powder (213 mg, 81% yield). Tr-HPLC/MS: 7.8 min (> 98%, 2-100% ACN in 9 min). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>) δ (ppm) 167.9 (C<sub>6'</sub>, CO), 167.1 and 166.4 (C<sub>7'</sub>, CO, rotamers), 166.3 (C-Bz, CO), 153.3 (C<sub>4'</sub>, C), 151.9 (C<sub>2'</sub>, C), 150.6 (C<sub>6'</sub>, C), 146.1 (C<sub>8'</sub>, CH), 135.0 (C<sub>ortho</sub>-TBDPS, CH), 133.4 (C<sub>ph</sub>-Bz, C), 130.5 (C<sub>meta</sub>-Bz, CH), 130.4 (C<sub>ph</sub>-TBDPS), 129.7 (C<sub>para</sub>-TBDPS, CH), 129.0 (C<sub>ortho</sub>/para-Bz, CH), 128.0 (C<sub>meta</sub>-TBDPS, CH), 125.0 (C<sub>5'</sub>, C), 61.2 and 60.9 (C<sub>2'</sub>, CH<sub>2</sub>, rotamers), 50.3 and 47.9 (C<sub>5'</sub>, CH<sub>2</sub>, rotamers), 49.3 and 49.2 (C<sub>3'</sub>, CH<sub>2</sub>, rotamers), 46.7 and 43.2 (C<sub>8'</sub>, CH<sub>2</sub>, rotamers), 40.0 (C<sub>11'</sub>, CH<sub>2</sub>), 39.4 (C<sub>10'</sub>, CH<sub>2</sub>), 26.6 (C-tBu-TBDPS, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>36</sub>H<sub>43</sub>N<sub>8</sub>O<sub>4</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 679.3173, found 679.3171. HPLC/MS ESI<sup>+</sup>/ESI<sup>-</sup> 678.81 [M+H]<sup>+</sup>; 676.97 [M-H]<sup>-</sup>.

**2-((((Thymin-1-yl)acetyl)-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))amino-N-[2-aminoethyl]acetamide (5d).**

Following the general procedure, compound **4d** (689 mg, 1.03 mmol) afforded **5d** as a white powder (498.9 mg, 85% yield). Tr-HPLC/MS: 7.7 min (> 95%, 2-100% ACN in 9 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ (ppm) 11.27 (bs, 1H, H<sub>3</sub>), 8.11 (t, 1H, rotamer 1, J = 5.8 Hz, H<sub>9</sub>), 7.87 (t, 1H, rotamer 2, J = 5.8 Hz, H<sub>9</sub>), 7.65-7.57 (m, 4H, H<sub>ortho</sub>-TBDPS), 7.53 – 7.42 (m, 6H, H<sub>meta</sub>/para-TBDPS), 7.22 (s, 1H, rotamer 1, H<sub>6</sub>), 7.10 (s, 1H, rotamer 2, H<sub>6</sub>), 4.56 (s, 2H, rotamer 1, H<sub>8</sub>), 4.47 (s, 2H, rotamer 2, H<sub>8</sub>), 4.12 (s, 2H, rotamer 1, H<sub>5</sub>), 4.06 (s, 2H, rotamer 2, H<sub>5</sub>), 3.88 (m, 2H, rotamer 1, H<sub>2</sub>), 3.78 (m, 2H, rotamer 2, H<sub>2</sub>), 3.57 (t, 2H, rotamer 1, J = 5.8 Hz, H<sub>3</sub>), 3.48 (t, 2H, rotamer 2, J = 5.8 Hz, H<sub>3</sub>), 3.25 (m, 2H, rotamer 1, H<sub>10</sub>), 3.06 (m, 2H, rotamer 2, H<sub>10</sub>), 3.06 (m, 2H, rotamer 1, H<sub>11</sub>), 2.99 (m, 2H, rotamer 2, H<sub>11</sub>), 1.75 (d, 3H, H<sub>7</sub>), 0.96 (m, 9H, tBu-TBDPS). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>) δ (ppm) 168.7 (C<sub>7'</sub>, CO), 168.1 (C<sub>6'</sub>, CO), 164.9 (C<sub>4'</sub>, CO), 151.5 (C<sub>2'</sub>, CO), 142.4 (C<sub>6'</sub>, CH), 135.6 (C<sub>ortho</sub>-TBDPS, CH), 133.3 (C<sub>ph</sub>-TBDPS, C), 130.5 (C<sub>para</sub>-TBDPS, CH), 128.5 (C<sub>meta</sub>-TBDPS, CH), 108.5 (C<sub>5'</sub>, C), 62.3 and 61.6 (C<sub>2'</sub>, CH<sub>2</sub>, rotamers), 51.3 and 49.9 (C<sub>5'</sub>, CH<sub>2</sub>, rotamers), 49.8 and 48.9 (C<sub>3'</sub>, CH<sub>2</sub>, rotamers), 48.4 (C<sub>8'</sub>, CH<sub>2</sub>) 40.0 and 39.3 (C<sub>11'</sub>, CH<sub>2</sub>, rotamers), 39.5 and 39.3 (C<sub>10'</sub>, CH<sub>2</sub>, rotamers), 27.2 (C-tBu-TBDPS, CH<sub>3</sub>), 19.2 (C-tBu-TBDPS, C), 12.5 (C<sub>7</sub>, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>29</sub>H<sub>40</sub>N<sub>5</sub>O<sub>5</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 566.2791, found 566.2793. HPLC/MS ESI<sup>+</sup>/ESI<sup>-</sup> 565.99 [M+H]<sup>+</sup>; 564.01 [M-H]<sup>-</sup>.

**2-((((Uracil-1-yl)acetyl)-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))amino-N-[2-aminoethyl]acetamide (5e).**

Following the general procedure, compound **4e** (159 mg, 0.24 mmol) afforded **5e** as a white powder (92.1 mg, 69% yield). Tr-

HPLC/MS: 7.5 min (> 95%, 2-100% ACN in 9 min). <sup>1</sup>H NMR (400 MHz, DMSO-d<sub>6</sub>) δ (ppm) 11.26 (1H, s, H<sub>3</sub>), 8.17 (t, 1H, rotamer 1, J = 5.8 Hz, H<sub>9</sub>), 7.88 (t, 1H, rotamer 2, J = 5.8 Hz, H<sub>9</sub>), 7.67 – 7.57 (m, 4H, H<sub>ortho</sub>-TBDPS), 7.51 – 7.41 (m, 6H, H<sub>meta</sub>/para-TBDPS), 7.40 (m, 1H, rotamer 1, H<sub>6</sub>), 7.25 (m, 1H, rotamer 2, H<sub>6</sub>), 5.52 (dd, 1H, H<sub>5</sub>), 4.58 (s, 2H, rotamer 1, H<sub>8</sub>), 4.52 (s, 2H, rotamer 2, H<sub>8</sub>), 4.12 (s, 2H, rotamer 1, H<sub>5</sub>), 3.90 (s, 2H, rotamer 2, H<sub>5</sub>), 3.77 (t, 2H, rotamer 1, J = 5.8 Hz, H<sub>2</sub>), 3.66 (t, 2H, rotamer 2, J = 5.8 Hz, H<sub>2</sub>), 3.53 (t, 2H, rotamer 1, J = 5.8 Hz, H<sub>3</sub>), 3.49 (t, 2H, rotamer 2, J = 5.8 Hz, H<sub>3</sub>), 3.25 (m, 2H, rotamer 1, H<sub>10</sub>), 3.04 (m, 2H, rotamer 2, H<sub>10</sub>), 3.03 (m, 2H, rotamer 1, H<sub>11</sub>), 2.97 (m, 2H, rotamer 2, H<sub>11</sub>), 0.99 (d, 9H, tBu-TBDPS). <sup>13</sup>C NMR (100 MHz, DMSO-d<sub>6</sub>) δ (ppm) 168.2 (C<sub>6'</sub>, CO), 167.6 (C<sub>7'</sub>, CO), 164.2 (C<sub>4'</sub>, CO), 151.4 (C<sub>2'</sub>, CO), 146.9 (C<sub>6'</sub>, CH), 135.5 (C<sub>ortho</sub>-TBDPS, CH), 133.3 (C<sub>ph</sub>-TBDPS, C), 130.4 (C<sub>para</sub>-TBDPS, CH), 128.4 (C<sub>meta</sub>-TBDPS, CH), 101.1 (C<sub>5'</sub>, CH), 61.2 (C<sub>2'</sub>, CH<sub>2</sub>), 49.9 (C<sub>3'</sub>, CH<sub>2</sub>), 49.8 (C<sub>5'</sub>, CH<sub>2</sub>), 48.5 (C<sub>8'</sub>, CH<sub>2</sub>), 39.7 (C<sub>11'</sub>, CH<sub>2</sub>), 39.4 (C<sub>10'</sub>, CH<sub>2</sub>), 27.2 (C-tBu-TBDPS, CH<sub>3</sub>), 19.2 (C-tBu-TBDPS, C). HRMS ESI<sup>+</sup> calcd for C<sub>28</sub>H<sub>38</sub>N<sub>5</sub>O<sub>5</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 552.2637, found 552.2637. HPLC/MS ESI<sup>+</sup>/ESI<sup>-</sup> 551.99 [M+H]<sup>+</sup>; 549.92 [M-H]<sup>-</sup>.

**Preparation of compounds 6a-f.**

**2-((((N<sup>2</sup>-(isobutanoyl)guanin-9-yl)acetyl)-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))amino-N-[2-((((N<sup>4</sup>-(4-tert-butylbenzoyl)cytosin-1-yl)acetyl)-ethyl))amino-N-[(methylacetate)glycinamidyl]acetamide (6a)**

A mixture of compound **5a** (50 mg, 0.08 mmol, 1.0 equiv.) and paraformaldehyde (2.5 mg, 0.08 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 550 μL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocyanacetate (6.8 μL, 0.08 mmol, 1.0 equiv.) and N<sup>4</sup>-(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (25 mg, 0.08 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 100 W (100 °C) for 45 min. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 10% MeOH) to afford **6a** as a brown solid (36.3 mg, 43% yield). Tr-HPLC: 23.8 min (95%, 10-100% ACN in 40 min). <sup>1</sup>H NMR major conformer (600 MHz, 298 K, CD<sub>3</sub>OD) δ (ppm): 7.94 (d, 2H, J = 8.3 Hz, H<sub>ortho</sub>-tBuBz), 7.77 (d, 1H, J = 7.0 Hz, H<sub>6</sub> of C), 7.70 (d, 4H, J = 7.1, H<sub>ortho</sub>-TBDPS), 7.58 (2H, d, J = 8.3 Hz, H<sub>meta</sub>-tBuBz), 7.38-7.45 (m, 2H, H<sub>para</sub>-TBDPS & m, 4H, H<sub>meta</sub>-TBDPS), 7.30 (s, 1H, H<sub>8</sub> of G), 7.16 (d, 1H, J = 7.0 Hz, H<sub>5</sub> of C), 5.08 (s, 2H, H<sub>8</sub> of G), 4.73 (s, 2H, H<sub>8</sub> of C), 4.14 (bs, 2H, H<sub>5</sub> of C), 4.12 (bs, 2H, H<sub>5</sub> of G), 4.02 (t, 2H, H<sub>2</sub> of G), 3.96 (s, 2H, H<sub>10</sub>), 3.72 (2H, H<sub>3</sub> of G), 3.63 (s, 3H, H<sub>12</sub> & m, 2H, H<sub>3</sub> of C), 3.51 (m, 2H, H<sub>2</sub> of C), 2.59 (m, 1H, CH-iPrCO), 1.37 (s, 9H, CH<sub>3</sub>-tBuBz), 1.18 (d, 6H, J = 6.8 Hz, CH<sub>3</sub>-iPrCO), 1.11 (s, 9H, CH<sub>3</sub>-TBDPS).

**2-((((N<sup>2</sup>-(isobutanoyl)guanin-9-yl)acetyl)-(2-(tert-butyl(diphenylsilyl)oxy)ethyl))amino-N-[2-((((Uracil-1-yl)acetyl)-ethyl))amino-N-[(methylacetate)glycinamidyl]acetamide (6b)**

A mixture of compound **5a** (411.9 mg, 0.62 mmol, 1.0 equiv.) and paraformaldehyde (20.6 mg, 0.62 mmol, 1.1 equiv.) in anhydrous isopropanol (0.5 M, 1.2 mL) was placed in a sealed vessel (0.5 – 2



mL) and stirred at rt for 1 h. Methyl isocyanacetate (56.7  $\mu$ L, 0.62 mmol, 1.0 equiv.) and uracil-1-acetic acid **2e** (106.2 mg, 0.62 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 100 W (100  $^{\circ}$ C) for 45 min. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 10% MeOH) to afford **6b** as a white solid (205.5 mg, 35% yield). Tr-HPLC: 20.5 min (90%, 10-100% ACN in 40 min). HRMS ESI<sup>+</sup> calcd for C<sub>44</sub>H<sub>56</sub>N<sub>11</sub>O<sub>11</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 942.3925, found 942.3925.

**2-[[[[(N<sup>6</sup>-(benzoyl)adenin-9-yl)acetyl]-2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-[[[[(N<sup>4</sup>-(4-tert-butylbenzoyl)cytosin-1-yl)acetyl]-ethyl]]]amino-N-[[methylacetate]glycinamidyl]acetamide (6c)**

A mixture of compound **5c** (507.6 mg, 0.75 mmol, 1.0 equiv.) and paraformaldehyde (22.4 mg, 0.75 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.4 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocyanacetate (68  $\mu$ L, 0.75 mmol, 1.0 equiv.) and N<sup>4</sup>-(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (246 mg, 0.75 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 100 W (100  $^{\circ}$ C) for 45 min. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 10% MeOH) to afford **6c** as a white solid (250.8 mg, 30% yield). Tr-HPLC: 24.7 min (98%, 10-100% ACN in 40 min). <sup>1</sup>H NMR (400 MHz, 298 K, 15 mM, CDCl<sub>3</sub>)  $\delta$  (ppm): 9.43 (bs, 1H, H<sub>N6</sub> of A), 8.69 (s, 1H, H<sub>2</sub> of A), 8.69 (bs, 1H, H<sub>N4</sub> of C), 8.33 (bs, 1H, H<sub>N1'</sub> of C), 8.23 (bs, 1H, H<sub>N9'</sub> of C), 7.95 (d, 2H, J=7.6 Hz, H<sub>ortho</sub>-Bz), 7.92 (s, 1H, H<sub>8</sub> of A), 7.69 (dd, 4H, J=7.6 and 2.0 Hz, H<sub>ortho</sub>-TBDPS), 7.62 (d, 2H, J=7.2 Hz, H<sub>ortho</sub>-tBuBz), 7.55 (d, 1H, J=7.0 Hz, H<sub>6</sub> of C), 7.42 (m, 2H, H<sub>para</sub>-TBDPS), 7.39 (m, 4H, H<sub>meta</sub>-TBDPS), 7.38 (2H, H<sub>meta</sub>-tBuBz), 7.37 (m, 1H, H<sub>para</sub>-Bz), 7.20 (t, 2H, J=7.6 Hz, H<sub>meta</sub>-Bz), 7.00 (bs, 1H, H<sub>5</sub> of C), 5.21 (s, 2H, H<sub>8'</sub> of A), 4.59 (s, 2H, H<sub>8'</sub> of C), 4.05 (d, 2H, J=5.6 Hz, H<sub>10'</sub>), 4.05 (bs, 2H, H<sub>5'</sub> of A), 4.01 (bs, 2H, H<sub>5'</sub> of C), 3.94 (t, 2H, J=5.3 Hz, H<sub>2'</sub> of A), 3.63 (bs, 2H, H<sub>3'</sub> of A), 3.54 (s, 3H, H<sub>12'</sub>), 3.48 (bs, 2H, H<sub>3'</sub> of C), 3.42 (bs, 2H, H<sub>2'</sub> of C), 1.34 (s, 9H, tBu-tBuBz), 1.11 (s, 9H, tBu-TBDPS). <sup>13</sup>C{<sup>1</sup>H} NMR (100 MHz, 298 K, 15 mM, CDCl<sub>3</sub>)  $\delta$  (ppm): 171.6 (C<sub>6'</sub> of C, CO), 170.8 (C<sub>11'</sub>, CO), 168.0; 167.9 & 167.8 (C<sub>6'</sub> of A, CO; C<sub>7'</sub> of C, CO & C<sub>7'</sub> of A, CO), 166.4 (C-tBuBz, CO), 165.2 (C-Bz, CO), 162.8 (C<sub>4</sub> of C, C), 157.0 (C-tBuBz, C), 156.5 (C<sub>2</sub> of C, C), 152.7 (C<sub>2</sub> of A, CH), 152.5 (C<sub>4</sub> of A, C), 151.4 (C<sub>6</sub> of C, CH), 149.5 (C<sub>6</sub> of A, C), 145.4 (C<sub>8</sub> of A, CH), 135.7 (C<sub>ortho</sub>-TBDPS, CH), 133.8 (C-Bz, C), 132.7 (C-TBDPS, C), 132.3 (C<sub>para</sub>-Bz, CH), 130.3 (C<sub>para</sub>-TBDPS, CH), 130.0 (C-tBuBz, C), 128.6 (C<sub>meta</sub>-Bz, CH), 128.2 (C<sub>ortho</sub>-Bz, CH & C<sub>meta</sub>-TBDPS, CH), 127.7 (C<sub>ortho</sub>-tBuBz, CH), 125.9 (C<sub>meta</sub>-tBuBz, CH), 123.0 (C<sub>5</sub> of A, C), 96.5 (C<sub>5</sub> of C, CH), 61.6 (C<sub>2'</sub> of A, CH<sub>2</sub>), 52.4 (C<sub>12'</sub>, CH<sub>3</sub>), 50.7 & 50.0 (C<sub>5'</sub> of A, CH<sub>2</sub> & C<sub>5'</sub> of C, CH<sub>2</sub>), 50.5 (C<sub>8'</sub> of C, CH<sub>2</sub> & C<sub>3'</sub> of A, CH<sub>2</sub>), 48.9 (C<sub>3'</sub> of C, CH<sub>2</sub>), 44.4 (C<sub>8'</sub> of A, CH<sub>2</sub>), 41.6 (C<sub>10'</sub>, CH<sub>2</sub>), 36.3 (C<sub>2'</sub> of C, CH<sub>2</sub>), 35.3 (C-tBu-tBuBz, C), 31.2 (C-tBu-tBuBz, CH<sub>3</sub>), 27.1 (CtBu-TBDPS, CH<sub>3</sub>), 19.3 (CtBu-TBDPS, C). HRMS ESI<sup>+</sup> calcd for C<sub>58</sub>H<sub>67</sub>N<sub>12</sub>O<sub>10</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 1119.4868, found 1119.4867.

**2-[[[[(N<sup>6</sup>-(benzoyl)adenin-9-yl)acetyl]-2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-[[[[(N<sup>2</sup>-**

**(isobutanoyl)guanin-9-yl)acetyl]-ethyl]]]amino-N-[[methylacetate]glycinamidyl]acetamide (6d)**

A mixture of compound **5c** (100 mg, 0.15 mmol, 1.0 equiv.) and paraformaldehyde (4.9 mg, 0.15 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 294  $\mu$ L) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocyanacetate (13.4  $\mu$ L, 0.15 mmol, 1.0 equiv.) and N<sup>2</sup>-(isobutanoyl)-9-(carboxymethyl) guanine **2a** (41 mg, 0.15 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 100 W (100  $^{\circ}$ C) for 45 min. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 10% MeOH) to afford **6d** as a brown solid (39.6 mg, 25% yield). Tr-HPLC/MS: 8.8 min (< 80%, 2-100% ACN in 9 min). HRMS ESI<sup>+</sup> calcd for C<sub>52</sub>H<sub>61</sub>N<sub>14</sub>O<sub>10</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 1069.4460, found 1069.4459.

**2-[[[[(Thymin-1-yl)acetyl]-2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-[[[[(Thymin-1-yl)acetyl]-ethyl]]]amino-N-[[methylacetate]glycinamidyl]acetamide (6e)**

A mixture of compound **5d** (498.9 mg, 0.88 mmol, 1.0 equiv.) and paraformaldehyde (26.5 mg, 0.88 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 1.7 mL) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocyanacetate (80  $\mu$ L, 0.88 mmol, 1.0 equiv.) and thymin-1-acetic acid **2d** (163.2 mg, 0.88 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 100 W (100  $^{\circ}$ C) for 45 min. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 10% MeOH) to afford **6e** as a white solid (303.2 mg, 40% yield). Tr-HPLC: 20.2 min (96%, 10-100% ACN in 40 min). HRMS ESI<sup>+</sup> calcd for C<sub>41</sub>H<sub>53</sub>N<sub>8</sub>O<sub>11</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 861.3599, found 861.3598.

**2-[[[[(Uracil-1-yl)acetyl]-2-(tert-butyl(diphenylsilyl)oxy)ethyl]]]amino-N-[2-[[[[(N<sup>4</sup>-(4-tert-butylbenzoyl)cytosin-1-yl)acetyl]-ethyl]]]amino-N-[[methylacetate]glycinamidyl]acetamide (6f)**

A mixture of compound **5e** (80 mg, 0.15 mmol, 1.0 equiv.) and paraformaldehyde (4.3 mg, 0.15 mmol, 1.0 equiv.) in anhydrous isopropanol (0.5 M, 289  $\mu$ L) was placed in a sealed vessel (0.5 – 2 mL) and stirred at rt for 1 h. Methyl isocyanacetate (13.2  $\mu$ L, 0.15 mmol, 1.0 equiv.) and N<sup>4</sup>-(4-tert-butylbenzoyl)-1-(carboxymethyl)cytosine **2b** (47.8 mg, 0.15 mmol, 1.0 equiv.) were then added dropwise. The resulting solution was irradiated at 100 W (100  $^{\circ}$ C) for 45 min. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by column chromatography (DCM with 10% MeOH) to afford **6f** as a white solid (34.4 mg, 25% yield). Tr-HPLC: 23.6 min (95%, 10-100% ACN in 40 min). HRMS ESI<sup>+</sup> calcd for C<sub>50</sub>H<sub>62</sub>N<sub>9</sub>O<sub>11</sub>Si<sup>+</sup> (M+H)<sup>+</sup> 992.4334, found 992.4333.

**Preparation of compounds 7a-c and 7e.**

**2-[[[[(Guanin-9-yl)acetyl]-2-(hydroxyethyl)]amino-N-[2-[[[[(Cytosin-1-yl)acetyl]-ethyl]]]amino-N-[(carbamoyl)glycinamidyl]acetamide (7a)**



PNA dimer **6a** (60 mg) was dissolved in THF/MeOH 1:3 (v:v) (400  $\mu$ L) and 28% aqueous ammonia was added (600  $\mu$ L). The flask was sealed and the mixture was stirred 6 h at 50 °C. The solvents were then removed under reduced pressure and the crude material was taken up in THF (500  $\mu$ L). A solution of TBAF (1 M) in THF was added and the reaction mixture was stirred 1 h at room temperature. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by reverse phase column chromatography (0 to 100% of 0.05 M TEAB buffer (pH 7.5) in 50% ACN) to give PNA dimer **7a** (12 mg, 33%). Tr-HPLC: 16.1 min (98%, 0-100% TEAB (0.05 M) ACN/H<sub>2</sub>O 50/50 in 55 min). <sup>1</sup>H NMR major conformer (400 MHz, 323K, 15 mM, D<sub>2</sub>O)  $\delta$  (ppm): 7.69 (s, 1H, H<sub>8</sub> of G), 7.15 (d, 1H, *J*=7.3 Hz, H<sub>6</sub> of C), 5.68 (d, 1H, *J*=7.3 Hz, H<sub>5</sub> of C), 5.23 (s, 2H, H<sub>8'</sub> of G), 4.58 (s, 2H, H<sub>8'</sub> of C), 4.14 (s, 1H, H<sub>5'</sub> of G), 4.11 (s, 1H, H<sub>5'</sub> of C), 3.90 (m, 2H, H<sub>10'</sub>), 3.86 (t, 2H, *J*=5.3 Hz, H<sub>2'</sub> of G), 3.70 (m, 2H, H<sub>3'</sub> of G), 3.65 (m, 2H, H<sub>3'</sub> of C), 3.51 (t, 2H, *J*=5.3 Hz, H<sub>2'</sub> of C). <sup>13</sup>C{<sup>1</sup>H} NMR major conformer (100 MHz, 323K, 15 mM, D<sub>2</sub>O)  $\delta$  (ppm): 176.7 (C<sub>11'</sub>, C), 174.2 (C<sub>6'</sub> of G, C), 173.8 (C<sub>6'</sub> of C, C), 172.8 (C<sub>7'</sub> of C, C), 172.1 (C<sub>7'</sub> of G, C), 169.3 (C<sub>4</sub> of C, C), 160.5 (C<sub>2</sub> of C, C), 154.4 (C<sub>4</sub> of G, C), 149.7 (C<sub>6</sub> of C, CH), 143.0 (C<sub>8</sub> of G, CH), 118.4 (C<sub>5</sub> of G, C), 97.7 (C<sub>5</sub> of C, CH), 61.2 (C<sub>2'</sub> of G, CH<sub>2</sub>), 53.2 (C<sub>3'</sub> of G, CH<sub>2</sub>), 52.6 (C<sub>5'</sub> of G, CH<sub>2</sub>), 52.3, (C<sub>8'</sub> of C, CH<sub>2</sub>), 51.9 (C<sub>5'</sub> of C, CH<sub>2</sub>), 50.7 (C<sub>3'</sub> of C, CH<sub>2</sub>), 47.5 (C<sub>8'</sub> of G, CH<sub>2</sub>), 44.7 (C<sub>10'</sub>, CH<sub>2</sub>), 39.6 (C<sub>2'</sub> of C, CH<sub>2</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>23</sub>H<sub>32</sub>N<sub>13</sub>O<sub>8</sub><sup>+</sup> (M+H)<sup>+</sup> 618.2497, found 618.2494.

**2-[[[Guanin-9-yl)acetyl]-(2-hydroxyethyl)]amino-N-2-[[[Uracil-1-yl)acetyl]-ethyl]]amino-N-[[carbamoyl]glycinamidyl]acetamide (7b)**

PNA dimer **6b** (30 mg) was dissolved in THF/MeOH 1:3 (v:v) (1.2 mL) and 28% aqueous ammonia was added (1.8 mL). The flask was sealed and the mixture was stirred 6 h at 50 °C. The solvents were then removed under reduced pressure and the crude product was taken up in THF (1.3 mL). A solution of TBAF (1 M) in THF (205  $\mu$ L) was added and the reaction mixture was stirred 1 h at room temperature. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by reverse phase column chromatography (0 to 100% of 0.05 M TEAB (pH 7.5) in 50% ACN) to give PNA dimer **7b** (4.7 mg, 23% yield). Tr-HPLC: 18.2 min (> 95%, 0-100% TEAB (0.05 M) ACN/H<sub>2</sub>O 50/50 in 45 min). <sup>1</sup>H NMR major conformer (600 MHz, 323 K, D<sub>2</sub>O)  $\delta$  (ppm): 7.68 (s, 1H, H<sub>8</sub> of G), 7.27 (d, 1H, *J*=7.8 Hz, H<sub>6</sub> of U), 5.60 (d, 1H, *J*=7.8 Hz, H<sub>5</sub> of U), 5.21 (s, 2H, H<sub>8'</sub> of G), 4.66 (s, 2H, H<sub>8'</sub> of U), 4.13 (s, 2H, H<sub>5'</sub> of G), 4.12 (s, 2H, H<sub>5'</sub> of U), 3.90 (s, 2H, H<sub>10'</sub>), 3.85 (m, 2H, H<sub>2'</sub> of G), 3.70 (m, 2H, H<sub>3'</sub> of G), 3.61 (t, *J*=5.3 Hz, 2H, H<sub>3'</sub> of U), 3.50 (t, *J*=5.3 Hz, 2H, H<sub>2'</sub> of U). <sup>13</sup>C{<sup>1</sup>H} NMR major conformer (150 MHz, 323 K, D<sub>2</sub>O)  $\delta$  (ppm): 176.9 (C<sub>11'</sub>, CO), 174.3 (C<sub>6'</sub> of G, CO), 173.8 (C<sub>6'</sub> of U, CO), 172.1 (C<sub>7'</sub> of G, CO & C<sub>7'</sub> of U, CO), 169.1 (C<sub>4</sub> of U, CO), 161.3 & 156.5 (C<sub>2</sub> of G, C & C<sub>6</sub> of G, CO), 154.7 (C<sub>4</sub> of G, C & C<sub>2</sub> of U, CO), 150.0 (C<sub>6</sub> of U, CH), 143.4 (C<sub>8</sub> of G, CH), 118.4 (C<sub>5</sub> of G, C), 103.9 (C<sub>5</sub> of U, CH), 61.5 (C<sub>2'</sub> of G, CH<sub>2</sub>), 53.5 (C<sub>3'</sub> of G, CH<sub>2</sub>), 53.0 (C<sub>5'</sub> of G, CH<sub>2</sub>), 52.2 (C<sub>5'</sub> of U, CH<sub>2</sub>), 51.5 (C<sub>8'</sub> of U, CH<sub>2</sub>), 50.8 (C<sub>3'</sub> of U, CH<sub>2</sub>), 47.4 (C<sub>8'</sub> of G, CH<sub>2</sub>), 44.9 (C<sub>10'</sub>, CH<sub>2</sub>), 39.8 (C<sub>2'</sub> of U, CH<sub>2</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>23</sub>H<sub>31</sub>N<sub>12</sub>O<sub>9</sub><sup>+</sup> (M+H)<sup>+</sup> 619.2337, found 619.2332.

**2-[[[Adenin-9-yl)acetyl]-(2-hydroxyethyl)]amino-N-2-[[[cytosin-1-yl)acetyl]-ethyl]]amino-N-[[carbamoyl]glycinamidyl]acetamide (7c)**

PNA dimer **6c** (50 mg) was dissolved in THF/MeOH 1:3 (v:v) (5 mL) and 28% aqueous ammonia was added (3 mL). The flask was sealed and the mixture was stirred 6 h at 50 °C. The solvents were then removed under reduced pressure and the crude product was taken up in THF (3 mL). A solution of TBAF (1 M) in THF (440  $\mu$ L) was added and the reaction mixture was stirred 1 h at room temperature. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by reverse phase column chromatography (0 to 100% of 0.05 M TEAB buffer (pH 7.5) in 50% ACN) to give PNA dimer **7c** (8.5 mg, 32%). Tr-HPLC: 15.9 min (98%, 0-100% TEAB (0.05 M) ACN/H<sub>2</sub>O 50/50 in 45 min). <sup>1</sup>H NMR major conformer, 400 MHz, 298 K, 8.3 mM, D<sub>2</sub>O)  $\delta$  (ppm): 8.12 (s, 1H, H<sub>2</sub> of A), 8.02 (s, 1H, H<sub>8</sub> of A), 6.89 (d, 1H, *J*=7.3 Hz, H<sub>6</sub> of C), 5.52 (d, 1H, *J*=7.3 Hz, H<sub>5</sub> of C), 5.42 (s, 2H, H<sub>8'</sub> of A), 4.44 (s, 2H, H<sub>8'</sub> of C), 4.14 (s, 1H, H<sub>5'</sub> of A), 4.10 (s, 1H, H<sub>5'</sub> of C), 3.89 (m, 2H, H<sub>10'</sub>), 3.88 (m, 2H, H<sub>2'</sub> of A), 3.75 (m, 2H, *J*=5.2 Hz, H<sub>3'</sub> of A), 3.64 (m, 2H, H<sub>3'</sub> of C), 3.51 (m, 2H, H<sub>2'</sub> of C). <sup>13</sup>C{<sup>1</sup>H} NMR (major conformer, 100 MHz, 298 K, 8.3 mM, D<sub>2</sub>O)  $\delta$  (ppm): 175.6 (C<sub>11'</sub>, C), 172.9 (C<sub>6'</sub> of A, C), 172.2 (C<sub>6'</sub> of C, C), 171.4 (C<sub>7'</sub> of C, C), 170.8 (C<sub>7'</sub> of A, C), 167.9 (C<sub>4</sub> of C, C), 159.3 (C<sub>2</sub> of C, C), 156.8 (C<sub>6</sub> of A, C), 153.7 (C<sub>2</sub> of A, CH), 150.6 (C<sub>4</sub> of A, C), 148.0 (C<sub>6</sub> of C, CH), 144.3 (C<sub>8</sub> of A, CH), 119.5 (C<sub>5</sub> of A, C), 96.5 (C<sub>5</sub> of C, CH), 60.1 (C<sub>2'</sub> of A, CH<sub>2</sub>), 52.3 (C<sub>3'</sub> of A, CH<sub>2</sub>), 51.6 (C<sub>5'</sub> of A, CH<sub>2</sub>), 51.1, (C<sub>8'</sub> of C, CH<sub>2</sub>), 50.7 (C<sub>5'</sub> of C, CH<sub>2</sub>), 49.7 (C<sub>3'</sub> of C, CH<sub>2</sub>), 46.7 (C<sub>8'</sub> of A, CH<sub>2</sub>), 43.5 (C<sub>10'</sub>, CH<sub>2</sub>), 38.3 (C<sub>2'</sub> of C, CH<sub>2</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>23</sub>H<sub>32</sub>N<sub>13</sub>O<sub>7</sub><sup>+</sup> (M+H)<sup>+</sup> 602.2543, found 602.2542.

**2-[[[Thymin-9-yl)acetyl]-(2-hydroxyethyl)]amino-N-2-[[[thymin-1-yl)acetyl]-ethyl]]amino-N-[[carboxyl]glycinamidyl]acetamide (7e).**

PNA dimer **6e** (50 mg) was dissolved in THF (3 mL) and TBAF 1 M in THF (440  $\mu$ L) was added. The reaction mixture was stirred 1 h at room temperature. The solvent was removed under reduced pressure and the crude material was dissolved in MeOH (2.5 mL), NaOH 0.2 N (2.5 mL) was added and the reaction mixture was stirred 1 h at rt. The solution was neutralized with HCl 0.2 N. The reaction mixture was concentrated to dryness under reduced pressure and the residue was purified by reverse phase column chromatography (0 to 100% of 0.05M TEAB buffer (pH 7.5) in 50% ACN) to give PNA dimer **7e** (15 mg, 45% yield). Tr-HPLC: 18.8 min (99%, 0-100% TEAB (0.05M) ACN/H<sub>2</sub>O 50/50 in 45 min). <sup>1</sup>H NMR (400 MHz, 298 K, 27 mM, D<sub>2</sub>O)  $\delta$  (ppm): Conformer 1 (major): 7.48 (q, 1H, *J*=1.2 Hz, H<sub>6</sub> of T<sub>2</sub>), 7.19 (q, 1H, *J*=1.2 Hz, H<sub>6</sub> of T<sub>1</sub>), 4.86 (ABq, 2H,  $\Delta\nu$ =7.0 Hz, H<sub>8'</sub> of T<sub>1</sub>), 4.65 (ABq, 2H,  $\Delta\nu$ =6.8 Hz, H<sub>8'</sub> of T<sub>2</sub>), 4.16 (s, 2H, H<sub>5'</sub> of T<sub>2</sub>), 4.15 (s, 2H, H<sub>5'</sub> of T<sub>1</sub>), 3.93 (s, 2H, H<sub>10'</sub>), 3.81 (t, 2H, *J*=5.2 Hz, H<sub>2'</sub> of T<sub>1</sub>), 3.63 (t, 2H, *J*=5.2 Hz, H<sub>3'</sub> of T<sub>1</sub>), 3.59 (m, 2H, H<sub>3'</sub> of T<sub>2</sub>), 3.51 (m, 2H, H<sub>2'</sub> of T<sub>2</sub>), 1.76 (d, 3H, *J*=1.0 Hz, H<sub>7</sub> of T<sub>2</sub>), 1.75 (d, 3H, *J*=1.0 Hz, H<sub>7</sub> of T<sub>1</sub>). Conformer 2 : 7.41 (q, 1H, *J*=1.2 Hz, H<sub>6</sub> of T<sub>2</sub>), 7.40 (q, 1H, *J*=1.2 Hz, H<sub>6</sub> of T<sub>1</sub>), 4.88 (ABq, 2H, H<sub>8'</sub> of T<sub>1</sub>), 4.61 (ABq,  $\Delta\nu$ =7.2 Hz, 2H, H<sub>8'</sub> of T<sub>2</sub>), 4.30 (s, 2H, H<sub>5'</sub> of T<sub>2</sub>), 4.06 (s, 2H, H<sub>5'</sub> of T<sub>1</sub>), 3.97 (s, 2H, H<sub>10'</sub>), 3.78 (t, 2H, *J*=5.2 Hz, H<sub>2'</sub> of T<sub>1</sub>), 3.58 (m, 2H, H<sub>3'</sub> of T<sub>1</sub>), 3.58 (m, 2H, H<sub>3'</sub> of T<sub>2</sub>), 3.43 (m, 2H, H<sub>2'</sub> of T<sub>2</sub>), 1.84 (d, 3H, *J*=1.0 Hz, H<sub>7</sub> of T<sub>2</sub>), 1.82 (d, 3H, *J*=1.0 Hz, H<sub>7</sub> of T<sub>1</sub>). <sup>13</sup>C{<sup>1</sup>H} NMR (major conformer, 100 MHz, 298 K, 27 mM, D<sub>2</sub>O)  $\delta$  (ppm): 177.3 (C<sub>11'</sub>, CO), 173.8 (C<sub>6'</sub> of T<sub>1</sub>, CO), 173.7 (C<sub>6'</sub> of T<sub>2</sub>, CO), 172.4 (C<sub>7'</sub> of T<sub>1</sub>, CO), 171.6

(C<sub>7</sub> of T<sub>2</sub>, CO), 169.4 (C<sub>4</sub> of T<sub>2</sub>, C), 169.3 (C<sub>4</sub> of T<sub>1</sub>, C), 154.8 (C<sub>2</sub> of T<sub>1</sub> & T<sub>2</sub>, CO), 146.4 (C<sub>6</sub> of T<sub>1</sub> & T<sub>2</sub>, CH), 113.0 (C<sub>5</sub> of T<sub>1</sub>, C), 112.7 (C<sub>5</sub> of T<sub>2</sub>, C), 61.2 (C<sub>2</sub> of T<sub>1</sub>, CH<sub>2</sub>), 53.0 (C<sub>3</sub> of T<sub>1</sub>, CH<sub>2</sub>), 52.5 (C<sub>5</sub> of T<sub>1</sub>, CH<sub>2</sub>), 52.0 (C<sub>8</sub> of T<sub>1</sub>, CH<sub>2</sub>), 51.8 (C<sub>8</sub> of T<sub>2</sub>, CH<sub>2</sub>), 51.5 (C<sub>5</sub> of T<sub>2</sub>, CH<sub>2</sub>), 50.6 (C<sub>3</sub> of T<sub>2</sub>, CH<sub>2</sub>), 44.7 (C<sub>10</sub>, CH<sub>2</sub>), 39.1 (C<sub>2</sub> of T<sub>2</sub>, CH<sub>2</sub>), 13.9 (C<sub>7</sub> of T<sub>2</sub>, CH<sub>3</sub>), 13.7 (C<sub>7</sub> of T<sub>1</sub>, CH<sub>3</sub>). HRMS ESI<sup>+</sup> calcd for C<sub>24</sub>H<sub>33</sub>N<sub>8</sub>O<sub>11</sub><sup>+</sup> (M+H)<sup>+</sup> 609.2269, found 609.2271.

### MD Simulations

The model systems - PNA dimers in various spatial arrangements (see S57, Supporting Information) were constructed using the crystal structures<sup>35, 36, 43</sup> (Protein Data Bank entries 176D, 1PNN, 2K4G). All simulated systems were surrounded by TIP3P water molecules<sup>49</sup> which extended to a distance of approximately 10 Å (in each direction) from the PNA atoms. This gives a periodic box size of ~32 Å, ~36 Å, ~28 Å for a simulated system consisting of ~3.200 atoms. Force constants for PNA dimers were derived using the Generalized Amber Force Field (GAFF<sup>50</sup>) by means of the ACEMD protocol. Charges were calculated according to the AM1-BCC method<sup>51</sup>. New \*.inpcrd (initial coordinates) and \*.prmtop (molecular topology, force field) files for the whole simulated system, were created using the *tleap* module from the AMBER software package<sup>52</sup>.

Prior to production MD simulations, all systems were energy-minimized using the *pmemd* module of AMBER 14. Production MD simulation runs were performed with the *pmemd.cuda.MPI* module of AMBER 14, which runs exclusively on GPUs at the equivalent speed of tens of standard processor cores<sup>53</sup>. For production MD runs (lasting for 900-1200 ns) CUDA programmable NVIDIA GeForce GTX 465 GPUs equipped with 352 cores were used. MD trajectories were computed at a rate of ~330 ns per day (i.e. ~30 days of GPU time in total). The SPFP precision model was used<sup>53</sup>. Periodic boundary conditions (PBC) were applied. The particle mesh Ewald (PME) was used for calculation of electrostatic interactions<sup>54</sup>. 9 Å cutoff distance was applied for Lennard-Jones interactions. The temperature was maintained at 300 K via Langevin dynamics with a friction factor of 5. Further, the Monte Carlo barostat (a new addition to Amber 14), that samples rigorously from the isobaric-isothermal ensemble and does not necessitate computing the virial, was used for the production phase.

Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm. For water molecules, a special "three-point" SETTLE algorithm was used<sup>55</sup>. The hydrogen mass repartitioning scheme (the mass of the bonded heavy atoms to hydrogen is repartitioned among hydrogen atoms, leaving the total mass of the system unchanged) allowed a time step set to 4 fs<sup>56</sup>. By default in AMBER 14, partitioning is only applied to the solute since SHAKE on water is handled analytically (via the SETTLE algorithm).

Data were recorded every 400 ps. Trajectories analyses were performed using the *cpptraj* module of AMBER 14<sup>57</sup> and the VMD program. Figures were produced by means of the ChemBioOffice, ICM Molsoft and Gnuplot software packages.

### Acknowledgments

This work was supported by grant from the Aix-Marseille University. We wish to thank Dr G. Valette (Montpellier University, France) for his expertise in conducting mass spectrometry analysis. Ivan Barvik was supported by the Czech Science Foundation under contract #P205/12/G118.

### Notes and references

1. S. Priet, I. Zlatev, I. Barvik, K. Geerts, P. Leyssen, J. Neyts, H. Dutartre, B. Canard, J. J. Vasseur, F. Morvan and K. Alvarez, *J. Med. Chem.*, 2010, **53**, 6608-6617.
2. I. Zlatev, H. Dutartre, I. Barvik, J. Neyts, B. Canard, J. J. Vasseur, K. Alvarez and F. Morvan, *J. Med. Chem.*, 2008, **51**, 5745-5757.
3. P. E. Nielsen, M. Egholm, R. H. Berg and O. Buchardt, *Science*, 1991, **254**, 1497-1500.
4. P. Rathee, D. Rathee, A. Hooda, V. Kumar and S. Rathee, *Pharma Innovation*, 2012, **1**, 25-41.
5. R. Gambhari, *Exper Opin Ther Pat*, 2014, **24**, 267-294.
6. K. L. Dueholm, M. Egholm, C. Behrens, L. Christensen, H. F. Hansen, T. Vulpius, K. H. Petersen, R. H. Berg, P. E. Nielsen and O. Buchardt, *J. Org. Chem.*, 1994, **59**, 5767-5773.
7. A. Farese, N. Patino, R. Condom, S. Dalleu and R. Guedj, *Tetrahedron Lett.*, 1996, **37**, 1413-1416.
8. C. Di Giorgio, S. Pairot, C. Schwergold, N. Patino, R. Condom, A. Farese-Di Giorgio and R. Guedj, *Tetrahedron*, 1999, **55**, 1937-1958.
9. A. Domling, *Nucleosides & Nucleotides*, 1998, **17**, 1667-1670.
10. A. Domling, K. Z. Chi and M. Barrere, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 2871-2874.
11. W. Maison, I. Schlemminger, O. Westerhoff and J. Martens, *Bioorg. Med. Chem.*, 2000, **8**, 1343-1360.
12. W. Maison, I. Schlemminger, O. Westerhoff and J. Martens, *Bioorg. Med. Chem. Lett.*, 1999, **9**, 581-584.
13. P. Xu, T. Zhang, W. H. Wang, X. M. Zou, X. Zhang and Y. Q. Fu, *Synthesis-Stuttgart*, 2003, 1171-1176.
14. A. Domling, *Chem. Rev.*, 2006, **106**, 17-89.
15. A. Domling and I. Ugi, *Angew. Chem. Int. Ed.*, 2000, **39**, 3169-3210.
16. A. Domling, *Comb Chem High T Scr*, 1998, **1**, 1-22.
17. D. W. Will, D. Langner, J. Knolle and E. Uhlmann, *Tetrahedron*, 1995, **51**, 12069-12082.
18. C. Schwergold, G. Depecker, C. Di Giorgio, N. Patino, F. Jossinet, B. Ehresmann, R. Terreux, D. Cabrol-Bass and R. Condom, *Tetrahedron*, 2002, **58**, 5675-5687.
19. T. Shinozuka, Y. Yamamoto, T. Hasegawa, K. Salto and S. Naito, *Tetrahedron Lett.*, 2008, **49**, 1619-1622.
20. E. Riva, D. Comi, S. Borrelli, F. Colombo, B. Danieli, J. Borlak, L. Evensen, J. B. Lorens, G. Fontana, O. M. Gia, L. D. Via and D. Passarella, *Bioorg. Med. Chem.*, 2010, **18**, 8660-8668.
21. T. Kofoed, H. F. Hansen, H. Orum and T. Koch, *J. Pept. Sci.*, 2001, **7**, 402-412.

22. E. Uhlmann, D. W. Will, G. Breipohl, D. Langner and A. Ryte, *Angew. Chem. Int. Ed.*, 1996, **35**, 2632-2635.
23. P. J. Finn, N. J. Gibson, R. Fallon, A. Hamilton and T. Brown, *Nucleic Acids Res.*, 1996, **24**, 3357-3363.
24. S. Hanessian and P. Lavallee, *Can. J. Chem.*, 1975, **53**, 2975-2977.
25. S. Hatakeyama, H. Irie, T. Shintani, Y. Noguchi, H. Yamada and M. Nishizawa, *Tetrahedron*, 1994, **50**, 13369-13376.
26. H. Köster, K. Kulikowski, T. Liese, W. Heikens and V. Kohli, *Tetrahedron*, 1981, **37**, 363-369.
27. H. Schaller, G. Weimann, B. Lerch and H. G. Khorana, *J. Am. Chem. Soc.*, 1963, **85**, 3821-3827.
28. G. Breipohl, D. W. Will, A. Peyman and E. Uhlmann, *Tetrahedron*, 1997, **53**, 14671-14686.
29. D. Musumeci, G. N. Roviello, M. Valente, R. Sapio, C. Pedone and E. M. Bucci, *Peptide Science*, 2004, **76**, 535-542.
30. P. Clivio, D. Guillaume, M.-T. Adeline, J. Hamon, C. Riche and J.-L. Fourrey, *J. Am. Chem. Soc.*, 1998, **120**, 1157-1166.
31. M. Oleszczuk, S. Rodziewicz-Motowidlo and B. Falkiewicz, *Nucleosides Nucleotides & Nucleic Acids*, 2001, **20**, 1399-1402.
32. S.-M. Chen, V. Mohan, J. S. Kiely, M. C. Griffith and R. H. Griffey, *Tetrahedron Lett.*, 1994, **35**, 5105-5108.
33. M. Gaglione, G. Malgieri, S. Pacifico, V. Severino, B. D'Abrosca, L. Russo, A. Fiorentino and A. Messere, *Molecules*, 2013, **18**, 9147-9162.
34. T. A. Plöger and G. von Kiedrowski, *Helv. Chim. Acta*, 2011, **94**, 1952-1980.
35. S. Brown, S. Thomson, J. Veal and D. Davis, *Science*, 1994, **265**, 777-780.
36. W. He, E. Hatcher, A. Balaeff, D. N. Beratan, R. R. Gil, M. Madrid and C. Achim, *J. Am. Chem. Soc.*, 2008, **130**, 13264-13273.
37. M. Rance, O. W. Sørensen, G. Bodenhausen, G. Wagner, R. R. Ernst and K. Wüthrich, *Biochem. Biophys. Res. Commun.*, 1983, **117**, 479-485.
38. A. Bax and L. Lerner, *Science*, 1986, **232**, 960-967.
39. A. Bax and S. Subramanian, *Journal of Magnetic Resonance*, 1986, **67**, 565-569.
40. G. Bodenhausen and D. J. Ruben, *Chem. Phys. Lett.*, 1980, **69**, 185-189.
41. A. A. Bothner-By, R. L. Stephens, J. Lee, C. D. Warren and R. W. Jeanloz, *J. Am. Chem. Soc.*, 1984, **106**, 811-813.
42. C. Avitabile, L. Moggio, G. Malgieri, D. Capasso, S. Di Gaetano, M. Saviano, C. Pedone and A. Romanelli, *PLoS ONE*, 2012, **7**, e35774.
43. L. Betts, J. A. Josey, J. M. Veal and S. R. Jordan, *Science*, 1995, 1838-1841.
44. O. Almarsson, T. C. Bruice, J. Kerr and D. N. Zuckermann, *PNAS*, 1993, **90**, 7518-7522.
45. C. M. Topham and J. C. Smith, *Biophys. J.*, 2007, **92**, 769-786.
46. S. Sen and L. Nilsson, *J. Am. Chem. Soc.*, 2001, **123**, 7414-7422.
47. C. Avitabile, L. Moggio, G. Malgieri, D. Capasso, S. Di Gaetano, M. Saviano, C. Pedone and A. Romanelli, *PLoS One*, 2012, **7**, e35774.
48. D. S. Wishart, C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley and B. D. Sykes, *J. Biomol. NMR*, 1995, **6**, 135-140.
49. W. L. Jorgensen, J. Chandrasekhar, J. Madura and M. L. Klein, *J. Chem. Phys.*, 1983, 926-935.
50. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
51. A. Jakalian, D. B. Jack and C. I. Bayly, *J. Comput. Chem.*, 2002, **23**, 1623-1641.
52. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wires Comput Mol Sci*, 2013, **3**, 198-210.
53. S. Le Grand, A. W. Gotz and R. C. Walker, *Comput. Phys. Commun.*, 2013, **184**, 374-380.
54. T. E. Cheatham, J. L. Miller, T. Fox, T. A. Darden and P. A. Kollman, *J. Am. Chem. Soc.*, 1994, **117**, 4193-4194.
55. S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952-962.
56. K. A. Feenstra, B. Hess and H. J. C. Berendsen, *J. Comput. Chem.*, 1999, **20**, 786-798.
57. D. R. Roe and T. E. Cheatham, *J. Chem. Theor. Comput.*, 2013, **9**, 3084-3095.