J|A|C|S
ARTICLES
Published on Web 12/31/2002

Hydration of Hydroxypyrrole Influences Binding of ImHpPyPy-β-Dp Polyamide to DNA

Bernd Wellenzohn, Markus J. Loferer, Michael Trieb, Christine Rauch, Rudolf H. Winger, Erwin Mayer, and Klaus R. Liedl*

Contribution from the Institute of General, Inorganic and Theoretical Chemistry, University of Innsbruck, Innrain 52a, A-6020 Innsbruck, Austria

Received July 19, 2002; E-mail: Klaus.Liedl@uibk.ac.at.

Abstract: Ligands which are able to recognize DNA sequence specifically are of fundamental interest as transcription controlling drugs. Recently a polyamide ligand was developed (ImHpPyPy-β-Dp) which differentiates in a dimeric arrangement between all four possible base pair steps in the minor groove. This is a landmark for the design of DNA binding drugs because it was believed that such a recognition could only be possible in the major groove of DNA. Although the OH groups of the hydroxypyrrole (Hp) moieties of the ligands are responsible for this sequence discrimination, experiments showed that this OH group also reduces the absolute binding constant. We performed a free energy calculation by means of thermodynamic integration in order to find out the influence of this single hydroxyl on DNA binding. In our simulation, we found that the hydroxyl group reduces binding by about 1.3 kcal/mol, which is in excellent agreement with the experimentally determined value of 1.2 kcal/mol. In further MD simulations, the structural reasons for this reduction was estimated. The results of these simulations qualitatively agree with the X-ray structures, but in contrast, in the simulations both (ImHpPyPy-β-Dp and ImPyPyPy-β-Dp) ligand−DNA (d(CCAGTACTGG)$_2$) complexes exhibit only slight structural differences. This is consistent with a recently published second pair of similar polyamide DNA crystal structures. Thus, we believe that the explanations resulting from the X-ray structures must be modified. We attribute the large structural differences between the two polyamide DNA complexes to a buffer molecule which binds only in the case of the ImHpPyPy-β-Dp-DNA complex at the region of interest. We propose that the differential hydration of both ligands in the unbound state is responsible for the reduction of the binding constant. Additionally, we suggest an indirect readout of DNA, because of a lengthening of the Watson−Crick base pairs, which possibly contributes to the differentiation between T·A, A·T from G·C, C·G base pairs.

Introduction

DNA binding ligands are of extraordinary interest as transcription controlling drugs.[1−5] The recognition of DNA by such ligands and proteins is a real challenge because DNA consists only of four basic building blocks. Recognition through the so-called direct readout of DNA is achieved exclusively by contacts with the bases. The energetic contribution of one H-bond contact is estimated to be about 1.5 kcal/mol.[6] Because of the structure arising from the Watson−Crick base pairing, only the edges of the four bases are accessible for building such direct contacts. The resulting arrangement of H-bond donors and acceptors is different for all four possible base pairs (A·T, T·A, G·C, C·G) in the major groove. Thus, this H-bond donor and acceptor arrangement could be used as a recognition code for discriminating between them. In the minor groove, the situation is more complicated. On the basis of this H-bond donor and acceptor recognition site composition, it would only be possible to differentiate G·C from C·G but not A·T from T·A.[7] Recently, however, small polyamide molecules have shown the ability to bind sequence specifically[8−19] into the minor groove of B-DNA. Especially the ImHpPyPy-β-Dp polyamide (Im = imidazole,

(1) Krugh, T. R. *Curr. Opin. Struct. Biol.* **1994**, *4*, 351−364.
(2) Geierstanger, B.; Wemmer, D. E. **1995**, *24*, 463−493.
(3) Neidle, S. *Biopolymers (Nucleic Acid Sci.)* **1997**, *44*, 105−121.
(4) Chaires, J. B. *Curr. Opin. Struct. Biol.* **1998**, *8*, 314−320.
(5) Strey, H. H.; Podgornik, R.; Rau, D. C.; Parsegian, V. A. *Curr. Opin. Struct. Biol.* **1998**, *8*, 309−313.
(6) Lesser, D. E. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 7429−7430.
(7) Branden, S.; Tooze, J. DNA structures. Garland Publishing Inc.: New York and London, 1991; pp 79−84.
(8) Urbach, A. R.; Szewczyk, J. W.; White, S.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 11621−11629.
(9) Kielkopf, C. L.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Nat. Struct. Biol.* **1998**, *5*, 104−108.
(10) White, S.; Turner, J. M.; Szewczyk, J. W.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 260−261.
(11) Turner, D. *Curr. Opin. Struct. Biol.* **1996**, *6*, 299−304.
(12) Greenberg, W. A.; Baird, E. E.; Dervan, P. B. *Chem.−Eur. J.* **1998**, *4*, 796−805.
(13) Herman, D. M.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 1121−1129.
(14) Herman, D. M.; Baird, E. E.; Dervan, P. B. *Chem.−Eur. J.* **1999**, *5*(3), 975−983.
(15) Swalley, S. E.; Baird, E. E.; Dervan, P. B. *J. Am. Chem. Soc.* **1997**, *119*, 6953−6961.
(16) de Clairac, R. P. L.; Geierstanger, B. H.; Mrksich, M.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **1997**, *119*, 7909−7916.
(17) Wade, W. S.; Mrksich, M.; Dervan, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 8783−8794.
(18) Mrksich, M.; Wade, W. S.; Dwyer, T. J.; Geierstanger, B. H.; Wemmer, D. E.; Dervan, P. B. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 7586−7590.
(19) Hawkins, C. A.; de Clairac, R. P.; Dominey, R. N.; Baird, E. E.; White, S.; Dervan, P. B.; Wemmer, D. E. *J. Am. Chem. Soc.* **2000**, *122*(7), 5235−5243.

10.1021/ja0277778 CCC: $25.00

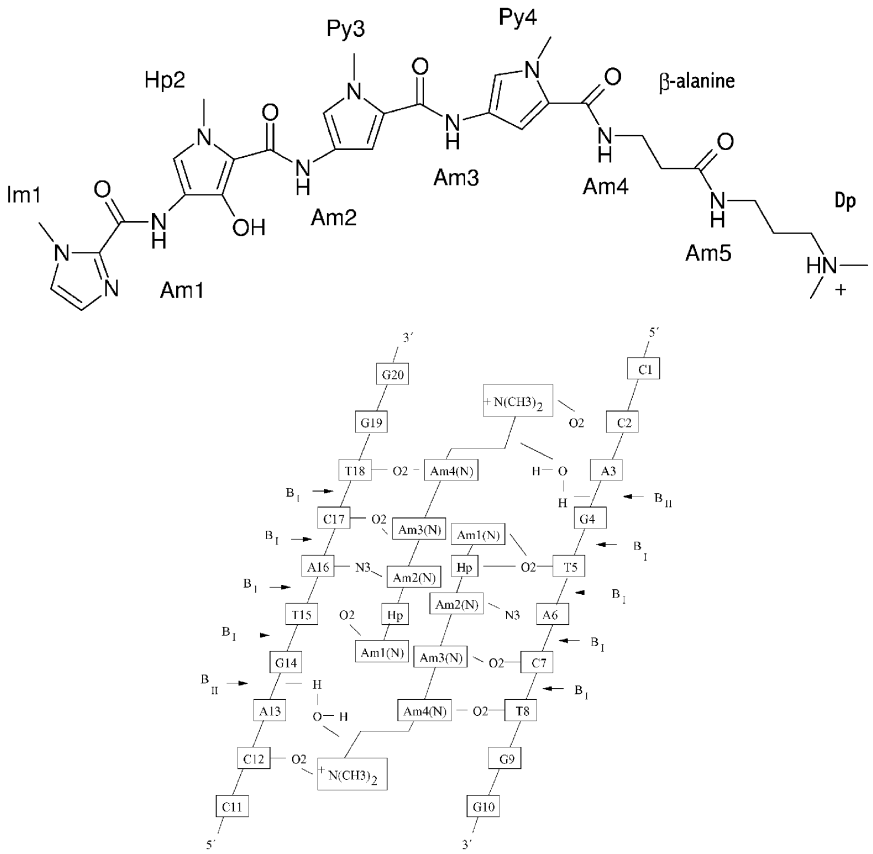


Figure 1. Chemical structure of the ImHpPyPy-*β*-Dp polyamide (Im = imidazole, Hp = hydroxypyrrole, Py = pyrrole, Am = amide, *β* = *β*-alanine, Dp = dimethylaminopropylamide). Numbering scheme is according to Kielkopf et al.[20] The phosphates which are frozen in a particular backbone conformation are indicated by arrows (described in the text).

Hp = hydroxypyrrole, Py = pyrrole, *β* = *β*-alanine; Dp = dimethylaminopropylamide) ligand is able to discriminate between all four base pairs in the minor groove (Figure 1).[20]

It has been shown that the OH group on Hp is responsible for the sequence specific recognition because in contrast to ImHpPyPy-*β*-Dp polyamide the ligand lacking the hydroxypyrrole OH group (ImPyPyPy-*β*-Dp) does not differentiate A·T from T·A. Recently, Kielkopf et al.[20] established the underlying structural basis for this by determining the crystal structure of the $d(CCAGTACTGG)_2$ decamer in the B form complexed with ImHpPyPy-*β*-Dp (see Figure 2). The polyamide binds in an antiparallel dimeric arrangement in the minor groove of DNA. The asymmetric C2 cleft of adenine and the double hydrogen bond acceptor potential of the thymine O2 are used as an explanation for the sequence specificity. Because of this double hydrogen bond acceptor potential of the thymine O2, the OH group of Hp is able to form a hydrogen bond, although Am1 (see Figure 1 for the nomenclature) is already bound.

Although the Hp OH group is essential for the sequence specificity experiments by means of quantitative DNase I, footprint titration showed that this OH reduces the binding affinity in a range of 1.2 kcal/mol.[20,15] Furthermore several other polyamide ligands show a nonadditive reduction in binding affinity because of the introduction of hydroxyl groups.[21,22] On the basis of the X-ray structures of both $d(CCAGTACTGG)_2$ complexes (ImHpPyPy-*β*-Dp and ImPyPyPy-*β*-Dp), a partial melting of a DNA base pair was observed. Additionally, in the case of the ImHpPyPy-*β*-Dp-DNA complex, the hydrogen bonds between the amides and the DNA are lengthened, and because of the complexation with hydroxypyrrole, one of the amide DNA bonds becomes bifurcated and therefore weaker. All these effects together should result in the 1.2 kcal/mol reduction of the binding constant. Recently, a second set of DNA ($d(CCAGATCTGG)_2$) polyamide (ImPyHpPy-*β*-Dp and ImPyPyPy-*β*-Dp) cocrystal structures was published.[21] Again, the hydroxyl group of hydroxypyrrole determines the sequence specificity but reduces the binding affinity. In these structures, the proposed partial melting of the target A·T base pair is not observed. So the structural basis of the reduction of the binding affinity upon addition of the hydroxyl group is still unclear.

Besides direct readout, the indirect readout has already shown to contribute to DNA sequence recognition and was proposed for the first time by Otwinowski et al.[23] Indirect readout is mediated on one hand by contacts with the backbone and with the nonspecific part of the bases and on the other hand by the sequence dependent energetic penalty which is needed to distort DNA from its low energy conformation of the unbound state. The role of bending, unwinding, and other recognition tools in the indirect readout have been investigated exten-

(20) Kielkopf, C. L.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B.; Rees, D. C. *Science* **1998**, *282*, 111−115.
(21) Kielkopf, C. L.; Bremer, R. E.; White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. D.; Rees, D. C. *J. Mol. Biol.* **2000**, *295*, 557−567.
(22) White, S.; Szewczyk, J. W.; Turner, J. M.; Baird, E. E.; Dervan, P. B. *Nature* **1998**, *391*, 468−471.
(23) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; Lawson, C. L.; Joachimiak, A.; Marmorstein, R. Q.; Luisi, B. F.; Sigler, P. B. *Nature* **1988**, *335*, 321−329.

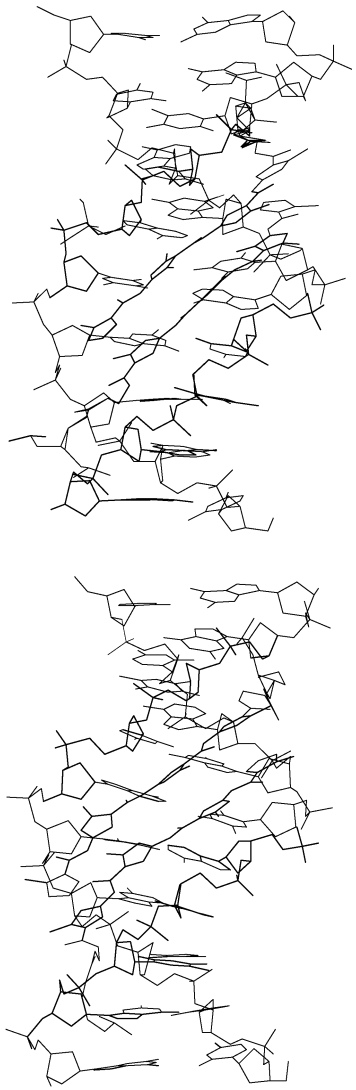


Figure 2. X-ray structure of d(CCAGTACTGG)$_2$ decamer in the B-form complexed by two ImHpPyPy-β-Dp polyamides (top). The structure has the NDB Code = BDD002.[20] The bottom graph shows the same DNA complexed with ImPyPyPy-β-Dp-polyamides (NDB Code = BDD003).[20] The overall structures of both complexes are similar.

sively.[3,24−42] Backbone conformations of B-DNA such as B_I/B_{II} could be another important element in recognition. A recently published[43] MD simulation of the ImHpPyPy-β-Dp-d(CCAGTACTGG)$_2$ complex indicates that complexation freezes the DNA backbone in a specific conformation (Figure 1). Additionally, simulations of the complexed and uncomplexed DNA propose that, in contrast to the X-ray structure, in solution the DNA double helices are curved.[44] The bending is mainly introduced by two TpG/CpA steps, one at each end, and the TpA step in the middle of the DNA sequence. Such steps are well-known flexible sites with a high bending capability.[45−57]

We performed a free energy simulation by means of thermodynamic integration in which we mutated the ImHpPyPy-β-Dp polyamides to ImPyPyPy-β-Dp. The resulting $\Delta\Delta G$ value of the respective mutations in the complexed and in the uncomplexed state gives the contribution of the hydroxypyrrole OH group to binding. As determined by the simulation, this group reduces the binding constant by about 1.3 kcal/mol for each hydroxypyrrole substitution (experimental value 1.2 kcal/mol). To clarify the underlying structural reasons for this reduction of the binding constants, we performed molecular dynamics simulations of both polyamide−DNA (ImHpPyPy-β-Dp and ImPyPyPy-β-Dp) complexes and a simulation of the uncomplexed DNA (starting from two different conformations) and compared them with the X-ray structures. In contrast to the X-ray structures, in our simulations all parameters such as H-bond distances and DNA helical parameters are symmetrical for both ligands and for the DNA, as expected for this symmetrical complex. We attribute the anomalies of the X-ray structures to side effects. For example, in the X-ray structures an interaction between a buffer molecule and the major groove of the DNA was observed. Nevertheless, our simulation results are in high qualitative agreement with the suggestions made from the X-ray structures. We also conclude that the H-bond distances of the ligand amides to DNA are lengthened in the presence of the OH group, but in our simulations, the effect is not that pronounced (1.0 Å versus 0.2 Å) as in the X-ray structures. The energetic penalty introduced by this weakening of the H bonds, which was estimated by quantum chemical ab initio calculations, is too small to explain the reduction in the binding constant. In addition, the Watson−Crick base pairing

(24) Dickerson, R. E.; Chiu, T. K. *Biopolymers (Nucleic Acid Sci.)* **1998**, *44*, 361−403.
(25) Dickerson, R. E. *Nucleic Acids Res.* **1998**, *26*, 1906−1926.
(26) Giese, K.; Pagel, J.; Grosschedl, R. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12845−12850.
(27) Chen, H.; Liu, X.; Patel, D. J. *J. Mol. Biol.* **1996**, *258*, 457−479.
(28) Shi, Y.; Berg, J. M. *Biochemstry* **1996**, *35*, 3845−3848.
(29) von Hippel, P. H. *Science* **1994**, *263*, 769−770.
(30) Strauss, J. K.; Roberts, C.; Nelson, M. G.; Switzer, C.; Maher, L. J., III. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9515−9529.
(31) Bareket-Samish, A.; Cohen, I.; Haran, T. E. *J. Mol. Biol.* **1998**, *277*, 1071−1080.
(32) Wenz, C.; Jeltsch, A.; Pingoud, A. *J. Biol. Chem.* **1996**, *271*(10), 5565−5573.
(33) Steitz, T. A. *Structural studies of protein-nucleic acid interaction*; Cambridge University Press: Cambridge, U.K., 1993.
(34) Lilley, D. M. J. *DNA − Protein: Structural Interactions*. Oxford University Press: Oxford, U.K., 1995.
(35) Travers, A. *DNA − Protein Interactions*. Chapman & Hall: London, U.K., 1993.
(36) Bewley, C. A.; Grinenborn, A. M.; Clore, G. M. *Annu. Rev. Biophys. Biophys. Chem.* **1998**, *27*, 105−131.
(37) Chen, Y.-Q.; Gosh, S.; Gosh, G. *Nat. Struct. Biol.* **1998**, *5*, 67−73.
(38) Ellenberger, T. *Curr. Opin. Struct. Biol.* **1994**, *4*, 12−21.
(39) Gehring, W. J.; Qian, Y. Q.; Furukubo-Tokunaga, M. K.; Schlier, F.; Resendez-Perez, D.; Affolter, M.; Otting, G.; Wüthrich, K. *Cell* **1994**, *78*, 211−223.
(40) Harrison, S. C.; Aggarwal, A. K. *Annu. Rev. Biochem.* **1990**, *59*, 933−969.
(41) Neidle, S. *Oxford Handbook of Nucleic Acid Structure*. Oxford University Press: Oxford, U.K., 1999.
(42) Pabo, C. O.; Sauer, R. T. *Annu. Rev. Biochem.* **1992**, *61*, 1053−1095.
(43) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2001**, *123*(21), 5044−5049.
(44) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Phys. Chem. B* **2001**, *105*(15).
(45) Nikolov, D. B.; Chen, H.; Halay, E. D.; Hoffmann, A.; Roeder, R. G.; Burley, S. K. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 4862−4867.
(46) Juo, Z. S.; Chiu, T. K.; Leiberman, P. M.; Baikalov, I.; Berk, A. J.; Dickerson, R. E. *J. Mol. Biol.* **1996**, *261*, 239−254.
(47) Calladine, C. R.; Drew, H. R. *Understanding DNA*. Academic Press: San Diego, CA, 1997.
(48) Kim, J. L.; Nikolov, D. B.; Burley, S. K. *Nature* **1993**, *365*, 520−527.
(49) Pardo, L.; Pastor, N.; Weinstein, H. *Biophys. J.* **1998**, *75*, 2411−2421.
(50) Tonelli, M.; Ragg, E.; Bianucci, A. M.; Lesiak, K.; James, T. L. *Biochemistry* **1998**, *37*, 11745−11761.
(51) Nagaich, A. K.; Bhattacharyya, D.; Brahmachari, S. K.; Bansal, M. *J. Biol. Chem.* **1994**, *10*, 7824−7833.
(52) Tzou, W.-S.; Hwang, M.-J. *Biophys. J.* **1999**, *77*, 1191−1205.
(53) El Hassan, M.; Calladine, C. *J. Mol. Biol.* **1996**, *259*, 95−103.
(54) Gavathiotis, E.; Sharman, G. J.; Searle, M. S. *Nucleic Acids Res.* **2000**, *28*, 728−735.
(55) Dlakic, M.; Harrington, R. E. *J. Biol. Chem.* **1995**, *270*(50), 29945−29952.
(56) DiGabriele, A. D.; Sanderson, M. R.; Steitz, T. E. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 1816−1820.
(57) Bolshoy, A.; Harrington, R. E.; Trifonov, E. N. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 2312−2316.

of the central A·T base pairs in our simulation are also stretched by the binding of the ImHpPyPy-*β*-Dp group, but again, the effect is not as pronounced as in the X-ray structures (0.6 Å versus 0.2 Å). Complexation with ImPyPyPy-*β*-Dp also leads to a reduction of the Watson−Crick hydrogen-bond index of the target T·A base pair, showing that a slight base pair lengthening occurs always when these ligands bind. Simulations of a second pair of polyamide DNA structures (d(CCAGATCTGG)$_2$ complexed with ImPyPyPy-*β*-Dp and with ImPyHpPy-*β*-Dp) confirm the previously mentioned base pair stretching. Thus, these distortions of the base pairs could be responsible for the differentiation between T·A/A·T and G·C/C·G because of an indirect readout of DNA, but it seems not to be sufficient to determine the reduction of the binding constant because of the hydroxyl group. On the basis of the hydration free energy, which we calculated by means of free energy calculations, we propose that hydration effects are responsible for the reduction in the binding constant. The additional hydroxyl group leads to a better hydration of the ImHpPyPy-*β*-Dp ligand compared with ImPyPyPy-*β*-Dp in the unbound state, which cannot be compensated by the additional hydrogen bond in the complexed state because of its bifurcated (and therefore weaker) character.

Methods

Simulation of DNA[58−61] and DNA−ligand complexes[62−66] have proven to be a valuable tool for a deeper understanding of structural and dynamical properties. The inclusion of the long-range interactions via the Ewald summation in form of the so-called particle mesh Ewald method allows the calculation of stable B-form DNA trajectories[67−69] in the nanosecond region. To take advantage of findings of previous extensive simulations,[70−73] protocols employed therein were directly adapted for our needs. We carried out free energy calculations (A), three molecular dynamics simulations of the d(CCAGTACTGG)$_2$ polyamide (ImPyPyPy-*β*-Dp and ImHpPyPy-*β*-Dp) complexes (B,C, and D), and two simulations of the unbound DNA (E and F), and another two molecular dynamics simulations of d(CCAGATCTGG)$_2$ complexed with ImPyPyPy-*β*-Dp and ImPyHpPy-*β*-Dp (G and H). Results of simulation B, E, and F are already published.[43,44] The simulation time in B is 5 ns, the simulation times in C, D, G, and H are 1 ns (the convergence of the results was verified), and the simulation times in E and F are 5 ns long.

Simulation A. The free energy calculation was made by means of thermodynamic integration, which is a standard method for calculating relative free energies. A detailed description of the method is given elsewhere.[43,44] All calculations were performed using the GIBBS module of the Amber5 package.[74] The OH groups of the hydroxypyrrole ligands were mutated to H and backward to OH. The charges of ImHpPyPy-*β*-Dp and of ImPyPyPy-*β*-Dp are obtained using the RESP[75] method. The ab initio electrostatic potentials for RESP were calculated using GAUSSIAN98[76] at HF/6-31G* level of theory. The force field parameters for the polyamides were selected in analogy to existing parameters in the force field. All parameters of the ImHpPyPy-*β*-Dp are already published, and the respective ImPyPyPy-*β*-Dp ligand needs only slight modifications. As a starting structure, the respective snapshots after 100 ps of MD simulation were used. The general simulation parameters were used for the DNA−ligand complexes, as described later in the text. For the unbound ligand, because of stability reasons a time step of 1 fs was applied. The total time for the mutation is 252 ps divided into 21 sampling windows. A doubling of the sample rate was performed in order to estimate the convergence of the results, showing that the phase space is sufficiently sampled. The estimation of the hydration free energy was performed on a model molecule consisting of the ImHpPyPy part. This model molecule was mutated both in a vacuum and in solution to ImPyPyPy. Because of the fact that this molecule is uncharged, no counterions for charge neutralization were needed.

Simulation B. As a starting structure for the simulation of the d(CCAGTACTGG)$_2$−(ImHpPyPy-*β*-Dp)$_2$ complex, the crystal structure (NDB ID = BDD002)[20] was used. Each strand of the DNA has nine PO_4^- anions, and each of the two polyamide ligands has one positive charge. To achieve electroneutrality, 16 Na^+ counterions were added using the program CION of the AMBER[74] package. Subsequently, solvation of the DNA with TIP3P Monte Carlo water boxes requiring a 12 Å solvent shell in all directions resulted in a system with the dimension 62.18 × 47.19 × 48.57 Å^3 containing 3914 water molecules. The corresponding Γ value (water/nucleotide) is 195.7. The simulation was carried out using the AMBER5[74] package with the all atom force field of Cornell et al.[77] and the modifications by Cheatham et al.[78] The force field parameters for the polyamides were selected in analogy to existing parameters in the force field. Charges were derived using the RESP[75] charge fitting procedure (multiconformational RESP). The ab initio electrostatic potential for RESP was calculated using GAUSSIAN98[76] at the HF/6-31G* level of theory.

(58) von Kitzing, E. *Methods Enzymol.* **1992**, *211*, 449−467.
(59) Beveridge, D. L.; Ravishanker, G. *Curr. Opin. Struct. Biol.* **1994**, *4*, 246−255.
(60) Beveridge, D. L.; Swaminathan, S.; Ravishanker, G.; Withka, J. M.; Srinivasan, J.; Prevost, C.; Louise-May, S.; Langley, D. R.; DiCapua, F. M.; Bolton, P. H. Molecular Dynamics Simulations on the Hydration, Structure and Motions of DNA Oligomers. The Macmillan Press Ltd.: London, 1993; pp 165−225.
(61) Louise-May, S.; Auffinger, P.; Westhof, E. *Curr. Opin. Struct. Biol.* **1996**, *6*, 289−298.
(62) Wellenzohn, B.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Am. Chem. Soc.* **2000**, *122*, 3927−3931.
(63) Wellenzohn, B.; Flader, W.; Winger, R. H.; Hallbrucker, A.; Mayer, E.; Liedl, K. R. *J. Biol. Chem.*, submitted.
(64) Perree-Fauvet, M.; Gresh, N. *J. Biomol. Struct. Dyn.* **1994**, *11*, 1203−1224.
(65) Ketterle, C.; Gabarro-Arpa, J.; Ouali, M. B.; Auclair, C.; Helissey, P.; Giorgi-Renault, S.; Bret, M. L. *J. Biomol. Struct. Dyn.* **1996**, *13*, 963−977.
(66) Bifulco, G.; Galeone, A.; Nicolaou, K.; Chazin, W. J.; Gomez-Paloma, L. *J. Am. Chem. Soc.* **1998**, *120*, 7183−7191.
(67) Cieplak, P.; Cheatham, T. E., III.; Kollman, P. A. *J. Am. Chem. Soc.* **1997**, *119*, 6722−6730.
(68) Young, M. A.; Nirmala, R.; Srinivasan, J.; McConnell, K. J.; Ravishanker, G.; Beveridge, D. L.; Berman, H. M. Analysis of helix bending in crystal structures and molecular dynamics simulations of DNA oligonucleotides. Proceedings of the 8th Conversation; Adenine Press: Albany, NY, 1994; pp 197−214.
(69) Cheatham, T. E., III.; Kollman, P. A. *J. Mol. Biol.* **1996**, *259*, 434−444.
(70) Young, M. A.; Jayaram, B.; Beveridge, D. L. *J. Am. Chem. Soc.* **1997**, *119*, 59−69.
(71) Young, M. A.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1997**, *73*, 2313−2336.
(72) de Souza, O. N.; Ornstein, R. L. *J. Biomol. Struct. Dyn.* **1997**, *14*, 607−611.
(73) de Souza, O. N.; Ornstein, R. L. *Biophys. J.* **1997**, *72*, 2395−2397.
(74) Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T., III.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER5*. University of California, San Francisco, 1997.
(75) Bayly, C.; Cieplak, P.; Cornell, W.; Kollman, P. *J. Phys. Chem.* **1993**, *97*, 10269−10280.
(76) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian98*, revision A.4. Gaussian, Inc.: Pittsburgh, PA, 1998.
(77) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(78) Cheatham, T. E., III.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845−862.

Simulation C. The first simulation of the d(CCAGTACTGG)$_2$−(ImPyPyPy-β-Dp)$_2$ complex is made using the same procedure as described in simulation B, but the OH group is substituted by H.

Simulation D. The second simulation of the d(CCAGTACTGG)$_2$−(ImPyPyPy-β-Dp)$_2$ complex is started from the X-ray structure of the respective complex with the pdb code = bdd003. The procedure is exactly the same as described in simulation B and C.

Simulation E. As a starting structure for the first simulation of the uncomplexed d(CCAGTAC TGG)$_2$−DNA, the coordinates of the crystal structure from the DNA complex (NDB ID = BDD002) were used. Each strand of the DNA has nine PO_4^- anions, so 18 Na^+ counterions were added to achieve electroneutrality. The solvation of the DNA resulted in a box of the dimensions 61.04 × 48.62 × 48.32 Å^3 containing 3998 water molecules. The corresponding Γ value is 199.9. The simulation was carried out using the AMBER5[74] package with the all atom force field of Cornell et al.[77] and the modifications by Cheatham et al.[78]

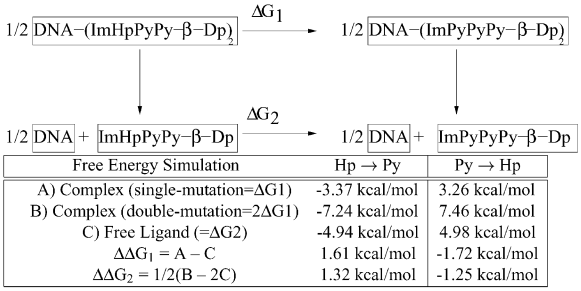
Simulation F. The starting coordinates for the second simulation of the free d(CCAGTACTGG)$_2$−DNA are from the crystal structure of the uncomplexed decamer (NDB ID = BD0023).[79] The same procedure as that for the simulation E was employed, resulting in a box with the dimensions 61.48 × 47.04 × 47.80 Å^3 and a Γ value of 193.9.

Simulation G. The simulation of the d(CCAGATCTGG)$_2$−(ImPyPyPy-β-Dp)$_2$ complex is started from the X-ray structure of the respective complex with the pdb code = dd0021.[21] The procedure is exactly the same as that described in simulation B and C.

Simulation H. The simulation of the d(CCAGATCTGG)$_2$−(ImPyHpPy-β-Dp)$_2$ complex is started from the X-ray structure of the respective complex with the pdb code = dd0020.[21] The procedure is exactly the same as that described in simulation B and C.

Minimization/Equilibration. First, 500 steps of minimization were carried out with harmonic restraints of 25 kcal mol^{-1} Å^{-2} on DNA, counterions, and ligands positions. During the following five 100-step minimizations, the restraints on the counterions were relaxed faster than those on DNA and the ligand. Finally, 500 steps of unrestrained minimization were carried out. For the equilibration, a similar procedure was applied. After heating the constant volume system during 10 ps from 50 to 300 K and keeping the DNA and ion positions constant, the harmonic restraints were reduced throughout the following 25 ps, faster on the counterions than on the oligonucleotide and ligand using constant-pressure and constant-temperature conditions. Finally, 5 ps unrestrained equilibration was carried out before the trajectory was generated for 2960 ps more (simulation). The temperature bath coupling was achieved by the Berendsen algorithm.[80]

General simulation parameters were kept constant during the whole simulation: 2 fs time step, SHAKE constraints of 0.000 05 Å on all bonds involving hydrogen atoms, 9 Å nonbonded cutoff, and 0.000 01 convergence criterion for the Ewald part of the nonbonded interactions. The structural information was collected every 50 steps (0.1 ps). The resulting trajectory was analyzed with the AMBER5 package, and snapshots were investigated with different visualization programs.[81,82] The molecular dynamics toolchest was used for the calculation of the helical parameters[83] and for visualization in the form of graphs.[84] All calculations were performed on an SGI octane.



Free Energy Simulation	Hp → Py	Py → Hp
A) Complex (single-mutation=ΔG1)	-3.37 kcal/mol	3.26 kcal/mol
B) Complex (double-mutation=2ΔG1)	-7.24 kcal/mol	7.46 kcal/mol
C) Free Ligand (=ΔG2)	-4.94 kcal/mol	4.98 kcal/mol
$\Delta\Delta G_1$ = A − C	1.61 kcal/mol	-1.72 kcal/mol
$\Delta\Delta G_2$ = 1/2(B − 2C)	1.32 kcal/mol	-1.25 kcal/mol

Figure 3. Schematic picture on the top shows the thermodynamical cycle (double mutation) used for calculating the influence of the Hp hydroxyl group on binding. The table summarizes the results obtained from these free energy calculations. In simulation A, only one of both Hp was mutated to Py, while, in simulation B, both bound ligands are changed. The resulting $\Delta\Delta G$ values are normalized to one OH group, showing that Hp reduces binding with respect to Py. The values 1.29 kcal/mol (mean of $\Delta\Delta G_2$) and 1.67 kcal/mol (mean of $\Delta\Delta G_1$) of reduction in the binding free energy for each hydroxypyrrole substitution qualitatively agree with the experimentally determined 1.2 kcal/mol. The two vertical reactions are identical and thus compensate each other.

Results

The effect of the hydroxypyrrole OH group on the binding constant was calculated by means of thermodynamic integration. The thermodynamic cycle and a synopsis of the results are shown in Figure 3. The ImHpPyPy-β-Dp ligand is mutated to ImPyPyPy-β-Dp one time in the complex and one time uncomplexed as free ligand. Because of the fact that the polyamide ligands bind in a 2:1 stoichiometry to DNA, mutation of one ligand (single mutation) and the simultaneous mutation of both ligands (double mutation) were performed. The results of the single mutation give the influence of the OH group in the mixed ImHpPyPy-β-Dp/ImPyPyPy-β-Dp complex, while the results of the double mutation give the respective value in the pure complex.

The results of the free energy calculation propose that one ImPyPyPy-β-Dp ligand binds 1.3 kcal/mol (double mutation, pure complex) better than the ImHpPyPy-β-Dp. Thus, the OH group reduces the affinity although it builds an H bond to DNA. Our simulation results are in high agreement with the experimental value of 1.2 kcal/mol. The results are independent of variation of the sampling rate underlining the reliability. For the mixed complex (single mutation), the change in $\Delta\Delta G$ is even higher.

To estimate the reasons for this reduction of the binding constant, molecular dynamics simulations of the two complexes and of uncomplexed DNA were performed. We performed a molecular dynamics simulation of the d(CCAGTACTGG)$_2$−(ImHpPyPy-β-Dp)$_2$ complex, two calculations of d(CCAGTACTGG)$_2$−(ImPyPyPy-β-Dp)$_2$ starting from different structures, and two simulations of the uncomplexed DNA (also starting from different conformations). Starting from different conformations helps to overcome the problem of insufficient sampling of the phase space. As starting structure, on one hand the respective X-ray structures were used and on the other hand they were generated by removing the ligands or the OH group of the DNA−(ImHpPyPy-β-Dp)$_2$ complex. Analyses of the simulations show that the presented results are independent of the starting structure. The rmsd values and the total energies of all simulations stay constant at a value of about 2.0 Å after a short equilibration period. In contrast to the X-ray structures, the DNA is curved. The bending is introduced by a positive

(79) Kielkopf, C. L.; Ding, S.; Kuhn, P.; Rees, D. C. *J. Mol. Biol.* **2000**, *296*, 787−801.
(80) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(81) Sayle, R.; Müller, A.; Bohne, A. *rasmol2.6ab9*. Molecular Modeling Group, German Cancer Research Center: Heidelberg, Germany, 1999.
(82) Laaksonen, L. *gOpenMol 1.21*. Centre for Scientific Computing, Espoo, Finland, 1996.
(83) Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1988**, *6*, 63−91.
(84) Ravishanker, G.; Swaminathan, S.; Beveridge, D. L.; Lavery, R.; Sklenar, H. *J. Biomol. Struct. Dyn.* **1989**, *6*, 669−699.

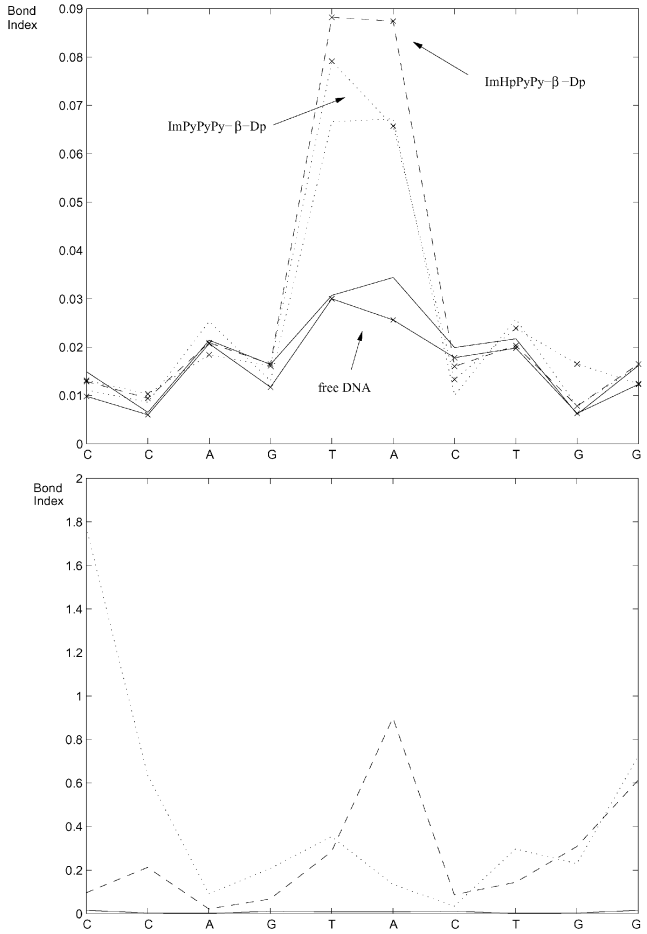


Figure 4. Graph on the top shows the mean values of the hydrogen-bond quality index [Å] of all five simulations. The dashed line indicates the ImHpPyPy-*β*-Dp complex, the dotted curve gives the values for the two simulations of the ImPyPyPy-*β*-Dp complex, and the solid line represents the values of the unbound DNA. The small x indicates the simulations, starting from the respective X-ray structure. The bottom graph shows the respective values of the X-ray structures (same line representation).

roll at the CpA/TpG steps and by the central TpA site.[44] In the X-ray structure, binding of ImHpPyPy-*β*-Dp leads to a melting of the central T·A base pair, while during binding of ImPyPyPy-*β*-Dp the hydrogen bonds stay intact. The difference in the Watson−Crick bond lengths between both complexes at the central GpTpApC recognition site is up to 0.6 Å. It was proposed that the energetic penalty of this melting may account for the reduction in binding affinity. To quantify the deviation from the ideal Watson−Crick bond, we employ a hydrogen bond quality index I_H which is defined as[85]

$$I_H = \sum_{D-H\cdots A}[(d_{DA} - d^0_{DA})^2 + (1 + \cos\gamma)^2]$$

where d_{DA} is the donor−acceptor distance, d^0_{DA} is the ideal donor−acceptor distance, and γ is the D−H···A angle. The summation over all Watson−Crick hydrogen bonds adopts a value of zero for ideal Watson−Crick bonding. The advantage of this index is that besides the bond lengths the bond angles are also considered, giving a better picture of the real strength of the base pairing. The hydrogen bond quality indices of the X-ray structures and the simulations are shown in Figure 4.

(85) Yan, S.; Shapiro, R.; Geacintov, N. E.; Broyde, S. *J. Am. Chem. Soc.* **2001**, *123*, 7054−7066.

Figure 4 shows that the Watson−Crick bonding of the two central A·T base pairs is weakened, as expected from the X-ray structure, but the effect is not that pronounced. The maximum difference in the Watson−Crick hydrogen bond lengths between both complexes is only 0.04 Å. This is an extremely small deviation (especially when compared with the 0.6 Å of the X-ray structure), and although the energetic penalty is hard to estimate, we propose that these small differences in the base pair are not sufficient to explain the previously mentioned reduction of the binding constant. We attribute the large deviations in the Watson−Crick bond index in the X-ray structures to a buffer molecule which binds from the major groove side to the central A·T base pair only in the ImHpPyPy-*β*-Dp complex. No evidence for a bound buffer molecule was found in the ImPyPyPy-*β*-Dp complex. Furthermore, the trend of the bond indices of the X-ray structures does not reflect the palindromic nature of the molecules, showing that effects such as binding of the buffer molecule influence the structures.

The change in the bond index between uncomplexed and complexed DNA is more dramatic than that between the two different complexes. The difference is mainly introduced by a lengthening of the Watson−Crick bonds (up to 0.2 Å) on complexation, but differences in the bond angles also occur. We propose that lengthening of the H bonds on complexation contributes to the recognition by indirect readout of the DNA. For G·C, C·G base pairs, the energy needed for disturbing the base pair should be higher (three hydrogen bonds instead of two) than in the case of A·T, T·A. Thus, we surmise that this indirect readout helps to differentiate A·T/T·A from G·C/C·G. The proof of the consistency with the pairing rules derived from Dervan et al. will be the aim of future research.

As mentioned in the Introduction, a second set of polyamide DNA cocrystal structures was recently published (d(CCA-GATCTGG)$_2$ complexes). As shown in Figure 5, the X-ray structure of the Hp complex exhibits no dramatic melting of the target base pairs. The two peaks of the hydrogen bond quality index of the X-ray structures are due to a too low hydrogen bond distance of C-02 to G-N2 of the two C·G base pairs. The respective values of 2.48 and 2.28 Å are extremely small compared to the 2.86 Å of the ideal Watson−Crick hydrogen bond. The hydrogen bond index describes the deviation from the ideal Watson−Crick base pairing, and therefore also, a lowering in the hydrogen bond distances leads to a large hydrogen bond index.

Nevertheless the X-ray structure of the ImPyPyPy-*β*-Dp complex and the simulations of both complexes show a small enlargement of the hydrogen bond quality index at the target A·T base pairs. This is attributed to lengthening (about 0.2 Å compared to the ideal bond length) T-N3 to A-N1. The difference between both complexes (Hp versus Py) again is not sufficient to explain the reduction in the binding affinity. Thus, these results agree with the suggestions and conclusions described previously. A more detailed analysis of these still ongoing simulations will be given elsewhere.

The X-ray studies of the d(CCAGATCTGG)$_2$ complexes show that the Hp OH group lengthens the hydrogen bond distances between the amides of the polyamide ligand and the DNA. In the case of ImHpPyPy-*β*-Dp, the H-bond distances are up to 1.0 Å larger than in the ImPyPyPy-*β*-Dp complex.

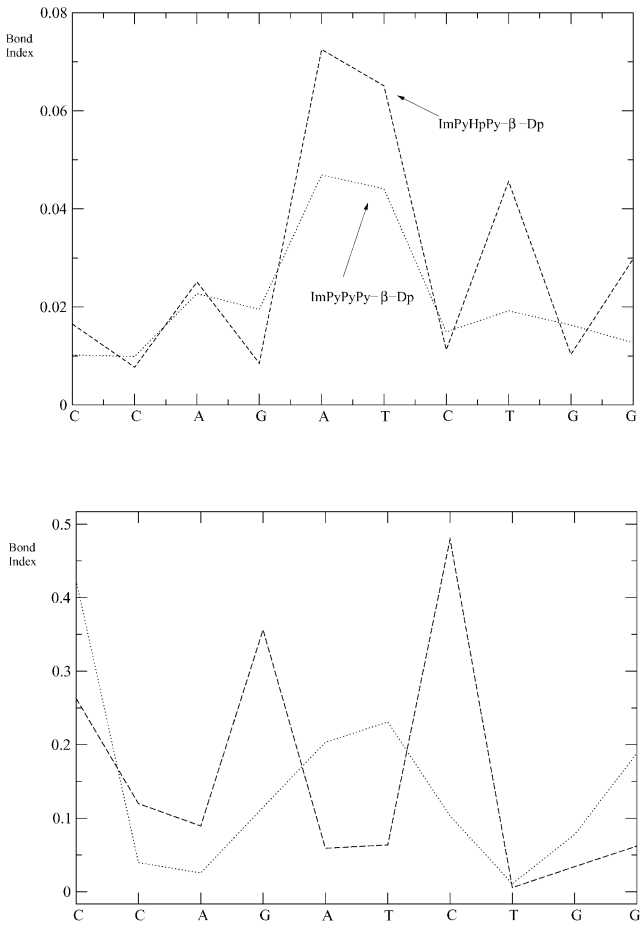


Figure 5. Graph on the top shows the mean values of the hydrogen-bond quality index of the simulations of d(CCAGATCTGG)$_2$ complexed with ImPyHpPy-β-Dp (dashed) and with ImPyPyPy-β-Dp (dotted). The bottom graph shows the respective values of the X-ray structures. The two peaks in the X-ray structure of the ImPyHpPy-β-Dp complex are due to a too low hydrogen-bond distance (discussed in the text) and thus do not represent a base pair melting.

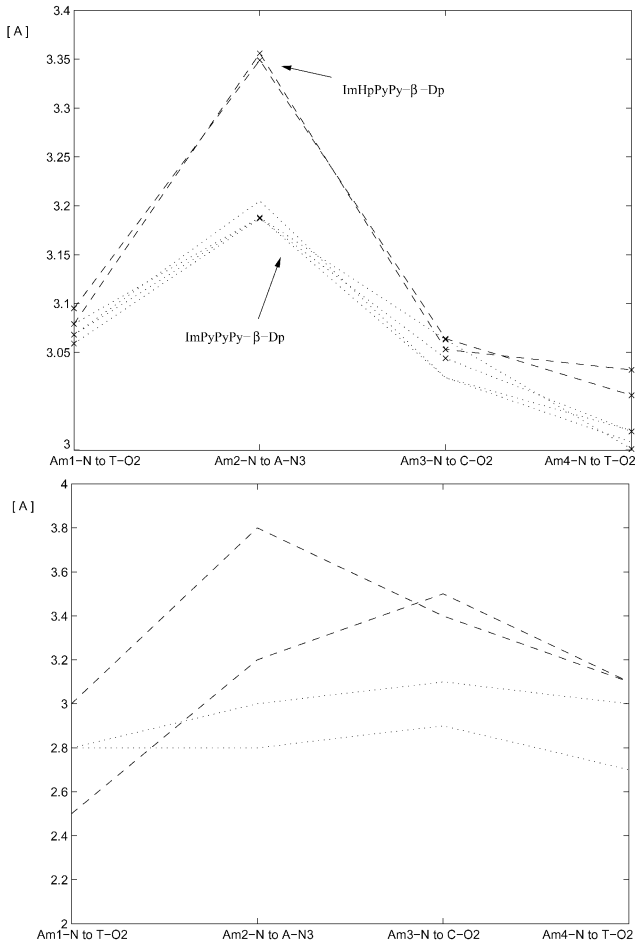


Figure 6. Distances [Å] between the hydrogen-bond acceptor atoms of the DNA and the amide donor nitrogens of the polyamide ligand are shown in these graphs. For each simulation or X-ray structures, two values exist because the ligand binds in a dimeric arrangement. The mean values of the simulation (top, small x indicates the simulations starting from the X-ray structures) indicate that in the ImHpPyPy-β-Dp (dashed line) complex the hydrogen bonds are longer than in the ImPyPyPy-β-Dp (dotted line) complex. The respective X-ray values (bottom) show in principle a similar trend but with different quantities.

Although the simulations show the same trend, the maximal difference is only in the range of 0.2 Å, as shown in Figure 6.

The X-ray structures are the starting structures of the simulation. It can be seen that although in the starting structure both ligands are not symmetrically (the curves of the two ligands do not coincide in Figure 6, bottom) bound, our simulations reflect the expected symmetric nature of the complex (the curves coincide). In the simulations, only one hydrogen bond is substantially elongated, namely Am2-N to A-N3. The OH group of the hydroxypyrrole is located next to this amide group. To determine the energetic contribution of this elongation (0.2 Å; see Figure 6), we performed quantum chemical calculations in which we varied the distance of a model compound. As a model molecule (Figure 7), we used formamide which interacts with its amide group with the 1-nitrogen of 4-amino-pyrimidine. The 4-amino-pyrimidine and the formamide mimic the interaction between adenine and the amide group of the polyamide ligands. The geometry of the interaction was taken from a snapshot of the simulation. This model molecule extremely simplifies the interaction, but a more complete model would not allow the estimation of the energy of only one hydrogen bond. Furthermore, we are only interested in the order of magnitude of the energy, and therefore, this model should be sufficient. Figure 7 shows the energy, calculated by means of density functional theory (B3LYP), as a function of the nitrogen−nitrogen distance.

It can be seen that the mean values of the Am2-N to A-N3 hydrogen bond distance of our simulations coincide with the minimum energy distance of the quantum chemical calculations (about 3.15 Å). The energetic penalty introduced by the hydrogen bond elongation is in the range of 0.15 kcal/mol, which is by far too small to explain the binding reduction. Thus, we believe that the structural suggestions based on the X-ray structures must be modified. Although in principle the simulation results agree with the suggestions from the X-ray structures, the structural differences between both complexes are not sufficient to explain the reduction in the binding constant.

We propose that hydration effects modulate the binding differences of the two polyamide ligands. In the complexed state, both DNA−polyamide complexes should be hydrated similarly because the solvent accessible surface of both complexes is the same. The differing hydroxyl group is hidden and thus not accessible for water. In contrast, in the unbound state we expect that the ImHpPyPy-β-Dp is better hydrated than ImPyPyPy-β-Dp because of the hydrophilicity of the OH group. In the

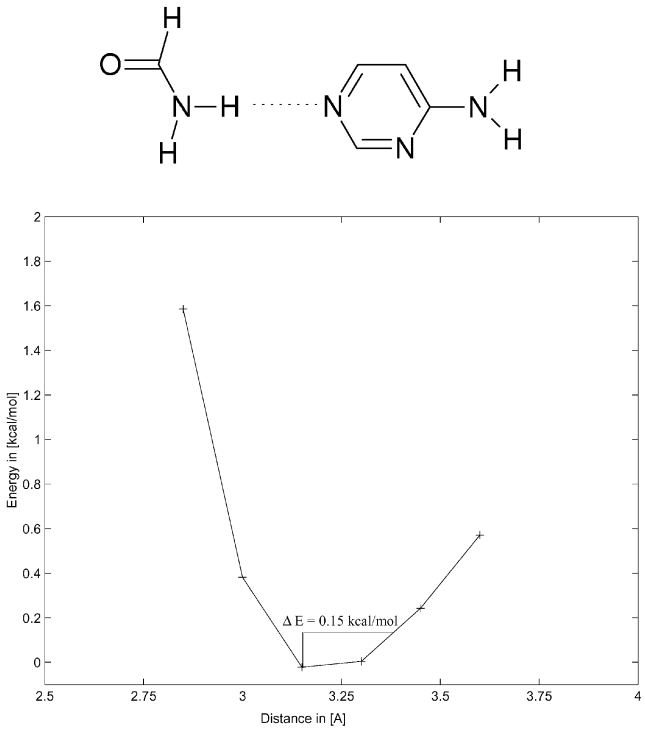


Figure 7. Graph shows the quantum mechanically calculated energy of our model compound (top) as function of the hydrogen-bond donor−acceptor distance (N−N distance) which is indicated by the dashed line. The maximum elongation of the hydrogen-bond distance found (deviation from 3.15 to 3.35 Å, shown in Figure 5) raised the energy by about 0.15 kcal/mol. The results were obtained by applying density function theory (B3LYP) using a 6-31+g(d) double-ζ basis set.

simulations of the free ligands, the first hydration shell (number of water nearer than 3.4 Å) of the ImHpPyPy-β-Dp ligand consists of one water molecule more than the ImPyPyPy-β-Dp counterpart raising up a difference of two water molecules up to the second solvation shell (nearer than 5 Å). The difference in the hydration free energies was estimated by means of ΔG calculations (see Methods section, simulation A) to be in the range of 3.9 kcal/mol. The strength of the additional H bond between the hydroxypyrrole OH and of Am4 to DNA should be reduced because of its bifurcated character (Figure 1). Several studies indicate that a bifurcated hydrogen bond is less stable than a two center hydrogen bond.[86,87] It is even assumed that three centered hydrogen bonds can lose up to one-half of their stability[88]. Thus, the additional but weak hydrogen bond is not able to compensate the better hydration, and therefore, we believe that the hydration of the hydroxypyrrole determines the reduction in the binding constant.

Singh et al.[89] performed similar calculations in which imidazole and pyrrole containing ligands were investigated. They also concluded that the desolvation energy is an unfavorable factor which contributes significantly to the binding affinities of the ligands. Nevertheless, the introduction of a hydrophilic group of course does not always reduce the binding constant. The interplay between unfavorable desolvation and favorable formation of specific hydrogen bonds is highly sensitive. For example, in the previously mentioned case of the imidazole versus pyrrole ligands, the impact of this substitution is highly dependent on the site at which it is introduced. In the case of the polyamide ligands, such a nonadditivity for Hp/Py pair substitutions[22,10] is also observed. For example, the addition of multiple Hp residues does not additively reduce the binding constant. The hydration presumably is also not additive; thus, this is at least consistent with our hypothesis of the hydration influence. To prove our suggestions, similar calculations on several different ligands are necessary and will be performed in future research.

Summary and Conclusion

We performed molecular dynamics simulations as well as free energy and quantum chemical calculations in order to explain why ImPyPyPy-β-Dp binds stronger to DNA than to ImHpPyPy-β-Dp. Although the OH group of hydroxypyrrole introduces sequence specificity and builds an H-bond contact to DNA, it reduces the binding constant by 1.2 kcal/mol. The free energy calculations agree optimally with these experimental findings. The discrepancies between the two complexes derived from X-ray structures are in principle observed in our simulation, but on the basis of our simulations, we see that the structural distortions are less emphasized. Thus, they are not sufficient to determine the reduction of the binding constant. The differences between X-ray and theoretical results are attributed to a buffer molecule cocrystallizing with the ImHpPyPy-β-Dp−DNA complex. Furthermore, we suggest that the stronger hydration of the ImHpPyPy-β-Dp ligand in the uncomplexed state is responsible for the reduction in the binding constant. Additionally, we propose an indirect readout of the DNA which could contribute to the differentiation between A·T and T·A.

Acknowledgment. This work was supported by a grant of the Austrian Science Fund (Grant No. P13845-TPH). We thank the referees for their contributions.

JA0277778

(86) Bureiko, S. F.; Golubev, N. S.; Pihlaja, K. *J. Mol. Struct.* **1999**, 297−301.
(87) Yang, J.; Christianson, L. A.; Gellman, S. H. *Org. Lett.* **1999**, *1*, 11−13.
(88) Giguerre, P. A. *J. Chem. Phys.* **1987**, *87*, 287−300.
(89) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673−7677.