New Parameters for the Refinement of Nucleic Acid-Containing Structures

GARY PARKINSON,[a]† JAROSLAV VOJTECHOVSKY,[a]† LESTER CLOWNEY,[a] AXEL T. BRÜNGER[b] AND HELEN M. BERMAN[a]

[a]*Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855-0939, USA, and* [b]*The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA*

Abstract

Structures at atomic resolution (up to 1.0 Å) which contain bases, sugars or the phosphodiester linkage, were selected from the Nucleic Acid Database or the Cambridge Structural Database to build a nucleic acid dictionary from X-ray refined structures. The dictionary consists of the average values for bond distances, bond angles and dihedral angles. The variance of the sample is used to provide information about the expected r.m.s. deviations of the refined parameters. A dictionary was constructed for refinement trials in *X-PLOR*. The dictionary includes RNA and DNA in C2'-*endo* and C3'-*endo* sugar pucker conformations, as well as values for the backbone dihedrals. Tests were performed on the dictionary using three structures: a B-DNA, a Z-DNA and a protein–DNA complex. During the course of refinement, all three structures showed significant improvements as measured by r.m.s. deviations and *R* factors when compared to the previous DNA dictionary.

1. Introduction

Refinement of macromolecular crystal and NMR structures requires knowledge of the geometry of the monomer components of the polymer chains, including bond distances, bond angles, dihedral angles and planarity. The use of molecular dynamics requires the additional knowledge of suitable energy constants for each geometric parameter. Equilibrium geometry and the energy constants can be determined from the statistical mean values and the sample standard deviations of a dependable set of high-resolution small-molecule crystal structures. In the case of proteins, this information was derived from the selection of suitable chemical fragments for 20 standard amino acids (Engh & Huber, 1991). These parameters are now in general use and have improved the refinements of protein structures.

Structures that contain nucleic acids, including protein–nucleic acid complexes, have been difficult to refine effectively with *X-PLOR* (Brünger, Kuriyan & Karplus, 1987) using current parameter dictionaries. As part of the Nucleic Acid Database Project (NDB) (Berman *et*

al., 1992), the standard geometries have been determined for all the nucleic acid components by a systematic analysis of well determined small molecules (Clowney *et al.*, 1996; Gelbin *et al.*, 1996). The variance of the sample is used to provide information about the expected r.m.s. deviations of the refined parameters. A dictionary was constructed for refinement trials in *X-PLOR*. The variance is used to calculate an applied force constant used during refinement. The scaling of the force constants is based on an iterative formulation and tested for three different crystallographic structures. The construction of the *X-PLOR* dictionary for nucleic acids, the scaling of the parameters for self consistency and the results of refinement will be presented.

2. Methods

2.1. Selection criteria

The base, the sugar and the phosphodiester backbone linkage were considered separately in determining average values of the geometric parameters and their standard deviations. The bond distances and bond angles for all three nucleic acid components and dihedral angles for the sugar and phosphodiester backbone linkage were chosen for parameterization.

The five standard bases, guanine, adenine, thymine, cytosine and uracil, were selected from the Cambridge Structural Database (CSD) (Allen *et al.*, 1979) for parameterization. Only structures without modifications and whose *R* factors were less than 0.06 and whose estimated standard deviations of the C—C bonds were not greater than 0.01 Å were considered for inclusion into the data set. A detailed discussion of the procedures used for analyses of these structures is given elsewhere (Clowney *et al.*, 1996).

The ribose and deoxyribose sugars associated with bases were selected from the CSD creating a mini database associated with NDB for further analysis. Only structures with *R* factors at least as good as 0.08 were included for the calculations of the mean values and the sample standard deviations. Although resolution values are not stated in the CSD, review of the original manuscripts shows that their structures are better than 1.0 Å resolution. There were statistically significant differences

---

† Both authors contributed equally on this paper.

between bond distances and bond angles of the ribose and deoxyribose sugars indicated by the t-test modified for two populations with different variances (Hamilton, 1964). The two sets contained hits for 80 ribose and 47 deoxyribose sugars. Additional analysis of the sugars revealed statistically significant differences between C2'-endo and C3'-endo conformations for external bond angles. A more detailed discussion of the derivation of these values is given elsewhere (Gelbin et al., 1996). Two sample sets containing 80 ribose sugars and 47 deoxyribose sugars were used to derive the values for bond distances and angles. The dihedral angles were also parameterized. The sample size for deriving dihedral values was 49 for C2'-endo ribose sugars, 27 for C2'-endo deoxyribose sugars and 24 for C3'-endo ribose sugars. The sample size of five for the C3'-endo deoxyribose sugar subset was considered as insufficient for the parameterization.

DNA and RNA structures containing the phosphodiester linkage were selected from NDB and were included if the R factor was below 0.08 and if the structure was refined by full-matrix least squares. This set contained structures with data between 0.8 to 1.0 Å resolution. No separation was made within the final set of ten structures for the calculations of the average bond distances, bond angles or their sample standard deviations. The same sample set was chosen for the parameterization of the dihedral angles. All three energetically favorable conformations of the phosphate backbone torsions α, γ and ζ were considered. This divided the resulting distributions of α, γ and ζ into three subsets. Because of the limited sample size, insufficient data existed for the analysis of torsion angles α and ζ in the trans conformation.

2.2. Dictionary and derivation of energy constants

The new topology and parameter files were developed from X-PLOR topology (toph11.dna) and parameter files (param11.dna) (Brooks et al., 1983; Brünger, Karplus & Petsko, 1989). The topology file was modified to include appropriate dihedrals. Two additional dihedrals, C5'—C4'—C3'—O3' (δ) and O4'—C1'—N9/1—C2/8 (χ), were added and one over-determined dihedral constraint, O5'—C5'—C4'—O4', was removed. The number of atom types was increased in order to reflect the unique bond types. The parameter file was modified to include the new equilibrium (X_eq) and energy constants [$k(x)$]. The derivation of energy constants [$k(x)$] for the new parameter file followed the work of Engh & Huber for their construction of an amino-acid dictionary (Engh & Huber, 1991). The equation used to determine an appropriate energy constant $k(x)$ was based on variance of the sample,

$$k(x) = C/\sigma(x)^2, \qquad (1)$$

where $\sigma(x)$ is the sample standard deviation for a particular type-based parameter x (bond distance, bond angle or dihedral angle) and where C is a constant applied

to provide consistency within the dihedral and improper dihedral force constants. The constant C was initially set equal to $kT_{298} = 0.592$ kcal mol^{-1}, corresponding to an assumption that the sample distribution follows the Boltzmann distribution at room temperature.

The refinement of the 2.5 Å resolution test structure of the Escherichia coli catabolite gene activator protein (CAP) complexed to the consensus DNA sequence (Parkinson, Gunasekera, Wilson, Ebright & Berman, 1996) using the new parameters indicated that the newly parameterized energy terms were over-weighted. Two new characteristics of the refinement were noted. The DNA energy constants were significantly higher than those for the protein, and even minimal shift in geometry caused high energy gradients. To correct this, it was necessary to consider that the constant C from (1) is
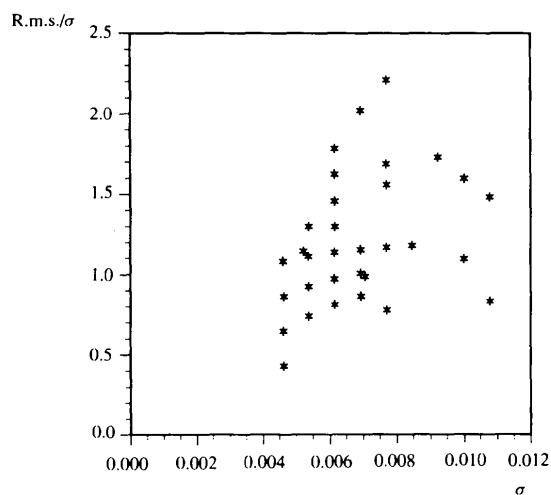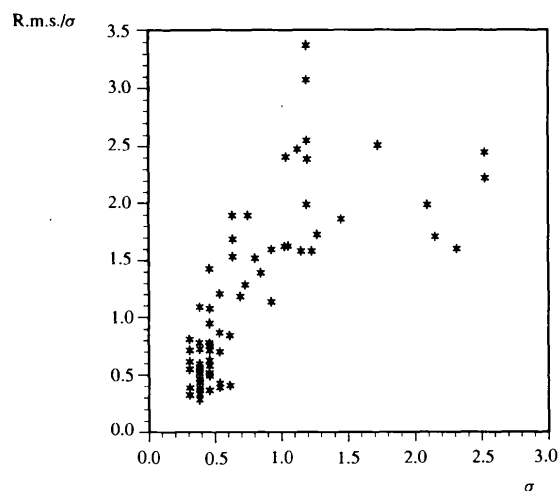
Fig. 1. A plot of the r.m.s.(x)/$\sigma(x)$ for each type-based parameter x against $\sigma(x)$. The r.m.s.'s were calculated from the structure of CAP–DNA14/17 complex after simulated annealing (a) bond-distance parameters and (b) bond-angle parameters.

actually a product of two factors C_1 and C_2. The first constant, $C_1 = kT_{298} = 0.592$, is based on the Boltzmann distribution at room temperature and is always included for the energy-constant calculation. The second scale C_2 is applied to balance the bond-distance, bond-angle and dihedral energy terms to each other and to the other energy terms in the *X-PLOR* energy function. Several cycles of simulated-annealing refinement were run with various estimated scales of C_2 applied to base, sugar and phosphodiester linkage parameters. The separation represented different structural features of nucleic acid components and their derivation from different sources of statistical data. The initial scales were estimated from their energy contributions and the resulting refinement r.m.s. deviations. However, these scales did not result in a completely balanced distribution. This can be seen from the graph of r.m.s./σ versus σ (Fig. 1) where the relationship is expected to be constant for an ideally balanced parameter set. To describe better the expected r.m.s distribution to σ, a set of new energy constants $k'(x)$ was calculated using (2). This takes into account the ratio of the refinement r.m.s. to the expected sample standard deviation σ,

$$k'(x) = \{[\text{r.m.s.}(x)/\sigma(x)]/\rho_{\text{AVE}}\}^{1/2}k(x), \qquad (2)$$

where $k(x)$ is the energy constant used in the previous cycle of simulated annealing and,

$$\rho_{\text{AVE}} = 1/N\sum_{x=1}^{N}\text{r.m.s.}(x)/\sigma(x),$$

for all N bond and angle parameters.

In this formula, the energy constant is increased for the parameters, where r.m.s.$(x)/\sigma(x)$ is higher than the overall average and *vice versa*. A plot of $k'(x)$ versus the original $k(x)$ ($C_2 = 1$, $C_1 = 0.592$) illustrates the clustering of the parameters. The slopes of the

linear regressions in the selected clusters were used to derive new sets of C_2 optimized for a balanced distribution of r.m.s. *versus* σ. Plots were made for each bond-distance and bond-angle type (Fig. 2) derived from (2). It showed three clusters corresponding to the parameters for bases, sugars and phosphates. The only exception is the group of phosphate bond distances that fitted into the cluster of sugar parameters, rather than phosphate angles. An analysis of r.m.s. and σ within each subset of data further suggested that the angles involving the connections between the base and the sugar belong with the sugar parameters. The external sugar-ring bond angles C2'—C3'—O3' and C4'—C3'—O3' belong with the phosphate parameters. The slopes of each linear regression are $C_2 = 0.188$ for base bonds and angles, $C_2 = 0.566$ for glycosidic bonds, sugar bonds, sugar angles, and phosphate bonds, and $C_2 = 1.548$ for phosphate angles, with correlation coefficient 0.953, 0.982 and 0.989, respectively. The energy constants for all bonds and angles were recalculated using the new scaling factors. A subsequent cycle of refinement using these values yielded a consistent relationship between r.m.s.(x) and $\sigma(x)$, showing that bond and angle energy terms were balanced among the different sources of data.

Dihedral angle energy constants were calculated from (1) assuming a Boltzman distribution. The dihedral angles were analyzed using the refinement of a B-DNA dodecamer (Vojtechovsky, Eaton, Gaffney, Jones & Berman, 1996) to 2.3 Å resolution. A graph of the new $k'(x)$ *versus* the old energy constants $k(x)$ was plotted using (2). All dihedral angles were internally consistent and no clustering was observed. As the B-DNA data set and refinement cannot be assumed as analogous to the CAP–DNA14/17 set, rescaling was not performed, thus leaving C_2 equal to 0.3. The selection for improper parameters and their energy constants was taken directly from the file param11.dna supplied with version 3.1 of *X-PLOR* (Brünger *et al.*, 1987). As was carried out for

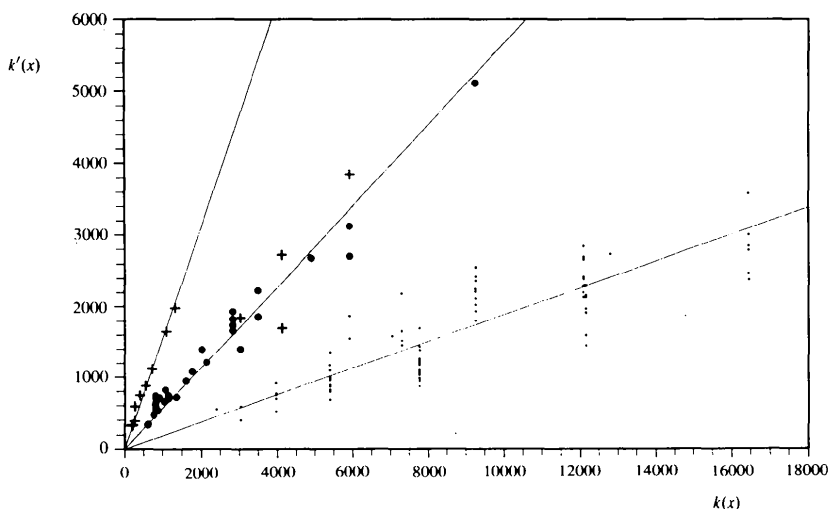Fig. 2. A plot of the energy constants $k'(x)$ as calculated from (2) against the original force constants $k(x)$ calculated from (1) with $C = kT_{298}$ for each type-based bond-distance and bond-angle parameter x. Points represent bond distances and bond angles from CSD. Circles represent bond distances and bond angles from the mini database. Crosses represent bond distances and bond angles from NDB.

the parameterization of protein residues (Engh & Huber, 1991), the energy constants were multiplied by three for scaling against the new parameter set. The same scaling was employed for all geometric parameters defining the H atoms. No alterations were made to the parameters of other terms used in the energy function.

3. Results and discussion

3.1. Implementation of parameters

The parameter and topology files for *X-PLOR* were appropriately modified. The number of atom types used in the topology file was extended in order to reflect the separation described in the selection criteria. Fig. 3 shows the nomenclature of the new atom types in the nucleic acid bases. There are two schemes for the sugar phosphate backbone, one for RNA (Fig. 4a) and one for DNA (Fig. 4b). The subroutine *DEOX* in the topology file was modified to assign the deoxyribose sugar atom types in the refinement of DNA.

The type-based bond-distance and bond-angle parameters, as well as their energy constants $k(x)$, equilibrium values x_eq and sample standard deviations $\sigma(x)$, are listed separately for nucleic acid bases (Table 1) and sugar-phosphate backbone (Table 2). For comparison,
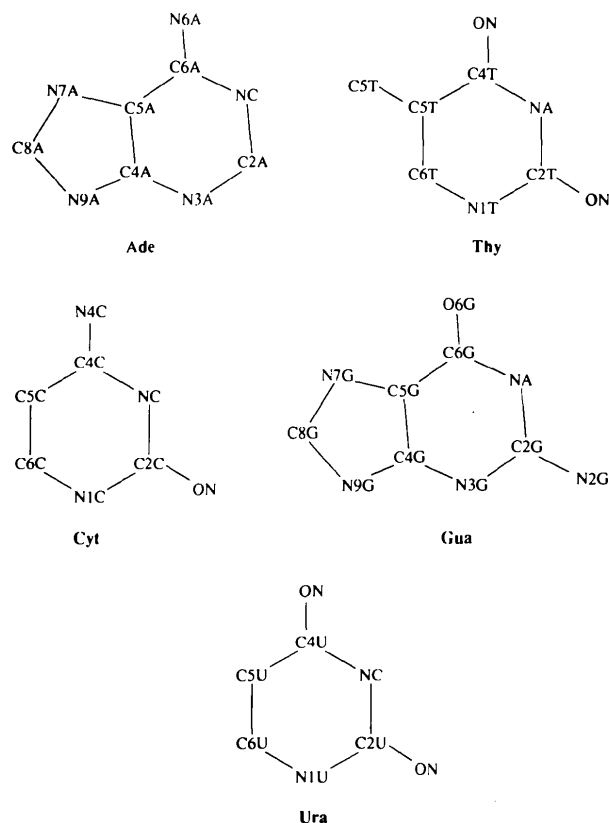
the equilibrium values from the previous parameter file, param11.dna, are shown under the column x_11. In many cases the equilibrium constants differ by more than several sample standard deviations.

Dihedral angle parameters are listed in Table 3. The equilibrium constants can be compared to the phase shifts and multiplicities used in the old parameter files (param11.dna). The differences in equilibrium values are particularly significant for sugar dihedral angles. The dihedrals of C3'-endo and C2'-endo were separated by a phase shift of 60° but this distribution could not be adequately modeled using the periodical potential. Therefore, a unique set of dihedral angle equilibrium values were constructed for each sugar conformation, C2'- and C3'-endo for both DNA and RNA, rather than using non-zero periodicities. The dihedrals of C3'-endo sugar conformation were accepted as the default in the parameter file for ribose and C2'-endo for deoxyribose sugars. The alternative C2'-endo and C3'-endo values are also provided. Backbone torsion angles α, γ and ζ can be represented as either unique target values, or, with the loss of accuracy, a phase shift plus a periodicity of 120°. As a default, the new parameter file uses an exact threefold approximation containing an appropriate phase shift and energy constant calculated from combining all three population states of the dihedral angles α, γ and ζ. This allows for refinement without manual intervention. Individual equilibrium values and their standard deviations are also provided for completeness.

To evaluate the self consistency of the new parameter file, *X-PLOR* energy minimization refinement was performed on five DNA nucleotides, with C2'-endo sugar pucker conformations (Table 4). Only bond, angle, dihedral and improper energy terms were included. The first column represents the original param11.dna, force constants k_11 and equilibrium constants x_11, while the second column shows the results using the new equilibrium constants x_eq and original force constants k_11. The r.m.s. deviations and maximum deviations are shown against bond distances, bond angles and dihedral angles. The use of the new equilibrium constants x_eq represents a marked improvement for distances, an order

Fig. 3. The nomenclature of the atom types used for the parameterization of the nucleic acid bases.
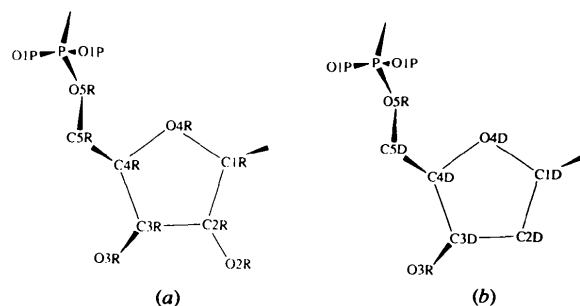

Fig. 4. The nomenclature of the atom types used for the parameterization of the RNA sugar-phosphate backbone (a) and DNA sugar-phosphate backbone (b).

Table 1. *The list of type-based bond-distance and bond-angle parameters, their energy constants k(x), equilibrium values x_eq and standard deviations σ(x) used for the parameterization of the nucleic acid bases compared to the original equilibrium values x_11 from param11.dna*

The symbol R/D is used in the sugar atom types for parameters that are the same for RNA and DNA.

	$k(x)$	x_eq	$\sigma(x)$	x_11
Cytosine				
C1R/D—N1C	2327	1.470	0.012	1.475
C2C—ON	1370	1.240	0.009	1.229
C4C—N4C	1370	1.335	0.009	1.333
N1C—C2C	1110	1.397	0.010	1.383
N1C—C6C	3083	1.367	0.006	1.365
C2C—NC	1734	1.353	0.008	1.358
NC—C4C	2265	1.335	0.007	1.339
C4C—C5C	1734	1.425	0.008	1.433
C5C—C6C	1734	1.339	0.008	1.350
Thymine				
C1R/D—N1T	1710	1.473	0.014	1.475
N1T—C2T	1734	1.376	0.008	1.383
C2T—NA	1734	1.373	0.008	1.388
NA—C4T	1734	1.382	0.008	1.388
C4T—C5T	1370	1.445	0.009	1.444
C5T—C6T	2265	1.339	0.007	1.343
C6T—N1T	2265	1.378	0.007	1.365
C2T—ON	1734	1.220	0.008	1.229
C4T—ON	1370	1.228	0.009	1.229
C5T—CH3E	3083	1.496	0.006	1.525
Adenine				
C1R/D—N9A	3351	1.462	0.010	1.475
NC—C2A	1370	1.339	0.009	1.324
C2A—N3A	1370	1.331	0.009	1.324
N3A—C4A	3083	1.344	0.006	1.354
C4A—C5A	2265	1.383	0.007	1.370
C5A—C6A	1370	1.406	0.009	1.404
C6A—NC	2265	1.351	0.007	1.339
C5A—N7A	3083	1.388	0.006	1.391
N7A—C8A	2265	1.311	0.007	1.304
C8A—N9A	1734	1.373	0.008	1.371
N9A—C4A	3083	1.374	0.006	1.374
C6A—N6A	1734	1.335	0.008	1.333
Guanine				
C1R/D—N9G	4137	1.459	0.009	1.475
NA—C2G	1734	1.373	0.008	1.381
C2G—N3G	1734	1.323	0.008	1.339
N3G—C4G	2265	1.350	0.007	1.354
C4G—C5G	2265	1.379	0.007	1.370
C5G—C6G	1110	1.419	0.010	1.419
C6G—NA	2265	1.391	0.007	1.388
C5G—N7G	3083	1.388	0.006	1.391
N7G—C8G	3083	1.305	0.006	1.304
C8G—N9G	2265	1.374	0.007	1.371
N9G—C4G	1734	1.375	0.008	1.374
C2G—N2G	1110	1.341	0.010	1.333
C6G—O6G	1370	1.237	0.009	1.229
Uridine				
C1R/D—N1U	4137	1.469	0.009	1.475
C2U—ON	1370	1.219	0.009	1.229
C4U—ON	1734	1.232	0.008	1.229
N1U—C2U	1370	1.381	0.009	1.383
N1U—C6U	1370	1.375	0.009	1.365
C2U—N3U	2265	1.373	0.007	1.388
N3U—C4U	1370	1.380	0.009	1.388
C4U—C5U	1370	1.431	0.009	1.444
C5U—C6U	1370	1.337	0.009	1.350
Cytosine				
C6C—N1C—C2C	2277	120.3	0.4	121.6
N1C—C2C—NC	744	119.2	0.7	118.6
C2C—NC—C4C	1458	119.9	0.5	120.5

	$k(x)$	x_eq	$\sigma(x)$	x_11
NC—C4C—C5C	2277	121.9	0.4	121.5
C4C—C5C—C6C	1458	117.4	0.5	117.0
C5C—C6C—N1C	1458	121.0	0.5	121.2
N1C—C2C—ON	1012	118.9	0.6	120.9
NC—C2C—ON	744	121.9	0.7	122.5
NC—C4C—N4C	744	118.0	0.7	119.8
C5C—C4C—N4C	744	120.2	0.7	120.1
C6C—N1C—C1R/D	764	120.8	1.2	121.2
C2C—N1C—C1R/D	909	118.8	1.1	117.6
Thymine				
C6T—N1T—C2T	1458	121.3	0.5	121.6
N1T—C2T—NA	1012	114.6	0.6	115.4
C2T—NA—C4T	1012	127.2	0.6	126.4
NA—C4T—C5T	1012	115.2	0.6	114.1
C4T—C5T—C6T	1012	118.0	0.6	120.7
C5T—C6T—N1T	1012	123.7	0.6	121.2
N1T—C2T—ON	569	123.1	0.8	120.9
NA—C2T—ON	1012	122.3	0.6	120.6
NA—C4T—ON	1012	119.9	0.6	120.6
C5T—C4T—ON	744	124.9	0.7	125.3
C4T—C5T—CH3E	1012	119.0	0.6	119.7
C6T—C5T—CH3E	1012	122.9	0.6	119.7
C6T—N1T—C1R/D	489	120.4	1.5	121.2
C2T—N1T—C1R/D	430	118.2	1.6	117.6
Adenine				
C6A—NC—C2A	1012	118.6	0.6	118.6
NC—C2A—N3A	1458	129.3	0.5	129.1
C2A—N3A—C4A	1458	110.6	0.5	111.0
N3A—C4A—C5A	744	126.8	0.7	127.7
C4A—C5A—C6A	1458	117.0	0.5	117.3
C5A—C6A—NC	1458	117.7	0.5	117.3
C4A—C5A—N7A	1458	110.7	0.5	110.4
C5A—N7A—C8A	1458	103.9	0.5	103.8
N7A—C8A—N9A	1458	113.8	0.5	113.9
C8A—N9A—C4A	2277	105.8	0.4	105.4
N9A—C4A—C5A	2277	105.8	0.4	106.2
N3A—C4A—N9A	569	127.4	0.8	126.0
C6A—C5A—N7A	744	132.3	0.7	132.4
NC—C6A—N6A	1012	118.6	0.6	119.8
C5A—C6A—N6A	569	123.7	0.8	123.5
C8A—N9A—C1R/D	339	127.7	1.8	128.8
C4A—N9A—C1R/D	339	126.3	1.8	125.8
Guanine				
C6G—NA—C2G	1012	125.1	0.6	125.2
NA—C2G—N3G	1012	123.9	0.6	123.3
C2G—N3G—C4G	1458	111.9	0.5	112.2
N3G—C4G—C5G	1458	128.6	0.5	127.7
C4G—C5G—C6G	1012	118.8	0.6	119.2
C5G—C6G—NA	1458	111.5	0.5	111.3
C4G—C5G—N7G	2277	110.8	0.4	110.4
C5G—N7G—C8G	1458	104.3	0.5	103.8
N7G—C8G—N9G	1458	113.1	0.5	113.9
C8G—N9G—C4G	2277	106.4	0.4	105.4
N9G—C4G—C5G	2277	105.4	0.4	106.2
N3G—C4G—N9G	1012	126.0	0.6	126.0
C6G—C5G—N7G	1012	130.4	0.6	130.0
NA—C2G—N2G	450	116.2	0.9	116.0
N3G—C2G—N2G	744	119.9	0.7	119.8
NA—C6G—O6G	1012	119.9	0.6	120.6
C5G—C6G—O6G	1012	128.6	0.6	128.8
C8G—N9G—C1R/D	651	127.0	1.3	128.8
C4G—N9G—C1R/D	651	126.5	1.3	125.8

Table 1 (cont.)

Uridine	$k(x)$	x_eq	$\sigma(x)$	x_11
C6U—N1U—C2U	1012	121.0	0.6	121.6
N1U—C2U—N3U	1012	114.9	0.6	115.4
C2U—N3U—C4U	1012	127.0	0.6	126.4
N3U—C4U—C5U	1012	114.6	0.6	114.1
C4U—C5U—C6U	1012	119.7	0.6	120.7
C5U—C6U—N1U	1458	122.7	0.5	121.2
N1U—C2U—ON	743	122.8	0.7	120.9
N3U—C2U—ON	743	122.2	0.7	120.6
N3U—C4U—ON	743	119.4	0.7	120.6
C5U—C4U—ON	1012	125.9	0.6	125.3
C6U—N1U—C1R/D	561	121.2	1.4	121.2
C2U—N1U—C1R/D	764	117.7	1.2	117.6

of magnitude improvement for angles and a 40-fold improvement for the dihedral angles. The r.m.s. deviations for distances are less than 0.001; for angles, less than 0.3; and for dihedrals, less than 0.75.

Three structures were then chosen to test the new energy and equilibrium geometry constants for crystallographic simulated-annealing and positional refinements: a B-DNA dodecamer using 10–2.25 Å resolution data, a Z-DNA hexamer using 10–1.35 Å resolution data (Parkinson, Arvanitis et al., 1996, and a CAP–DNA14/17 complex using 10–2.5 Å resolution data. There were no refinements carried out with A-form nucleic acids. For each of the test structures three sets of refinement were performed. The structures were first refined using the original parameter set provided in the X-PLOR package for DNA refinement, param11.dna. This provided the benchmark against which statistics would be compared. The second refinement used the new equilibrium geometry constants but with the original force constants from param11.dna (Table 5). This tested the integrity of the new equilibrium geometry independent of the scaled force constants. The third refinement shows the result of a full refinement using the new parameter set, containing both the new equilibrium geometry constants and new scaled force constants (Table 6). This test examines how much the model can be restrained without increasing the R factor. The 0.1% difference for CAP–DNA14/17 complex was negligible in comparison to the improvements of refinement r.m.s. deviations. There were no difficulties with the convergence of the newly parameterized models for any of the tested structures.

Using the new equilibrium geometry constants it was observed that bond-angle r.m.s. values for Z- and B-DNA improved substantially while r.m.s. improvements for bond distances were less dramatic. Dihedrals for B-DNA showed a moderate improvements while the addition of C3′-endo sugars for Z-DNA was an important contribution (Table 6). An artifact resulting from refinement using the earlier dictionary can be observed for the CAP–DNA14/17 complex from the r.m.s. values. The DNA component was over weighted in an attempt

Table 2. The list of type-based bond-distance and bond-angle parameters, their energy constants $k(x)$, equilibrium values x_eq and standard deviations $\sigma(x)$ used for the parameterization of the sugar-phosphate backbone compared to the original equilibrium values x_11 from param11.dna

The symbol R/D is used in the sugar atom types for parameters that are the same for RNA and DNA. The symbol N1/9 means either N9 of purine or N1 of pyrimidine.

	$k(x)$	x_eq	$\sigma(x)$	x_11
Backbone				
P—O1P	1159	1.485	0.017	1.480
P—O2P	1159	1.485	0.017	1.480
P—O5R	3351	1.593	0.010	1.610
P—O3R	2327	1.607	0.012	1.610
O5R—C5R/D	1309	1.440	0.016	1.430
RNA sugar				
C5R—C4R	1983	1.510	0.013	1.525
C4R—C3R	2769	1.524	0.011	1.525
C3R—C2R	2769	1.525	0.011	1.525
C2R—C1R	3351	1.528	0.010	1.525
O4R—C1R	2327	1.414	0.012	1.430
O4R—C4R	2327	1.453	0.012	1.430
O3R—C3R	1710	1.423	0.014	1.430
C2R—O2R	1983	1.413	0.013	1.430
DNA sugar				
C5D—C4D	5235	1.511	0.008	1.525
C4D—C3D	3351	1.528	0.010	1.525
C3D—C2D	3351	1.518	0.010	1.525
C2D—C1D	1710	1.521	0.014	1.525
O4D—C1D	1983	1.420	0.013	1.430
O4D—C4D	2769	1.446	0.011	1.430
O3R—C3D	1983	1.431	0.013	1.430

Angle	$k(x)$	x_eq	$\sigma(x)$	x_eq 11
Backbone				
O1P—P—O2P	1337	119.6	1.5	119.9
O5R—P—O1P	358	108.1	2.9	108.2
O5R—P—O2P	413	108.3	2.7	108.2
O3R—P—O5R	833	104.0	1.9	102.6
O2P—P—O3R	294	108.3	3.2	108.2
O1P—P—O3R	294	107.4	3.2	108.2
O5R—C5R/D—C4R/D	1535	110.2	1.4	112.0
P—O5R—C5R/D	1175	120.9	1.6	120.5
P—O3R—C3R/D	2089	119.7	1.2	120.5
RNA sugar				
O4R—C4R—C3R	561	105.5	1.4	111.0
C5R—C4R—C3R	489	115.5	1.5	111.0
C5R—C4R—O4R	561	109.2	1.4	111.0
C1R—O4R—C4R	1358	109.6	0.9	111.5
C4R—C3R—C2R	1100	102.7	1.0	111.0
C3R—C2R—C1R	1358	101.5	0.9	111.0
O4R—C1R—C2R	561	106.4	1.4	111.0
N1/9—C1R—C2R	430	113.4	1.6	111.0
O4R—C1R—N1/9	1100	108.2	1.0	111.0
C1R—C2R—O2R	334	110.6	3.0	111.0
C3R—C2R—O2R	358	113.3	2.9	111.0
C4R—C3R—O3R	445	110.6	2.6	111.0
C2R—C3R—O3R	384	111.0	2.8	111.0
DNA sugar				
O4D—C4D—C3D	1100	105.6	1.0	111.0
C5D—C4D—C3D	489	114.7	1.5	111.0
C5D—C4D—O4D	430	109.4	1.6	111.0
C1D—O4D—C4D	561	109.7	1.4	111.5
C4D—C3D—C2D	1100	103.2	1.0	111.0
C3D—C2D—C1D	561	102.7	1.4	111.0
O4D—C1D—C2D	909	106.1	1.1	111.0

Table 2 (cont.)

DNA sugar	k(x)	x_eq	σ(x)	x_11
N1/9—C1D—C2D	430	114.2	1.6	111.0
O4D—C1D—N1/9	1719	107.8	0.8	111.0
C4D—C3D—O3R	622	110.3	2.2	111.0
C2D—C3D—O3R	413	110.6	2.7	111.0

Table 3. *The list of type-based dihedral angle parameters, their energy constants k(x), equilibrium values x_eq and standard deviations σ(x) used for the parameterization of the nucleic acid compared to the original periodical potentials*

The symbol R/D is used in the sugar atom types for parameters that are the same for RNA and DNA. The symbol N1/9 means either N9 of purine or N1 of pyrimidine.

Dihedral angle	k(x)	x_eq	σ(x)	x_11
Backbone				
O3R—P—C5R/D	6.1	285.3	9.8	0.0 (3)
	4.0	81.0	12.1	0.0 (2)
P—O5R—C5R/D—C4R/D	3.4	183.5	13.0	0.0 (3)
O5R—C5R/D—C4R/D—C3R/D	17.9	52.5	5.7	0.0 (3)
	14.2	179.4	6.4	0.0 (3)
	3.8	292.9	12.3	0.0 (3)
C4R/D—C3R/D—O3R—P	7.9	214.0	8.6	0.0 (3)
C3R/D—O3R—P—O5R	25.3	289.2	4.8	0.0 (3)
	3.9	80.7	14.3	0.0 (3)
C2'-endo sugar				
RNA				
C5R—C4R—C3R—O3R	24.3	147.3	4.9	0.0 (3)
O4R—C4R—C3R—O3R	20.7	268.1	5.3	0.0 (3)
O4R—C1R—C2R—C3R	50.4	35.2	3.4	0.0 (3)
C1R—C2R—C3R—C4R	74.4	324.6	2.8	0.0 (3)
C2R—C3R—C4R—O4R	29.7	24.2	4.4	0.0 (3)
C3R—C4R—O4R—C1R	17.9	357.7	5.7	0.0 (3)
C4R—O4R—C1R—C2R	21.6	339.2	5.2	0.0 (3)
C5R—C4R—C3R—C2R	34.7	263.4	4.1	0.0 (3)
O3R—C3R—C2R—O2R	33.0	319.7	4.2	0.0 (3)
DNA				
C5D—C4D—C3D—O3R	36.4	145.2	4.0	0.0 (3)
O4D—C4D—C3D—O3R	31.5	265.8	4.3	0.0 (3)
O4D—C1D—C2D—C3D	24.3	32.8	4.9	0.0 (3)
C1D—C2D—C3D—C4D	45.0	326.9	3.6	0.0 (3)
C2D—C3D—C4D—O4D	28.8	22.6	4.5	0.0 (3)
C3D—C4D—O4D—C1D	15.7	357.7	6.1	0.0 (3)
C4D—O4D—C1D—C2D	14.7	340.7	6.3	0.0 (3)
C5R—C4D—C3D—C2D	34.7	262.0	4.1	0.0 (3)
DNA/RNA				
C4R/D—O4R/D—C1R/D—N1/9	13.0	217.7	6.7	0.0 (3)
O4R/D—C1R/D—N1—C2	1.7	229.8	18.4	0.0 (2)
O4R/D—C1R/D—N9—C4	1.0	237.0	24.3	0.0 (2)
C3'-endo sugar				
RNA				
C5R—C4R—C3R—O3R	30.1	81.0	4.4	0.0 (3)
O4R—C4R—C3R—O3R	33.1	201.8	4.2	0.0 (3)
O4R—C1R—C2R—C3R	24.3	335.4	4.9	0.0 (3)
C1R—C2R—C3R—C4R	74.4	35.9	2.8	0.0 (3)
C2R—C3R—C4R—O4R	60.7	324.7	3.1	0.0 (3)
C3R—C4R—O4R—C1R	22.4	20.5	5.1	0.0 (3)
C4R—O4R—C1R—C2R	15.7	2.8	6.1	0.0 (3)
C5R—C4R—C3R—C2R	60.7	204.0	3.1	0.0 (3)
O3R—C3R—C2R—O2R	28.8	44.3	4.5	0.0 (3)
DNA/RNA				
C4R/D—O4R/D—C1R/D—N1/9	13.8	241.4	6.5	0.0 (3)
O4R/D—C1R/D—N1—C2	13.4	195.7	6.6	0.0 (2)
O4R/D—C1R/D—N9—C4	3.0	193.3	14.0	0.0 (2)

Table 4. *The comparison of the self consistency of the parameter file*

		Bond distances (Å)		Bond angles (°)		Dihedral angles (°)	
	Energy const. Equilibrium	k_11	k_11	k_11	k_11	k_11	k_11
Residue	const.	x_11	x_eq	x_11	x_eq	x_11	x_eq
Cyt	R.m.s.	0.004	<0.001	3.227	0.283	30.195	0.715
	Max. dev.	0.015	0.002	12.189	1.310	52.350	2.454
Gua	R.m.s.	0.004	<0.001	2.938	0.186	18.223	0.696
	Max. dev.	0.015	0.003	12.247	0.562	52.379	2.258
Ade	R.m.s.	0.004	<0.001	3.015	0.196	18.190	0.692
	Max. dev.	0.016	0.002	12.259	0.604	52.363	2.260
Thy	R.m.s.	0.004	<0.001	3.146	0.199	29.736	0.754
	Max. dev.	0.014	0.002	10.170	0.584	52.332	2.295
Ura	R.m.s.	0.004	<0.001	3.236	0.235	29.735	0.751
	Max. dev.	0.016	0.004	12.255	0.578	52.419	2.306
	1	2	3	4	5	6	7

Table 5. *A comparison of refinements using equilibrium constants from param11.dna and the new equilibrium constants x_eq*

Only energy constants k_11 from param11.dna were used in the refinement.

	B-DNA Dodecamer		Z-DNA hexamer		CAP-DNA14/17 complex	
Energy const.	k_11	k_11	k_11	k_11	k_11	k_11
Equilibrium const.	x_11	x_eq	x_11	x_eq	x_11	x_eq
R factor	16.6	16.7	18.0	18.3	20.9	20.9
Final r.m.s.						
Bonds	0.015	0.014	0.013	0.011	0.019	0.015
Angles	3.45	2.41	2.86	1.94	3.95	3.28
Dihedrals	27.40	22.88	31.14	12.07	31.16	24.83
Final energy						
Bonds	56.6	54.1	10.8	8.7	596	441
Angles	199.6	100.6	35.7	16.7	1841	1360
Dihedrals	264.2	71.9	69.9	4.7	1593	1014

to reduce the overall r.m.s. deviations leading to poor protein geometry. Overall, the results revealed a dramatic improvement in refinement r.m.s. statistics for nucleic acid-containing structures. The improvement over the previous DNA dictionary will probably have a more significant effect than the improvement observed for the implementation of the protein amino-acids dictionary (Engh & Huber, 1991).

A selection criteria based on a separation into C2'- and C3'-endo sugar pucker for ribose and deoxyribose sugars was also examined. It was expected that statistically significant differences would exist between the two sample sets for both bond distances and for bond angles. Several exocyclic bond angles were found to be statistically different. After extensive scaling, it was determined that this separation based on C2'- and C3'-endo sugar pucker conformation for the derivation of equilibrium constants was unnecessary, although structurally correct. No significant differences were observed in the final refined protein–DNA complex at 2.5 Å resolution, after using either of the parameter sets.

Table 6. *Comparisons of full refinement*

Structure	B-DNA (10–2.25 Å)		Z-DNA (10–1.35 Å)		CAP–DNA14/17 (10–2.5 Å)			
Energy constants	k_11	k_eq	k_11	k_eq	k_11		k_eq^*	
Equilibrium constants	x_11	x_eq	x_11	x_eq	x_11		x_eq^*	
R factor (%)†	16.6	16.3	18.0	18.0	20.9		21.0	
Geometry					C§	D¶	C§	D¶
Bonds (Å)	0.015	0.009	0.013	0.009	0.019	0.023	0.016	0.015
Angles (°)	3.450	1.410	2.860	1.270	3.750	3.950	2.180	2.110
Dihedrals (°)	27.40	19.63‡	31.14	8.330‡	30.00	33.50	23.30	22.80

* Parameter file parhcsdx.pro (Engh & Huber, 1991) used for protein parameterization. † Structures were refined to reduce the r.m.s. deviations while maintaining a consistent R factor. ‡ C2' and C3'-*endo* sugar puckering included in parameters. § Combined protein–DNA statistics. ¶ DNA statistics alone.

3.2. *Specific recommendations for refinement*

The topology file is arranged such that the default assignments for the sugar-ring pucker are C3'-*endo* for RNA and C2'-*endo* for DNA. It will however be necessary to individually check the sugar-ring pucker during refinement. This can be achieved by checking r.m.s. deviations for the particular dihedral angles. From our experience, the dihedral angles for sugar pucker tend towards the correct target values even when inappropriate values are applied during refinement. Alternative sugar dihedral angles can be applied using a restraints dihedral assignment. Example files arestraint.inp, and brestraint.inp, will be distributed with *X-PLOR* (Brünger, 1992; Brünger, unpublished work) and are available upon request from ATB. The values that can be put into those example files are supplied in the parameter file and annotated. In the case of backbone dihedrals, the periodical potentials for α, γ and ζ are automatically applied to ensure the possibility of three minima. In the latter stage of refinement the user may wish to apply the more precise single target equilibrium constants and energy constants. These additional values are provided in the parameter file.

For high-resolution structures, it is possible to use parameters derived for the bond distances and bond angles in C2'-*endo* and C3'-*endo* sugar conformations. Parameters suitable for the refinement of such high-resolution structures are available at URL http://ndbserver.rutgers.edu. These parameters were successfully used for the refinement of a Z-DNA structure with 1.35 Å data.

The weighting of the energy constants during refinement is related to the resolution, the quality of the data and the refinement strategy. Dihedral energy terms are particularly sensitive, especially in the final steps of the refinement, which emphasizes the need for limiting the dihedral angle constraints in the case of sufficient crystallographic data. A 20% weighting of the dihedral angle energy constant was found to be appropriate for the B-DNA at 2.3 Å resolution and 50% weighting for the CAP–DNA14/17 complex, yielding a balanced contribution between the dihedrals and other energies. Caution is suggested when refining unusual structures and non-standard regions, *i.e.* bulges, loops, *etc.* Additionally, the refinement of protein–DNA complexes requires the balancing of overall energy contributions between the protein and the nucleic acid. Weighting of the specific terms included in the potential energy function can be easily adjusted using the 'constraint interaction' term. The r.m.s. deviations can be used to assist in the assignment, as they should correspond to the sample standard deviations. The application of additional restraints such as base planarity and hydrogen bonding for the refinement of DNA duplexes may be necessary. This is particularly the case for low-resolution structures, during the initial refinement cycles, or for poor models.

References

Allen, F. H., Bellard, S., Brice, M. D., Cartright, B. A., Doubleday, A., Higgs, H., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R. & Watson, D. G. (1979). *Acta Cryst.* B35, 2331–2339.

Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S.-H., Srinivasan, A. R. & Schneider, B. (1992). *Biophys. J.* 63, 751–759.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* 4, 187–217.

Brünger, A. T. (1992). *X-PLOR, A System for X-ray Crystallography and NMR.* New Haven: Yale University Press.

Brünger, A. T., Karplus, M. & Petsko, G. A. (1989). *Acta Cryst.* A45, 60–61.

Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science,* 235, 458–460.

Clowney, L., Westbrook, J., Jain, S. C., Srinivasan, N., Srinivasan, A. K., Olson, W. K. & Berman, H. M. (1996). Submitted.

Engh, R. A. & Huber, R. (1991). *Acta Cryst.* A47, 392–400.

Gelbin, A., Westbrook, J., Jain, S. C., Srinivasan, N., Srinivasan, A. K., Olson, W. K. & Berman, H. M. (1996). Submitted.

Hamilton, W. (1964). *Statistics in Physical Science.* New York: Ronald Press.

Parkinson, G. N., Arvanitis, G., Lessinger, L., Ginell, S., Jones, R., Gaffney, B. & Berman, H. M. (1996). *Biochemistry.* In the press.

Parkinson, G. N., Gunasekera, A. H., Wilson, C., Ebright, R. & Berman, H. M. (1996). *Biochemistry.* In the press.

Vojtechovsky, J., Eaton, M. D., Gaffney, B., Jones, R. & Berman, H. M. (1996). In preparation.