

Improving the Accuracy of NMR Structures of DNA by Means of a Database Potential of Mean Force Describing Base–Base Positional Interactions

John Kuszewski,[†] Charles Schwieters,[‡] and G. Marius Clore^{*,†}

Contribution from the Laboratory of Chemical Physics, Building 5, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892-0510, and Computational Bioscience and Engineering Laboratory, Building 12A, Center for Information Technology, National Institutes of Health, Bethesda, Maryland 20892-5624

Received January 3, 2001. Revised Manuscript Received February 21, 2001

Abstract: NMR structure determination of nucleic acids presents an intrinsically difficult problem since the density of short interproton distance contacts is relatively low and limited to adjacent base pairs. Although residual dipolar couplings provide orientational information that is clearly helpful, they do not provide translational information of either a short-range (with the exception of proton–proton dipolar couplings) or long-range nature. As a consequence, the description of the nonbonded contacts has a major impact on the structures of nucleic acids generated from NMR data. In this paper, we describe the derivation of a potential of mean force derived from all high-resolution (2 Å or better) DNA crystal structures available in the Nucleic Acid Database (NDB) as of May 2000 that provides a statistical description, in simple geometric terms, of the relative positions of pairs of neighboring bases (both intra- and interstrand) in Cartesian space. The purpose of this pseudopotential, which we term a DELPHIC base–base positioning potential, is to bias sampling during simulated annealing refinement to physically reasonable regions of conformational space within the range of possibilities that are consistent with the experimental NMR restraints. We illustrate the application of the DELPHIC base–base positioning potential to the structure refinement of a DNA dodecamer, d(CGCGAAT-TCGCG)₂, for which NOE and dipolar coupling data have been measured in solution and for which crystal structures have been determined. We demonstrate by cross-validation against independent NMR observables (that is, both residual dipolar couplings and NOE-derived interproton distance restraints) that the DELPHIC base–base positioning potential results in a significant increase in accuracy and obviates artifactual distortions in the structures arising from the limitations of conventional descriptions of the nonbonded contacts in terms of either Lennard-Jones van der Waals and electrostatic potentials or a simple van der Waals repulsion potential. We also demonstrate, using experimental NMR data for a complex of the male sex determining factor SRY with a duplex DNA 14mer, which includes a region of highly unusual and distorted DNA, that the DELPHIC base–base positioning potential does not in any way hinder unusual interactions and conformations from being satisfactorily sampled and reproduced. We expect that the methodology described in this paper for DNA can be equally applied to RNA, as well as side chain–side chain interactions in proteins and protein–protein complexes, and side chain–nucleic acid interactions in protein–nucleic acid complexes. Further, this approach should be useful not only for NMR structure determination but also for refinement of low-resolution (3–3.5 Å) X-ray data.

Introduction

Determining the structures of nucleic acids by NMR has presented an intrinsically difficult problem for many years.¹ The principal source of geometric information for any NMR structure determination resides in short ($\leq 5\text{--}6$ Å) interproton distance restraints derived from nuclear Overhauser enhancement (NOE) measurements.² The success of NMR in determining 3D

structures of globular proteins resides in the fact that short interproton distance restraints between residues that are far apart in the sequence are conformationally highly restrictive.³ In contrast, DNA is essentially a linear molecule, and short interproton distance contacts are limited to adjacent base pairs.¹ Moreover, even in RNA, which can adopt a tertiary structure, the number of contacts between nucleotides that are far apart in the sequence is limited.^{1d} In addition to this fundamental limitation, NMR structure determination of nucleic acids is further hindered by the fact that the density of protons is much less than that in proteins.^{1,2} Long-range orientational restraints^{3,4} derived from dipolar couplings measured in dilute liquid crystalline media⁵ can potentially lead to improvements in the

* To whom correspondence should be addressed. E-mail: clore@speck.niddk.nih.gov.

[†] National Institute of Diabetes and Digestive and Kidney Diseases.

[‡] Center for Information Technology.

(1) (a) Pardi, A.; Hare, D. R.; Wang, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8785–8789. (b) Clore, G. M.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 5978–5984. (c) Lane, A. N. *Prog. Nucl. Magn. Reson. Spectrosc.* **1993**, *25*, 481–505. (d) Varani, G.; Nicholson, J. K.; Wilson, I. D. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51–127. (e) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387. (f) Moolova, E. T.; Pardi, A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 298–302.

(2) (a) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986. (b) Clore, G. M.; Gronenborn, A. M. *CRC Crit. Rev. Biochem. Mol. Biol.* **1989**, *24*, 479–564.

(3) Clore, G. M.; Gronenborn, A. M. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *95*, 5891–5898.

accuracy of NMR structures of proteins,⁶ nucleic acids,^{7,8} protein–protein complexes,⁹ and protein–nucleic acid complexes¹⁰ by providing information that is qualitatively different from that afforded by the NOE data. Unfortunately, dipolar couplings corresponding to fixed length internuclear vectors do not contain either short-range or long-range translational information which is key for accurate NMR structure determination of nucleic acids. Moreover, although ¹H–¹H dipolar couplings can provide both distance and orientational information, just as the NOEs, they are limited to short interproton distances (typically less than 3.5–5 Å, depending on the density of coupling partners). In proteins, for example, structures calculated using target functions in which the nonbonded contacts are represented by a purely repulsive van der Waals term are invariably expanded since, by way of entropic considerations, there are many more expanded than compacted structures that satisfy the experimental NMR data.¹¹ The same is true of nucleic acids. Since the packing density in proteins is fairly constant,¹² this problem is readily resolved by the inclusion of a radius of gyration restraint.¹¹ Unfortunately, such a simple solution is not applicable to nucleic acids since they are not globular and the interbase packing density is highly variable, depending on the conformation of the DNA (i.e., A, B, Z, or

single-stranded). As a result, inclusion of a radius of gyration restraint will simply distort the overall structure of linear DNA (i.e., it will result in bending).¹³ What is clear is that the description of the nonbonded interactions employed in the target function has a significant influence on the base–base packing and consequently on the resulting structures.

One approach for incorporating a more realistic description of the nonbonded contacts in nucleic acids is to include a full empirical energy function comprising Lennard-Jones van der Waals and electrostatic terms¹⁴ into the target function. Indeed, many NMR structures of nucleic acids are calculated in this manner.¹ However, inclusion of such terms is problematic. First, the Lennard-Jones potential offers a rather poor approximation of the interactions between large π orbitals such as those of the nucleic acid bases. Second, electrostatic calculations tend to be even more approximate, particularly when refinements are carried out in vacuo with no counterions or water present. With careful adjustment of the Lennard-Jones and electrostatic terms together with the inclusion of counterions and water, molecular dynamics simulations can reproduce some of the features observed in nucleic acid crystal structures, but only to an approximate degree.¹⁴ The careful balance, however, of the various terms in the empirical energy function is immediately lost upon the inclusion of experimental NMR restraints. Indeed, it is often the case that NMR nucleic acid structures calculated with a full empirical energy term tend to be somewhat compressed and display some local structural features intermediate between A and B DNA.¹

In this paper, we present an alternative approach for improving the quality of NMR-derived nucleic acid structures. Rather than attempting to model the physics of nucleic acids in solution, we apply knowledge from high-resolution crystal structures to the problem of finding physically reasonable conformations within the range of possibilities that are consistent with the experimental NMR restraints.

In a series of previous papers, we described the implementation of a database potential of mean force comprising one-, two-, three-, and four-dimensional potential surfaces describing the likelihood for various combinations of torsion angles derived from a database of high-resolution protein and nucleic acid crystal structures.¹⁵ The aim of the database torsion angle potential is to bias sampling during simulated annealing refinement to conformations that are likely to be energetically possible by effectively limiting the choice of dihedral angles to those that are known to be physically realizable.¹⁵ Nucleic acid structures can be described by six sugar–phosphate backbone torsion angles (α , β , γ , δ , ϵ , and ζ) and one glycosidic bond torsion angle per base.¹⁶ However, very small changes in torsion angles can produce large-scale changes in nucleic acid structure. Further, while the incorporation of the database torsion angle

(4) (a) Tjandra, N.; Omichinski, J. G.; Gronenborn, A. M.; Clore, G. M.; Bax, A. *Nature Struct. Biol.* **1997**, *4*, 732–738. (b) Clore, G. M.; Gronenborn, A. M.; Tjandra, N. *J. Magn. Reson.* **1998**, *131*, 159–162. (c) Clore, G. M.; Gronenborn, A. M.; Bax, A. *J. Magn. Reson.* **1998**, *133*, 216–221. (d) Tjandra, N.; Marquardt, J.; Clore, G. M. *J. Magn. Reson.* **2000**, *142*, 393–396.

(5) (a) Tjandra, N.; Bax, A. *Science* **1997**, *278*, 1111–1114. (b) Clore, G. M.; Starich, M. R.; Gronenborn, A. M. *J. Am. Chem. Soc.* **1998**, *120*, 10571–10572. (c) Prestegard, J. H. *Nature Struct. Biol.* **1998**, *Suppl. 5*, 517–522. (d) Hansen, M. R.; Mueller, L.; Pardi, A. *Nature Struct. Biol.* **1999**, *5*, 1065–1074. (e) Koenig, B. W.; Hu, J.-S.; Ottinger, M.; Bose, S.; Hendl, R. W.; Bax, A. *J. Am. Chem. Soc.* **1999**, *121*, 1385–1386. (f) Rückert, M.; Otting, G. *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797. (g) Fleming, J.; Gray, D.; Prasanna, S.; Matthews, S. *J. Am. Chem. Soc.* **2000**, *122*, 5224–5225.

(6) (a) Bewley, C. A.; Gustafson, K. R.; Boyd, M. R.; Covell, D. G.; Bax, A.; Clore, G. M.; Gronenborn, A. M. *Nature Struct. Biol.* **1998**, *5*, 571–578. (b) Cai, M.; Huang, Y.; Zheng, R.; Wei, S.-Q.; Guirlando, R.; Lee, M. S.; Craigie, R.; Gronenborn, A. M.; Clore, G. M. *Nature Struct. Biol.* **1998**, *5*, 903–909. (c) Clore, G. M.; Starich, M. R.; Bewley, C. A.; Cai, M.; Kuszewski, J. *J. Am. Chem. Soc.* **1999**, *121*, 6513–6514. (d) Drohat, A. C.; Tjandra, N.; Baldisseri, D. M.; Weber, D. J. *Protein Sci.* **1999**, *8*, 800–809. (e) Fischer, M. W. F.; Losonczy, J. A.; Weaver, J. L.; Prestegard, J. H. *Biochemistry* **1999**, *38*, 9013–9022. (f) Al-Hashimi, H. M.; Bolton, P. J.; Prestegard, J. H. *J. Magn. Reson.* **2000**, *142*, 153–158. (g) Delaglio, F.; Kontaxis, G.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 2142–2143. (h) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W. Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *296*, 1265–1273. (i) Mueller, G. A.; Choy, W. Y.; Yang, D.; Forman-Kay, J. D.; Venters, R. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *300*, 197–212. (j) Fowler, C. A.; Tian, F.; Al-Hashimi, H. M.; Prestegard, J. H. *J. Mol. Biol.* **2000**, *304*, 447–460.

(7) Tjandra, N.; Tate, S.; Ono, A.; Kainosho, M.; Bax, A. *J. Am. Chem. Soc.* **2000**, *122*, 6190–6200.

(8) (a) Vermeulen, A.; Zhou, H.; Pardi, A. *J. Am. Chem. Soc.* **2000**, *122*, 9638–9647. (b) Mollova, E. T.; Hanse, M. R.; Pardi, A. *J. Am. Chem. Soc.* **2000**, *122*, 11561–11562. (c) Trantirek, L.; Urbasek, M.; Stefl, R.; Feigon, J.; Sklenar, V. *J. Am. Chem. Soc.* **2000**, *122*, 10454–10455.

(9) (a) Garrett, D. S.; Seok, Y.-J.; Peterkofsky, A.; Gronenborn, A. M.; Clore, G. M. *Nature Struct. Biol.* **1999**, *6*, 166–173. (b) Clore, G. M. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 9021–9025. (c) Bewley, C. A.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 6009–6016. (d) Wang, G.; Louis, J. M.; Sondej, M.; Seok, Y.-J.; Peterkofsky, A.; Clore, G. M. *EMBO J.* **2000**, *19*, 5635–5649.

(10) (a) Bayer, P.; Varani, L.; Varani, G. *J. Biomol. NMR* **1999**, *14*, 149–155. (b) Huang, K.; Louis, J. M.; Donaldson, L.; Lim, F.-L.; Sharricks, A. D.; Clore, G. M. *EMBO J.* **2000**, *19*, 2615–2628.

(11) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Am. Chem. Soc.* **1999**, *121*, 2337–2338.

(12) (a) Abagayan, R. A.; Totrov, M. M. *J. Mol. Biol.* **1997**, *267*, 678–685. (b) Ratnaparkhi, G. S.; Ramachandran, S.; Udgaonkar, J. B.; Varadarajan, R. *Biochemistry* **1998**, *37*, 6958–6966. (c) Tsai, J.; Taylor, R.; Clothia, C.; Gerstein, M. *J. Mol. Biol.* **1999**, *290*, 253–266. (d) Skolnick, J.; Kolinski, A.; Ortiz, A. R. *J. Mol. Biol.* **1997**, *265*, 217–241.

(13) Unpublished data.

(14) (a) Beveridge, D. L.; McConnell, K. J. *Curr. Opin. Struct. Biol.* **2000**, *10*, 182–196. (b) Olson, W. K.; Zhurkin, V. B. *Curr. Opin. Struct. Biol.* **2000**, *10*, 286–297. (c) Levitt, M. *Cold Spring Harbor Symp. Quantum Biol.* **1993**, *47*, 251–262. (d) Feig, M.; Pettitt, B. M. *J. Phys. Chem. B* **1997**, *101*, 7361–7363. (e) Feig, M.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 134–149. (f) Langley, D. R. *J. Biomol. Struct. Dyn.* **1998**, *16*, 487–509. (g) Cheetham, T. E.; Cieplak, P.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845–862. (h) Darden, T.; Perera, L.; Li, L.; Pedersen, L. *Structure* **1999**, *7*, R55–R60. (i) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86–104. (j) MacKerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000**, *21*, 105–120.

(15) (a) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *Protein Sci.* **1996**, *5*, 1067–1080. (b) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1997**, *125*, 171–177. (c) Kuszewski, J.; Clore, G. M. *J. Magn. Reson.* **2000**, *146*, 249–254.

(16) Saenger, W. *Principles of Nucleic Acid Structures*; Springer-Verlag: New York, 1984.

potential ensures that the torsion angles lie within allowed (i.e., populated in high-resolution crystal structures) regions of torsion angle space, it does not ensure optimal packing of base–base interactions. In the present paper, we have therefore extended the formalism of the potential of mean force to translate positional information describing the spatial relationships of adjacent bases and base pairs from a database of high-resolution DNA crystal structures into an energy surface. The resulting potential surface constitutes an additional term in the overall target function that is minimized during simulated annealing.

Instead of torsion angles, the geometric analysis of segments of nucleic acid structures involves a description of the relative positions of pairs of bases in Cartesian space. By rotating and translating one base at a time into a standard orientation, and applying the same rotations and translations to the rest of the nucleic acid structure, we can define the Cartesian position of any base relative to the first. By extracting many examples of pairs of bases of a particular type from the database, we can define which regions of space around one base type are commonly populated by particular atoms of a second base type. This type of analysis can take into account the effects of many types of nonbonded interactions (e.g., base stacking and hydrogen bonding) without attempting to model the underlying physics. For example, since hydrogen bonds must be reasonably short and linear to be energetically favorable, the DNA crystal structure database should yield many examples of A•T base pairs with the N1 atom of adenine in one spot a few ångströms away from the N3 atom of thymine. Since the database would be expected to yield a large number of examples in one small area, there would be a corresponding minimum in the potential of mean force at that position. Simulated annealing calculations that included this potential of mean force would therefore feel forces to move the N1 atom of the adenine into the proper position relative to the thymine. By analogy to our previous torsion angle work, we refer to this new potential of mean force as the DELPHIC (for *database elucidated likelihood phor* internal coordinates) base–base positioning potential.

This sort of local Cartesian coordinate analysis has already been used by several groups to examine and evaluate the structures of proteins and protein–nucleic acid complexes. The “quality control” module of the program WHATIF¹⁷ makes use of a similar metric to examine the quality of the overall packing of a protein structure. The atoms that are in contact with each residue are examined in order to determine how common their three-dimensional positions are in relation to the given residue, and from this an overall “packing quality” score is calculated. The “commonness” of a particular relative orientation is determined by reference to a database of highly accurate protein crystal structures. A similar approach is used by the program X-CITE¹⁸ to predict binding sites in proteins. More recently, Pabo and Nekludova¹⁹ have looked at the relative Cartesian positions of the backbone atoms of protein residues that interact with DNA moieties in order to find common features of protein–DNA recognition.

To our knowledge, however, this analysis has never been used to refine protein or nucleic acid structures. The DELPHIC base–base positioning potential is complementary to our DELPHIC torsion angle potential, which guides combinations of torsion angles in proteins or nucleic acids into commonly observed values. By simultaneously ensuring that experimental NMR restraints (NOEs, scalar couplings, dipolar couplings, etc.) are

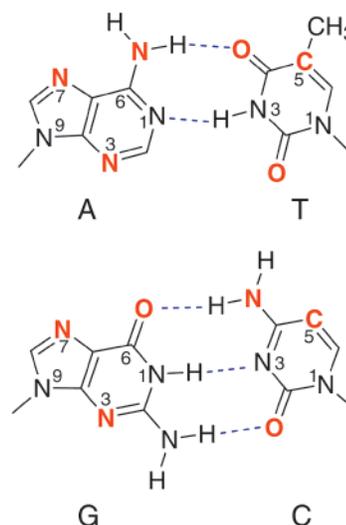


Figure 1. Definition of the orienting atoms for each base type. The three orienting atoms for each base are displayed in red. The three orienting atoms I, J, and K (cf. Figure 2) of adenine are N7, N6, and N3, respectively; for guanine, N7, O6, and N3, respectively; for thymine, C6, O4, and O2, respectively; and for cytosine, C6, N4, and O2, respectively.

satisfied, that torsion angles which are close in primary sequence are in reasonable conformations, and that the nonbonded interactions such as hydrogen bonding and base–base contacts are also in reasonable conformations, we find that we can greatly improve the accuracy of NMR structure determination. We illustrate the application of these methods to the solution structure determination of a self-complementary DNA dodecamer 5'-d(CGCGAATTCGCG)₂ for which extensive NOE and dipolar coupling data have been measured in solution⁷ and for which several crystal structures are available.^{20,21} In addition, we also present calculations using experimental NMR data on a complex of the male sex determining factor SRY with a 14mer duplex DNA which includes a region of DNA that is highly unusual and distorted.

It is well known that crystal structures of DNA are subject to crystal packing forces²² which can have a major impact on global structure. Thus, for example, the crystal structures^{20,21} of the DNA dodecamer are not symmetric, despite the fact that the sequence is palindromic and the structure is clearly symmetric in solution (as judged by the presence of only a single set of resonances).^{7,23} Similarly, a large number of crystal structures of A DNA^{22c} and Z DNA^{22d} have been determined, despite the fact that under physiological conditions, the conformation of DNA free in solution is found invariably to be in the B form.¹ It is important, however, to emphasize that these effects do not, in any way, undermine the application of the DELPHIC base–base positioning potential. This is because the potential deals specifically with nearest-neighbor interactions between adjacent base pairs and the database is sufficiently large to include all possible base–base interactions that are likely to exist in solution.

(20) Dickerson, R. E.; Drew, H. R. *J. Mol. Biol.* **1981**, *149*, 761–786.

(21) Shui, X.; McFail-Isom, L.; Hu, G. G.; Williams, L. D. *Biochemistry* **1998**, *37*, 8341–8355.

(22) (a) Tereshko, V.; Subirana, J. A. *Acta Crystallogr.* **1999**, *D55*, 810–819. (b) Berman, H. M. *Biopolymers* **1997**, *44*, 23–44. (c) Wahl, M. C.; Sundaralingam, M. *Biopolymers* **1997**, *44*, 45–63. (d) Ho, P. S.; Moers, B. H. M. *Biopolymers* **1997**, *44*, 65–90.

(23) Hare, D. R.; Wemmer, D. E.; Chou, S. H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319–336.

(17) Vriend, G.; Sander, C. *J. Appl. Crystallogr.* **1993**, *26*, 47–60.

(18) Laskowski, R.; Thornton, J.; Humblet, C.; Singh, J. *J. Mol. Biol.* **1996**, *259*, 175–201.

(19) Pabo, C. O.; Nekludova, L. *J. Mol. Biol.* **2000**, *301*, 597–624.

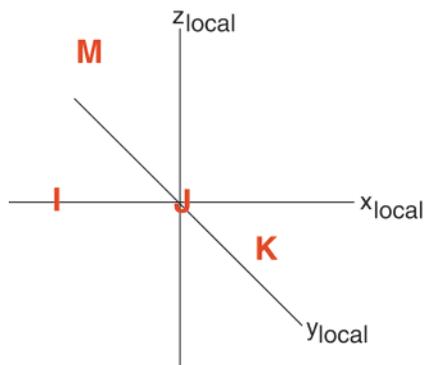


Figure 2. Defining the relative geometry of two bases. The three orienting atoms, I, J, and K, of the first base have been rotated and translated into a standard geometry, such that atom J is at the origin, atom I is along the negative x axis, and atom K is in the xy plane. The position of atom M, one of the three orienting atoms of the second base, after applying the same rotations and translations is also indicated.

Theory

Defining the Relative Geometry of Two Bases. From each DNA base type (A, T, C, and G), three atoms, close to the functional groups of the base, are chosen whose coordinates are used to define the position of the overall base. Since each base constitutes a rigid planar group, these three atoms are sufficient to uniquely define the position of a given base. The identities of the three orienting atoms for each base type are shown in Figure 1.

The position of one base in relation to another is defined by rotating and translating the orienting atoms of the first base into a standard geometry and then applying these same rotations and translations to the second base. The relative geometry of the second base is then defined by the coordinates of its translated and rotated orienting atoms. For clarity, we label the three orienting atoms of the first base as I, J, and K and the orienting atoms of the second base as M, N, and P.

We define the standard geometry to have atom J at the origin, atom I along the negative x axis, and atom K in the xy plane; that is,

$$\begin{aligned}\mathbf{x}_{\text{local}} &= \mathbf{J} - \mathbf{I} \\ \mathbf{z}_{\text{local}} &= \mathbf{x}_{\text{local}} \times (\mathbf{K} - \mathbf{J}) \\ \mathbf{y}_{\text{local}} &= \mathbf{z}_{\text{local}} \times \mathbf{x}_{\text{local}}\end{aligned}\quad (1)$$

where \mathbf{I} , \mathbf{J} , and \mathbf{K} are the coordinate vectors of the orienting atoms of the first base (see Figure 1).

Rotation into the standard geometry is equivalent to rotations that superimpose $\mathbf{x}_{\text{local}}$, $\mathbf{y}_{\text{local}}$, and $\mathbf{z}_{\text{local}}$ with the values they would have in the standard geometry:

$$\begin{aligned}\mathbf{x}_{\text{target}} &= (1, 0, 0) \\ \mathbf{y}_{\text{target}} &= (0, 1, 0) \\ \mathbf{z}_{\text{target}} &= (0, 0, 1)\end{aligned}\quad (2)$$

The rotation of the two bases is accomplished by application of direction cosines (defined as the cosine of the angle between

Table 1. Breakdown of DNA Database Used To Create the DELPHIC Base–Base Positioning Potential^a

A. Number of Structures in Each Structure Class	
A DNA	51
B DNA	52
Z DNA	42
protein–DNA complexes	48
intercalating drug–DNA complexes	63
major/minor groove binding drug–DNA complexes	13
“unusual” DNA structures	22
total	291
B. Total Number of Valid Residue Pairs in Each Structure Class	
A DNA	2718
B DNA	4002
Z DNA	1181
protein–DNA complexes	4696
intercalating drug–DNA complexes	2487
major/minor groove binding drug–DNA complexes	985
“unusual” DNA structures	957
total	17026

^a The structures are taken from the NDB²⁵ and represent all structures present as of May 2000 that have been solved at a resolution of 2 Å or less with R factors better than 25%.

two vectors) which, from the law of cosines,²⁴ is given by

$$\cos \theta_{\mathbf{ab}} = [(\mathbf{a} - \mathbf{b})^2 - a^2 - b^2] / [-2ab] \quad (3)$$

where \mathbf{a} and \mathbf{b} are arbitrary vectors, a and b are their respective lengths, and $\theta_{\mathbf{ab}}$ is the angle between them. The various direction cosines between the local and target vectors define the rotation matrix, \mathbf{RM} :

$$\mathbf{RM} = \begin{pmatrix} \cos \theta_{\mathbf{x}_{\text{local}} \mathbf{x}_{\text{target}}} & \cos \theta_{\mathbf{x}_{\text{local}} \mathbf{y}_{\text{target}}} & \cos \theta_{\mathbf{x}_{\text{local}} \mathbf{z}_{\text{target}}} \\ \cos \theta_{\mathbf{y}_{\text{local}} \mathbf{x}_{\text{target}}} & \cos \theta_{\mathbf{y}_{\text{local}} \mathbf{y}_{\text{target}}} & \cos \theta_{\mathbf{y}_{\text{local}} \mathbf{z}_{\text{target}}} \\ \cos \theta_{\mathbf{z}_{\text{local}} \mathbf{x}_{\text{target}}} & \cos \theta_{\mathbf{z}_{\text{local}} \mathbf{y}_{\text{target}}} & \cos \theta_{\mathbf{z}_{\text{local}} \mathbf{z}_{\text{target}}} \end{pmatrix} \quad (4)$$

Thus, the standardized coordinates for any orienting atom from the second base (e.g., atom M) can be obtained by first translating it by the amount needed to move atom J onto the origin and then applying the rotation matrix:

$$\begin{pmatrix} M_x \text{ (standardized)} \\ M_y \text{ (standardized)} \\ M_z \text{ (standardized)} \end{pmatrix} = \mathbf{RM} \begin{pmatrix} M_x - J_x \\ M_y - J_y \\ M_z - J_z \end{pmatrix} \quad (5)$$

Assembling the Database. All 291 DNA crystal structures available in the Nucleic Acid Database (NDB) database²⁵ as of May 2000 solved at a resolution of ≤ 2.0 Å with an R factor $\leq 25\%$ comprised the database used to define the DELPHIC base–base positioning potential surfaces. The breakdown into different classes of structures (i.e., A DNA, B DNA, protein–DNA, etc.) is given in Table 1 and shows that all the different conformations that can be adopted by DNA and that have been observed crystallographically are well represented. Atoms with thermal B factors greater than 25 Å² or occupancies less than 95% were excluded.

Residue pairings were included in the potential surface calculation only if all six orienting atoms (three from each base) had known coordinates. In addition, since the vast majority of base–base interactions in DNA involve bases that are neighbors in primary sequence (defined in terms of Watson–Crick base

(24) Pearson, C. *Handbook of Applied Mathematics*, 2nd ed.; van Nostrand: New York, 1983.

(25) (a) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. *Biophys. J.* **1992**, *63*, 751–759. (b) <http://www.ndbserver.rutgers.edu/NDB>.

pairs), only pairings that were close in primary sequence were included in the potential surface calculation. Residue pairs that are considered “close” in primary sequence for the purpose of generating the DELPHIC base–base positioning potential surfaces are as follows: a base with its 5′ and 3′ neighbors, a base with its Watson–Crick partner, and a base with the 5′ and 3′ neighbors of its Watson–Crick partner.

The database contained a total of 17 026 valid residue pairs that were close in primary sequence (see Table 1 for breakdown into the different classes of DNA structures). These were divided on the basis of base type and position (e.g., from an adenine to the first orienting atom of a cytosine that is 5′ to the adenine’s Watson–Crick partner) into 204 different potential surface types (i.e., $4 \times [(4 \times 4) + 1] \times 3$). Each potential surface had a mean of 248 examples from the database.

Translating the Databases into Potential Energies. For each of the 204 potential surface types, its entries are translated into a potential of mean force in a manner very similar to that used in our previous work.¹⁵ Specifically, we define a 20 Å per side cube in the standard coordinate space over which the DELPHIC base–base positioning potential surface is to be calculated. We subdivide this “valid space” into $0.2 \times 0.2 \times 0.2$ Å³ cubes, and the number of entries from the database that are found in each cube is recorded. These database counts are translated into a potential of mean force using the relationship

$$E(x, y, z) = -\ln P(x, y, z) \quad (6)$$

where

$$P(x, y, z) = [N(x, y, z)/V(x, y, z)]/[N(\text{total})/V(\text{total})] \quad (7)$$

and

$N(x, y, z)$ = number of examples found in
cube centered at point (x, y, z)

$V(x, y, z)$ = volume of cube centered at point (x, y, z)

$N(\text{total})$ = number of examples of this pair type in
the entire database

$V(\text{total})$ = volume of space over which the
potential is defined (8)

For cubes that have no examples found in the database, the volume of the cube is gradually expanded, including the examples found in its neighboring cubes, until the number of examples found in the expanded cube is greater than 5. The potential of mean force is then calculated using the volume of this expanded cube and the number of examples for $V(x, y, z)$ and $N(x, y, z)$.

Finally, these raw potentials of mean force are replaced by sums of fitted three-dimensional Gaussian functions as described previously.^{15c} Each potential surface type was fitted with up to 128 independent three-dimensional Gaussians. The motivation for replacing the raw potentials with fitted Gaussian functions is twofold:^{15c} (a) to reduce the memory requirements of the DELPHIC base–base positioning potentials and (b) to smooth out the potential surface and its derivatives. An example of a potential surface is shown in Figure 3.

Calculating the DELPHIC Base–Base Positioning Energies during Simulated Annealing. The energy for the DELPHIC base–base positioning potential is given by

$$E_{\text{DELPHIC-position}} = k_{\text{base-base}} \sum E_{\text{DELPHIC-position}(i)} \quad (9)$$

where $k_{\text{base-base}}$ is a unitless force constant or scale factor, and $E_{\text{DELPHIC-position}(i)}$ is the sum of each of the Gaussians fitted to the potential surface type appropriate for the four orienting atoms of restraint i :

$$E_{\text{DELPHIC-position}(i)} = \sum \text{Gaussian}(\text{width}_j, \text{center}_j, \text{orienting atom } 1_j, \text{orienting atom } 2_j, \text{orienting atom } 3_j, \text{orienting atom } 4_j) \quad (10)$$

The expression for the Gaussian term uses the standardized coordinates for the four orienting atoms, which are calculated on the fly using eqs 1–5.

The atomic forces are calculated from the partial derivative of $E_{\text{DELPHIC-position}(i)}$ with respect to the standardized Cartesian coordinates of the orienting atoms. These forces are then rotated back into the main coordinate frame using the rotation matrix in eq 5.

The computational cost of including the DELPHIC base–base positioning potential is minimal. In the case of a target function comprising terms for covalent geometry, experimental NMR restraints (NOEs, dipolar couplings, torsion angles), a noncrystallographic symmetry restraint, and a quartic van der Waals repulsion term, the addition of the DELPHIC base–base positioning potential typically increases the CPU time per structure by only 4%.

Methods

To ascertain the effectiveness of the DELPHIC base–base positioning potential, we examined the effect of its inclusion on the solution structure determination of the DNA dodecamer 5′d(CGCGAATTCGCG)₂. The experimental data set used in the current calculations is that recently described by Tjandra et al. (PDB accession code 1DUF).⁷ There are 162 NOE-derived interproton distance restraints comprising 50 intrasidue, 108 sequential, and 4 interstrand restraints. These interproton distance restraints were derived from the relative cross-peak intensities in a 2D ¹H–¹H NOE spectrum recorded at 750 MHz in D₂O with a mixing time of 100 ms, using the intrasidue H1′–H2′′ distance (2.3 Å) as an internal reference.⁷ The upper and lower bounds for the distance restraints (used for the corresponding square-well restraining potential in the target function) were derived by applying a tolerance of ±15% on the derived distances.⁷ There are 408 dipolar couplings derived from measurements in a dilute liquid crystalline bicelle medium.⁷ The dipolar couplings comprised 94 C–H (ribose), 24 C–H(base), 4 C–H(methyl), and 10 N–H(imino) couplings measured with an accuracy of ±2 Hz, 64 C–H(ribose) and 12 C–H(base) couplings measured with an accuracy of ±4 Hz; and 200 ¹H–¹H dipolar couplings (of which the sign could be determined for 74) which were converted into approximate ranges corresponding to strong, medium, and weak/absent intensities in a 2D ¹H–¹H COSY spectrum.⁷ The values for the axial component and rhombicity of the alignment tensor are –16 Hz for one-bond C–H vectors (–7.7 Hz for one-bond N–H vectors) and 0.26, respectively.⁷ Twenty ϵ (C4′–C3′–O3′–P) torsion angle restraints (ranging from –160° to –180° with error limits of ±20°) and 24 δ (C5′–C4′–C3′–O3′) torsion angle restraints (–145 ± 35°) were derived from ¹H–¹H and ¹H–³¹P J couplings.^{7,26} In addition, broad range torsion angle restraints, encompassing both A and B DNA conformations,^{1b,7,16} were employed for the α (–70 ± 50°), β (180 ± 50°), γ (60 ± 35°), and ζ (–85 ± 50°) sugar–phosphate torsion angles. These sugar–phosphate backbone torsion angle restraints are fully consistent with the ³¹P NMR spectrum of the dodecamer which spans a very narrow region,⁷ characteristic of regular, undistorted B DNA.²⁷ Finally, six distance restraints per base pair were employed to describe Watson–Crick base pairing: for G·C base pairs, $r_{\text{N1-N3}} = 2.87$ Å, $r_{\text{H1-N3}} = 1.86$ Å, $r_{\text{O6-N4}} = 2.81$ Å, $r_{\text{N2-O2}}$

(26) (a) Bax, A.; Lerner, L. *J. Magn. Reson.* **1988**, *79*, 429–438. (b) Sklenar, V.; Bax, A. *J. Am. Chem. Soc.* **1987**, *109*, 7525–7526.

(27) Roontga, V. A.; Jones, C. R.; Gorenstein, D. G. *Biochemistry* **1990**, *29*, 5245–5258.

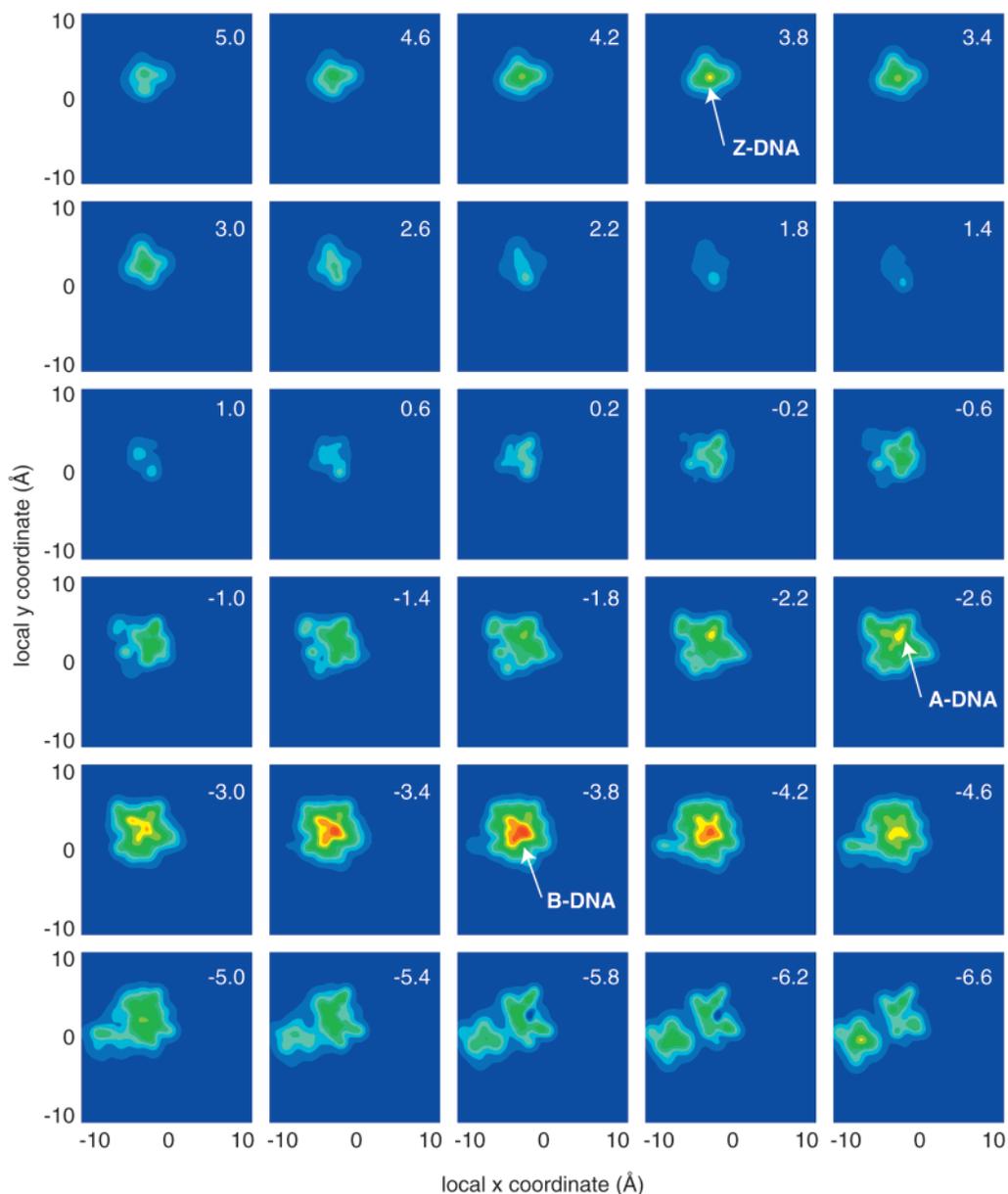


Figure 3. Example of a DELPHIC base–base positioning potential surface. The DELPHIC base–base positioning potential describing the interaction between the N7, O6, and N3 atoms of a guanine base and the C5 atom of a thymine base immediately 5′ to the guanine in primary sequence is shown. Slices of the 3D potential surface are shown every 0.4 Å along the local z coordinate, between -6.6 and $+5.0$ Å. The relative potential energy (calculated with $k_{\text{base-base}} = 1.0$) at each point is color coded, with red representing -11 kcal mol $^{-1}$ and blue representing $+2$ kcal mol $^{-1}$. The locations of the minima associated with A, B, and Z DNA structures are indicated. The minimum around $z = -6.6$ Å corresponds to unusual DNA structures. The regions from $z = +5$ to $+10$ Å and from -6.6 to -10 Å do not contain any minima and are therefore not displayed.

$= 2.81$ Å, $r_{\text{O6-N3}} = 3.58$ Å, and $r_{\text{N2-N3}} = 3.63$ Å; for A•T base pairs, $r_{\text{N1-N3}} = 2.92$ Å, $r_{\text{N1-H3}} = 1.87$ Å, $r_{\text{N6-O4}} = 2.89$ Å, $r_{\text{H2-O2}} = 2.94$ Å, $r_{\text{N1-O4}} = 3.69$ Å, and $r_{\text{N1-O2}} = 3.67$ Å. The O6–N3 and N2–N3 distance restraints for G•C base pairs and the N1–O4 and N1–O2 distance restraints for A•T base pairs serve to prevent unduly large shearing of the bases within each base pair.^{10b}

All calculations were carried out using the NIH version²⁸ of XPLOR.²⁹ All simulated annealing and minimization was carried out in torsion angle space. Torsion angle dynamics employed a sixth-order predictor–corrector integrator with automatic time step selection (which varies during the course of the calculation).³⁰ Bond lengths and angles were constrained to idealized covalent geometry. The target function

(28) The NIH version of XPLOR (as well as the simulated annealing protocol) is available by anonymous ftp on portal.niddk.nih.gov in the directory /pub/clore/xplor_nih.

(29) Brünger, A. T. *XPLOR Manual*; Yale University: New Haven, CT, 1993.

(30) Schwieters, C.; Clore, G. M., to be published.

comprised harmonic terms for the covalent geometry (i.e., bonds, angles, and improper torsion angles used to define planarity of bases and chirality; note that the bonds and angles are held fixed by constraints with the exception of the C4′–C3′ bond and the angles involving both the C4′ and C3′ atoms in the sugar rings which are restrained to idealized values), square-well potentials for the interproton distance and torsion angle restraints,³¹ a harmonic potential for the Watson–Crick hydrogen-bonding distance restraints, a harmonic potential for the C–H and N–H dipolar couplings,^{4b} a square-well potential for the ^1H – ^1H dipolar couplings,^{4d} a harmonic potential for a noncrystallographic symmetry restraint to ensure that the structure of the palindromic DNA remains symmetric, a harmonic potential for the base pair planarity restraints used to prevent undue buckling while allowing propeller twisting to occur,³³ a quartic van der Waals repulsion term,³¹

(31) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *229*, 317–324. (b) Nilges, M.; Gronenborn, A. M.; Brünger, A. T.; Clore, G. M. *Protein Eng.* **1988**, *2*, 27–38.

the DELPHIC torsion angle database potential term,^{15c} and the DELPHIC base–base positioning database potential term. The simulated annealing protocol employed is essentially identical (with some very minor variations) to that described previously,³² with the difference that torsion angle rather than Cartesian coordinate dynamics are employed, and that the target function includes terms for the dipolar couplings and the DELPHIC base–base positioning potential. The final values for the various force constants are as follows: 1000 kcal mol⁻¹ Å⁻² for bonds (one per sugar); 500 kcal mol⁻¹ rad⁻² for angles (associated with both the C4' and C3' atoms of the sugar ring), 500 kcal mol⁻¹ rad⁻² for improper torsions; 30 kcal mol⁻¹ Å⁻² for the distance restraints, 200 kcal mol⁻¹ Å⁻² for the torsion angle restraints; 200 kcal mol⁻¹ Å⁻² for the noncrystallographic symmetry restraint; 20 kcal mol⁻¹ Å⁻² for the base pair planarity restraints,³³ except for the two penultimate base pairs (2 and 11), where a force constant of 50 kcal mol⁻¹ Å⁻² was employed, and for the end base pairs (1 and 12), where a force constant of 80 kcal mol⁻¹ Å⁻² was employed; 4 kcal mol⁻¹ Å⁻⁴ for the quartic van der Waals repulsion term with a scale factor of 0.78 for the van der Waals radii (the radii used are those from the CHARMM PARNAHIER1 DNA parameters³⁴); 1 kcal mol⁻¹ Hz⁻² for the dipolar couplings, with the exception of the low-precision C–H base and sugar dipolar couplings, which had final force constants of 0.125 and 0.2 kcal mol⁻¹ Hz⁻², respectively; 1 for the DELPHIC torsion angle database potential term; and 0–0.9 for the DELPHIC base positioning potential term. We note that the convergence power of the simulated annealing protocol in torsion angle space is very high, and identical results are obtained irrespective of starting coordinates.³⁵ The total CPU time per structure on a 1998 DEC Alpha 600 MHz workstation is ~10 min.

Twenty structures were calculated for each set of conditions (e.g., different values of the force constant for the DELPHIC base–base positioning potential; presence or absence of NOE distance restraints; presence or absence of dipolar coupling restraints). The restrained regularized mean coordinates were obtained by averaging the coordinates of the individual structures within each ensemble, best-fit to base pairs 1–12, and subjecting the resulting coordinates to restrained regularization. This involved first regularizing the covalent geometry by minimization in Cartesian space against a target function comprising only bond, angle, and improper torsion terms, followed by minimization in torsion angle space against the complete target function.

In addition to structures on the DNA dodecamer, we also carried out a number of calculations using experimental NMR restraints³⁶ on a complex of the male sex determining factor SRY with a 14mer duplex DNA to illustrate the impact of the DELPHIC base–base positioning potential on unusual DNA structures. The simulated annealing protocol employed was the same as that described for the DNA dodecamer.

DNA structural parameters were analyzed using the program COMPDNA.³⁷ Structures were visualized using the program VMD-XPLOR.³⁸

Results and Discussion

Nomenclature of Structures. The reference structures employed in this study comprise two X-ray structures and two previous NMR structures. The X-ray structures are 1BNA,²⁰ solved at room temperature and a resolution of 2.5 Å, and 355D,²¹ solved at –136 °C and a resolution of 1.4 Å. The two

NMR structures are those reported by Tjandra et al.⁷ which were calculated with Lennard-Jones van der Waals and electrostatic terms (using the CHARMM PARNAHIER1 DNA parameters³⁴) in the target function: 1DUF is the structure calculated with the complete NOE and dipolar coupling data set, and LJ-nodipo is the structure calculated only on the basis of NOE data.

To assess the validity and usefulness of the DELPHIC base–base positioning database potential, we carried out a series of calculations. Structures calculated with the complete set of NOE interproton distance and dipolar coupling data set (i.e., the same experimental restraints as those used to calculate 1DUF) are referred to as fullxx, those with only NOE distance restraints as nodipox, and those with only dipolar coupling restraints as dipox. For each of these data sets, a series of ensembles (comprising at least 20 structures) was calculated with the force constant for the DELPHIC base–base positioning potential ranging from 0 (i.e., no DELPHIC base–base positioning potential) to 0.9. The last two characters, identified by xx, in the structure notation indicates the force constant for the DELPHIC base–base positioning potential. Thus, for example, full03 would indicate the structures calculated with a force constant of 0.3 and the complete NOE and dipolar coupling data set. When the structures are denoted in $\langle \rangle$ brackets, we refer to an ensemble average; otherwise, the structures refer to the restrained regularized mean structures of an ensemble of simulated annealing structures.

These series of calculations permit four issues to be directly assessed: (a) what is the impact of dipolar couplings on the resulting structures?; (b) what is the impact of the description of the nonbonded contacts on the resulting structures?; (c) what is the impact of the DELPHIC base–base positioning potential on accuracy?; and (d) what is the optimal force constant for

(36) The complex solved comprised an 85-residue portion of human SRY (corresponding to residues 57–140 of the natural sequence plus an N-terminal methionine) which includes the DNA binding HMG-box domain and a 14mer duplex DNA, 5'd(CCTGCACAAACACC)-5'd(GGTG-TTTGTGCAGG). (In the numbering scheme employed here, residues 2 and 85 of the protein construct correspond to residues 57 and 140, respectively, of the natural sequence.) The structures of the SRY–DNA complex were calculated on the basis of 2793 experimental NMR restraints (Clore, G. M.; Murphy, E. C., in preparation). The breakdown of the restraints is as follows: 1566 NOE-derived interproton distance restraints comprising 1130 restraints within the protein (378 intraresidue, 340 sequential, and 412 nonsequential interresidue restraints), 268 restraints within the DNA (90 intraresidue, 161 sequential intrastrand, and 17 interstrand restraints), and 168 intermolecular restraints (all NOE-derived interproton distance restraints were classified into approximate ranges corresponding to strong, medium, weak, and very weak NOEs; 191 distance restraints for hydrogen bonds, including 106 used to maintain Watson–Crick base pairing and prevent unduly large shearing of the base pairs [Huang, K.; Louis, J. M.; Donaldson, L.; Lim, F.-L.; Sharrocks, A. D.; Clore, G. M. *EMBO J.* **2000**, *19*, 2615–2628]; 433 torsion angle restraints [294 within the protein and 139 within the DNA]; 70 ³J_{H_N} coupling constant restraints; 165 ¹³Ca/β chemical shift restraints; and 368 residual dipolar coupling restraints [274 within the protein and 94 within the DNA]). The residual dipolar couplings within the DNA comprise 7 ¹D_{NH}, 32 ¹D_{CH}, and 55 ¹D_{HH} dipolar couplings. The torsion angle restraints within the DNA comprise loose restraints^{1b} for the sugar–phosphate backbone of only those residues for which the ³¹P chemical shifts lie in the range –4.0 to –4.9 ppm;²⁷ the ³¹P shifts of A12 and C25 lie downfield of this envelope between –3.0 and –3.5 ppm, while those of A9, T21, and G22 lie upfield between –5.0 and –6.0 ppm (Clore, G. M.; Murphy, E. C., unpublished data). Two hundred structures each were calculated with and without the DELPHIC base–base positioning potential. All the structures satisfy the experimental restraints within their errors, display good nonbonded contacts, and have minimal deviations from idealized covalent geometry (~0.003 Å for bonds and ≤0.8° for angles and improper torsions).

(37) (a) Gorin, A. A.; Zhurkin, V. B.; Olson, W. K. *J. Mol. Biol.* **1995**, *247*, 34–48. (b) Olson, W. K.; Gorin, A. A.; Lu, X.-J.; Hock, L. M.; Zhurkin, V. B. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11163–11168.

(38) (a) Schwieters, C. S.; Clore, G. M. *J. Magn. Reson.* **2000**, in press. (b) Available on-line at <http://vmd-xplor.cit.nih.gov/>. (c) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.

(32) Omichinski, J. G.; Pedone, P. V.; Felsenfeld, G.; Gronenborn, A. M.; Clore, G. M. *Nature Struct. Biol.* **1997**, *4*, 122–132.

(33) The planarity restraints are used to prevent undue buckling of Watson–Crick base pairs while permitting propeller twisting to take place unhindered. This is achieved by applying a planarity restraint for each Watson–Crick base pair to ensure that the N1, C6, and C2 atoms of the purine base lie in approximately the same plane as the N3 atom of the pyrimidine base, and the N3, C2, and C4 atoms of the pyrimidine base lie in approximately the same plane as the N1 atom of the purine base.

(34) Nilsson, L.; Karplus, M. *J. Comput. Chem.* **1986**, *7*, 591–616.

(35) Initial coordinates were obtained by subjecting the coordinates of regular A DNA to 10 ps of torsion angle dynamics at 3000 K against a target function comprising only covalent geometry restraints, a quartic van der Waals repulsion term, and the DELPHIC torsion angle database. The atomic rms differences between the resulting structures on one hand and A and B DNA on the other range from 8 to 20 Å.

Table 2. Root-Mean-Square Deviations between Observed and Calculated Dipolar Couplings and Dipolar Coupling *R* factors^a

structure	rms deviation (Hz) between observed and calculated dipolar couplings (% dipolar coupling <i>R</i> factors)							<i>D</i> _{HH}	
	all (408)	<i>D</i> _{CH} ^{ribose} ^b (94)	<i>D</i> _{CH} ^{ribose} ^c (64)	<i>D</i> _{CH} ^{base} ^b (24)	<i>D</i> _{CH} ^{base} ^c (12)	<i>D</i> _{CH} ^{methyl} ^b (4)	<i>D</i> _{NH} ^{imino} ^b (10)	absolute (126)	sign (74)
full03	2.72	2.27 (11.4)	5.62 (28.1)	2.78 (13.9)	1.89 (9.26)	0.98 (4.9)	1.56 (15.9)	1.26	0.93
dipo03	2.58	2.17 (10.9)	5.24 (26.2)	2.81 (14.1)	2.01 (10.0)	0.79 (4.0)	1.50 (15.2)	1.25	0.88
nodipo03	3.82	4.33 (21.6)	7.21 (36.1)	3.19 (15.9)	3.21 (16.0)	1.05 (5.3)	1.98 (20.1)	1.55	1.46
full00	2.69	2.27 (11.4)	5.63 (28.2)	2.67 (13.3)	2.42 (12.1)	0.96 (4.8)	1.41 (14.3)	1.34	0.85
dipo00	2.64	1.99 (10.0)	5.37 (26.9)	2.76 (13.8)	3.27 (16.3)	0.84 (4.2)	1.39 (14.1)	1.41	0.80
nodipo00	4.01	4.31 (21.6)	7.78 (38.9)	3.19 (15.9)	3.36 (16.8)	1.48 (7.4)	1.93 (19.6)	1.73	1.33
1DUF (NMR)	2.73	2.37 (11.8)	5.63 (28.2)	2.74 (13.7)	1.82 (9.1)	0.96 (4.8)	1.43 (14.5)	1.22	0.80
LJ-nodipo (NMR)	5.46	5.81 (29.1)	11.1 (55.7)	3.78 (18.9)	3.06 (15.3)	1.61 (8.0)	1.91 (19.4)	1.82	1.32
1BNA (X-ray)	4.86	6.19 (31.0)	8.76 (43.8)	3.95 (19.8)	3.38 (16.9)	0.83 (4.2)	1.64 (16.7)	1.49	2.05
355D (X-ray)	5.79	7.38 (36.9)	10.6 (52.8)	3.83 (19.2)	3.19 (16.0)	1.05 (5.5)	1.78 (18.0)	1.66	2.69

^a The dipolar coupling *R* factor (given in parentheses) is defined as the ratio of the rms deviation between observed and calculated values to the expected rms deviation if the vectors were randomly oriented. The latter is given by $\{2D_a^2[4 + 3\eta^2]/5\}^{1/2}$, where D_a is the magnitude of the axial component of the alignment tensor and η the rhombicity.⁴⁰ The value of D_a^{CH} , D_a^{NH} , and η are -16 Hz, -7.7 Hz, and 0.26 , respectively.⁷ These values⁷ were obtained from the distribution of dipolar couplings,^{4c} followed by a grid search.^{4b} ^b Measured with an accuracy of ± 2 Hz. ^c Measured with an accuracy of ± 4 Hz.

Table 3. Pairwise Atomic Root-Mean-Square Differences^a

	atomic rms difference (Å): base pairs 1–9 and base pairs 1–12									
	1BNA (X-ray)	355D (X-ray)	full03	dipo03	nodipo03	full00	dipo00	nodipo00	1DUF (NMR)	LJ-nodipo (NMR)
1BNA (X-ray)	–	0.88	0.69	0.72	0.81	1.02	1.02	1.13	0.90	1.49
355D (X-ray)	0.90	–	1.15	1.13	1.27	1.39	1.36	1.50	1.21	1.65
full03	1.31	1.56	–	0.34	0.49	0.62	0.72	0.79	0.71	1.43
dipo03	1.37	1.55	0.46	–	0.69	0.74	0.72	0.79	0.51	1.23
nodipo03	1.39	1.70	0.62	0.92	–	0.85	0.97	0.89	0.94	1.54
full00	1.60	1.86	0.84	1.01	1.12	–	0.47	0.62	0.89	1.86
dipo00	1.69	1.90	0.99	0.92	1.34	0.63	–	0.69	1.05	1.74
nodipo00	1.70	1.93	0.93	1.02	1.13	0.65	0.69	–	1.05	1.74
1DUF (NMR)	1.49	1.62	0.79	0.61	1.21	1.24	1.11	1.27	–	1.22
LJ-nodipo (NMR)	2.24	2.24	1.74	1.63	2.08	2.53	2.24	2.29	1.55	–

^a The values above the diagonal relate to base pairs 1–9, and those below the diagonal relate to base pairs 1–12.

the DELPHIC base–base positioning potential? The latter two issues are readily assessed by cross-validation, that is, by looking at the agreement between observed and calculated dipolar couplings in the structures calculated without dipolar couplings, and between calculated and target interproton distance restraints in the structures calculated without NOE distance restraints.

Summaries of the agreement with the dipolar couplings and the pairwise atomic root-mean-square (rms) differences for the various structures are presented in Tables 2 and 3, respectively. None of the structures (with the exception of the set calculated without NOE interproton distance restraints) exhibited NOE interproton distance violations greater than 0.5 Å. Similarly, none of the structures exhibited torsion angle restraint violations greater than 1° , and the majority have no torsion angle violations whatsoever. Best-fit superpositions of various structures are shown in Figure 4. Figure 5 shows the effect of $k_{\text{base–base}}$ on the radius of gyration (R_{gyr}), Figures 6–8 illustrate various atomic rms differences, including coordinate precision,³⁹ as well as the agreement between observed and calculated dipolar couplings and NOE restraints, as a function of $k_{\text{base–base}}$, and Figures 9–14 display the variation in helical twist, roll, helical rise, slide, and propellor twist as a function of base pair step (or base pair in the case of the propellor twist) for various structures.

Impact of the Description of the Nonbonded Contacts. In the paper by Tjandra et al.,⁷ all structures were calculated with

Lennard-Jones and electrostatic terms in the target function. It was observed that, in the absence of dipolar couplings, the resulting structure, LJ-nodipo (precision of 0.4 Å; N. Tjandra, personal communication), is compacted by 3 – 4 Å relative to B DNA (Figure 4f) and has a low R_{gyr} value of 12.79 Å (Figure 5). In addition, LJ-nodipo exhibits a distinct bend with a bend angle of 7.5° (measured by the angle between the average of the base normals for base pairs 2–4 and 9–11). The introduction of dipolar couplings (1DUF, precision of 0.06 Å; N. Tjandra, personal communication) corrects this problem, resulting in an essentially straight structure (bend angle $\sim 3.5^\circ$) (Figure 4e) with a R_{gyr} of 13.43 Å (Figure 5), without affecting the agreement with the interproton distance restraints to any significant degree. The rms difference between these two structures is 1.6 Å (for all 12 base pairs). From these data, one would therefore conclude that the dipolar couplings have a large effect on global structure. If these calculations, however, are repeated with the Lennard-Jones and electrostatic terms omitted and replaced by a simple quartic van der Waals repulsion term,³¹ a quite different perspective emerges. First, the structures are 2 – 2.5 Å longer than regular B DNA (Figure 4b,c) with R_{gyr} values ranging from 14.0 to 14.4 Å (Figure 4); second, the overall rms difference between the structures calculated with NOE interproton distance and dipolar coupling restraints (full00, precision 0.13 Å), without dipolar coupling restraints (nodipo00, precision 0.53 Å), and with dipolar coupling restraints but no NOE interproton distance restraints (dipo00, precision 0.45 Å) is rather small, namely only ~ 0.6 Å, and therefore within the errors of these coordinates (Figure 4b); and third, all the structures are essentially straight,

(39) Coordinate precision is defined as the average atomic rms difference (for all heavy atoms) between the individual simulated annealing structures and the mean coordinate positions (obtained after best-fitting the individual structures to base pairs 1–12).

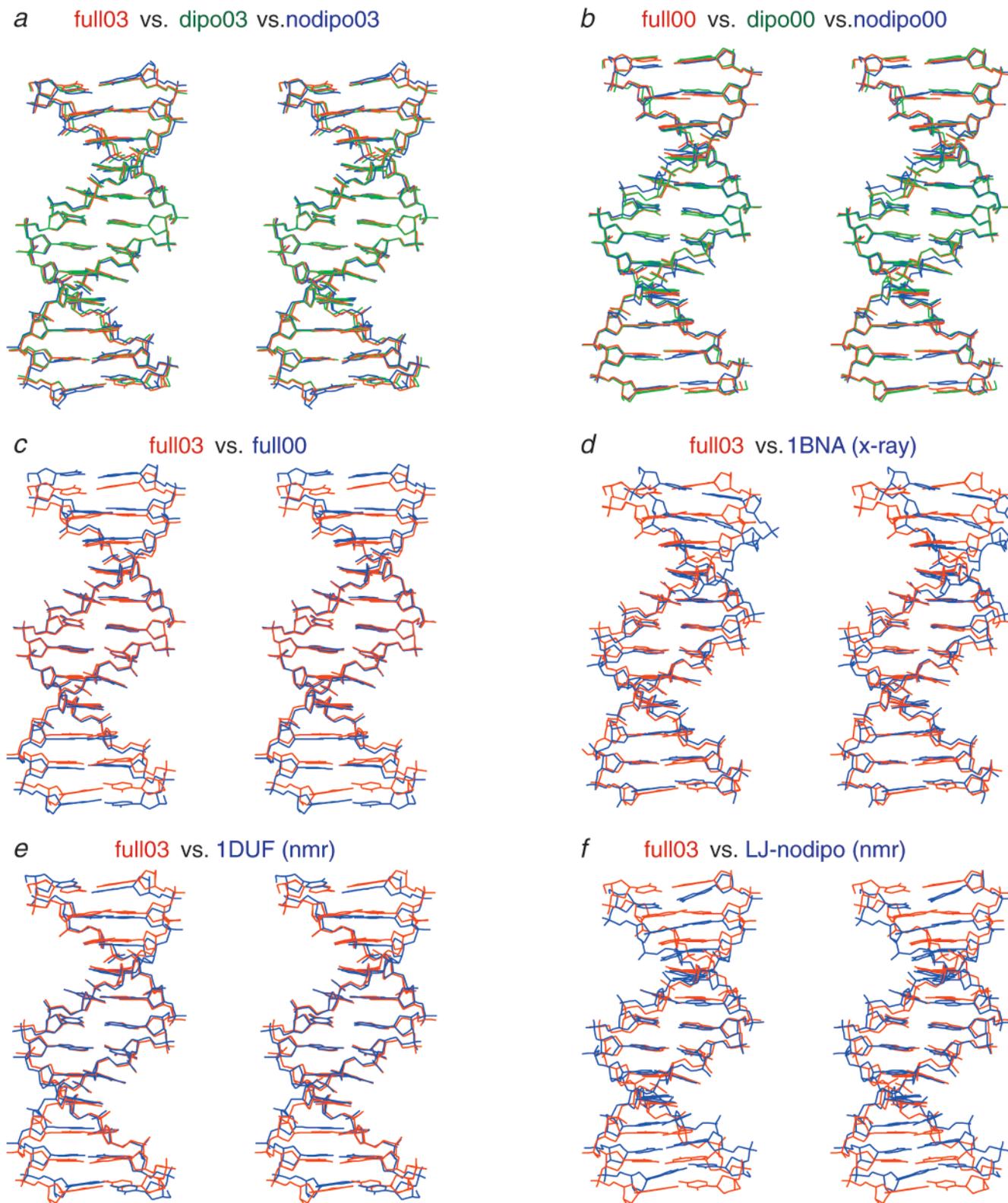


Figure 4. Stereoviews illustrating best-fit superpositions of the various structures. The structure nomenclature is given in the text. Best-fitting is carried out using all 12 base pairs, except in the case of the superposition involving the 1BNA crystal structure (d), where best-fitting is carried out with respect to base pairs 1–9. (Note the kink between base pairs 10 and 11 in 1BNA which results in an asymmetric structure.)

with a bend angle ranging from 1.5° to 3.5° . Thus, in the case of a target function comprising only a quartic van der Waals repulsion term to describe the nonbonded contacts, the impact of dipolar couplings on the global structure in this particular instance is rather small, although there are some significant differences in local helical parameters (Figures 9–13). If we

then consider the structures calculated with the full complement of NOE interproton distance and dipolar coupling restraints, we find that the rms difference between the structure calculated with a van der Waals repulsion term (full00, precision 0.13 Å) and that calculated with the Lennard-Jones and electrostatic terms (1DUF, precision 0.06 Å) is 1.2 Å, yet both structures

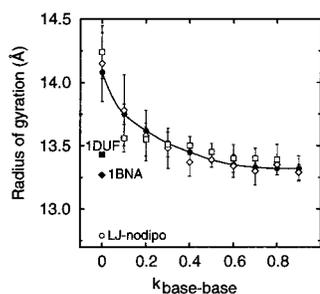


Figure 5. Effect of the force constant, $k_{\text{base-base}}$, for the DELPHIC base–base positioning potential on the radius of gyration, R_{gyr} , for the restrained regularized mean structures calculated with both NOE interproton distance and dipolar coupling restraints (●), with only dipolar couplings (◇), and with only NOE interproton distance restraints (□). The radii of gyration for the X-ray structure, 1BNA (◆), and the structures calculated with Lennard-Jones and electrostatic terms in the target function on the basis of both NOE and dipolar coupling restraints (1DUF; ■) and only NOE restraints (LJ-nodipo; ○) are also indicated for comparison. The vertical bars represent the standard deviations in R_{gyr} observed in each ensemble of simulated annealing structures.

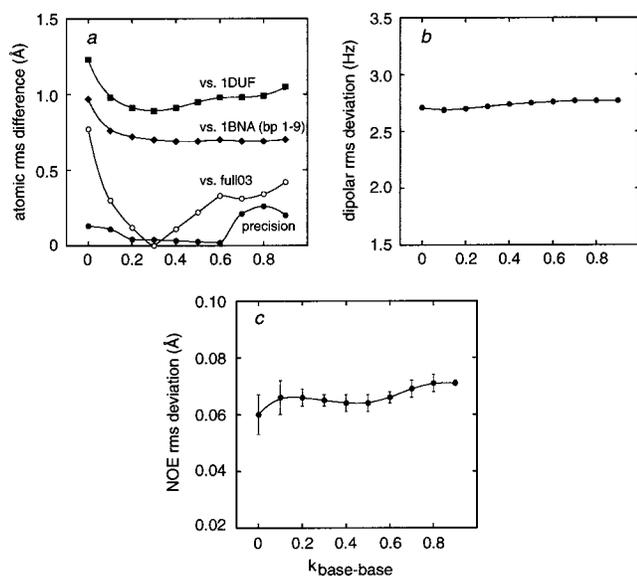


Figure 6. Effect of the force constant, $k_{\text{base-base}}$, for the DELPHIC base–base positioning potential in the case of structures calculated with the full NOE and dipolar coupling data set. (a) The coordinate precision and the atomic rms differences of the restrained regularized mean structures to full03, 1BNA, and 1DUF are shown in (a), the rms deviation between observed and calculated dipolar couplings is shown in (b), and the rms deviation between observed and calculated NOE interproton distance restraints is shown in (c). Vertical bars in (b) and (c) represent the standard deviations in the values observed for each ensemble of simulated annealing structures. The coordinate precision for the structures calculated with different values of $k_{\text{base-base}}$ is defined as the average atomic rms difference (for all heavy atoms) between the individual simulated annealing structures in a given ensemble and the corresponding mean coordinate positions (obtained after best-fitting the individual structures to base pairs 1–12).

satisfy the experimental NOE and dipolar coupling data (Table 2) equally well. This atomic rms difference is far larger than the small differences (~ 0.3 Å) observed upon changes in the magnitude of the alignment tensor over a range of -15.5 to -16.5 Hz for D_a^{CH} and 0.22 to 0.3 for the rhombicity.⁷

The results lead to two very important conclusions. First, the description of the nonbonded contacts clearly has a large influence on the resulting structures (irrespective of the presence or absence of dipolar couplings). Second, for the purposes of

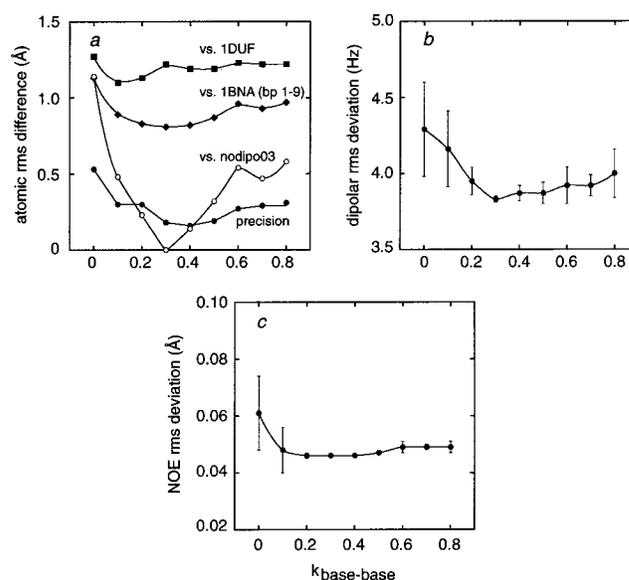


Figure 7. Effect of the force constant, $k_{\text{base-base}}$, for the DELPHIC base–base positioning potential in the case of structures calculated with only NOE-derived interproton distance restraints. (a) The coordinate precision and the atomic rms differences of the restrained regularized mean structures to nodipo03, 1BNA, and 1DUF are shown in (a), the rms deviation between observed and calculated dipolar couplings is shown in (b), and the rms deviation between observed and calculated NOE interproton distance restraints is shown in (c). Vertical bars in (b) and (c) represent the standard deviations in the values observed for each ensemble of simulated annealing structures.

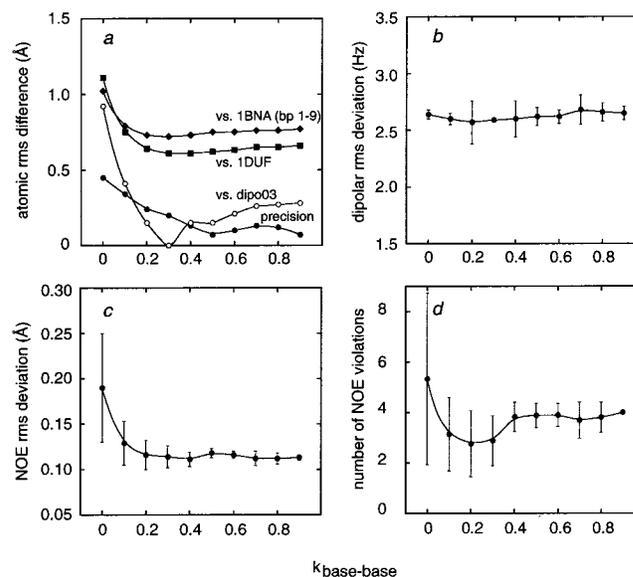


Figure 8. Effect of the force constant, $k_{\text{base-base}}$, for the DELPHIC base–base positioning potential in the case of structures calculated with only dipolar coupling restraints (a). The coordinate precision and the atomic rms differences of the restrained regularized mean structures to dipo03, 1BNA, and 1DUF are shown in (a), the rms deviation between observed and calculated dipolar couplings is shown in (b), the rms deviation between observed and calculated NOE interproton distance restraints is shown in (c), and the number of NOE interproton distance violations greater than 0.5 Å is shown in (d). Vertical bars in (b), (c), and (d) represent the standard deviations in the values observed for each ensemble of simulated annealing structures.

NMR structure determination, it is clearly desirable to develop a description of the nonbonded contacts that does not have an intrinsic tendency to distort the structures. That is to say, the global structure should not be systematically compacted or

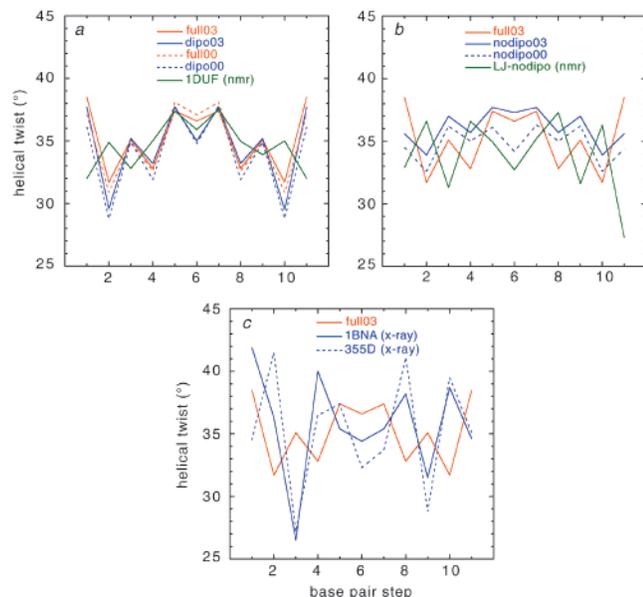


Figure 9. Variation in helical twist as a function of base pair step for the various structures.

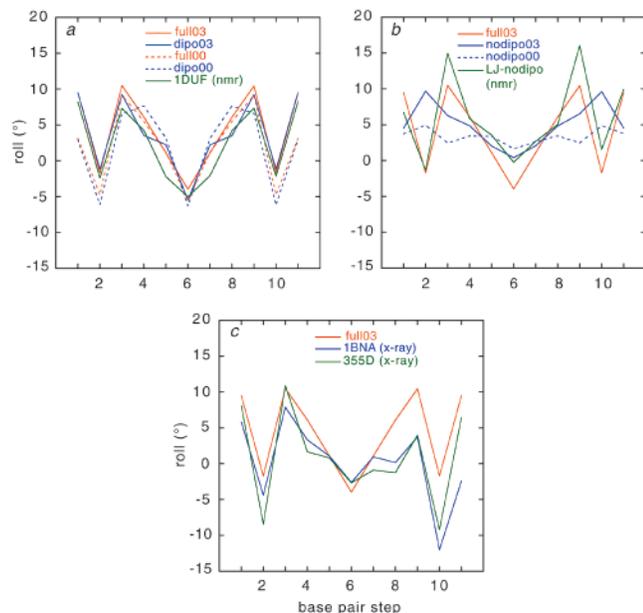


Figure 10. Variation in base pair roll as a function of base pair step for the various structures.

expanded and should not be bent unless warranted by the experimental NMR restraints.

Impact of the Inclusion of the DELPHIC Base–Base Positioning Potential in the Target Function. The impact of the DELPHIC base–base positioning potential is clearly illustrated by the various structure superpositions shown in Figure 4, as well as the plots displayed in Figures 5–8.

The DELPHIC base–base positioning potential does not introduce artifactual bending of the DNA since all the fullxx, dipox, and nodipox structures are essentially straight with bend angles of 0.5–4°. Increasing the value of $k_{\text{base–base}}$, independent of the presence or absence of dipolar couplings, progressively compacts the structures relative to structures calculated with purely a van der Waals repulsion term up to a limiting R_{gyr} value of ~13.3 Å. This value is the same as that observed in the 1BNA crystal structure of the dodecamer (Figure 5). This is effectively achieved by reducing the helical rise to

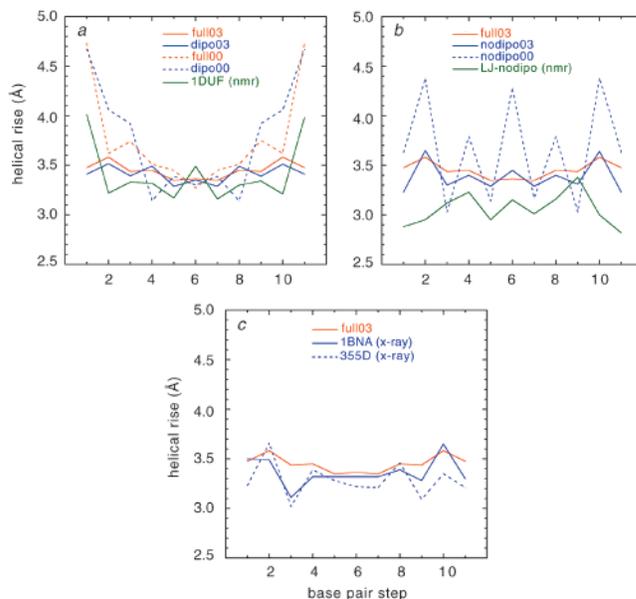


Figure 11. Variation in helical rise as a function of base pair step for the various structures.

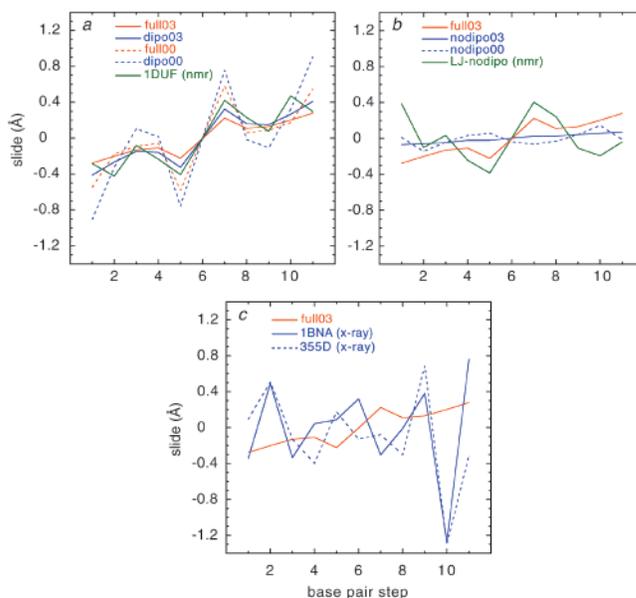


Figure 12. Variation in base pair slide as a function of base pair step for the various structures.

values characteristic of B DNA. In contrast to the Lennard–Jones and electrostatic terms, the DELPHIC base–base positioning potential does not overly compress the DNA and hence does not distort the resulting structures (as is the case of the LJ-nodipo structure, for example; cf. Figure 4a,e).

The question therefore arises as to what is the optimal force constant, $k_{\text{base–base}}$, for the DELPHIC base–base positioning potential. This can be assessed using three criteria: (a) cross-validation against dipolar couplings (i.e., structures calculated only from NOE interproton distance restraints); (b) cross-validation against NOE interproton distance restraints (i.e., structures calculated with only dipolar coupling restraints); and (c) comparison with the 1BNA and 355D crystal structures and the 1DUF NMR structure. With regard to the crystal structures, we note that neither 1BNA nor 355D is symmetric;^{20,21} this is entirely due to crystal packing forces²² since in solution the structure is perfectly symmetric (i.e., only one set of resonances is observed).^{7,23} Specifically, in both crystal structures there is

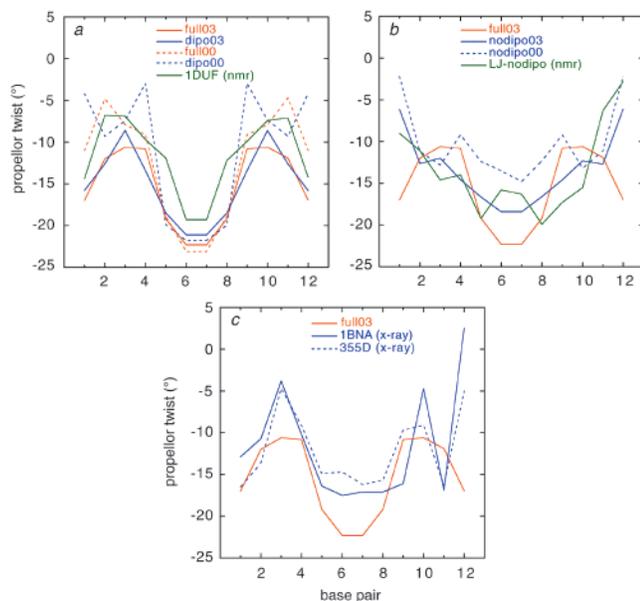


Figure 13. Variation in base pair propellor twist as a function of base pair for the various structures.

a large kink between base pairs 10 and 11, manifested by unduly large values for the role (Figure 10) and slide parameters (Figure 12) at base pair step 10. This results in an apparent overall bend angle of $\sim 10^\circ$. Consequently, rms comparisons with the crystal structure are restricted to base pairs 1–9. It should also be emphasized that in contrast to the case of proteins, crystal structures of free DNA cannot be regarded as an absolute gold standard, precisely because of the effects of crystal packing.

First, we note that the agreement between observed and calculated dipolar couplings for the structures calculated on the basis of only the NOE interproton distance restraints is significantly better for the structure calculated with a van der Waals repulsion term (nodipo00, 4.01 Hz for base pairs 1–12 and 3.29 Hz for base pairs 2–11) than for the structure calculated with the Lennard-Jones and electrostatic terms (LJ-nodipo, 5.46 Hz for base pairs 1–12 and 5.26 Hz for base pairs 2–11). Moreover, the nodipo00 structure agrees better with the dipolar couplings than either of the two crystal structures (1BNA, 4.86 Hz for base pairs 1–12 and 3.59 Hz for base pairs 2–11; 355D, 5.79 Hz for base pairs 1–12 and 4.53 Hz for base pairs 2–11). As an aside, it is perhaps not surprising that the 1BNA crystal structure agrees better with the dipolar couplings than 355D, since the former was solved at room temperature, while the latter was solved at -137°C and consequently is likely to be frozen out in one conformation, perhaps subject to a higher degree of distortion from crystal packing forces.

In the case of structures calculated with both NOE interproton distance and dipolar coupling restraints (fullxx), the agreement with the dipolar couplings and NOEs is only minimally affected as $k_{\text{base-base}}$ increases (Figure 6b,c), and in all cases the experimental restraints are satisfied within their experimental errors. However, the agreement with the measured dipolar couplings in the case of structures calculated with only NOE interproton distance restraints (nodipox) improves as $k_{\text{base-base}}$ increases up to a value of 0.3 (3.82 Hz for base pairs 1–12 and 3.31 Hz for base pairs 2–11) (Figure 7b). As $k_{\text{base-base}}$ is increased further, there is a minimal deterioration in the agreement between observed and calculated dipolar couplings (Figure 7b). The same pattern of behavior is observed for the structures calculated with only dipolar couplings (Figure 8): namely, the agreement between observed and calculated NOE

interproton distance restraints (as measured by both the rms difference between observed and calculated values, and the number of interproton distance violations greater than 0.5 \AA) improves, with optimal agreement for $k_{\text{base-base}} = 0.3$ (Figure 8c,d). Thus, these two sets of cross-validation results provide direct and independent evidence that the introduction of the DELPHIC base–base positioning potential results in an increase in accuracy, as judged by their agreement with the cross-validated NMR observables.

It is also worth noting that the rms difference between the structures calculated with $k_{\text{base-base}}$ values ranging from 0.1 to 0.9 is less than $\sim 0.4 \text{ \AA}$ (for base pairs 1–12) and that there is no significant difference between the coordinates calculated with $k_{\text{base-base}}$ ranging from 0.2 to 0.4 (i.e., they differ by $\leq 0.2 \text{ \AA}$). Thus, the actual value of $k_{\text{base-base}}$ employed, within obvious limits, is not that critical. The precision of the structures is high (e.g., 0.04, 0.17, and 0.18 \AA for full03, dipo03, and nodipo03, respectively), and there is no question that the introduction of the DELPHIC base–base positioning potential does result in an increase in precision (Figures 6–8). However, it is also clear from the data that the main contributing factor to the very high degrees of precision observed actually lies in the combination of NOE interproton distance restraints and dipolar couplings (cf. Figure 6a versus Figures 7a and 8a).

For all three sets of calculations, the introduction of the DELPHIC base–base positioning potential reduces the atomic rms difference between the calculated structures and the 1BNA crystal structure, and once again the optimal value of $k_{\text{base-base}}$ is ~ 0.3 (Figures 6a, 7a, and 8a, and Table 3). Thus, the atomic rms difference (base pairs 1–9) between full00, dipo00, and nodipo00 and 1BNA is 1.02, 1.02, and 1.13 \AA , respectively, compared to 0.69, 0.72, and 0.81 \AA for full03, dipo03, and nodipo03, respectively. For comparison, the atomic rms difference between 1DUF (the NMR structure calculated with the Lennard-Jones and electrostatic terms on the basis of the full NOE and dipolar coupling data set) and 1BNA is 0.9 \AA (base pairs 1–9), and that between the two crystal structures, 1BNA and 355D, is 0.9 \AA . Interestingly, all the NMR structures are closer to 1BNA than 355D, presumably reflecting increased distortions arising from crystal packing in the low-temperature 355D crystal structure. Nevertheless, the introduction of the DELPHIC base–base positioning potential still reduces the atomic rms difference to 355D. Thus, the atomic rms shift (base pairs 1–9) from 355D to full03, dipo03, and nodipo03 ranges from 1.1 to 1.3 \AA , compared to $1.4\text{--}1.5 \text{ \AA}$ from 355D to full00, dipo00, and nodipo00. For comparison, the atomic rms shifts from 355D to 1DUF and LJ-nodipo are 1.2 and 1.7 \AA , respectively.

The DELPHIC base–base positioning potential also reduces the atomic rms difference between the calculated structures and the 1DUF NMR structure, even in the case of structures calculated without dipolar couplings (Figures 6a, 7a, and 8a). Interestingly, however, the atomic rms difference between dipo03 and 1DUF (0.6 \AA for base pairs 1–12) is slightly less than those between full03 and 1DUF (0.8 \AA) and nodipo03 and 1DUF (1.2 \AA). The atomic rms differences, however, between full03 and nodipo03, full03 and dipo03, and dipo03 and nodipo03 are only 0.6, 0.5, and 0.9 \AA , respectively. Indeed, a best-fit superposition of full03, dipo03, and nodipo03 indicates that all three structures are very similar, indeed (Figure 4a). Moreover, the agreement of full03 with the experimental NOE and dipolar coupling data is essentially identical to that of 1DUF. Thus, the difference between 1DUF and full03 is entirely attributable to the description of the nonbonded contacts.

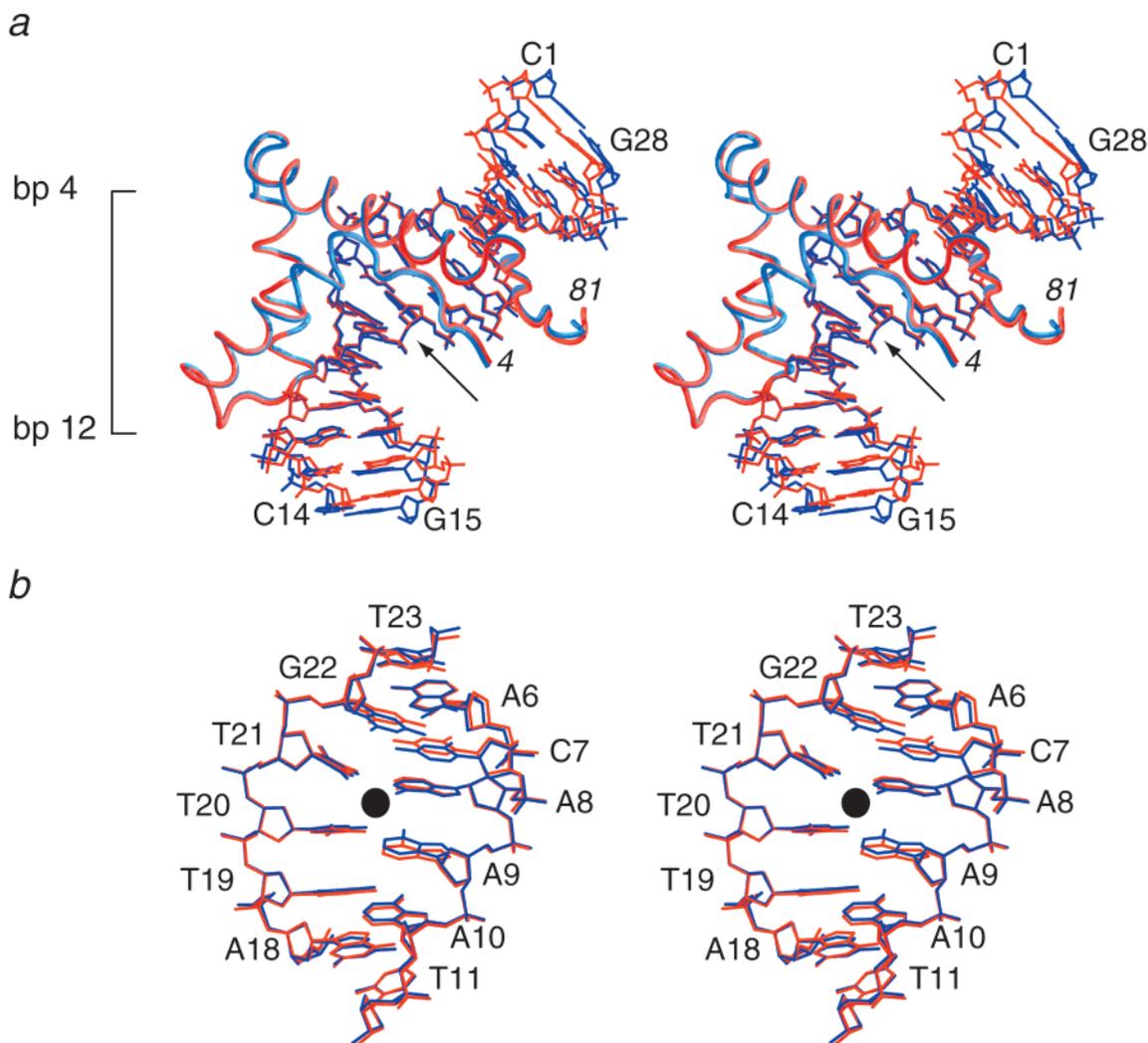


Figure 14. Comparison of the DNA structure in the SRY–DNA complex calculated with (red) and without (blue) the base–base positioning potential. (a) Stereoview providing a global comparison displaying base pairs 1–14 with the protein backbone depicted as a tube; the coordinates are best-fit to residues 4–81 of the protein and base pairs 4–12 of the DNA. The site of partial intercalation of the side chain of Ile13 between base pairs 8 and 9 is indicated by an arrow. The overall bend angle of the DNA is $\sim 60^\circ$. (b) Detailed comparison of the DNA structure around the site of intercalation (indicated by the solid circle); the coordinates are best-fit to base pairs 6–11. The structures shown are the restrained regularized mean structures obtained by restrained regularization of the mean coordinates derived from an ensemble of 200 simulated annealing structures calculated on the basis of 2793 experimental NMR restraints.³⁶ The precision of the coordinates for the structures calculated with the base–base positioning potential is 0.21 Å for the DNA and 0.24 Å for the protein backbone plus DNA; the corresponding values for the structures calculated without the base–base positioning potential are 0.31 and 0.33 Å, respectively. The value of $k_{\text{base–base}}$ is 0.25.

A summary of the DELPHIC base–base positioning energy ($E_{\text{DELPHIC–position}}$) for the various structures calculated for $k_{\text{base–base}} = 0.3$ is provided in Table 4. It can be seen that the values for $E_{\text{DELPHIC–position}}$ for the full03, dipo03, and nodipo03 structures and the two X-ray structures, 1BNA and 355D, are comparable (-955 to -985 kcal mol $^{-1}$), even though the global features of the two X-ray structures are clearly subject to crystal packing forces, as manifested by the presence of distinct asymmetry. The values for $E_{\text{DELPHIC–position}}$ for the 1DUF (-810 kcal mol $^{-1}$) and full00 (-830 kcal mol $^{-1}$) structures, however, are significantly lower. At the very least, this indicates that the base–base interactions observed in full03 (as well as dipo03 and nodipo03) are closer to the overall ensemble of interactions observed in the crystallographic DNA database than those in full00 and 1DUF. Taken together with the cross-validation results discussed above, this suggests that full03 probably represents a more accurate picture of the “true” solution structure than 1DUF.

Structural Characteristics of Structures Calculated with the DELPHIC Base–Base Positioning Potential. The structural features of the calculated structures as a function of base pair step (or base pair) are compared in Figure 9–13, and provide a means of assessing both the impact of the description of the nonbonded contacts and the presence or absence of dipolar couplings on various local helical parameters. The dodecamer has five different types of base pair steps: CpG at steps 1, 3, 9, and 11; GpC at steps 2 and 10; GpA at step 4 and the equivalent TpC at step 8; ApA at step 5 and the equivalent TpT at step 7; and ApT at step 6.

Although there are small quantitative differences in the actual values of the helical twist, it is evident from Figure 9a,b that the variation in helical twist follows the same trends for the structures calculated with the DELPHIC base–base positioning potential (full03, dipo03, nodipo03) and with only the hard van der Waals repulsion term (full00, dipo00, nodipo00). The main differences relative to 1DUF (Figure 9a) occur at the first and

Table 4. Values of the DELPHIC Base–Base Position Energy ($E_{\text{DELPHIC-position}}$) for the Various Structures

structure	$E_{\text{DELPHIC-position}}$ (kcal mol ⁻¹) ^a
full03 ^b	-955
dipo03 ^b	-961
nodipo03 ^b	-985
full00	-826
dipo00	-778
nodipo00	-905
1DUF (NMR)	-809
LJ-nodipo (NMR)	-854
1BNA (X-ray)	-951
355D (X-ray)	-986

^a Calculated with the force constant $k_{\text{base-base}} = 0.3$. ^b The values for $E_{\text{DELPHIC-position}}$ for the (full03), (dipo03), and (nodipo03) ensembles of simulated annealing structures are -948 ± 1 , -955 ± 5 , and -982 ± 1 kcal mol⁻¹, respectively.

second base pair steps, where 1DUF appears overwound and underwound, respectively, relative to the other six structures. The LJ-nodipo structure (Figure 9b) follows a trend similar to that observed in 1DUF, but the variations are somewhat larger. One can therefore conclude that, while the quantitative values of helical twist are influenced by the presence or absence of dipolar couplings, the description of the nonbonded contacts in terms of hard-sphere van der Waals repulsion versus Lennard-Jones and electrostatics is also a significant contributory factor, independent of the presence or absence of the DELPHIC base–base positioning potential. Thus, for example, the values for the helical twist for full03 are almost identical to those of full00, and likewise for dipo03 and dipo00 (Figure 9a). There are somewhat larger differences between nodipo03 and nodipo00 (Figure 9b), reflecting the fact that the structures are less well restrained by the NOE data than by the dipolar coupling data. In contrast to the NMR structures, there are much larger variations in helical twist observed for the 1BNA and 355D X-ray structures (Figure 9c). The trends, however, observed for the two X-ray structures, with the notable exception of base pair steps 3/9 (CpG/CpG) and 4/8 (GpA/TpC), are similar to those observed for full03.

As in the case of helical twist, base roll in the presence of dipolar couplings is only minimally influenced by the description of the nonbonded contacts (Figure 10a). In the absence of dipolar coupling data, however, the differences are much larger (Figure 10b). Thus, while full03 and LJ-nodipo follow similar trends, one finds that nodipo03 exhibits significantly larger roll at base pair steps 2/9 (GpC) (Figure 10b). Nodipo00, on the other hand, exhibits minimal variation in roll angles, although the trends follow those of nodipo03 (Figure 10b). With the exception of asymmetry, both crystal structures follow trends similar to those observed for full03 (Figure 10c).

Helical rise is heavily influenced by the description of the nonbonded contacts both quantitatively and qualitatively (Figure 11). Thus, the values and trends of the helical rise for the full03, dipo03, and nodipo03 structures are very similar and clearly different from those of the full00, dipo00, and nodipo00 structures, where the helical rise overall is systematically increased and in addition shows much greater variability (cf. the nodipo00 structure, for example) (Figure 11a,b). With the exception of base pair steps 1/11 (CpG) and 6 (ApT), the trends in helical rise for 1DUF are similar to those of full03 and dipo03, but the actual values are systematically smaller by 0.1–0.2 Å (Figure 11a). In the case of LJ-nodipo, the values of the helical rise are systematically smaller by 0.3–0.6 Å than those of full03 (Figure 11b). The helical rise observed for the two X-ray structures is very similar to that of full03, with the exception

of base pair step 3 (and in the case of 355D base pair step 9 as well), where the rise is significantly less for the X-ray structures (Figure 11c).

Slide, unlike helical rise, is heavily influenced by the dipolar couplings (Figures 12a,b). The patterns of slide for full03, dipo03, and 1DUF are qualitatively and quantitatively similar (Figure 12a). The pattern of slide for full00 and dipo00 is similar to that of full03, but the variations are somewhat larger, particularly for dipo00 (Figure 12a). In the absence of dipolar couplings, the overall trend for nodipo03 and nodipo00 is similar to that for full03, but the range of variation is very much reduced, such that slide falls within a ± 0.15 Å range (Figure 12b). However, for LJ-nodipo, large variations in slide are observed, and the trends bear no similarity to those observed for either 1DUF or full03 (Figure 12b). With the exception of base pair steps 10 and 11, the values of the slide parameter for the two X-ray structures lie within a range of ± 0.4 Å (Figure 12c), comparable to the NMR structures.

The final parameter we have plotted is base pair propellor twist. Full03, dipo03, and 1DUF show very similar trends, but the magnitude of the propellor twist is 3–5° smaller for 1DUF relative to those for full03 and dipo03 (Figure 13a). Propellor twist for the central four base pairs are essentially the same for full03, full00, and dipo00, but the magnitude of the propellor twist for the outer four bases is systematically reduced (Figure 13a). With the exception of the first (and last base pair), nodipo03 exhibits patterns of propellor twist similar to those of full03 (Figure 13b); nodipo00, on the other hand, exhibits rather little variation in propellor twist, with the exception of the first and last base pair (Figure 13b). LJ-nodipo, however, displays significantly less propellor twist for the two central base pairs 6 and 7 (Figure 13b). The variation in propellor twist for the first 10 base pairs of the two X-ray structures follows the trends observed for full03 (Figure 13c).

From these data, we can conclude that the description of the nonbonded contacts and the presence or absence of dipolar couplings has a significant impact on local helical parameters. Moreover, the nonbonded contacts have a much larger quantitative effect on these parameters than do small changes in the magnitude of the dipolar coupling alignment tensor employed in the calculations.⁷

Impact of the DELPHIC Base–Base Positioning Potential in the Case of Unusual DNA Structures. The purpose of the DELPHIC base–base positioning potential is to ensure that structures are generated that not only are consistent with the experimental NMR restraints but also reflect the large range of base–base interactions that are known to occur from high-resolution crystal structures, thereby circumventing deficiencies arising from the usual descriptions of the nonbonded contacts, either in terms of a Lennard-Jones potential or a simple van der Waals repulsive potential. While it is clear from the above calculations on a B DNA dodecamer that the DELPHIC base–base positioning potential improves the quality of NMR structures of DNA that conform to a prevalent conformational motif (in this case B DNA), it is still important to show that uncommon conformers are not washed out by this procedure. In this regard, we would emphasize that the force constant used for the DELPHIC base–base positioning potential ensures that the experimental restraints are the principal driving force governing the conformational space that is sampled. Thus, for example, if the experimental data were to indicate that a particular nucleotide was in a syn conformation with the base flipped out of the DNA helix, the DELPHIC base–base

positioning potential would not hinder in any way this conformation from being generated by simulated annealing.

To illustrate the impact of the DELPHIC base–base positioning potential on an unusual DNA structure, we have carried out a series of calculations based on 2793 experimental NMR restraints (including residual dipolar couplings) on a specific complex of the male sex determining factor SRY with a 14mer duplex DNA.³⁶ SRY is an architectural minor groove DNA binding protein that severely distorts and bends the DNA.⁴¹ Earlier structural work on a complex of SRY and a DNA octamer indicated that the minor groove was expanded and the major groove compressed, the DNA helix was underwound and partial intercalation of an Ile side chain between two A.T base pairs was present.⁴²

Figure 14 provides a comparison of the restrained regularized mean coordinates derived from two ensembles comprising 200 simulated annealing structures each, calculated with ($k_{\text{base–base}} = 0.25$) and without the DELPHIC base–base positioning potential. The precision of the coordinates for the structures calculated with the DELPHIC base–base positioning potential is 0.21 Å for the DNA and 0.24 Å for the protein backbone plus DNA; the corresponding values for the structures calculated without the base–base positioning potential are 0.31 and 0.33 Å, respectively. Of the 14 base pairs, the protein contacts only base pairs 3–13. Of these 11 base pairs, extensive contacts are made only with base pairs 4–12. Partial intercalation of the side chain of Ile13 occurs between base pairs 8 and 9. This location is at the center of the bend. The overall bend angle is $\sim 60^\circ$.

Figure 14a provides an overall superposition of the two restrained regularized mean structures, best-fit to residues 4–81 of the protein and base pairs 4–12 of the DNA. It is evident that the DNA between base pairs 4–12, the region that displays significant distortions, is essentially identical for the structures calculated with and without the DELPHIC base–base positioning potential with an atomic rms difference of 0.42 Å, which is within the error of the coordinates. Differences, however, are clearly evident for the first three base pair steps (i.e., base pairs 1/2, 2/3, and 3/4) and for the last two base pair steps (i.e., base pairs 12/13 and 13/14) of the 14mer, which are in a B conformation. As expected from the calculations on the B DNA dodecamer, the helical rise for these five base pair steps at the ends of the 14mer is significantly increased in the structures calculated without the DELPHIC base–base positioning potential (3.8–4.4 Å) compared to those calculated with it (3.3–3.6 Å). As a result, the overall atomic rms difference for base pairs 1–14 between the DNA calculated with and without the DELPHIC base–base positioning potential is 0.9 Å.

A detailed comparison of the two DNA structures calculated with and without the DELPHIC base–base positioning potential around the site of intercalation is provided in Figure 14b, which shows a blow-up of base pairs 6–11. The atomic rms difference for base pairs 6–11 between the two structures is only 0.24 Å. This portion of the DNA structure is far removed from either canonical A or B DNA. The sugar pucker conformations of A8, T19, and T20 are C3'-endo, while the other sugars are C2'-endo. The minor groove for base pairs 7–10 is expanded and shallow, with a width of 11–12 Å and a depth of 0–2 Å, while the major groove is compressed and deep, with a width of 9–10

Å and a depth of 8–10 Å. The DNA is severely underwound between base pairs 8 and 11, with a helical twist of 23–25°, while the twist for the remaining base pair steps ranges from 30 to 38°. The helical rise is also increased between base pairs 9/10 and 10/11, with values ranging from 4 to 4.5 Å, while that between base pairs 8 and 9 is reduced to ~ 3 Å. (For comparison, the average width and depth of the minor groove in regular B DNA are ~ 6 Å and ~ 4.5 Å, respectively, and in regular A DNA ~ 11 Å and ~ -0.5 Å, respectively; the average width and depth of the major groove in B DNA are ~ 11 Å and ~ 4 Å, respectively, and in A DNA ~ 2.4 Å and ~ 10 Å, respectively; the average helical twist in B and A DNA is 36° and 33°, respectively, and the average helical rise in B and A DNA is ~ 3.3 Å and ~ 2.6 Å, respectively.¹⁶) Thus, it is evident that in the case of a segment of DNA where the experimental NMR restraints are indicative of an unusual and distorted structure which deviates significantly from either A or B DNA, the DELPHIC base–base positioning potential does not hinder such an unusual conformation from being readily attained during simulated annealing.

Concluding Remarks

In this paper we have presented a DELPHIC base–base positioning potential of mean force derived from high-resolution DNA crystal structures which aims to provide a statistical description of the range of base–base interactions observed in DNA and can readily be employed in simulated annealing refinement. We have shown that the incorporation of the DELPHIC base–base positioning potential results in (a) a clear-cut improvement in the accuracy of the resulting structures as judged by complete cross-validation using both dipolar couplings and NOE interproton distance restraints as independent observables (cf. Figures 7 and 8); (b) base–base interactions that are both consistent with the experimental NMR restraints and observed in a large database of DNA crystal structures (cf. Table 4, Figures 4 and 9–13); and (c) the elimination of artifactual distortions in the structures arising from the limitations of conventional descriptions of the nonbonded contacts in terms of either Lennard-Jones van der Waals and electrostatic potentials or a simple van der Waals repulsion potential (cf. Figure 4). In addition, we have shown that the incorporation of the DELPHIC base–base positional potential does *not* in any way preclude unusual DNA structures, as illustrated by calculations on the SRY–DNA complex in which the DNA is bent, underwound, and highly distorted with an expanded minor groove and compressed major groove, and includes partial intercalation of an Ile side chain between two base pairs (Figure 14).

We expect that the application of a DELPHIC base–base positioning potential to NMR structure determination of RNA should be equally successful. Moreover, the application of the same methodology to describe side chain–side chain interactions in proteins and protein–protein complexes, and protein side chain–nucleic acid interactions in protein–nucleic acid complexes, should be valuable not only to NMR structure determination but as an aid to X-ray structure determination in cases where only relatively low resolution data (e.g., 3–3.5 Å) are available.

We have also shown that both local helical parameters and global long-range structure are sensitive not only to the presence or absence of dipolar couplings, but also to the description of the nonbonded contacts. Indeed, different structures can be

(40) Clore, G. M.; Garrett, D. S. *J. Am. Chem. Soc.* **1999**, *121*, 9008–9012

(41) Bewley, C. A.; Gronenborn, A. M.; Clore, G. M. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 105–131.

(42) Werner, M. H.; Juth, J.; Gronenborn, A. M.; Clore, G. M. *Cell* **1995**, *81*, 705–714.

(43) (a) Yip, P.; Case, D. A. *J. Magn. Reson.* **1989**, *83*, 643–648. (b) James, T. L. *Methods Enzymol.* **1994**, *239*, 416–439.

generated which satisfy the experimental NOE and dipolar coupling data equally well but differ in terms of overall length (e.g., full03 and 1DUF versus full00; cf. Figure 4b and e), in terms of overall atomic rms differences (cf. Table 3 and Figure 6), and in terms of various local helical parameters (Figures 9–13). More importantly, the effect of the description of the nonbonded contacts is much larger than that of small changes in the magnitude of the alignment tensor employed for the dipolar couplings.

An obvious concern that can potentially be raised regarding the DELPHIC base–base positioning potential is the possibility that its use may prevent one from observing conformations that have not been previously observed in the crystallographic database. We have already discussed this issue extensively with regard to the DELPHIC torsion angle database potential of mean force,^{15a} and the results presented on the SRY–DNA complex, which includes a region of highly distorted DNA (Figure 14), indicate that this is not a problem. In applying the DELPHIC base–base positioning potential, it is clearly important to ensure that the forces employed for the experimental and empirical restraints are appropriately balanced, since the aim is to ensure that one samples conformations that are both consistent with the experimental NMR restraints and represented in the crystallographic database. With the optimal force constants for the DELPHIC base–base positioning potential (cf. $k_{\text{base–base}} = 0.2\text{--}0.4$, see Figures 6–8) and the experimental NMR restraints set out in this paper, there is no question that if the experimental NMR restraints are indicative of base–base interactions that are not observed in the database, those interactions will be satisfactorily sampled and reproduced.

Another issue that deserves mention, particularly in the context of nucleic acid structures determined solely on the basis of NOE-derived interproton distance restraints, is the possible untoward effects of spin diffusion. Depending on the choice of mixing time, the relative intensities of NOE cross-peaks may vary due to spin diffusion. If the NOE intensities are interpreted conservatively in terms of generous interproton distance ranges to take into account spin diffusion effects,^{1b,2b} or if the structures are directly refined against NOE intensities using complete cross-relaxation matrix analysis,⁴³ spin diffusion is generally not an issue. However, if the NOE intensities are interpreted in a less conservative manner and narrow interproton distance ranges are employed, it is likely that some of the interproton distance restraints will be systematically underestimated.^{43b} While such systematic underestimation of interproton distance

restraints may not have any effect on the precision of the calculated structures, it will invariably result in a reduction in accuracy. At a local structural level, the use of both DELPHIC torsion angle and base–base positional potentials of mean force should partially compensate for such systematic errors in interproton distance restraints and, in addition, may highlight errors in the interproton distance restraints by introducing interproton distance violations. From the perspective of global structure, however, the effects of different nonbonded contact terms in the presence of systematically underestimated interproton distance restraints may be difficult to predict. In this regard, it should always be borne in mind that the DELPHIC database potentials are refinement tools aimed at increasing accuracy and as such should not perturb the structure extensively relative to that obtained with a purely repulsive van der Waals potential. One could, however, envision a situation under a very special set of circumstances, where the presence of a number of critical underestimated interproton distance restraints may introduce distortions in global structure, such as DNA bending, with one set of nonbonded contact terms but not another. Under these circumstances, the different solutions can be distinguished only by using additional experimental data, for example in the form of long-range orientational restraints, derived from residual dipolar couplings.

Finally, it must be borne in mind that the high levels of precision achieved by the use of dipolar couplings and/or the DELPHIC base–base positioning potential should not be confused with true accuracy. Thus, although the precision of the full03 (Figure 6a) and 1DUF structures (N. Tjandra, personal communication) is less than 0.1 Å, and the precision of the dipo03 and nodipo03 structures is ~ 0.2 Å (Figures 7a and 8a), mutual comparisons between these structures and between these structures and the two X-ray structures (Figures 6–8 and Table 2) suggest that the accuracy probably lies somewhere in the range of $\sim 0.5\text{--}0.8$ Å. At the present time, this probably represents the highest degree of accuracy that can realistically be achieved in NMR structure determinations of DNA.

Acknowledgment. We thank Ad Bax, Liz Murphy, Nico Tjandra, and Victor Zhurkin for useful discussions. This work was supported in part by the AIDS Targeted Antiviral Program of the Office of the Director of the National Institutes of Health (to G.M.C.). The coordinates of full03 and dipo03 have been deposited in the Protein Data Bank (accession code 1GIP).

JA010033U