Steric Constraints Dependent on Nucleobase Pair Orientation Vary in Different DNA Polymerase Active Sites

Frank Streckenbach, Gopinath Rangam, Heiko M. Möller, and Andreas Marx*[a]

The integrity of the genome relies heavily on the ability of DNA polymerases to efficiently catalyze selective DNA synthesis in a template-directed manner during DNA replication, repair and recombination. DNA polymerases involved in DNA replication achieve selective information transfer to the offspring according to the Watson–Crick rule with intrinsic error rates as low as one mistake within one million synthesized nucleotides.[1] On the other hand, recently discovered DNA polymerases involved in DNA repair, translesion synthesis (TLS) and somatic hypermutation exhibit strikingly low fidelity.[2]

Geometric factors are widely cited to significantly contribute to DNA polymerase selectivity. Thus, replicative high-fidelity DNA polymerases are believed to mostly select the canonical nucleotide based on the close steric complementarity of the nascent base pair to the active site of the enzyme.[1] In contrast, TLS DNA polymerases possess flexible and more solvent-accessible active sites.[1,2] This model of varied active-site tightness was probed by functional studies of DNA polymerases with tailored nucleotide analogues. Kool et al. have employed increasingly larger thymidine analogues to investigate sterics and polar effects within the active site acting on the nucleobase.[3] They used nonpolar nucleobase surrogates with a limited ability to form hydrogen bonds. Moreover, the size of these compounds increased incrementally through halogen substitution of the oxygens of thymidine. We have developed and utilized 4'-alkyl-modified thymidines

with increasing steric bulk for probing steric effects on DNA polymerase function (Figure 1 A).[4] We employed alkyl groups to minimize the potential effects of the modification on hydro-
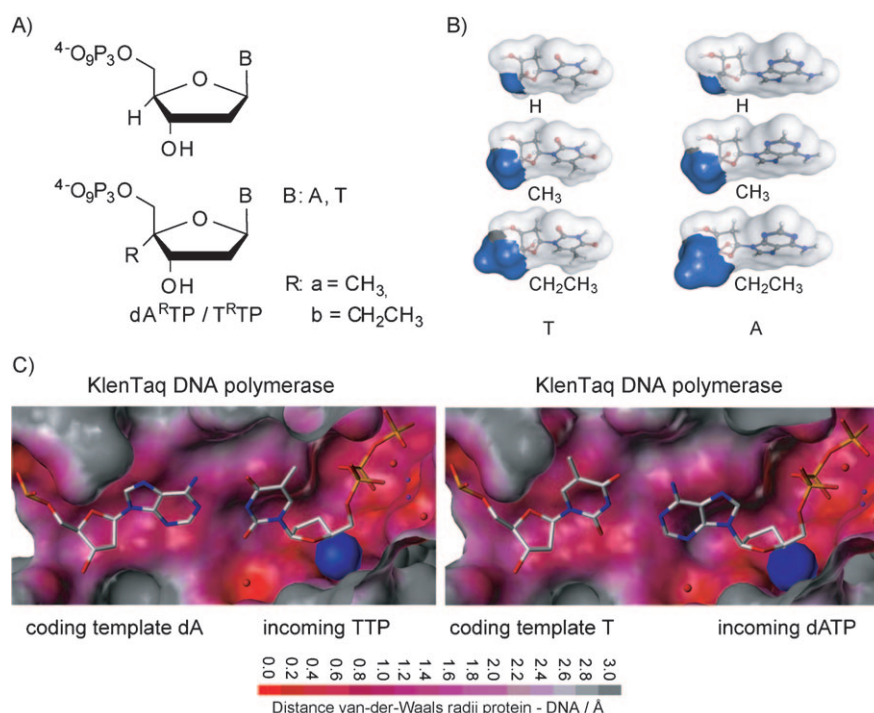

Figure 1. A) Size-augmented 2'-deoxynucleoside-5'-triphosphates as steric probes. B) Models showing the increasing steric demand of the 4'-residues (blue) for the six nucleoside derivatives visualized by a Connolly surface by using SYBYL 7.2 MOLCAD module. C) Graphic representation of the distances of the van der Waals radii between DNA and protein for KlenTaq DNA polymerase. The separated surfaces were calculated from PDB files (1QSY for dATP, 1QTM for TTP) by using SYBYL 7.2 (Tripos). View of the nascent base pair (incoming and templating nucleotides). The blue sphere illustrates the modification position. The distance–colour scale is given in the figure.

gen bonding patterns and nucleotide confirmation. Both approaches suggest varied contributions of steric effects on DNA polymerase selectivity.

One long-known but poorly understood feature of DNA polymerases is that enzyme selectivity can also vary depending on the orientation of the nascent nucleobase pair.[1d,5,6] Only a few mechanistic insights into the origin of this observation have been obtained in the past.[6] Similarly to mismatch processing,[7] varied conformations of the enzyme–substrate complex imposed by subtle degrees of geometric difference of the nascent nucleotide pairs might cause the observed effects. Herein, we describe the first investigations and insights along these lines by employing new steric probes.

[a] F. Streckenbach, Dr. G. Rangam, Prof. Dr. H. M. Möller, Prof. Dr. A. Marx
Department of Chemistry and Konstanz Research School Chemical Biology
Universität Konstanz
Universitätsstrasse 10, 78457 Konstanz (Germany)
Fax: (+49) 7531-88-5140
E-mail: andreas.marx@uni-konstanz.de

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/cbic.200900123.

Until now, only 4'-alkylated thymidines were employed in functional studies of DNA polymerases (Figure 1). Figure 1 C illustrates the position of the modification as a blue sphere inside the active site of a KlenTaq DNA polymerase (PDB ID: 1QSY, 1QTM). A graphical representation of the two complementary nascent base pairs dA-TTP/T-dATP is depicted. In order to investigate steric effects on DNA polymerases as a function of nucleotide pair orientation, we synthesized the complementary deoxyadenosines (dA). Indeed, by employing the T- and dA-derived steric probes in functional DNA polymerase studies, varied actions of DNA polymerases on these steric probes were observed.

First, we set out to synthesize the respective steric probes **dA^{R}TP** (Figure 1). We developed a route that allows the efficient synthesis of the corresponding **T^{R}TP** analogues as well. Our approach is based on a strategy we recently developed for the synthesis of 4'-methyl- and 4'-ethyl-2'-deoxyuridines.[8] Following a similar strategy allowed for the synthesis of the target compounds (Scheme 1). Starting from the intermediates **2a/b**, which are readily available on multigram scale, we successfully introduced the respective nucleobase by Vorbrüggen glycosylation to yield **3a/b-A/T**.[9] Next, saponification and de-

oxygenation of the 2'-hydroxyl group followed by cleavage of the protection groups yielded **1a/b-A/T**. We note that, for the successful hydrogenolysis of the benzyl ethers in **4a/b-A**, basic conditions were required in order to suppress side reactions that occur to a significant extent in the absence of NaOH. The corresponding triphosphates were synthesized employing standard conditions.[10]

In order to gain insight into potential effects of these modifications on the sugar conformations, we performed conformational analysis based on $^3J_{H,H}$ values deduced from 1D ^{1}H NMR spectra recorded in D_2O as described by Altona and co-workers (Table S1 in the Supporting Information).[11] Unmodified nucleosides were found to adopt about 60–70 % of the southern conformation [(S)-type], while **1a/b-A/T** were found to adopt approximately 50 % of the (S)-type conformations; this indicates that 4'-alkylation had only a moderate impact on sugar puckering in solution in general. We also note that the observed effects on the conformational equilibria are independent of the nature of the nucleobase.

Next, we tested the effect of probes **dA^{R}TP** and **T^{R}TP** on the exonuclease-deficient variant of the Klenow fragment of *Escherichia coli* DNA polymerase I (KF−), a relatively high-fidelity enzyme extensively used as a model in investigations of intrinsic DNA polymerase mechanism and function. In addition, DNA polymerase IV of *Sulfolobus solfataricus* P2 (Dpo4), a member of the error-prone, Y-family, TLS polymerases, was investigated. To monitor polymerase function, we used a gel-based single nucleotide insertion assay, in which a 24 nt primer was designed to hybridize with a corresponding 36 nt template strand that codes for the insertion of the respective canonical dNTP adjacent at the primer 3' end. To gain initial insights, we studied nucleotide insertion qualitatively with different concentrations of the respective dNTP analogues (Figure 2). In accordance with earlier studies, we found that **T^{R}TPs** were substrates for KF− and were incorporated opposite the canonical dA. The same was found for the corresponding **dA^{R}TPs**. We note that, under the applied conditions (concentrations in the range of K_M), the gel analysis showed higher levels of incorporation for the modified substrates, probably reflecting the higher k_{cat} values (Table 1). For both **T^{R}TPs** and **dA^{R}TPs**, incorporation opposite noncanonical templates was significantly diminished as expected. Steady-state kinetic analysis[11] of the matched cases showed that **dA^{H}TP** was more efficiently processed than was **T^{H}TP** in absolute values (see Table 1). Interestingly, the opposite was observed for the size-augmented nucleotides. To some extent, size augmentations were better tolerated in the thymidines **T^{R}TP** (relative efficiencies 4'-alkyl/4'-H) than they were in their **dA^{R}TP** counterparts (see the relative efficiencies in Table 1). This effect was most significant for the bulkiest, 4'-ethylated analogues. However, for each nucleotide, the differences between 4'-methylation and 4'-ethylation were small. A partly different picture was observed for Dpo4. Again, alkylated **T^{R}TPs** were processed more efficiently than were the corresponding **dA^{R}TP** analogues in terms of relative efficiency, while unmodified **dA^{H}TP** again yielded a higher absolute value. When comparing the relative incorporation efficiencies of **dN^{Me}TPs** by both enzymes, it was evident that the error-
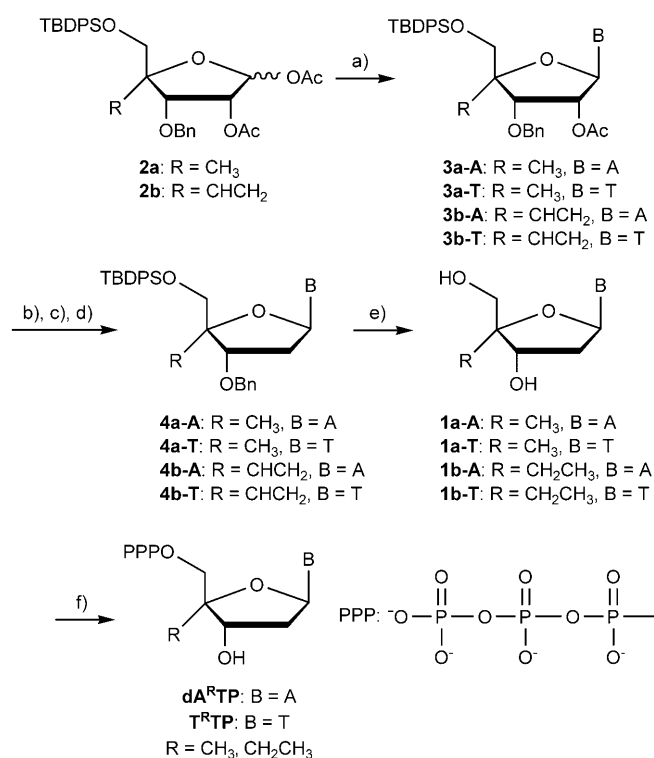
Scheme 1. Synthesis of modified nucleotides. a) Thymine, N,O-bis(trimethylsilyl)acetamide, TMSOTf, CH$_3$CN, reflux, 2 h, 62 % (Me)/61 % (Et); N^6-Benzoyladenine, TMSOTf, CH$_3$CN, 0 °C, 1 h, 55 % (Me)/48 % (Et); b) NaOMe, MeOH, RT 1–12 h; c) PhOCSCl, DMAP, CH$_3$CN, RT, 1 h; d) nBu$_3$SnH, AIBN, toluene, reflux, 1 h, 72 % (T^{Me})/56 % (T^{Et})/73 % (A^{Me})/78 % (A^{Et}) after three steps; e) 10 % Pd/C, H$_2$, THF/EtOH, reflux, 24 h, followed by TBAF, THF, RT, 7 h, 61 % (T^{Me})/63 % (T^{Et}); 10 % Pd/C, H$_2$, THF/EtOH, 1 N NaOH, reflux, 5–10 d, 94 % (A^{Me})/98 % (A^{Et}); f) i) POCl$_3$ in PO(OMe)$_3$, (proton sponge for T) 0 °C -> RT, 2–4 h, ii) Bu$_3$N, (Bu$_3$NH)$_2$H$_2$P$_2$O$_7$ in DMF, 0 °C, 15 min, iii) TEAB buffer, 40 min, 25 % (T^{Me}TP)/ 41 % (T^{Et}TP)/9 % (dA^{Me}TP)/11 % (dA^{Et}TP). For details, see the Supporting Information.

A)
24-mer 5'-GTG GTG CGA AAT TTC TGA CAG ACA
36-mer 3'-CAC CAC GCT TTA AAG ACT GTC TGT **N**CT GTC TGC GTG
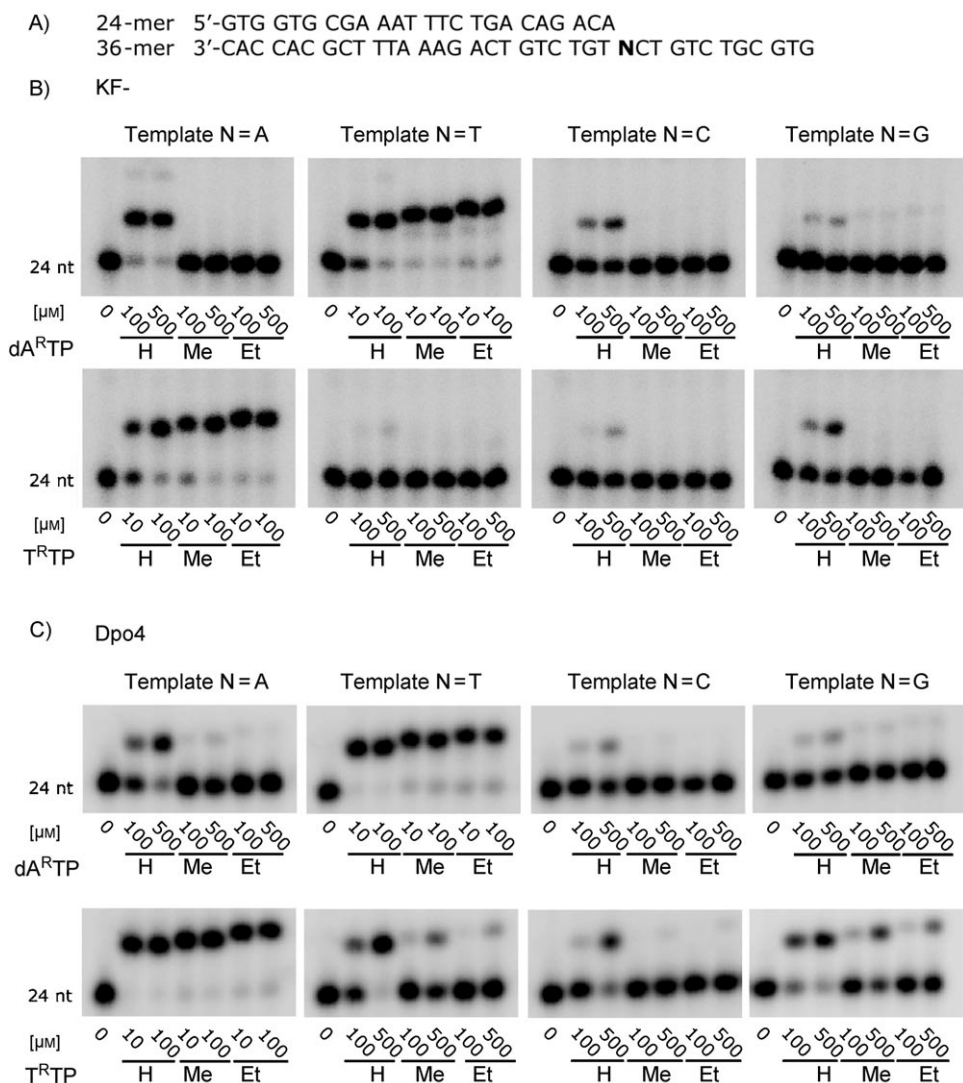
B) KF−

C) Dpo4

Figure 2. Nucleotide insertion catalyzed by KF− and Dpo4: A) Primer/template sequence. B) Reactions catalyzed by KF−. C) Reactions catalyzed by Dpo4. Conditions: Primer/template complex (150 nм), KF−/Dpo4 (10 nм), 37 °C, 15 min matched/45 min mismatched. In each case, the nucleotide concentrations [μM] are shown in the figure. Further experimental details are described in the Supporting Information.

unmodified **dA**[H] is more efficient than was **T**[H]. However, the opposite was found for the size-augmented probes. It appears that the bulkier purine **dA**[R] nucleotides were subject to more steric constraints within the active site of the enzymes than were the **T**[R] pyrimidines. These different properties might well be the cause for the observation that the selectivity of nucleotide incorporation might vary depending on the nascent nucleotide pair orientation. This assumption is further corroborated by the finding that the size-augmented **T**[R]**TPs** are more efficiently incorporated opposite noncanonical template nucleotides than their adenosine counterparts did. Structural investigations show that the sugar moiety of the incoming dNTP is fully embedded within the substrate binding pocket of DNA polymerases and is an integral part of substrate recognition processes.[1,7,13] Hence, editing of the incoming dNTP sugar might provide the enzyme with additional paths to achieve canonical base-pair formation through the indirect readout of aberrant sugar conformations.

In conclusion, through the development of novel steric probes and their employment in functional DNA polymerase studies, we have gained new insights into the complex mechanisms that govern DNA polymerase substrate recognition. These findings provide experimental evidence that packing of the nucleotide sugar residue varies within the orientation of the nascent nucleotide pair as well as in different polymerase families. These findings also support the model of varied active site tightness as a cause for the varied selectivity observed among different DNA polymerases.

prone Dpo4 processed the bulkier nucleotides with higher relative efficiencies than did KF−.

Interestingly, Dpo4 was able to incorporate the **T**[R] residues opposite noncanonical template dG and T with measurable efficiencies under single completed hit conditions.[12] The corresponding **dA**[R]**TP** analogues were processed significantly more selective, and only minor amounts of incorporation products were observed opposite noncanonical templates.

The results presented are interesting in many respects. First, the observation that the error-prone TLS DNA polymerase Dpo4 more efficiently incorporated the size-augmented nucleotides than the more selective KF− is consistent with the model that high-fidelity DNA polymerases form more rigid binding pockets that tolerate less geometric deviation while low-fidelity enzymes exhibit more flexibility, which leads to decreased fidelity. This is in agreement with earlier findings. Secondly, for both enzymes, we found that the incorporation of

Acknowledgements

Table 1. Steady-state kinetic analyses of nucleotide insertion by KF— and Dpo4[a]

DNA polymerase	Template N[b]	Nucleotide	K_M [μM]	k_{cat} [min^{-1}]	k_{cat}/K_M [mM^{-1} min^{-1}]	Relative efficiency (alkyl/H)
KF—	dA	TTP	0.05 ± 0.01	6.04 ± 0.15	121 000	1
		T^{Me}TP	3.50 ± 0.44	10.9 ± 0.6	3 110	0.03
		T^{Et}TP	2.45 ± 0.19	14.8 ± 0.6	6 080	0.05
	T	dATP	0.02	8.28 ± 0.34	414 000	1
		dA^{Me}TP	1.29 ± 0.09	11.0 ± 0.2	8 530	0.02
		dA^{Et}TP	1.90 ± 0.31	10.1 ± 0.5	5 320	0.01
Dpo4	dA	TTP	5.36 ± 0.66	18.5 ± 0.8	3 450	1
		T^{Me}TP	4.79 ± 0.28	10.1 ± 0.2	2 110	0.61
		T^{Et}TP	12.5 ± 0.59	6.72 ± 0.14	538	0.16
	T	dATP	3.96 ± 0.48	22.6 ± 0.8	5 710	1
		dA^{Me}TP	11.8 ± 2.2	9.17 ± 0.78	777	0.14
		dA^{Et}TP	11.4 ± 1.4	5.21 ± 0.26	457	0.08
	dG	TTP	807 ± 123	8.32 ± 0.47	10.3	1
		T^{Me}TP	274 ± 33	0.81 ± 0.05	2.96	0.29
		T^{Et}TP	233 ± 36	0.29 ± 0.01	1.24	0.12
	T	TTP	877 ± 122	5.80 ± 0.27	6.61	1
		T^{Me}TP	372 ± 68	0.71 ± 0.06	1.91	0.29
		T^{Et}TP	160 ± 24	0.19 ± 0.01	1.19	0.18

[a] The data shown represent averages derived from experiments repeated three times. [b] Primer-template substrates used: 5′-GTG GTG CGA AT TTC TGA CAG ACA primer (24 nt); 3′-CAC CAC GCT TTA AAG ACT GTC TGT **N**CT GTC TGC GTG template (36 nt).

[1] a) H. Echols, M. F. Goodman, *Annu. Rev. Biochem.* **1991**, *60*, 477–511; b) M. F. Goodman, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10493–10495; c) T. A. Kunkel, *J. Biol. Chem.* **2004**, *279*, 16895–16898; d) T. A. Kunkel, K. Bebenek, *Annu. Rev. Biochem.* **2000**, *69*, 497–529; e) E. T. Kool, J. C. Morales, K. M. Guckian, *Angew. Chem.* **2000**, *112*, 1046–1068; *Angew. Chem. Int. Ed.* **2000**, *39*, 990–1009; f) E. T. Kool, *Annu. Rev. Biochem.* **2002**, *71*, 191–219; g) W. Yang, R. Woodgate, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 15591–15598.

[2] a) U. Hübscher, G. Maga, S. Spadari, *Annu. Rev. Biochem.* **2002**, *71*, 133–163; b) M. F. Goodman, *Annu. Rev. Biochem.* **2002**, *71*, 17–50; c) A. Marx, D. Summerer, *ChemBioChem* **2002**, *3*, 405–407; d) S. Prakash, R. E. Johnson, L. Prakash, *Annu. Rev. Biochem.* **2005**, *74*, 317–353.

[3] a) T. W. Kim, J. C. Delaney, J. M. Essigmann, E. T. Kool, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15803–15808; b) S. Mizukami, T. W. Kim, S. A. Helquist, E. T. Kool, *Biochemistry* **2006**, *45*, 2772–2778; c) T. W. Kim, L. G. Brieba, T. Ellenberger, E. T. Kool, *J. Biol. Chem.* **2006**, *281*, 2289–2295;

d) H. O. Sintim, E. T. Kool, *J. Am. Chem. Soc.* **2006**, *128*, 396–397; e) H. O. Sintim, E. T. Kool, *Angew. Chem.* **2006**, *118*, 2008–2013; *Angew. Chem. Int. Ed.* **2006**, *45*, 1974–1979; f) A. P. Silverman, Q. Jiang, M. F. Goodman, E. T. Kool, *Biochemistry* **2007**, *46*, 13874–13881.

[4] a) D. Summerer, A. Marx, *Angew. Chem.* **2001**, *113*, 3806–3808; *Angew. Chem. Int. Ed.* **2001**, *40*, 3693–3695; b) M. Strerath, D. Summerer, A. Marx, *ChemBioChem* **2002**, *3*, 578–580; c) M. Strerath, J. Cramer, T. Restle, A. Marx, *J. Am. Chem. Soc.* **2002**, *124*, 11230–11231; d) J. Cramer, M. Strerath, A. Marx, T. Restle, *J. Biol. Chem.* **2002**, *277*, 43593–43598; e) F. Di Pasquale, D. Fischer, D. Grohmann, T. Restle, A. Geyer, A. Marx, *J. Am. Chem. Soc.* **2008**, *130*, 10748–10757; f) K.-H. Jung, A. Marx, *Cell. Mol. Life Sci.* **2005**, *62*, 2080–2091; g) J. Cramer, G. Rangam, A. Marx, T. Restle, *ChemBioChem* **2008**, *9*, 1243–1250.

[5] T. A. Kunkel, K. Bebenek, *Biochim. Biophys. Acta, Gene Struct. Expression* **1988**, *951*, 1–15.

[6] D. T. Minnick, L. Liu, N. D. F. Grindley, T. A. Kunkel, C. M. Joyce, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1194–1199.

[7] S. J. Johnson, L. S. Beese, *Cell* **2004**, *116*, 803–816.

[8] G. Rangam, N. Z. Rudinger, H. M. Müller, A. Marx, *Synthesis* **2005**, 1467–1472.

[9] a) H. Vorbrüggen, K. Krolikiewicz, B. Bennua, *Chem. Ber.* **1981**, *114*, 1234; b) H. Vorbrüggen, G. Höfle, *Chem. Ber.* **1981**, *114*, 1256.

[10] a) M. Yoshikawa, T. Kato, T. Takenishi, *Bull. Chem. Soc. Jpn.* **1969**, *42*, 3505–3508; b) J. Ludwig, *Bioact. Mol.* **1987**, *3*, 201–204; c) T. Kovacs, L. Ötvös, *Tetrahedron Lett.* **1988**, *29*, 4525–4528.

[11] a) F. A. A. M. De Leeuw, C. J. Altona, *J. Comput. Chem.* **1983**, *4*, 428–437; b) L. J. Rinkel, C. J. Altona, *J. Biomol. Struct. Dyn.* **1987**, *4*, 621–649.

[12] S. Creighton, L. B. Bloom, M. F. Goodman, *Methods Enzymol.* **1995**, *262*, 232–256.

[13] P. J. Rothwell, G. Waksman, *Adv. Protein Chem.* **2005**, *71*, 401–440.