Base Pairing

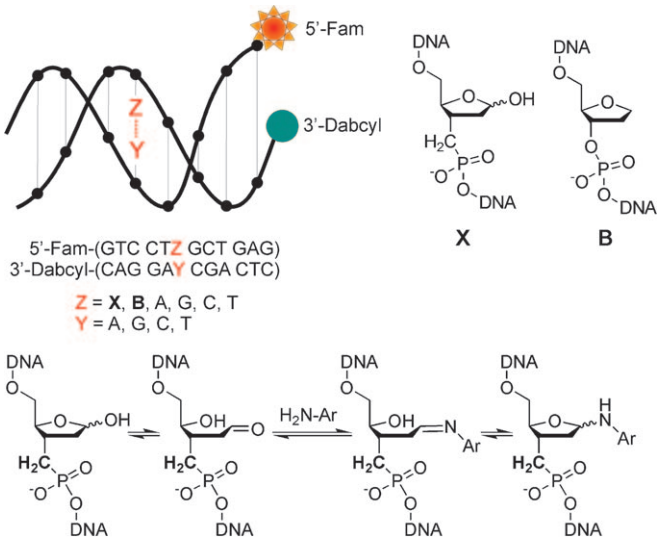
DOI: 10.1002/anie.201005300

A Parallel Screen for the Discovery of Novel DNA Base Pairs**

*Oezlem Yaren, Markus Mosimann, and Christian J. Leumann**

The genetic information in DNA is encoded in the sequence of the four nucleobases, and its readout is based on the highly specific Watson–Crick pairing A–T and G–C. Over the years there has been growing interest in expanding this genetic alphabet by adding novel, orthogonal base pairs. Such additional base pairs could be used in biotechnology to introduce nonnatural amino acids into engineered proteins,[1–3] and to create novel tools for identifying aberrant or exogenous, disease-related nucleic acids.[4–8] The degree of freedom in designing novel base pairs, however, is limited by the structural scaffold of the double helix. The most logical approach is based on structurally isomorphic bases with an alternative Watson–Crick-like hydrogen-bonding network.[6,9,10] Alternatively, nonisomorphic aromatic heterocycles can serve as base replacements; they recognize each other by means other than classical hydrogen bonding and can be replicated and transcribed with high fidelity and similar kinetic properties.[11–16] This latter approach provides access to a larger structural space for potential base replacements. Unfortunately this approach is hampered by a lack of design rules which makes it necessary to screen a large number of potential candidates. From each potential base analogue the corresponding nucleoside must first be synthesized and incorporated into oligonucleotides before its coding properties can be evaluated. This is associated with multistep syntheses for each nucleoside candidate which is a time-consuming procedure.

To address these drawbacks we set out to design an assay for the rapid parallel screening of novel potential base candidates out of a library of heterocyclic amines. In a first step we identified selective, high-affinity binders to natural nucleobases. The assay is based on a dodecamer DNA duplex which carries an abasic site **X** in the center and a fluorescence-quencher pair on one end of the duplex (Scheme 1). **X** structurally deviates from a natural DNA abasic site by the replacement of O(3′) by a CH_2 group which renders it chemically stable towards base- or heat-induced strand cleavage at the 3′-end.[17] These duplexes were then incubated with a library of heterocyclic amines in a parallel fashion. During this treatment the amines become covalently attached to **X** through the formation of a hemiaminal, resulting in *exo*-amino nucleosides that are structurally similar to the natural nucleosides. The thermal stabilities of these duplexes, which reflect the relative affinity of each amine to its target, were then determined by parallel fluorescence measurements of the melting temperature T_m (see the Supporting Information).



Scheme 1. Working principle of the assay; see text for details. Dabcyl: 4-[4-(dimethylamino)phenylazo]benzoic acid; Fam: 6-carboxyfluorescein.

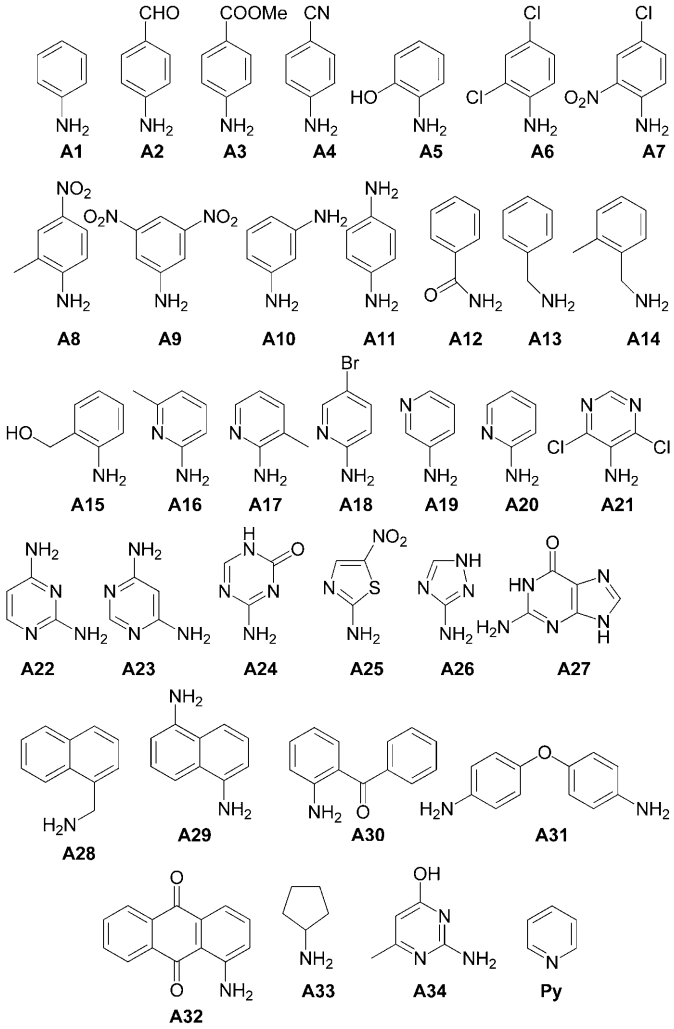
To validate the assay we searched a library of 34 randomly chosen, commercially available heterocyclic amines (Scheme 2) for individuals that bind selectively with high affinity to each of the four natural nucleobases. Duplexes (0.6 μM) were incubated with a 1000-fold excess of the amines in buffer at pH 8 for two days at 55 °C to form the hemiaminals. Each sample was then directly subjected to fluorescence T_m analysis. The resulting T_m data are graphically represented in Figure 1; numeric values are given in the Supporting Information.

As can be seen from Figure 1, all of the T_m values are between 35 and 45 °C. The abasic unit **X** in the absence of any amine ($T_m = 38$ or 39 °C) as well as the matched ($T_m = 54$ or 58 °C) and mismatched ($T_m = 40$–48 °C) duplexes (legend to Figure 1) served as controls. While, not unexpectedly, none of the amines outperforms the matched natural bases, a series of amines display strong binding to one of the four natural bases. The three highest affinity pairs found are **A29**–C, **A26**–G, and **A19**–C. All three amines show considerable selectivity as the T_m values of these complexes are 3–4 °C lower than those of complexes with the other three natural bases. All three novel base pairs are more stable than a classical natural mismatch (the purine pair A–G and the wobble pair T–G are not considered as classical mismatches). **A20** on the other hand is

[*] O. Yaren, Dr. M. Mosimann, Prof. C. J. Leumann
Department of Chemistry and Biochemistry, University of Bern
Freiestrasse 3, 3012 Bern (Switzerland)
Fax: (+41) 31-631-3422
E-mail: leumann@ioc.unibe.ch

[**] Financial support from the Swiss National Science Foundation (grant no. 200020-115913) is gratefully acknowledged.

Supporting information for this article (for example experimental details) is available on the WWW under http://dx.doi.org/10.1002/anie.201005300.



Scheme 2. The library of heterocyclic amines.

an interesting example of a universal base since it does not discriminate between complementary bases. Interestingly there are also amines that destabilize the double helix relative to the abasic site **X**. An example is the **A21**–T pair. As expected, pyridine, which cannot form a covalent bond with **X**, compared to the abasic site **X** alone. This demonstrates the importance of covalently anchoring the heterocyclic amines to the abasic site.

A series of control experiments were performed to confirm the proper working of the assay. Given the large excess of amines in the assay it was necessary to exclude that differential stabilization of duplexes just occurred by random binding of the heterocycles to the DNA. We therefore incubated the native duplex containing a matched base pair (Z–Y = A–T) with a subset of the library and found that neither of the heterocyclic amines affected the T_m of the duplex significantly (see the Supporting Information), thus ruling out interference by intercalation or groove binding. We then investigated whether the hemiacetal function of **X** is involved in covalently binding the amines. For this we prepared the control oligonucleotide d(TTT**X**TTT) and treated it with aniline (**A1**, Scheme 2) under the same

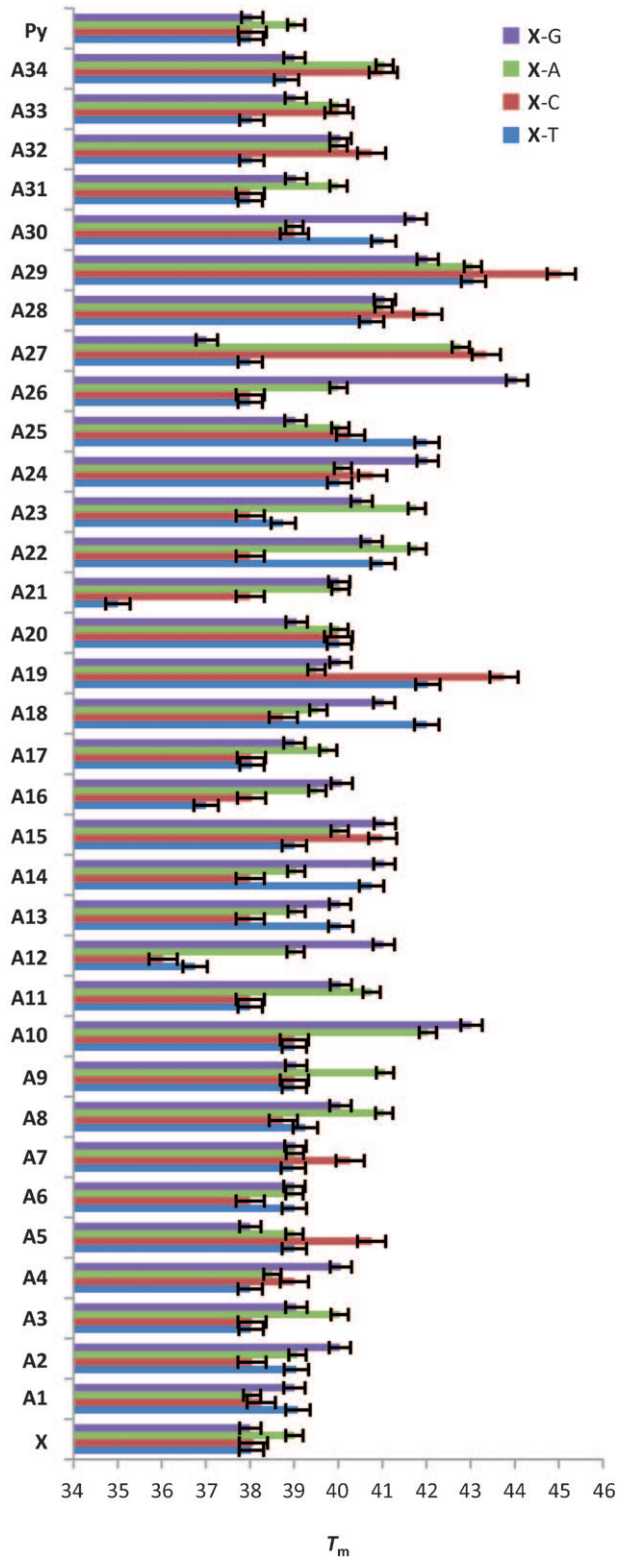
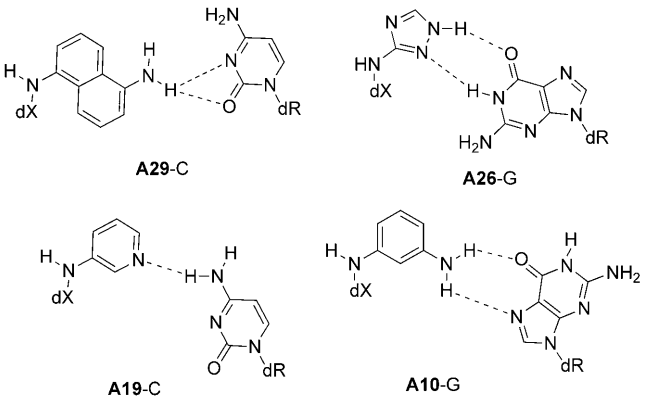


Figure 1. Graphical representation of the T_m as a function of the heterocyclic amines **A1**–**A34**. Experiments were performed in duplicate. T_m data of matched duplexes: T–A 54, C–G 58, A–T 54, G–C 58 °C; T_m of mismatched duplexes: T–C and C–T 41, A–C 40, A–G 47 °C, G–T 48 °C.

conditions. By ESI^- mass spectrometry we were able to prove formation of the corresponding hemiaminal (see the Supporting Information). In an additional experiment we wanted to verify the importance of the formation of the covalent bond between **X** and the amines for the proper working of the assay. We repeated the experiments with the whole amine library with the duplex containing an abasic site **B** instead of **X** opposite T (Z–Y = **B**–T, Scheme 1). Also in this case all measured T_m values are within a narrow range of (38 ± 0.5) °C and are thus not influenced by the amines (see the Supporting Information). Both experiments highlight the importance of the chemical reactivity at the anomeric center and thus verify that hemiaminal formation occurs under the assay conditions.

It is difficult to come up with a structural rationalization for the three high-affinity base pairs identified in the screen, as the amines can adopt either the α or the β configuration at the anomeric center of **X**, and as some of the amines contain multiple nonsymmetric amino functions, leading to additional structural variety. The structures depicted in Scheme 3 for the four most stable base pairs seem possible but are of course purely hypothetical.



Scheme 3. Possible structures for the four most stable base pairs identified in the screen (dR = 2′-deoxyribose, dX = C(3′)-methylene-2′-deoxyribose).

Several points must be taken into consideration in the interpretation of the results of this assay. First, the amines in this library show great variation in nucleophilicity at the amino function which will certainly influence the equilibrium position between hemiacetal and hemiaminal and the stability of the glycosidic bond. It is therefore likely that occupancy at the abasic site varies in each case and may not be 100%, despite the large excess of the amines in the assay. Thus, the measured T_m values do not necessarily reflect the maximum possible stabilization of a duplex but may be lower. This can cause false negative but not false positive results which would be worse.

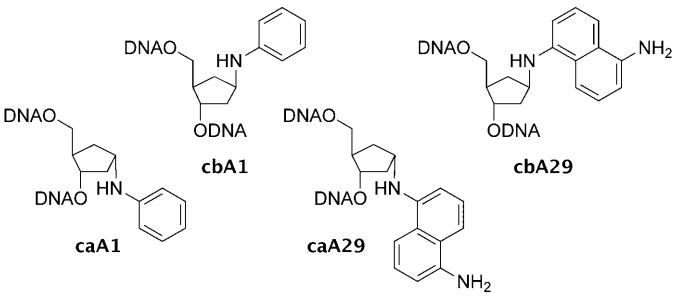
In an attempt to gauge the extent of hemiaminal formation we measured T_m values at different pH values (5.5–8.0) for amines **A10, A19, A26**, and **A29** as examples, reasoning that lowering the pH facilitates this reaction. Within the limits imposed by fluorescein (weak fluorescence below pH 6) we found no pH dependence of T_m (see the Supporting Information), indicating either pH-insensitivity of the reaction in this pH range or quantitative formation of hemiaminals for these cases.

This assay relies on the abasic-site analogue **X**, which deviates from a natural abasic site by the replacement of oxygen by CH_2 at C(3′). Thus the question arises of how much this change affects the T_m value. In previous work by ISIS it was shown that C(3′)-methylene nucleosides stabilize DNA duplexes only slightly.[18] Therefore this particular modification should not have a major impact on the outcome of the assay.

From earlier work on ribozymes it is known that nucleobases that are critical for catalysis can be dissected into an abasic residue **B** and the base. While the abasic residue itself leads to an inactive ribozyme, activity can be restored at least in part by addition of the corresponding free base or a similar unit (base rescue).[19–21] It is believed that the base fills the corresponding structural gap in the ribozyme and thus restores activity. As already mentioned, base rescue does not work in our screen as we observe no differences in T_m upon addition of any base to the **B**-containing duplex (see the Supporting Information). This is not surprising as the requirements for catalysis (kinetics) does not necessarily require high affinity, whereas duplex stability (thermodynamics) needs high affinity (covalent attachment) of the base to the gap.

Although the glycosidic bond between a sugar and a primary heterocyclic amine is less stable than that of a natural nucleoside in DNA and RNA, a variety of such *exo*-amino-nucleosides from natural or synthetic origins are known.[22,23] If a potential candidate is chemically too unstable to be useful, the corresponding derivative with a carbocyclic sugar replacing the 2′-deoxyribofuranose unit can be considered as a stable alternative. Such nucleosides are easily accessible by synthesis,[24] are structurally isomorphous to natural nucleosides, and are known to base-pair with complementary nucleosides much in the same way as their natural congeners.[25,26] At the same time this is a way to determine which anomeric form is responsible for the observed affinities.

To prove this we synthesized the Fam-labeled oligonucleotides with Z representing the carbocyclic nucleosides of the amines **A1** and **A29** in the α and β configuration (Scheme 4 and Table S2 in the Supporting Information) and measured T_m values with complementary oligonucleotides containing all



Scheme 4. Oligonucleotides (5′-Fam-(GTC CTZ GCT GAG), Scheme 1) with carbocyclic nucleosides containing the amines **A1** and **A29** in α or β configuration.

four natural bases (Y = A, T, C, G) in opposite position (Table 1).

For **cβA1** we find T_m data that are slightly higher (ca. +1 °C) than those found in the assay. In this case the discrimination of complementary bases is minimal as was also found in the assay. With **cαA1**, no consistent transitions were found in the fluorescence T_m curves. In the case of amine **A29** the situation is different. Here the α-nucleoside (**cαA29**) leads to T_m values that are again in the same range as those found in the assay with selectivities that are slightly less pronounced but that follow essentially those found in the assay. Interestingly, the corresponding β form (**cβA29**) leads to T_m values that are lower by 5–6 °C, indicating that in this case the α form is responsible for the T_m values found in the assay. For all cases the T_m values could also be confirmed by UV-melting curves.

Table 1: Fluorescence T_m data [°C] of duplexes (Scheme 1) containing the carbocyclic *exo*-aminonucleosides with the amines **A1** and **A29** (Scheme 4).

Y	Z[a]			
	cαA1[b]	**cβA1**	**cαA29**	**cβA29**
A	–	40	44	39[c]
C	–	39	45	39
G	–	41	43	40
T	–	40	44	39

[a] 0.6 μM duplex in 3.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris, pH 8.0. [b] The combination with **cαA1** was not detected. [c] T_m value from UV-melting curve (260 nm).

In reproducing the T_m data and, by and large, the base-pairing selectivities these data clearly show that the assay is working properly. In addition it appears that successful candidates may have either the α or the β configuration at the base-carrying center.

In summary, we have established a viable screen for identifying novel DNA base pairs. Undeniably the assay proposed here has been performed with a limited number of amines that cover a limited structural space. The assay can be further optimized and its scope widened in different ways. First fluorescence melting curves may be performed in multiparallel fashion which increases the throughput considerably and enables the use of larger libraries of amines. In terms of novel base pairs it is also conceivable to screen for completely nonnatural homo base pairs by using a duplex with two opposing abasic sites (Z = Y = **X**).

Received: August 25, 2010
Revised: November 15, 2010
Published online: January 21, 2011

Keywords: combinatorial chemistry · DNA recognition · fluorescent probes · nucleosides · parallel synthesis

[1] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537 – 539.
[2] I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada, S. Yokoyama, *Nat. Methods* **2006**, *3*, 729 – 735.
[3] A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727 – 733.
[4] M. Kimoto, R. Kawai, T. Mitsui, S. Yokoyama, I. Hirao, *Nucleic Acids Res.* **2009**, *37*, e14.
[5] D. A. Malyshev, Y. J. Seo, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 14620 – 14621.
[6] Z. Yang, F. Chen, S. G. Chamberlin, S. A. Benner, *Angew. Chem.* **2010**, *122*, 181 – 184; *Angew. Chem. Int. Ed.* **2010**, *49*, 177 – 180.
[7] S. Hoshika, F. Chen, N. A. Leal, S. A. Benner, Z. Yang, S. G. Chamberlin, *Angew. Chem.* **2010**, *122*, 5686 – 5689; *Angew. Chem. Int. Ed.* **2010**, *49*, 5554 – 5557.
[8] A. T. Krueger, H. Lu, A. H. F. Lee, E. T. Kool, *Acc. Chem. Res.* **2007**, *40*, 141 – 150.
[9] J. A. Piccirilli, T. Krauch, S. E. Moroney, S. A. Benner, *Nature* **1990**, *343*, 33 – 37.
[10] C. Switzer, S. E. Moroney, S. A. Benner, *J. Am. Chem. Soc.* **1989**, *111*, 8322 – 8323.
[11] M. Berger, A. K. Ogawa, D. L. McMinn, Y. Q. Wu, P. G. Schultz, F. E. Romesberg, *Angew. Chem.* **2000**, *112*, 3069 – 3071; *Angew. Chem. Int. Ed.* **2000**, *39*, 2940 – 2942.
[12] S. Matsuda, J. D. Fillo, A. A. Henry, P. Rai, S. J. Wilkens, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. G. Schultz, G. Spraggon, F. E. Romesberg, *J. Am. Chem. Soc.* **2007**, *129*, 10466 – 10473.
[13] M. Kimoto, I. Hirao, *Methods Mol. Biol.* **2010**, *634*, 355 – 369.
[14] A. T. Krueger, E. T. Kool, *Curr. Opin. Chem. Biol.* **2007**, *11*, 588 – 594.
[15] Y. J. Seo, S. Matsuda, F. E. Romesberg, G. T. Hwang, P. Ordoukhanian, *J. Am. Chem. Soc.* **2009**, *131*, 5046 – 5047.
[16] Y. J. Seo, G. T. Hwang, P. Ordoukhanian, F. E. Romesberg, *J. Am. Chem. Soc.* **2009**, *131*, 3246 – 3252.
[17] M. Mosimann, P. A. Küpfer, C. J. Leumann, *Org. Lett.* **2005**, *7*, 5211 – 5214.
[18] H. An, T. Wang, M. A. Maier, M. Manoharan, B. S. Ross, P. D. Cook, *J. Org. Chem.* **2001**, *66*, 2789 – 2801.
[19] Y. I. Kuzmin, C. P. Da Costa, M. J. Fedor, *J. Mol. Biol.* **2004**, *340*, 233 – 251.
[20] A. Peracchi, L. Beigelman, N. Usman, D. Herschlag, *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 11522 – 11527.
[21] A. T. Perrotta, T. S. Wadkins, M. D. Been, *RNA* **2006**, *12*, 1282 – 1291.
[22] R. J. Moss, C. R. Petrie, R. B. Meyer, L. D. Nord, R. C. Willis, R. A. Smith, S. B. Larson, G. D. Kini, R. K. Robins, *J. Med. Chem.* **1988**, *31*, 786 – 790.
[23] C. V. Varaprasad, Q. Habib, D. Y. Li, J. F. Huang, J. W. Abt, F. Rong, Z. Hong, H. Y. An, *Tetrahedron* **2003**, *59*, 2297 – 2307.
[24] O. R. Ludek, C. Meier, *Synthesis* **2003**, 2101 – 2109.
[25] J. Sagi, A. Szemzö, J. Szécsi, L. Ötvös, *Nucleic Acids Res.* **1990**, *18*, 2133 – 2140.
[26] a) K. H. Altmann, M. O. B'Vierre, A. De Mesmaeker, H. E. Moser, *Bioorg. Med. Chem.* **1995**, *5*, 431 – 436; b) Y. Xu, K. Kino, H. Sugiyama, *J. Biomol. Struct. Dyn.* **2002**, *20*, 437 – 446.

