

DOI: 10.1002/cmdc.201200536

# Artemisinin–Polypyrrole Conjugates: Synthesis, DNA Binding Studies and Preliminary Antiproliferative Evaluation

Louise La Pensée,<sup>[a]</sup> Sunil Sabbani,<sup>[a]</sup> Raman Sharma,<sup>[a]</sup> Inder Bhamra,<sup>[a]</sup> Emma Shore,<sup>[a]</sup> Amy E. Chadwick,<sup>[b]</sup> Neil G. Berry,<sup>[a]</sup> James Firman,<sup>[b]</sup> Nuna C. Araujo,<sup>[e]</sup> Lília Cabral,<sup>[e]</sup> Maria L. S. Cristiano,<sup>[e]</sup> Cerys Bateman,<sup>[b]</sup> Omar Janneh,<sup>[c]</sup> Adelina Gavrilă,<sup>[c]</sup> Yi Hang Wu,<sup>[c]</sup> Afthab Hussain,<sup>[c]</sup> Stephen A. Ward,<sup>[d]</sup> Paul A. Stocks,<sup>[d]</sup> Rick Cosstick,<sup>[a]</sup> and Paul M. O'Neill\*<sup>[a]</sup>

Artemisinin-based combination therapies (ACTs) are currently the recommended treatment for uncomplicated and severe cases of malaria.<sup>[1]</sup> Additionally, artemisinins, as well as a number of other sesquiterpene lactones (SLs), are currently in phase I–II clinical trials against breast, colorectal and non-small-cell lung cancers.<sup>[2]</sup> As outlined by the iron-dependent activation hypothesis,<sup>[3]</sup> the activity of artemisinin (ART) is dependent on the endoperoxide bridge.<sup>[4]</sup> The peroxide is cleaved by endogenous sources of Fe<sup>II</sup> to generate highly reactive carbon-centred radicals (CCRs), which are believed to react with critical cellular targets.<sup>[3]</sup> ART demonstrates selectivity towards rapidly proliferating cancer cell lines that possess a high intracellular iron content required to sustain their characteristic high rates of multiplication.<sup>[5]</sup> Iron activation links this particular potency of ART towards rapidly proliferating cancer cell lines; differentiation between healthy and cancerous cells by variation of iron concentration provides a strategy for selective cytotoxicity by ART and its derivatives.<sup>[4]</sup> The mechanism by which ART exerts its cytotoxic activity still remains elusive. ART acts by disruption of proliferation,<sup>[6,7]</sup> oxidative stress,<sup>[8]</sup> anti-angiogenesis,<sup>[9]</sup> NF- $\kappa$ B signalling,<sup>[10]</sup> apoptosis<sup>[4]</sup> and interfering with iron uptake and metabolism.<sup>[6]</sup> ART also induces DNA breakage,<sup>[11]</sup> and it has been reported that artesunate-mediated DNA damage contributes to its therapeutic efficacy.<sup>[12]</sup>

Natural antibiotics like netropsin (NET) and distamycin (DST) containing amide-linked *N*-methyl pyrrole units, show high affinity for the minor groove of DNA at adenine and thymine (AT) tracts.<sup>[13]</sup> Within the context of the hybrid approach, polypyrrole derivatives of NET and DST have been exploited as DNA-targeted delivery agents, such that conjugation to cytotoxic moieties (e.g., alkylators, intercalators and free radical generators) has provided hybrids with enhanced cytotoxic activity.<sup>[14]</sup>

Previously, we explored the antitumour and antimalarial activity of ART–acridine hybrids,<sup>[15]</sup> resulting in a two- to four-fold increase in activity against leukaemia and breast cancer cell lines compared with dihydroartemisinin (DHA) alone. These results encouraged us to prepare ART hybrids with di- and tripyrrole polyamides as DNA recognition elements. The effect that polypyrrole chains have on enhancing DNA cleavage potency, promoting cellular uptake,<sup>[16]</sup> and their use in the development of novel DNA-directed anticancer treatments,<sup>[17]</sup> make polypyrrole minor groove binders good candidates for enhancing the antiproliferative effects of artemisinin derivatives. We hypothesised that tethering the minor groove binding motif to ART might lead to enhanced cytotoxicity by directing the endoperoxide moiety to DNA. Studying the DNA binding affinity, cytotoxicity and the mechanism of cell death caused by the ART portion would also give insight into the relationship between the binding affinity and potential for DNA damage by this novel class of DNA-directed drugs.

In this paper, we describe the synthesis, DNA binding/modelling study, and preliminary biological evaluation of a series of hybrid ART–polypyrrole minor groove binders (1–7). The conjugates examined vary in the length of the pyrrole chain (2 or 3 pyrrole units), the nature of the linker (flexible or rigid), and length of the linker (oligomethylene ( $n > 1$ ); R = *p*-aryl, *p*-benzyl).<sup>[18]</sup>

Work within our group has developed efficient methods for modifying ART, allowing the introduction of synthetically useful functional groups at the C-10 position of the artemisinin framework. Using this experience, we selected and synthesised alkoxy-,<sup>[19]</sup> phenoxy-<sup>[20]</sup> and carba-<sup>[21]</sup> linked ART carboxylic acids to undergo amide coupling with the amine terminus of polypyrrole chains.

Treatment of DHA (**8**) with hydroxy esters **9a–d** provided the required phenoxy or alkoxy esters **10a–d**. Mild hydrolysis and neutralising work-up provided carboxylic acid intermedi-

[a] Dr. L. La Pensée,<sup>+</sup> Dr. S. Sabbani,<sup>+</sup> Dr. R. Sharma, Dr. I. Bhamra, E. Shore, Dr. N. G. Berry, Prof. R. Cosstick, Prof. P. M. O'Neill  
Department of Chemistry, University of Liverpool  
Crown Street, Liverpool, L69 7ZD (UK)  
E-mail: p.m.oneill1@liv.ac.uk

[b] Dr. A. E. Chadwick, J. Firman, C. Bateman  
MRC Centre for Drug Safety Science  
Department of Pharmacology & Therapeutics, University of Liverpool  
Sherrington Buildings, Ashton Street, Liverpool, L69 3GE (UK)

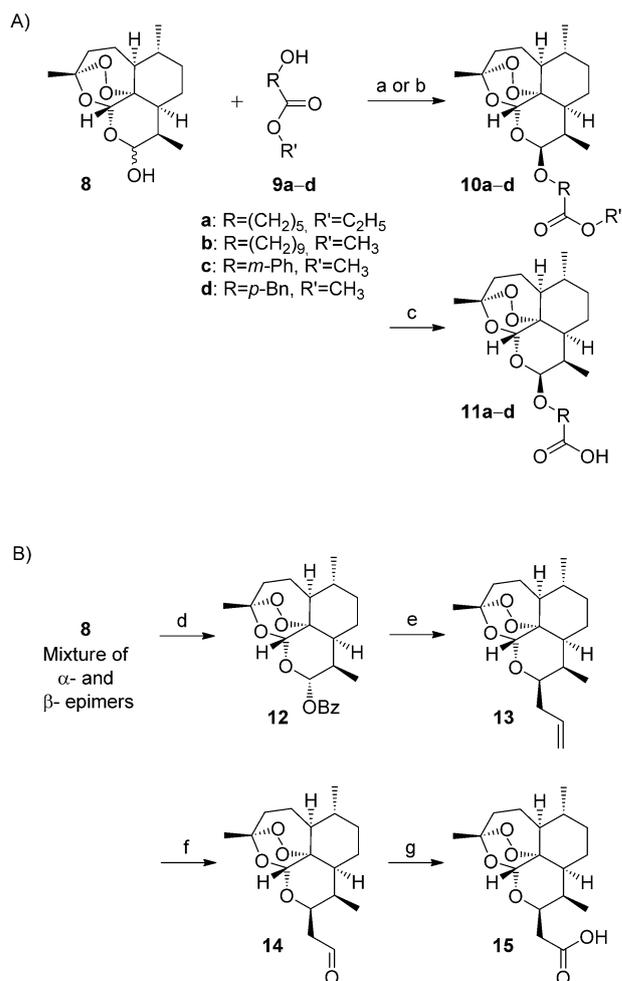
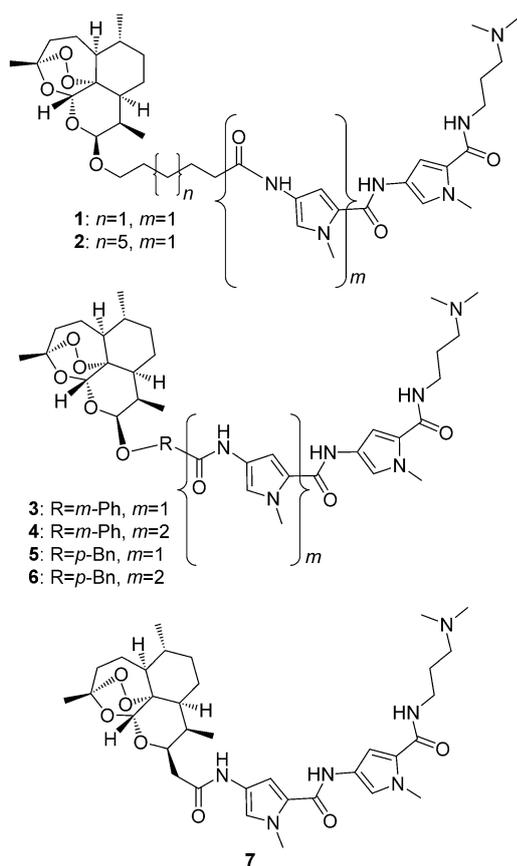
[c] Dr. O. Janneh, A. Gavrilă, Y. H. Wu, A. Hussain  
Department of Biomolecular & Sports Sciences, Coventry University  
Priory Street, Coventry, CV1 5FB (UK)

[d] Prof. S. A. Ward, Dr. P. A. Stocks  
Liverpool School of Tropical Medicine  
Pembroke Place, Liverpool, L3 5QA (UK)

[e] Dr. N. C. Araujo, L. Cabral, Prof. M. L. S. Cristiano  
Centre of Marine Sciences (CCMAR), University of the Algarve  
Campus de Gambelas, 8005-139, Faro (Portugal)

[†] These authors contributed equally to this work.

 Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cmdc.201200536>.



ates **11a–d** (Scheme 1A). Carba-linked intermediate **15** was synthesised in four steps as previously described (Scheme 1B).<sup>[15,21]</sup>

*N*-Methylpyrrole **16** was treated with trichloroacetyl chloride to give **17**, which was then nitrated to give minor groove binder building block **18**.<sup>[22]</sup> To synthesise the terminal unit of the carboxamide chain (**19**), compound **18** was reacted with 3-dimethylaminopropyl amine (Scheme 2A). This terminal group is often selected as a DNA recognition moiety as it allows a positive charge to be distributed across the dimethylamine terminus allowing electrostatic interactions with DNA. The di- and tripyrrole units were then built from the terminal unit (**19**) by reduction of the nitro group to the reactive amine and subsequent reaction with **18** (Scheme 2B).

A number of amide coupling conditions were trialed and found that *N,N,N,N*-tetramethyl-*O*-(1*H*-benzotriazol-1-yl)uronium hexafluorophosphate (HBTU) and ethyldiisopropylamine (EDIPA) were the most suitable carboxylic acid activating agents. The nitro reduction of polypyrrole units **20** and **21** gave reactive amines **20a** and **21a**, which were then coupled overnight with activated forms of carboxylic acids **11a–d** and **15** to give target ART–polyamide conjugates **1–7** (Scheme 3).

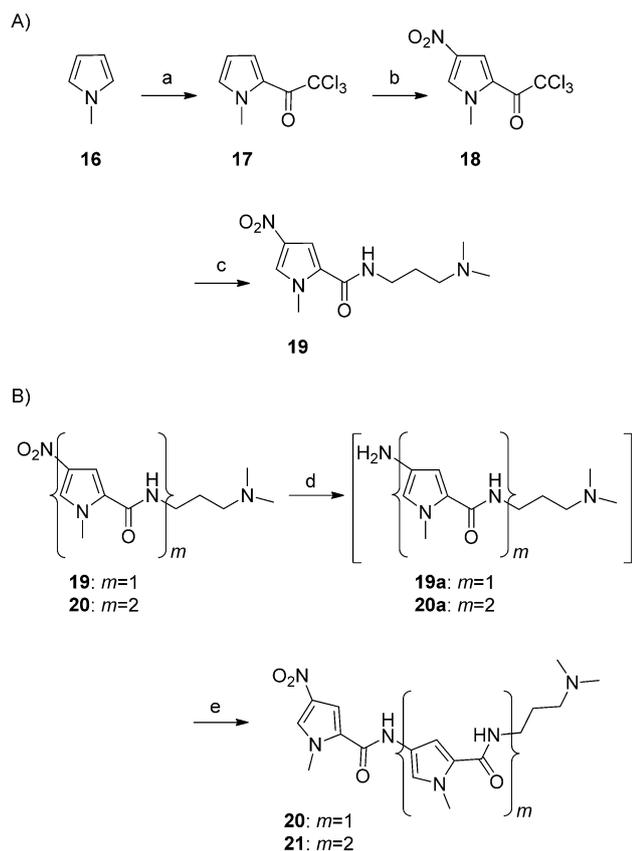
Thermal denaturation of the self-complementary d(CGCGAATTCGCG)<sup>[23]</sup> duplex in the presence of **1–7** and NET was carried out to ascertain the effect of the ART moiety on the DNA binding affinity of the polypyrrole portions (Table 1). The difference in the thermal denaturation temperatures ( $\Delta T_m$ )

**Scheme 1.** Synthesis of A) carboxylic acid intermediates **11a–d** and B) carba-linked intermediate **15**. *Reagents and conditions:* a)  $\text{BF}_3 \cdot \text{Et}_2\text{O}$ ,  $\text{Et}_2\text{O}$ , RT, overnight, 40–50%; b) TMSOTf,  $\text{AgClO}_4$ ,  $\text{NEt}_3$ ,  $\text{CH}_2\text{Cl}_2$ ,  $-78^\circ\text{C}$ , 3 h, 50–80%; c) 5% KOH in MeOH, RT, 4 days; or 2.5% KOH in MeOH/ $\text{H}_2\text{O}$  (1:1), 3–4 h, 75–98%; d) BzCl, Py,  $\text{CH}_2\text{Cl}_2$ ,  $0^\circ\text{C}$ , 16 h, 91%; e) Allyltrimethyl silane,  $\text{ZnCl}_2$ , sieves (4 Å),  $\text{C}_2\text{H}_2\text{Cl}_2$ ,  $0^\circ\text{C}$ , 3 h, 83%; f) 1.  $\text{O}_3$ , MeOH,  $-78^\circ\text{C}$ , 1 h; 2.  $\text{PPh}_3$ , MeOH,  $-78^\circ\text{C}$  → RT, 18 h, 69%; g)  $\text{NaClO}_2$ , 2-methyl-2-butene,  $\text{NaH}_2\text{PO}_4$ ,  $t\text{BuOH}/\text{H}_2\text{O}$  (5:1), RT, 2 h, 62%.

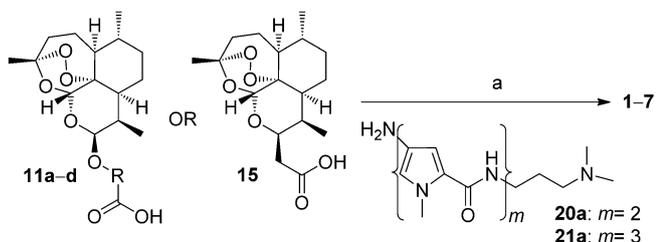
**Table 1.**  $\Delta T_m$  values of compounds **1–8** with netropsin (NET) and dipyrrole **20** as controls.

Compd	$\Delta T_m$ <sup>[a]</sup> [ $^\circ\text{C}$ ]	Compd	$\Delta T_m$ <sup>[a]</sup> [ $^\circ\text{C}$ ]
NET	$9.2 \pm 0.0$	<b>4</b>	$4.8 \pm 0.5$
<b>20</b>	$5.1 \pm 0.5$	<b>5</b>	$3.0 \pm 1.0$
<b>1</b>	$3.7 \pm 0.1$	<b>6</b>	$3.5 \pm 0.2$
<b>2</b>	$4.7 \pm 0.1$	<b>7</b>	$3.4 \pm 0.3$
<b>3</b>	$4.7 \pm 0.4$	<b>8</b>	$3.9 \pm 0.3$

[a]  $\Delta T_m = T_m(\text{duplex} + \text{drug}) - T_m(\text{duplex})$ ; data represent the mean  $\pm$  standard deviations (SD) of three experiments. Experiments were carried out at a duplex concentration of  $1 \mu\text{M}$  with a duplex/drug ratio of 1:1 in  $\text{H}_2\text{O}$  buffered with 1 M NaCl, 0.1 M  $\text{MgCl}_2$ , 10 mM Tris-HCl (pH 7). The absorbance was measured at 260 nm every  $0.5^\circ\text{C}$  whilst heating from 18– $90^\circ\text{C}$  at increments of  $0.5^\circ\text{C min}^{-1}$ . Further experimental details are given in the Supporting Information.



**Scheme 2.** Synthesis of polypyrrole minor groove binding components. *Reagents and conditions:* a)  $\text{Cl}_3\text{CCOCl}$ ,  $\text{CH}_2\text{Cl}_2$ , RT, 3 h, 90%; b) 70%  $\text{HNO}_3$ ,  $\text{Ac}_2\text{O}$ ,  $-40^\circ\text{C}$ , 0.5 h, 45%; c) 3-Dimethylaminopropylamine, THF,  $0^\circ\text{C} \rightarrow \text{RT}$ , 1 h, 86%; d)  $\text{H}_2$ , 1 atm,  $\text{PtO}_2$  (or 10% Pd/C), MeOH, 16 h (or 3–4 h); e) 18, DMF,  $0^\circ\text{C} \rightarrow \text{RT}$ , 1 h, 20: 60%, 21: 45% (two steps).



**Scheme 3.** Synthesis of conjugates 1–7. *Reagents and conditions:* a) 1. HBTU, EDIPA, DMF,  $0^\circ\text{C}$ , 0.5 h; 2. amine 20a or 21a, DMF, RT, overnight, 28–71%.

of the DNA duplex with and without the drug were measured as described in Table 1.

All the conjugates exhibited a stabilising effect on the duplex, and this observation demonstrates the suitability of the polyamide unit as a DNA delivery moiety. Conjugation with ART had some detrimental effects on binding when compared with both NET and 20; however, affinity for the duplex by the conjugates was evident, with 2, 3 and 4 demonstrating the greatest duplex stabilisation. Structurally, conjugate 2 has the longest and most flexible linker giving the minor groove binder portion “freedom” to bind closely into the DNA minor

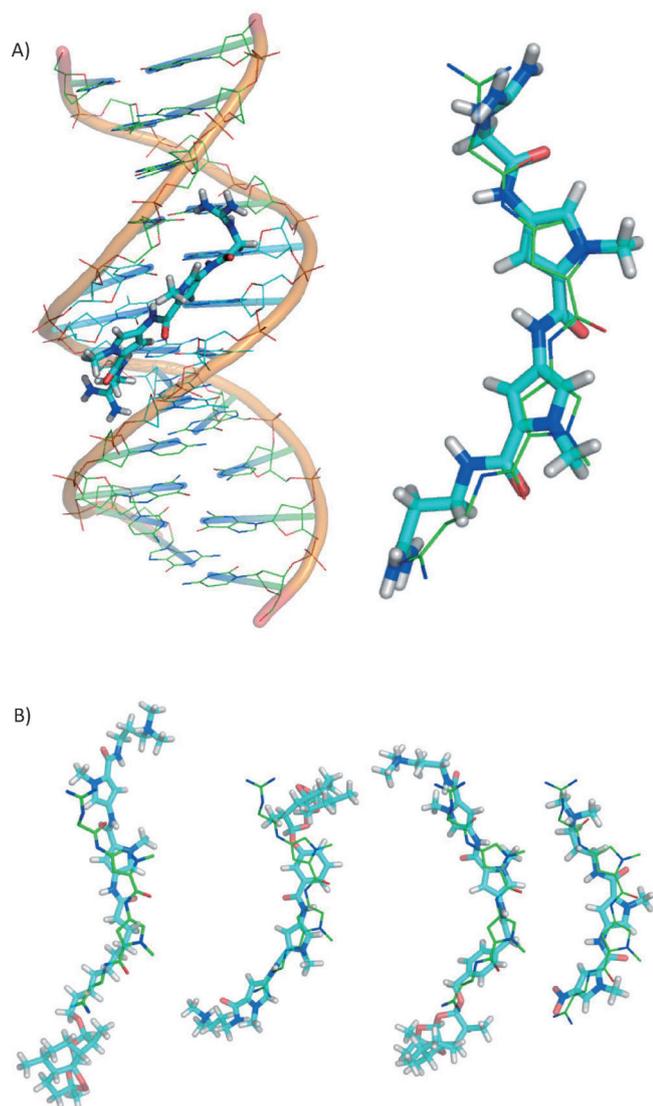
groove; this binding is supported by modelling studies conducted herein (see Supporting Information).

Molecular modelling was performed to rationalise the relative binding affinities of NET and compounds 2, 3, 5 and 20 to the minor groove of the d(CGCGAATTCGCG) duplex. Molecular docking was initially employed to predict binding modes, and then more rigorous molecular dynamics–molecular mechanics/Poisson–Boltzmann surface area (MM/PBSA) simulations were performed. These ligands were chosen on the basis that they span the range of experimentally determined  $\Delta T_m$  values. Interestingly, compounds 3 and 5 varied only in the nature of the linker from the ART moiety to the dipyrrole head group, yet their binding affinities and  $\Delta T_m$  values were markedly different ( $\Delta T_m = 4.7^\circ\text{C}$  for 3 cf.  $\Delta T_m = 3.0^\circ\text{C}$  for 5). The protonation state of the ligands was adjusted so that it was consistent with that found at the physiologically relevant pH 7.4. Full details of how the ligands were prepared for docking and molecular dynamics simulation are discussed in the Computational Methods part of the Experimental Section.

The X-ray crystal structure of d(CGCGAATTCGCG) duplex in complex with NET (PDB code: 1D86)<sup>[24]</sup> was used as the basis for all structure-based modelling work. This 2.2 Å-resolution structure represents the only example of the structural characterisation of an unmodified d(CGCGAATTCGCG)–NET complex currently described in the literature. For further information on the binding modes of NET characterised in co-crystals with various B-DNA duplexes, see the Supporting Information.

The molecular docking program GOLD<sup>[25]</sup> was employed to initially predict the binding pose and strength of interaction of the ligands in complex with the dodecameric DNA duplex. GOLD has been shown to accurately reproduce the crystallographically determined binding poses of minor groove binders to DNA, and a diverse range of ligands to RNA targets, ranking amongst the top methods in each study.<sup>[26,27]</sup> Previous studies have shown that the binding pose of NET in the crystal structure PDB code 1D86 can be accurately reproduced by using the GOLDScore scoring function<sup>[25,26]</sup> within GOLD 3.2.<sup>[26]</sup> Docking validation was repeated for the current version (GOLD 5.0.1) used in this study. Full computational details of all docking calculations are discussed in the Computational Methods part of the Experimental Section. Figure 1A shows that GOLD 5.0.1 accurately reproduces the native binding pose of NET in the minor groove of d(CGCGAATTCGCG) in the co-crystal structure (PDB code: 1D86). The root-mean-square (RMS) deviation from the experimental binding pose for the top three scoring poses as predicted by the GOLDScore scoring function was found to be within 1.1 Å, which is concordant with the performance of GOLD 3.2 for this system.<sup>[26]</sup> This validation gave further assurance that the binding poses of compounds 2, 3, 5 and 20, for which crystal structures are not determined, can be accurately predicted by GOLDScore as implemented in GOLD 5.0.1.

After removal of the native ligand, NET, from the crystal structure PDB code 1D86, compounds 2, 3, 5 and 20 were docked into the minor groove, and the top three scoring poses were examined for each compound. In all cases, at least two out of the three top scoring poses were in agreement



**Figure 1.** A) Best scoring docked pose of netropsin (NET) as predicted by GOLD 5.0.1 in the context of d(CGCGAATTCGCG) DNA (PDB code: 1D86<sup>[24]</sup>). The DNA section shown in blue line representation indicates the AATT recognition region. Also shown is a close up of the best docked pose (cyan) in comparison to native crystallographic binding mode (green). B) The best docked pose (cyan) for compounds **2**, **3**, **5** and **20** (left to right) in comparison with the native crystallographic binding mode of NET (green). Images created using PyMol.<sup>[28]</sup>

with the top scoring pose. Figure 1 shows the position of top poses for compounds **2**, **3**, **5** and **20** relative to the native binding pose of NET with d(CGCGAATTCGCG). From Figure 1A, it can be seen that all compounds are predicted to adopt a crescent shape and display a coarse overlap with the native binding pose of NET. Compounds **2**, **5** and **20** are predicted to bind with their propyl ammonium groups pointing towards the guanidinium terminus of NET and the ART groups towards the propyl amidinium of NET.

Compound **3**, however, is predicted to bind in the opposite configuration with the ART moiety being positioned towards the guanidinium group of NET. This might be of limited consequence due to the palindrome nature of the

d(CGCGAATTCGCG) DNA duplex. The amide and pyrrole groups that form the core of all the molecules are generally shifted relative to their position in the native (class II) NET binding pose seen in crystal structure PDB code 1D86.

From Figure 2, it can be seen that each ligand is predicted to interact with the minor groove of the d(CGCGAATTCGCG) duplex in a unique manner. Compound **2** is predicted to bind to the length of the minor groove with the pyrrole–amide motif sitting more towards the 5'-(C1) end of strand 1 (C1→G12), much “higher” than those in the native NET pose (Figure 2A). Compound **3** is predicted to bind lower in the minor groove, with the pyrrole–amide motif located much more towards 3' (strand 1) end of AATT region than found in the native binding pose of NET (Figure 2B). Furthermore, conjugate **3** is predicted to bind lower than NET in a class I complex, which is centrally located, as exemplified by the crystal structure of NET in complex with the d(CGCGAATTC<sup>5Br</sup>GCG) duplex (PDB code: 6BNA; see Figure S1 in the Supporting Information).<sup>[29]</sup> Compound **5** is predicted to be located in a similar position to **2**, with the pyrrole–amide motif again being located much “higher” than in the native NET pose seen in the crystal structure PDB code 1D86 (Figure 2C). Compound **20** is predicted to be located centrally in AATT region of the minor groove, in a very similar position to the class I binding mode of NET found in 6BNA (Figure 2D).<sup>[29]</sup> This similarity is exemplified by the significant overlap of the pyrrole–amide motifs in class I NET and **20** (see Figure S2 in the Supporting Information). Full details of each binding modes are reported in the Computational Methods section of the Supporting Information.

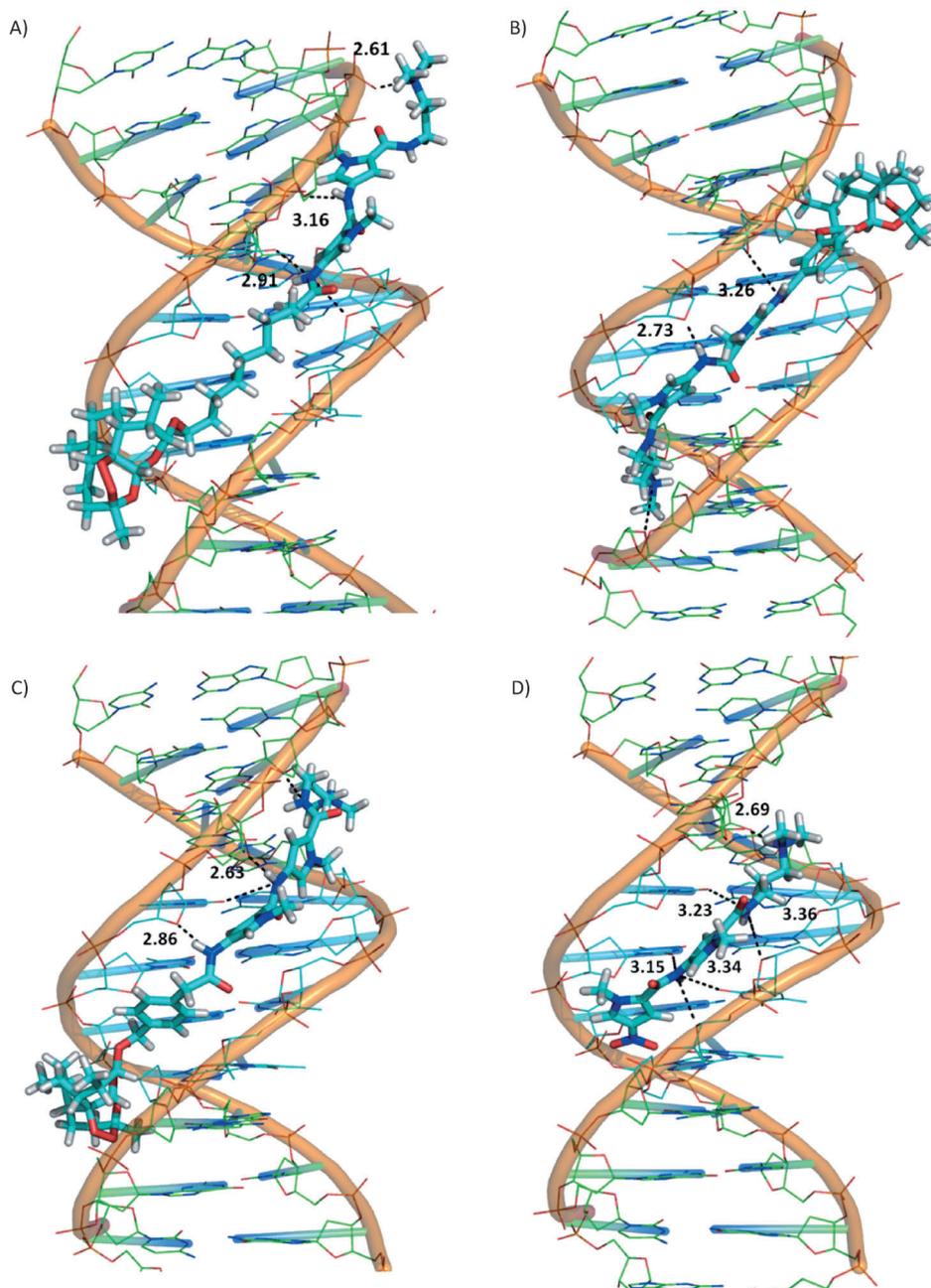
The GOLDScore for compounds **2**, **3**, **5** and **20** was found to correlate poorly with the  $\Delta T_m$  values ( $R^2=0.1$ ; Table 2). The binding strength or potency of a ligand can be normalised by

**Table 2.** Correlation of  $\Delta T_m$  values and GOLDScore results for ligands docked into the minor groove of d(CGCGAATTCGCG) DNA duplex (PDB code: 1D86<sup>[24]</sup>).<sup>[a]</sup>

Compd	$\Delta T_m$ [°C]	$\Delta T_m$ (HAC) [°C]	HAC	Goldscore	LigEff (HAC)
NET	9.20	0.30	31	88.64	2.86
<b>20</b>	5.10	0.19	27	77.16	2.86
<b>2</b>	4.70	0.08	56	96.95	1.73
<b>3</b>	4.70	0.09	53	93.58	1.77
<b>5</b>	3.00	0.05	55	104.72	1.90

[a]  $\Delta T_m$ /HAC = heavy atom count (HAC)-normalised  $\Delta T_m$ .

size of molecule or other molecular properties to give ligand efficiency indices; this can help to mitigate against size effects. Efficiency indices have been shown to be particularly useful in lead identification of the drug discovery process.<sup>[30–33]</sup> The concept of calculated ligand efficiency has been successfully employed in structure-based virtual screening, showing improved correlation against experimental values.<sup>[34–36]</sup> In the present docking study, the GOLDScore was found to increase with size of molecule and, in an effort to quantify this possible size effect, the heavy atom count (HAC) was calculated for each of



**Figure 2.** The best docked pose for ligands **2**, **3**, **5** and **20** (A–D) in the context of d(CGCGAATTCGCG) duplex DNA (PDB code: 1D86<sup>[24]</sup>). The DNA section shown in blue line representation indicates the AATT recognition region. The docked binding poses are shown in stick representation, with nitrogen atoms in dark blue and carbon atoms in cyan. For the ligand and duplex H-bond donors and acceptors, interatomic distances within 3.5 Å are shown as dashed lines, with those that are concordant with the directionality of a H-bond being further annotated with the interatomic distances (Å). Images created using PyMol.<sup>[28]</sup>

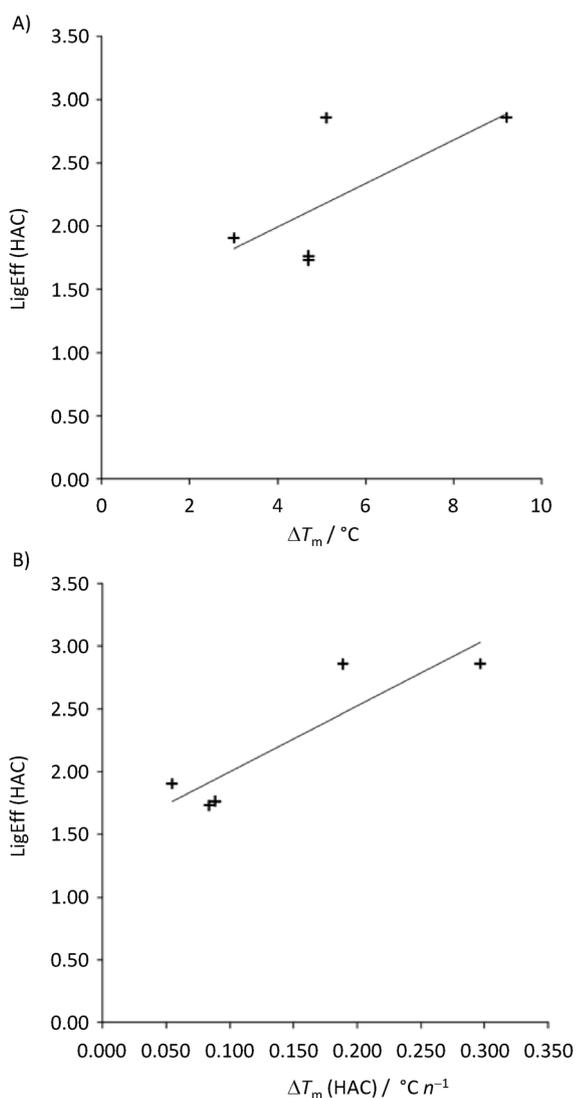
the ligands. The HAC values were found to positively correlate with the GOLDScore ( $R^2=0.43$ ).

GOLDScore normalised by HAC, LigEff(HAC), was calculated and found to correlate moderately ( $R^2=0.46$ ) with  $\Delta T_m$  values (Figure 3A). Furthermore, LigEff(HAC) was found to correlate better ( $R^2=0.81$ ) with HAC-normalised  $\Delta T_m$  values (Figure 3B).

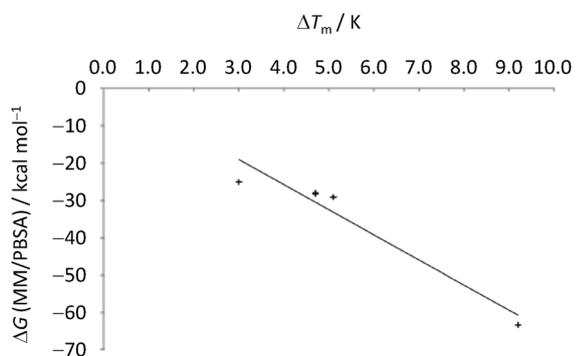
Molecular dynamics simulations and MM/PBSA calculations were performed for NET and compounds **2**, **3**, **5** and **20** in

order to assess their predicted binding modes and energies of interaction. Simulations were performed as specified in the Experimental Section. In particular, starting configurations for each of the ligand–d(CGCGAATTCGCG) complexes were taken as the best scoring docking poses for each ligand. Simulations were performed using explicit solvation and periodic boundary conditions and were followed by single trajectory MM/PBSA calculations, as recommended in the literature for accurate calculation of interaction energies of DNA minor groove binders.<sup>[37]</sup> Analysis of the representative binding poses for NET shows that NET has shifted up by half a base unit to accommodate bifurcated H-bonds as supported by NMR studies.<sup>[38]</sup> In general, the strongest binders (NET and **20**) change least both in absolute terms from the starting structure and fluctuation throughout the simulation. Ligands **2**, **3** and **5** are found to deviate more from their minimised starting configurations and fluctuate more substantially throughout the simulations (even after a substantial 16 ns equilibration phase). Fluctuations in the DNA duplexes for each simulation are less pronounced, although NET and **20** are still found to distort their DNA duplexes to a lesser degree and fluctuate less throughout the simulations. A detailed analysis of the simulation trajectories and discussion of the prevalent binding modes for each ligand–DNA duplex complex is discussed in the Supporting Information.

Figure 4 shows a plot of the calculated MM/PBSA interaction free energies against the experimentally determined  $\Delta T_m$  values. It can be seen that there is an excellent correlation between the calculated and experimental values, with a clear differentiation achieved between the predicted binding energies of NET and **20**. Furthermore, the  $\Delta G(\text{MM/PBSA})$  value for NET is in agreement with that found in a previous MM/PBSA study.<sup>[26]</sup>  $\Delta G(\text{MM/PBSA})$  of interaction is itself constructed from individual energies of com-



**Figure 3.** A) Heavy atom count (HAC)-normalised GOLDscore ligand efficiency (LigEff (HAC)) against  $\Delta T_m$  ( $R^2=0.46$ ). B) LigEff (HAC) against HAC-normalised  $\Delta T_m$  ( $\Delta T_m$  (HAC)), where  $n$  is the number of HAC ( $R^2=0.81$ ). GOLDscore scoring function values were calculated for compounds **2**, **3**, **5**, **20** and NET when docked into the minor groove of d(CGCGAATTCGCG) duplex DNA (PDB code: 1D86<sup>[24]</sup>).



**Figure 4.**  $\Delta G$  (MM/PBSA) values against experimental  $\Delta T_m$  ( $R^2=0.93$ ). Each  $\Delta G$  value was calculated for unrelated snapshots from 20 ns of production simulation.

plexation for van der Waals interactions ( $\Delta E_{vdW}$ ), electrostatic interactions ( $\Delta E_{Elec}$ ) and electrostatic and nonpolar solvation terms ( $\Delta E_{Solv(PB)}$  and  $\Delta E_{Solv-Cav}$  respectively).

Table 3 shows a decomposition of the contributions of the individual energetic terms to  $\Delta G$ (MM/PBSA) of binding for each ligand, and correlation of these individual contributions

Compd	$\Delta T_m$ [°C]	$\Delta G_{(MM/PBSA)}$	$\Delta E_{vdW}$	$\Delta E_{Elec}$ [kcal mol <sup>-1</sup> ]	$\Delta E_{Solv}$ (PB)	$\Delta E_{Solv-Cav}$
NET	9.20	-63.38	-69.28	-1168.12	1178.50	-4.47
<b>20</b>	5.10	-29.10	-47.56	-548.10	569.82	-3.26
<b>2</b>	4.70	-28.30	-50.55	-460.35	486.40	-3.81
<b>3</b>	4.70	-27.88	-50.91	-381.67	408.51	-3.81
<b>5</b>	3.00	-25.06	-63.19	-514.65	557.90	-5.12
<i>R</i>	-	-0.97	-0.50	-0.90	0.89	0.06
<i>R</i> <sup>2</sup>	-	0.93	0.25	0.82	0.80	0.00

[a] *R*: correlation coefficient; *R*<sup>2</sup>: coefficient of determination.

to the experimental  $\Delta T_m$  values. The electrostatic interaction terms and electrostatic component of solvation terms correlate well with the  $\Delta T_m$  value ( $R^2=0.82$  and  $0.80$ , respectively), although  $\Delta E_{Elec}$  is favourable and whereas  $\Delta E_{Solv(PB)}$  is unfavourable (due to the ligand and DNA duplex losing their individual solvent stabilisation on complexation). In contrast, the van der Waals and nonpolar solvation terms do not correlate with  $\Delta T_m$  ( $R^2=0.25$  and  $0.00$ , respectively). The van der Waals term is, however, significant in predicting the correct strength of binding for **5**, with the large van der Waals energy of interaction for this molecule mitigating against the unfavourable  $\Delta E_{Elec}$  and  $\Delta E_{Solv(PB)}$  values. From the decomposition of the  $\Delta G$ (MM/PBSA), it is apparent that binding is principally driven by differential solvation of both the ligand and DNA duplex on complexation and extent of polar interactions with the DNA duplex; these electrostatic terms only partially accounted for in the docking calculations discussed earlier.

As noted, an additional aim of this work was to determine whether the cytotoxic or antimalarial activity of ART can be augmented by conjugation to a DNA binding element. ART-polyamide hybrids **1–7** were evaluated for their in vitro activity against human leukaemia (HL-60) and colon adenocarcinoma (HT-29) cancer cell lines in a methylthiazol tetrazolium (MTT) assay, with DHA and NET as reference agents (Table 4). Further, the mechanism of cell death caused by conjugates **3** and **5** was examined against HL-60 cells, selecting mitochondrial depolarisation and DNA fragmentation as early and late events in the apoptotic pathway, evaluated using a tetramethylrhodamine ethyl ester (TMRE) assay and flow cytometric analysis after staining with propidium iodide (PI), respectively (for full details, see the Supporting Information).

Against HT-29, compounds **2**, **3** and **5** were more active than DHA and NET (Table 4). Compound **2** expressed potent nanomolar activity versus HL60 cells, in contrast to the other

**Table 4.** Cytotoxic (HL-60 and HT-29 cells) and antimalarial (*P. falciparum* 3D7) activities of compounds 1–8 with netropsin (NET) and dihydroartemisinin (DHA) as controls.<sup>[a]</sup>

Compd	IC <sub>50</sub> <sup>[b]</sup> [μM]		IC <sub>50</sub> <sup>[c]</sup> [nM] (3D7)
	(HL-60)	(HT-29)	
DHA	0.50 ± 0.10	12.92 ± 2.31	2 ± 0.5
NET	> 100	11.00 ± 1.87	ND
1	4.41 ± 0.49	22.87 ± 1.56	65 ± 9
2	0.11 ± 0.03	6.26 ± 0.45	40 ± 6
3	1.19 ± 0.09	7.38 ± 0.52	50 ± 12
4	6.35 ± 2.93	39.94 ± 3.93	126 ± 11
5	1.42 ± 0.66	8.78 ± 1.47	24 ± 8
6	4.60 ± 1.40	58.76 ± 2.77	119 ± 14
7	11.67 ± 4.38	48.01 ± 3.16	120 ± 18

[a] All values are the mean ± standard deviations (SD) of three experiments. [b] Cells were cultured in RPMI 1640 with 10% fetal bovine serum and 1% w/v L-glutamine. Cultures were incubated in humidified air with 5% CO<sub>2</sub> at 37 °C and kept below a density of 1 × 10<sup>6</sup>. Compounds were dissolved in DMSO before being added at various concentrations (0.01–100 μM) to the cells. Plates were incubated for 24 and 48 h. Cell viability was quantified using an MTT assay, and the absorbance of the samples was measured at 570 nm by a plate reader. Results were expressed as a percentage of the vehicle-only control, and IC<sub>50</sub> values were calculated using GraFit (version 4.0). [c] Parasites were maintained in continuous culture using the method of Jensen and Trager.<sup>[39]</sup> Antimalarial activity was assessed with an adaptation of the 48 h sensitivity assay of Desjardins et al.<sup>[40]</sup> using [<sup>3</sup>H]hypoxanthine incorporation as an assessment of parasite growth.

conjugates that were between two- and 20-fold less potent than DHA. It is interesting to note that NET is active against HT-29 but inactive against HL-60 cells. Compounds 3, 5 and DHA were further examined in HL-60 cells to determine the mechanism of cell death. Although these conjugates induced mitochondrial membrane depolarisation, no significant impact on the cellular DNA content was seen. These results are strongly indicative of 3 and 5 inducing cell death by mitochondrial membrane depolarisation and apoptotic death in HL-60 cell lines.

In the case of HT-29 cells, the addition of a minor groove binding element in compounds 2, 3 and 5 enhanced the anti-proliferative activity. The primary mechanism by which these compounds cause cell death in HT-29 would also appear to be by mitochondrial depolarisation and, as such, localisation of the compounds to cellular DNA is not advantageous.<sup>[41]</sup> Further investigations into the mechanism of cell death employed by these compounds in HT-29 cell lines is required to see if their improved activity is a result of a shift in the mechanism of action.

The human malaria parasite *Plasmodium falciparum* possesses AT-rich DNA (>80% cf. to 60% in man).<sup>[42]</sup> We proposed the AT-specific ART conjugates have the potential for enhanced and specific antimalarial activity. The conjugates were evaluated for their in vitro antimalarial activity against the chloroquine-sensitive 3D7 strain of *P. falciparum*, however, none of the conjugates were more active than DHA and were, in some cases, up to 60-fold less active. Nevertheless, the antimalarial activities of all these hybrid compounds 1–7 were in the nanomolar range.

In summary, conjugate 2 emerges as the most active molecule being twofold more active than DHA against HT-29 cells and exhibiting around five times the potency of DHA against HL-60 cells. Conjugate 2 also exhibits one of the strongest affinities for the DNA duplex d(CGCGAATTCGCG). Structurally, these results are not surprising as conjugate 2 has the longest flexible linker, allowing for the greatest steric freedom to both portions of the hybrid. Modelling studies have accurately predicted the experimental  $\Delta T_m$  data, and decomposition of these calculated free energies of binding has shown that the binding affinity of these hybrid compounds depends on multiple factors, in particular electrostatic interactions and differential solvation of the ligand and DNA duplexes on complexation. Moreover, these factors are intrinsically linked to the varying degrees of conformational change the ligands and DNA duplex undergoes on binding. Further mechanistic studies are needed in HT-29 cells to clarify whether ART maintains mitochondrial membrane depolarisation as a main mode of action or whether the presence of the DNA binding structure has in fact enhanced activity by causing a dual mode of action.

## Experimental Section

### Organic synthesis

Full details on the synthesis of intermediate compounds, including characterisation data, general remarks and instrument particulars, can be found in the Supporting Information. 10β-(2-Carboxyethyl)-deoxyartemisinin (13) was synthesised according to Jones and co-workers.<sup>[15]</sup>

**Artemisinin conjugates 1–7:** Nitrodipyrrole 20 or 21 (1.04 equiv with respect to ART carboxylic acid) was reduced according to the procedure described for the synthesis of compounds 20 or 21 from 19 (3–4 h reaction time; see the Supporting Information for details). The crude reaction mixture was filtered through celite and concentrated in vacuo. The residue was redissolved in anhyd DMF (30 mL), and the solution was concentrated to approximately half the original volume in vacuo. Resulting amine 20a or 21a was cooled to 0 °C and stored under an atmosphere of N<sub>2</sub>. In another flask, the appropriate ART-derived carboxylic acid (1.0 equiv) was dissolved in anhyd DMF, and the solution was cooled to 0 °C. This solution was then treated with HBTU (1.1 equiv) and DIPEA (2.0 equiv) and stirred at 0 °C under N<sub>2</sub> for 1.0 h. The ice bath was removed, and this solution containing the activated ART-derived acid was transferred to the 20a or 21a solution in DMF. The resulting reaction mixture was stirred at RT for 23 h. The solvent was removed in vacuo, and the crude material was redissolved in a minimum amount of MeOH, and partitioned between EtOAc (50 mL) and H<sub>2</sub>O (50 mL). The aqueous layer was extracted with EtOAc (3 × 50 mL) and CH<sub>2</sub>Cl<sub>2</sub> (2 × 50 mL). The organic layers were separately washed with saturated NaHCO<sub>3</sub> (1 × 50 mL) and brine (1 × 50 mL), and the combined organic layers were dried (MgSO<sub>4</sub>), filtered and concentrated in vacuo. Purification by flash chromatography (column flush = MeOH/EtOAc (1:1) then MeOH/EtOAc (1:1) + 3% NH<sub>4</sub>OH) gave near-pure material, which was redissolved in CH<sub>2</sub>Cl<sub>2</sub> and filtered to removed dissolved silica gel. Concentration of the filtrate in vacuo gave the desired ART hybrid (1–7).

Compound 1 (187 mg, 42% yield): <sup>1</sup>H NMR (400 MHz, CDCl<sub>3</sub>): δ = 7.77 (m, 2H), 7.62 (s, 1H), 7.15 (d, *J* = 1.7 Hz, 1H), 7.08 (d, *J* = 1.7 Hz, 1H), 6.59 (d, *J* = 1.7 Hz, 1H), 6.45 (d, *J* = 1.7 Hz, 1H), 5.40 (s, 1H),

4.76 (d,  $J=3.2$  Hz, 1H), 3.89 (s, 6H), 3.81 (dt,  $J=9.6, 6.7$  Hz, 1H), 3.46 (dd,  $J=11.6, 5.9$  Hz, 2H), 3.44 (dt,  $J=9.6, 6.4$  Hz, 2H), 2.61 (m, 1H), 2.45 (t,  $J=6.2$  Hz, 2H), 2.40 (m, 1H), 2.33 (m, 2H), 2.28 (s, 6H), 2.05–1.20 (m, 26H), 1.43 (s, 3H), 0.95 (d,  $J=6.2$  Hz, 3H), 0.90 ppm (d,  $J=7.3$  Hz, 3H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=171.3, 162.2, 159.4, 124.3, 123.5, 122.1, 121.6, 119.5, 119.1, 104.5, 104.0, 103.4, 102.3, 88.3, 81.6, 68.8, 59.4, 53.0, 45.9, 45.9, 44.9, 39.8, 37.9, 37.3, 37.0, 37.0, 36.9, 35.1, 31.4, 30.0, 29.8, 29.7, 29.6, 29.6, 26.6, 26.5, 26.3, 26.2, 25.1, 24.9, 20.8, 13.4$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=3310, 2937, 1651, 1099, 944, 873, 826$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{38}\text{H}_{58}\text{N}_6\text{O}_8$ : 727.4394, found: 727.4415.

Compound 2 (153 mg, 34%):  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta=7.78$  (s, 1H), 7.76 (s, 1H), 7.71 (s, 1H), 7.16 (d,  $J=1.7$  Hz, 1H), 7.07 (d,  $J=1.7$  Hz, 1H), 6.63 (d,  $J=1.7$  Hz, 1H), 6.45 (d,  $J=1.7$  Hz, 1H), 5.40 (s, 1H), 4.74 (d,  $J=3.4$  Hz, 1H), 3.89 (s, 6H), 3.74 (1H, dt,  $J=9.7, 6.6$  Hz), 3.44 (m, 2H, ), 3.36 (dt,  $J=9.8, 6.6$  Hz, 1H), 2.60 (m, 1H), 2.43 (t,  $J=6.2$  Hz, 2H), 2.40 (m, 1H), 2.33 (m, 2H), 2.28 (s, 6H), 2.05–1.20 (m, 16H), 1.43 (s, 3H), 0.95 (d,  $J=6.2$  Hz, 3H), 0.88 ppm (d,  $J=7.3$  Hz, 3H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=171.1, 162.2, 159.4, 124.4, 123.5, 122.0, 121.7, 119.6, 119.1, 104.6, 104.1, 103.4, 102.3, 88.4, 81.6, 68.7, 59.4, 52.9, 45.9, 45.9, 44.8, 39.9, 37.9, 37.1, 37.0, 37.0, 36.8, 35.0, 31.3, 29.8, 26.6, 26.3, 26.3, 25.9, 25.1, 24.9, 20.8, 13.4$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=3299, 2940, 1651, 1099, 937, 876, 822$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{42}\text{H}_{66}\text{N}_6\text{O}_8$ : 783.5020, found: 783.5056.

Compound 3 (283 mg, 68%):  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta=8.86$  (br s, 1H), 8.11 (br s, 1H), 7.63 (t,  $J=5.2$  Hz, 1H), 7.57–7.73 (m, 1H), 7.49 (d,  $J=7.7$  Hz, 1H), 7.23–7.31 (m, 1H), 7.20 (d,  $J=1.5$  Hz, 1H), 7.14 (dd,  $J=8.1, 1.6$  Hz, 1H), 6.97 (br s, 1H), 6.66 (s, 1H), 6.54 (d,  $J=1.6$  Hz, 1H), 5.68 (d,  $J=2.8$  Hz, 1H), 5.42 (s, 1H), 3.86 (s, 3H), 3.81 (s, 3H), 3.33–3.45 (m, 2H), 2.71–2.83 (m, 1H), 2.41 (t,  $J=6.6$  Hz, 2H), 2.34 (td,  $J=14.0, 3.2$  Hz, 1H), 2.24 (s, 6H), 1.27 (s, 3H), 1.20–2.00 (m, 11H), 1.01 (d,  $J=7.3$  Hz, 3H), 0.94 (d,  $J=5.3$  Hz, 3H), 0.90–0.97 ppm (m, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=165.2, 161.9, 158.9, 157.2, 135.9, 129.7, 123.5, 123.2, 121.5, 121.2, 120.6, 119.9, 118.6, 114.9, 104.9, 104.4, 103.0, 99.7, 88.3, 80.9, 58.3, 53.4, 45.1, 44.2, 38.7, 37.4, 36.5, 36.4, 36.2, 34.5, 30.8, 26.2, 25.7, 24.6, 24.4, 20.2, 12.7$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=2943, 1646, 1581, 1532, 1435, 1403, 1263, 1227, 1095, 1036, 979, 875$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{39}\text{H}_{53}\text{N}_6\text{O}_8$ : 733.3925, found: 733.3936.

Compound 4 (138 mg, 28%):  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta=9.23$  (br s, 1H), 8.68 (br s, 1H), 8.33 (br s, 1H), 7.56–7.77 (m, 2H), 7.45 (d,  $J=6.9$  Hz, 1H), 7.13–7.30 (m, 5H), 7.07 (d,  $J=7.2$  Hz, 1H), 6.55–6.73 (m, 2H), 5.55–5.97 (m, 1H), 5.39 (s, 1H), 3.76–3.89 (m, 9H), 3.33–3.42 (m, 2H), 2.66–2.85 (m, 1H), 2.39 (t,  $J=6.1$  Hz, 2H), 2.27–2.35 (m, 1H), 2.21 (s, 6H), 1.27 (s, 3H), 1.16–2.06 (m, 11H), 0.99 (d,  $J=7.2$  Hz, 3H), 0.94 (br s, 3H), 0.87–1.14 ppm (m, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=172.8, 161.9, 159.1, 157.0, 135.9, 130.9, 129.7, 123.4, 123.0, 121.8, 121.5, 120.7, 120.7, 120.1, 120.0, 119.1, 118.8, 118.8, 104.5, 103.5, 103.2, 99.5, 88.4, 80.8, 58.1, 53.4, 52.3, 45.0, 44.2, 38.6, 37.4, 36.5, 36.2, 34.4, 30.7, 26.1, 25.6, 24.5, 24.3, 22.5, 20.2, 12.7, -21.1$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=2935, 1647, 1582, 1529, 1434, 1403, 1255, 1096, 1035, 980, 876, 738$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{45}\text{H}_{59}\text{N}_8\text{O}_9$ : 855.4405, found: 855.4395.

Compound 5 (311 mg, 71%):  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta=7.97$  (br s, 1H), 7.93 (br s, 1H), 7.68 (t,  $J=5.0$  Hz, 1H), 7.25–7.29 (m, 4H), 7.15 (d,  $J=1.8$  Hz, 1H), 7.05 (d,  $J=1.8$  Hz, 1H), 6.59 (d,  $J=1.8$  Hz, 1H), 6.50 (d,  $J=1.8$  Hz, 1H), 5.46 (s, 1H), 4.88 (d,  $J=3.0$  Hz, 1H), 4.85 (d,  $J=12.6$  Hz, 1H), 4.49 (d,  $J=12.5$  Hz, 1H), 3.85 (s, 3H), 3.84 (s, 3H), 3.65 (s, 2H), 3.41 (q,  $J=5.9$  Hz, 2H), 2.59–2.71 (m, 1H), 2.43 (t,  $J=6.4$  Hz, 2H), 2.36 (dd,  $J=13.6, 3.9$  Hz, 1H), 2.26 (s, 6H), 2.01–

2.11 (m, 1H), 1.75–1.94 (m, 3H), 1.71 (quin,  $J=6.3$  Hz, 2H), 1.62 (dq,  $J=13.1, 3.0$  Hz, 1H), 1.46–1.56 (m, 2H), 1.44 (s, 3H), 1.19–1.38 (m, 2H), 0.94 (d,  $J=6.0$  Hz, 3H), 0.93 (d,  $J=7.3$  Hz, 3H), 0.89–0.92 ppm (m, 1H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=172.7, 168.6, 161.8, 158.9, 137.4, 133.9, 129.2, 127.7, 123.7, 123.1, 121.4, 121.3, 119.2, 118.7, 104.1, 103.7, 103.2, 101.2, 88.0, 81.1, 69.3, 58.5, 52.5, 45.2, 44.3, 43.4, 39.0, 37.4, 36.5, 36.5, 36.4, 34.5, 30.8, 26.1, 25.9, 24.6, 24.5, 22.5, 20.3, 13.0$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=2943, 1648, 1532, 1437, 1403, 1255, 1140, 1099, 1011, 826, 732$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{41}\text{H}_{57}\text{N}_6\text{O}_8$ : 761.4238, found: 761.4273.

Compound 6 (165 mg, 32%):  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta=8.36$  (br s, 1H), 8.18 (br s, 1H), 8.00 (br s, 1H), 7.67 (br s, 1H), 7.23–7.34 (m, 4H), 7.19 (br s, 1H), 7.11–7.16 (m, 1H), 6.95–7.07 (m, 1H), 6.56–6.69 (m, 3H), 5.46 (s, 1H), 4.76–5.00 (m, 2H), 4.35–4.54 (m, 1H), 3.76–3.90 (m, 9H), 3.58–3.68 (m, 2H), 3.39 (d,  $J=4.3$  Hz, 2H), 2.58–3.20 (m, 2H), 2.40 (t,  $J=6.2$  Hz, 2H), 2.23 (s, 6H), 2.09–2.17 (m, 1H), 1.75–2.04 (m, 3H), 1.56–1.73 (m, 3H), 1.42–1.53 (m, 5H), 1.16–1.36 (m, 2H), 0.80–1.00 ppm (m, 7H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=169.7, 169.0, 161.9, 159.0, 137.3, 134.8, 133.9, 129.2, 128.2, 127.6, 123.6, 123.2, 123.0, 121.6, 121.4, 121.4, 119.5, 119.3, 118.7, 104.2, 103.8, 103.2, 101.2, 88.0, 80.5, 69.3, 58.4, 52.5, 46.8, 45.2, 38.9, 37.4, 36.6, 36.6, 36.5, 36.4, 35.8, 34.5, 33.3, 29.6, 26.1, 26.0, 24.8, 24.6, 20.4, 13.0$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=2923, 1647, 1583, 1531, 1464, 1435, 1402, 1255, 1080, 1004, 775$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{47}\text{H}_{63}\text{N}_8\text{O}_9$ : 883.4718, found: 883.4720.

Compound 7 (159 mg, 35%):  $^1\text{H}$  NMR (400 MHz,  $\text{CDCl}_3$ ):  $\delta=8.91$  (br s, 1H), 8.14 (br s, 1H), 7.62 (t,  $J=5.1$  Hz, 1H), 7.21 (d,  $J=1.8$  Hz, 1H), 6.96 (d,  $J=1.8$  Hz, 1H), 6.70 (d,  $J=1.8$  Hz, 1H), 6.54 (d,  $J=1.9$  Hz, 1H), 5.48 (s, 1H), 4.92 (ddd,  $J=11.2, 6.2, 1.4$  Hz, 1H), 3.89 (s, 3H), 3.84 (s, 3H), 3.43 (q,  $J=5.9$  Hz, 2H), 2.56–2.71 (m, 2H), 2.46 (t,  $J=6.5$  Hz, 2H), 2.44–2.50 (m, 1H), 2.38–2.44 (m, 1H), 2.30 (s, 6H), 1.94–2.09 (m, 2H), 1.73 (quin,  $J=6.3$  Hz, 2H), 1.64–1.84 (m, 3H), 1.32 (s, 3H), 1.21–1.40 (m, 4H), 0.97 (d,  $J=5.4$  Hz, 3H), 0.95–1.00 (m, 1H), 0.90 ppm (d,  $J=7.5$  Hz, 3H);  $^{13}\text{C}$  NMR (100 MHz,  $\text{CDCl}_3$ ):  $\delta=169.2, 161.8, 159.0, 123.5, 123.1, 121.5, 121.5, 118.9, 118.5, 104.2, 103.1, 103.0, 90.2, 80.8, 70.1, 58.3, 51.7, 45.1, 43.4, 38.3, 37.4, 36.5, 36.4, 36.4, 34.2, 34.2, 30.3, 25.9, 25.8, 24.7, 24.7, 19.9, 12.0$  ppm; IR (neat):  $\tilde{\nu}_{\text{max}}=2939, 1647, 1583, 1530, 1437, 1404, 1263, 1207, 1095, 1044, 1013, 814$   $\text{cm}^{-1}$ ; HRMS (ES+):  $m/z$   $[M+H]^+$  calcd for  $\text{C}_{34}\text{H}_{51}\text{N}_6\text{O}_9$ : 655.3819, found: 655.3822.

## Computational methods

**Molecular docking studies:** All molecular docking was performed using the GOLD docking suite.<sup>[25]</sup> Ligands 2, 3, 5, 20 and NET were protonated to physiological pH (pH 7.4) using the “ionize molecule at pH” component within Pipeline Pilot,<sup>[43]</sup> 3D structures were created through minimisation using the Merck molecular force field (MMFF) 94<sup>[44]</sup> within the SPARTAN '08 1.0.0 package.<sup>[45]</sup> Preparation of all the input files and run configuration, including setting up the DNA duplex system, was performed in the Hermes visualise 1.4.1.<sup>[46]</sup> The d(CGCGAATTCGCG) duplex crystal structure (PDB code: 1D86<sup>[24]</sup>) was prepared by removing NET and water molecules to avoid potential interference with the docking. Hydrogen atoms were added to the DNA duplex according to the default heuristics. The binding site was defined as being all residues 6 Å from NET in the 1D86 structure.

A genetic algorithm (GA) employing GOLDScore as the fitness function was used for all docking calculations. Ten independent GA runs were performed for each ligand. Default settings were retained for the “ligand flexibility”, “fitness and search options” and the “GA settings”, however the “allow early termination” setting in

the "fitness and search option" was turned off to search a larger number of ligand conformations. In order to calculate HAC-normalised GOLDscore, the HAC value was calculated using the "element count (advanced)" component with the Pipeline Pilot.<sup>[43]</sup>

**Molecular dynamics (MD) simulations:** Ligand complexes for compounds **2**, **3**, **5**, **20** and NET were simulated at 300 K using MD. The starting configurations for each of the ligand–d(CGCGAATTCGCG) complexes were taken as the best scoring docking pose for that ligand. All MD simulations were carried out using the AMBER 11 MD package.<sup>[47]</sup> All ligands were modelled using the general amber force field (GAFF),<sup>[48]</sup> and the DNA duplex was modelled using the ff99SB force field.<sup>[49]</sup> Atomic charges for all ligands were obtained using electrostatic potential calculations at the Hartree–Fock ab initio level of theory<sup>[50]</sup> using the locally dense 6–31G\* basis set, similar to the previous MD study of NET.<sup>[51]</sup> Electrostatic potential calculations were performed using Gaussian 09 (revision B.01)<sup>[52]</sup> as in a previous MD study of NET.<sup>[51]</sup>

The LEaP module within AmberTools was used to prepare each ligand–DNA duplex system for MD simulation. All ligand–DNA complexes have a large net negative charge despite the ligands being cationic or dicationic at pH 7.4; this is due to the presence of 22PO<sub>4</sub><sup>−</sup> groups in the DNA duplexes. In order to neutralise the ligand–duplex complexes, an appropriate number of Na<sup>+</sup> counter ions was added to each system using TIP3P,<sup>[53]</sup> a water box of 12 Å distance from the edge atom of the DNA–ligand complex was constructed.

To achieve the most stable and representative simulations, previously described protocols were adapted to our specific requirements.<sup>[26,54,55]</sup> Prior to MD simulations, each system was subjected to a staged minimisation using the sander module in AMBER 11. During the first stage, a restrained minimisation was performed in order to relax the water around each ligand–duplex system. Restraints of 500.00 kcalmol<sup>−1</sup>Å<sup>2</sup> were applied to the DNA duplex and ligand. The steepest descent algorithm was run for 4000 steps, followed by the conjugate gradient method for 4000 steps. In the second minimisation stage, the entire system underwent 2000 steps of steepest descent and 3000 steps of conjugate gradient minimisation. For all minimisations and MD simulations, a nonbonded cut-off of 10 Å, periodic boundary conditions, and particle mesh Ewald treatment of long-range electrostatics were used.

For all MD simulations, hydrogen–heavy atom bonds were constrained by the SHAKE algorithm, and temperature was controlled by Langevin dynamics with a collision frequency of 1 ps<sup>−1</sup> and set to 300 K. The leapfrog algorithm was used to propagate the system, with a time step of 0.002 ps (as shown to be appropriate for a system that employs the SHAKE algorithm). A total of 36 ns of simulations was performed for each ligand–DNA duplex complex. Initially each system underwent a heating phase of 100 ps in the canonical ensemble (NVT), thereafter simulations were run in the constant-pressure canonical ensemble (NPT). Each ligand–DNA duplex system was equilibrated for 15.9 ns, and then a production phase of 20 ns was performed.

**Average-linkage clustering of MD trajectories:** The average-linkage clustering algorithm as implemented in "ptraj" with AmberTools was used to find representative binding poses for each ligand–DNA duplex system, as recommended for systems for which the optimal number of clusters is not known.<sup>[56]</sup> The representative binding pose is defined as the structure closest to the centroid of each cluster. In order to find the optimal cluster, a manual scan for the critical distance was employed from 1.0 to 2.4 Å. The critical distance is defined as the maximum intercluster

distance at which clusters are allowed to merge in an agglomerative algorithm. In order to judge the optimal number of clusters, the clustering quality metrics DBI, pSF and SSR/SST were monitored. The final decision on the optimal number of clusters involved simultaneous consideration of the clustering quality metrics and manual inspection of the superposed cluster centroids.

**MM/PBSA calculations:** MM/PBSA calculations were performed using the MM/PBSA.py tool within AmberTools, with default parameters on 1000 snapshots from the production phase of each simulation sampled at 2 ps intervals to ensure that snapshots were unrelated.

## Acknowledgements

The authors would like to thank the British Engineering and Physical Sciences Research Council (EPSRC) for a studentship to LP and the EU FP6 Malaria Initiative ANTIMAL for partial support of this work.

**Keywords:** artemisinin · binding studies · cytotoxicity · DNA · molecular modelling

- [1] *Guidelines for the Treatment of Malaria*, 2nd Edition, World Health Organisation (WHO), 2010; [http://whqlibdoc.who.int/publications/2010/9789241547925\\_eng.pdf](http://whqlibdoc.who.int/publications/2010/9789241547925_eng.pdf).
- [2] A. Ghantous, H. Gali-Muhtasib, H. Vuorela, N. A. Saliba, N. Darwiche, *Drug Discovery Today* 2010, 15, 668–678.
- [3] a) P. M. O'Neill, V. E. Barton, S. A. Ward, *Molecules* 2010, 15, 1705–1721; b) P. A. Stocks, P. G. Bray, V. E. Barton, M. Al-Helal, M. Jones, N. V. Araujo, P. Gibbons, S. A. Ward, R. H. Hughes, G. A. Biagini, J. Davies, R. K. Amewu, A. E. Mercer, G. Ellis, P. M. O'Neill, *Angew. Chem.* 2007, 119, 6394–6399; *Angew. Chem. Int. Ed.* 2007, 46, 6278–6283.
- [4] A. E. Mercer, J. L. Maggs, X. Sun, G. M. Cohen, J. Chadwick, P. M. O'Neill, B. K. Park, *J. Biol. Chem.* 2007, 282, 9372–9382.
- [5] I. Nakase, H. Lai, N. P. Singh, T. Sasaki, *Int. J. Pharm.* 2008, 354, 28–33.
- [6] T. Efferth, *Drug Resist. Updates* 2005, 8, 85–97.
- [7] T. Efferth, *Curr. Drug Targets* 2006, 7, 407–421.
- [8] T. Efferth, F. Oesch, *Biochem. Pharmacol.* 2004, 68, 3–10.
- [9] L. Anfosso, T. Efferth, A. Albini, U. Pfeffer, *Pharmacogenomics J.* 2006, 6, 269–278.
- [10] E. Aldieri, D. Atragine, L. Bergandi, C. Riganti, C. Costamagna, A. Bosia, D. Ghigo, *FEBS Lett.* 2003, 552, 141–144.
- [11] H. H. Cai, P. H. Yang, J. N. Chen, Z. H. Liang, Q. Y. Chen, J. Y. Cai, *Electrochim. Acta* 2009, 54, 3651–3656.
- [12] P. C. H. Li, E. Lam, W. P. Roos, M. Z. Zdzienicka, B. Kaina, T. Efferth, *Cancer Res.* 2008, 68, 4347–4351.
- [13] D. E. Wemmer, P. B. Dervan, *Curr. Opin. Struct. Biol.* 1997, 7, 355–361.
- [14] P. G. Baraldi, D. Preti, F. Fruttarolo, M. A. Tabrizi, R. Romagnoli, *Bioorg. Med. Chem.* 2007, 15, 17–35; L. Banfi, A. Basso, E. Bevilacqua, V. Gandolfo, G. Giannini, G. Guanti, L. Musso, M. Paravidino, R. Riva, *Bioorg. Med. Chem.* 2008, 16, 3501–3518; H. Tanaka, Y. Tanaka, M. Minoshima, S. Yamaguchi, S. Fuse, T. Doi, S. Kawauchi, H. Sugiyama, T. Takahashi, *Chem. Commun.* 2010, 46, 5942–5944.
- [15] M. Jones, A. E. Mercer, P. A. Stocks, L. J. I. La Pensée, R. Cosstick, B. K. Park, M. E. Kennedy, I. Piantanida, S. A. Ward, J. Davies, P. G. Bray, S. L. Rawe, J. Baird, T. Charidza, O. Jannah, P. M. O'Neill, *Bioorg. Med. Chem. Lett.* 2009, 19, 2033–2037.
- [16] S. Nishijima, K. Shinohara, T. Bando, M. Minoshima, G. Kashiwazaki, H. Sugiyama, *Bioorg. Med. Chem.* 2010, 18, 978–983.
- [17] X. Cai, P. J. Gray, Jr., D. D. Von Hoff, *Cancer Treat. Rev.* 2009, 35, 437–450.
- [18] C. Bailly, J. B. Chaires, *Bioconjugate Chem.* 1998, 9, 513–538.
- [19] P. M. O'Neill, L. P. Bishop, R. C. Storr, S. R. Hawley, J. L. Maggs, S. A. Ward, B. K. Park, *J. Med. Chem.* 1996, 39, 4511–4514.

- [20] P. M. O'Neill, A. Miller, L. P. D. Bishop, S. Hindley, J. L. Maggs, S. A. Ward, S. M. Roberts, F. Scheinmann, A. V. Stachulski, G. H. Posner, B. K. Park, *J. Med. Chem.* **2001**, *44*, 58–68.
- [21] J. Chadwick, A. E. Mercer, B. K. Park, R. Cosstick, P. M. O'Neill, *Bioorg. Med. Chem.* **2009**, *17*, 1325–1338.
- [22] R. Zhao, J. W. Lown, *Heterocycles* **1995**, *41*, 337–344; E. E. Baird, P. B. Dervan, *J. Am. Chem. Soc.* **1996**, *118*, 6141–6146.
- [23] D. S. Goodsell, M. L. Kopka, R. E. Dickerson, *Biochemistry* **1995**, *34*, 4983–4993.
- [24] M. Sriram, G. A. Van der Marel, H. L. P. F. Roelen, J. H. Van Boom, A. H. J. Wang, *Biochemistry* **1992**, *31*, 11823–11834.
- [25] GOLD docking suite: G. Jones, P. Willett, R. C. Glen, *J. Mol. Biol.* **1995**, *245*, 43–53. G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748. M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, R. D. Taylor, *Proteins: Struct. Funct. Genet.* **2003**, *52*, 609–623.
- [26] Minor groove binders study: H. K. Srivastava, M. Chourasia, D. Kumar, G. N. Sastry, *J. Chem. Inf. Model.* **2011**, *51*, 558–571.
- [27] Y. Li, J. Shen, X. Sun, W. Li, G. Liu, Y. Tang, *J. Chem. Inf. Model.* **2010**, *50*, 1134–1146.
- [28] The PyMol Molecular Graphics System, Version 1.3r1, Schrödinger, LLC, New York, USA, **2010**.
- [29] M. L. Kopka, C. Yoon, D. Goodsell, P. Pjura, R. E. Dickerson, *J. Mol. Biol.* **1985**, *183*, 553–563.
- [30] I. D. Kuntz, K. Chen, K. A. Sharp, P. A. Kollman, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9997–10002.
- [31] A. L. Hopkins, C. R. Groom, A. Alex, *Drug Discovery Today* **2004**, *9*, 430–431.
- [32] C. Abad-Zapatero, J. T. Metz, *Drug Discovery Today* **2005**, *10*, 464–469.
- [33] J. A. Wells, C. L. McClendon, *Nature* **2007**, *450*, 1001–1009.
- [34] C. Hetényi, U. Maran, A. T. García-Sosa, M. Karelson, *Bioinformatics* **2007**, *23*, 2678–2685.
- [35] A. T. García-Sosa, S. Sild, U. Maran, *QSAR Comb. Sci.* **2009**, *28*, 815–821.
- [36] A. T. García-Sosa, C. Hetényi, U. Maran, *J. Comput. Chem.* **2010**, *31*, 174–184.
- [37] H. Wang, C. A. Laughton, *Phys. Chem. Chem. Phys.* **2009**, *11*, 10722–10728.
- [38] D. J. Patel, *Proc. Natl. Acad. Sci. USA* **1982**, *79*, 6424–6428.
- [39] W. Trager, J. Jensen, *Science* **1976**, *193*, 673–675.
- [40] R. E. Desjardins, C. J. Canfield, J. D. Haynes, J. D. Chulay, *Antimicrob. Agents Chemother.* **1979**, *16*, 710–718.
- [41] A. E. Mercer, I. M. Copple, J. L. Maggs, P. M. O'Neill, B. K. Park, *J. Biol. Chem.* **2011**, *286*, 987–996.
- [42] M. J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R. W. Hyman, J. M. Carlton, A. Pain, K. E. Nelson, S. Bowman, I. T. Paulsen, K. James, J. A. Eisen, K. Rutherford, S. L. Salzberg, A. Craig, S. Kyes, M.-S. Chan, V. Nene, S. J. Shallom, B. Suh, J. Peterson, S. Angiuoli, M. Pertea, J. Allen, J. Selengut, D. Haft, M. W. Mather, A. B. Vaidya, D. M. A. Martin, A. H. Fairlamb, M. J. Fraunholz, D. S. Roos, S. A. Ralph, G. I. McFadden, L. M. Cummings, G. M. Subramanian, C. Mungall, J. C. Venter, D. J. Carucci, S. L. Hoffman, C. Newbold, R. W. Davis, C. M. Fraser, B. Barrell, *Nature* **2002**, *419*, 498–511.
- [43] Pipeline Pilot, Accelrys, Inc., San Diego, USA; <http://accelrys.com/products/pipeline-pilot/>.
- [44] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 490–519.
- [45] Spartan 08, Wavefunction, Inc., Irvine, USA, **2008**; <http://www.wavefun.com>.
- [46] Hermes Visualiser, version 1.4.1, Cambridge Crystallographic Database, Cambridge, UK, **2008**.
- [47] D. A. Case, T. A. Darden, T. E. Cheatham, III, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, P. A. Kollman, AMBER 11, University of California, San Francisco, USA, **2010**; <http://ambermd.org/>.
- [48] J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, *J. Comp. Chem.* **2004**, *25*, 1157–1174.
- [49] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, C. Simmerling, *Proteins: Struct. Funct. Bioinf.* **2006**, *65*, 712–725.
- [50] a) D. R. Hartree, *Proc. Cambridge Philos. Soc.* **1928**, *24*, 89–110; b) V. Fock, *Z. Phys.* **1930**, *61*, 126–148; c) U. C. Singh, P. A. Kollman, *J. Comput. Chem.* **1984**, *5*, 129–145.
- [51] W. L. Jorgensen, J. D. Madura, *Mol. Phys.* **1985**, *56*, 1381–1392.
- [52] Gaussian 09, Revision A1, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Menonucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski, D. J. Fox, Gaussian Inc., Wallingford, USA, **2009**.
- [53] B. Wellenzohn, R. H. Winger, A. Hallbrucker, E. Mayer, K. R. Liedl, *J. Am. Chem. Soc.* **2000**, *122*, 3927–3931.
- [54] M. A. Young, G. Ravishanker, D. L. Beveridge, *Biophys. J.* **1997**, *73*, 2313–2336.
- [55] M. A. Young, B. Jayaram, D. L. Beveridge, *J. Am. Chem. Soc.* **1997**, *119*, 59–69.
- [56] J. Shao, S. W. Tanner, N. Thompson, T. E. Cheatham III, *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.

Received: November 19, 2012

Published online on March 11, 2013