Top Curr Chem (2005) 253: 77–88
DOI 10.1007/b100443

Between Objectivity and Whim: Nucleic Acid Structural Biology

Loren Dean Williams (✉)

School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332–0400, USA
loren.williams@chemistry.gatech.edu

1 Introduction 78

2 Estimates of Errors 79

3 Isothermal Titration Calorimetry 79

4 Small Molecule Crystallography 79

5 Macromolecular Crystallography 79

6 Mixed and Partial Occupancies 81

7 Scattering Iso-Types 81

8 Compensating Parameters 83

9 Coordination Fingerprints 83

10 Revaluation of Published Structures 84

11 Mis-assigned Peaks 84
11.1 Z-DNA 84
11.2 B-DNA #1 85
11.3 B-DNA #2 85
11.4 DNA–Drug Complexes 86

12 Summary 87

References 87

Abstract Here we explore the subjectivity intrinsic to macromolecular crystallography, focusing on the hydration/counter-ion region of nucleic acids. Water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites. Many of these species give identical electron distributions (electron density maps). Such scattering iso-types allow one to construct different models that give similar fits of model to data. Even models with different electron densities can give similar fits of model to data because various parameters compensate. The geometries of the coordinating ligands of many species (magnesium excluded) are similar to each other, and are effectively identical within the

limitations of macromolecular crystallography. The observed distances are commonly occupancy-weighted averages. In sum, atom-type assignments and occupancies in the hydration/counter region cannot be unambiguously extracted from conventional X-ray diffraction experiments. We give several examples where incorrect atom-type assignments have been revealed by anomalous scattering experiments. Published structures and the database that contains them do not provide realistic representations of subjectivity. Reconsideration of macromolecular model-building protocols may be in order. Anomalous scattering provides information that can allow one to characterize the hydration/counter-ion region with greater accuracy than was previously possible.

Keywords Hydration · Water · Sodium · Potassium · Ammonium · Magnesium · Monovalent · Divalent · Cation · DNA · DNA–drug complexes · RNA · Uncertainty · Error

1 Introduction

The subjectivity inherent in macromolecular crystallography is not always apparent to end-users, or even practitioners of crystallography. Here we wish to focus on the dark side, and to discuss and illustrate errors and uncertainty. The focus is the hydration/counter-ion milieu of nucleic acids – including DNA and RNA oligonucleotides, DNA–drug complexes and DNA–protein complexes, and ribosomes. For nucleic acids and their complexes the hydration/counter-ion region is especially non-homogeneous because the high anionic charge induces association of many types of cations. Water molecules, monovalent and divalent cations, and polyamines compete for binding. Therefore many of the issues discussed here are relevant to nucleic acids, but not necessarily to proteins. The goal is to stimulate new approaches, and to promote cautious and reasoned interpretation of structural information. We believe reconsideration of model-building, reporting, and archiving protocols is in order.

The method scrutinized here is the assignment of water molecules and ions to isolated sum and difference electron density peaks ($2|F_o|-|F_c|$ and $|F_o|-|F_c|$ Fourier peaks) during the latter stages of refinement. The protocol is 'when in doubt, make it a water', with occupancy 1.00. If the thermal factor and geometry remain reasonable after refinement, then the water atom-type assignment and occupancy of 1.00 are considered to be validated. As explained below, advanced experimental approaches (anomalous scattering) demonstrate that these methods are not reliable. In the absence of anomalous scattering data, alternate atom-type assignments are possible, giving many models with reasonable thermal factors and geometry that fit the data equally well. Selecting one model, with a given composition of the hydration/counter-ion region, over another model with a different composition is essentially a subjective enterprise, and is not grounded in the experiment. In the concluding section we give some examples from our own work illustrating how the conventional approach can go very wrong.

2
Estimates of Errors

The macromolecular crystallographic method provides less information on errors, and less information to assess the fit of model to data, than is the norm for physical-chemical or biophysical fitting processes. To help the reader appreciate the distinctiveness of macromolecular crystallography, here we compare it with isothermal titration calorimetry (ITC) and small molecule crystallography.

3
Isothermal Titration Calorimetry

In this method one collects data (heats of injection) and conceives a model with parameters such as an equilibrium constant, a stoichiometric coefficient, and an enthalpy of binding. The parameters and the model are used to obtain calculated heats of injection. Differences between observed heats and the calculated heats are minimized by adjusting the parameters. During and after the fitting process one obtains a measure of global fit (χ^2), along with estimates of error of each parameter. The relationship between global fit and estimates of parameter error is not direct. The measure of global fit might be very good, even while estimate of error in one or more of the parameters is large, for example, if data on one side of the binding curve were absent.

4
Small Molecule Crystallography

In this method one collects data (observed structure factor amplitudes, $|F_o(hkl)|$) and establishes a preliminary model. The parameters of the model are the x, y, z coordinates of various atom-types, their thermal factors (which are generally anisotropic) and their occupancies. The parameters are used to obtain calculated structure factor amplitudes ($|F_c(hkl)|$). Differences between $|F_o(hkl)|$ and $|F_c(hkl)|$ are minimized by adjusting the parameters, and often by adding or subtracting atoms from the model. As with ITC, correctness of the model is indicated by global measures of fit (the R-factor) in addition to estimates of error in individual parameters.

5
Macromolecular Crystallography

As in the small molecule method, in the macromolecular method differences between $|F_o(hkl)|$ and $|F_c(hkl)|$ are minimized by adjusting the parameters of the model. Both methods yield global measures of fit (R-factor, R-free). How-

ever, realistic estimates of uncertainties of individual parameters (the x, y, z coordinates of each atom, the thermal factors, and the occupancies) are not obtained from the macromolecular method. The origins of this deficiency are beyond the scope of this discussion, but are related to very large numbers of parameters and very large numbers of data. Macromolecular structures (entries in the Nucleic Acid Database, NDB [1]) do not specify uncertainties of individual x, y, z coordinates, their thermal factors, or occupancies because those estimates of error are not available.

The protocol, after a reasonable model of the macromolecule along tight binding ligands is established, is to model the hydration/counter-ion region by adding water molecules and/or ions to corresponding sum and difference peaks of electron density (examples of sum electron peaks are shown in Fig. 1). However, the electron density in the hydration/counter-ion region of nucleic acid crystals is intrinsically uninformative and ambiguous. Ambiguity arises (i) from mixed and partial occupancies, (ii) from scattering iso-types, and (iii) from parameter-compensation. Each of these effects is explained and discussed below.

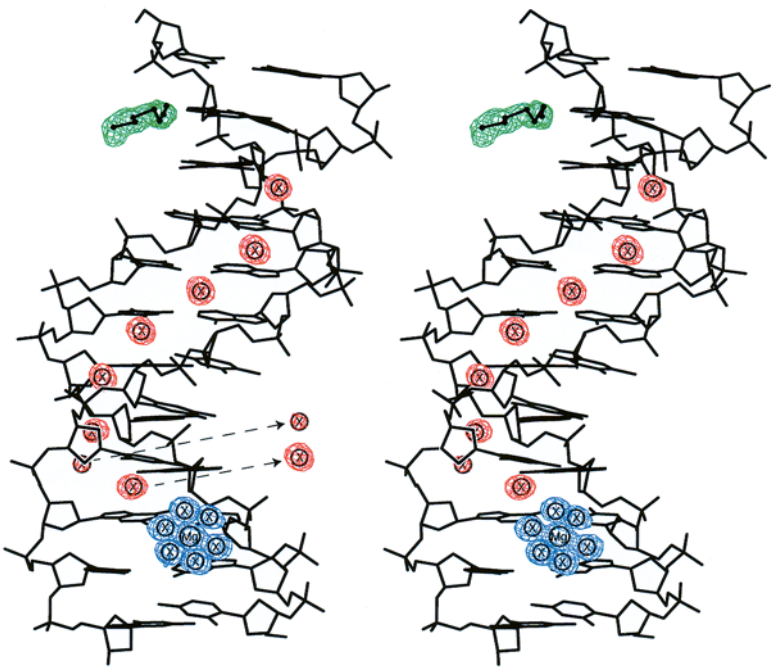

Fig. 1 Stereoview of sum ($2|F_o|-|F_c|$) electron density from a DNA duplex CGCGAATTCGCG at 1.4 Å resolution contoured at 1.0 σ (NDB entry bdl084). Only the electron density surrounding the hydrated magnesium ion (*blue net*), the partial spermine molecule (*green net*), and the minor groove solvent sites (*red net*) that interact with DNA bases is shown. For this structure, solvent sites were fitted to water molecules. The sum density from two minor groove solvent sites has been abstracted on the *left side*, to illustrate the lack of correspondence between the radius of a peak of electron density and the radius of the atom type

6
Mixed and Partial Occupancies

Crystallographic electron density and models are not restricted to formally correct chemical entities. The observed electron density is an ensemble average. Partially-occupied atoms and hybrid atom-types are 'observed'. This phenomenon is particularly acute in the hydration/counter-ion region of nucleic acids. Many species including water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites. It is possible for an electron density peak to arise from 40% one species, 30% another species, and 30% nothing.

7
Scattering Iso-Types

A large number of chemically distinct models give identical electron distributions (electron densities). This degeneracy arises because many of the species in the hydration/counter-ion region are *scattering iso-types* (our term). Scattering iso-types are defined as chemically distinct species (molecules or ions) with indistinguishable electron distributions (and X-ray scattering properties). An ammonium ion and a water molecule are scattering iso-types. The number of electrons in each is the same (10 electrons). The shape of the electron density is the same (spherical). A list of scattering iso-types is given in Table 1. Since partial and mixed occupancies are possible, the list of scattering iso-types is not restricted to formally correct chemical entities. The effective number of electrons of a potassium ion, when 55% occupied, is the same as that of a water molecule. In Table 1, occupancies that give equivalent numbers of elec-

Table 1 Scattering Iso-types: full and partial occupancy

Species	Occupancy	Radius (Å)	Number of electrons
H_2O	1.0	1.40	10
Na^+	1.0	0.95	10
Na_4^+	1.0	1.40	10
Mg^{2+}	1.0	0.65	10
$K_{0.6}^{+}$ [a]	0.6	1.33	10
$Rb_{0.4}^{+}$ [b]	0.28	1.48	10
$Tl_{0.11}^{+}$ [c]	0.11	1.49	10
$Ca_{0.4}^{+}$ [d]	0.4	1.69	10

[a] 60% occupied potassium ion.
[b] 28% occupied rubidium ion.
[c] 11% occupied thallium ion.
[d] 40% occupied calcium ion.

Table 2 Scattering iso-types: mixed occupancy H_2O/Na^+

Occupancy H_2O	Occupancy Na^+	Effective radius (Å)	Effective number of electrons
1.0	0.0	1.40	10
0.9	0.1	1.36	10
0.8	0.2	1.31	10
0.7	0.3	1.27	10
0.6	0.4	1.22	10
0.5	0.5	1.18	10
0.4	0.6	1.13	10
0.3	0.7	1.09	10
0.2	0.8	1.04	10
0.1	0.9	1.00	10
0.0	1.0	0.95	10

Table 3 Scattering iso-types: mixed occupancy H_2O/K^+

Occupancy H_2O	Occupancy K^+	Effective radius (Å)	Effective number of electrons
1.0	0.00	1.40	10
0.9	0.05	1.40	10
0.8	0.10	1.39	10
0.7	0.15	1.39	10
0.6	0.20	1.38	10
0.5	0.30	1.37	10
0.4	0.35	1.37	10
0.3	0.40	1.36	10
0.2	0.45	1.35	10
0.1	0.50	1.34	10
0.0	0.55	1.33	10

trons are indicated for common species. The list of possible scattering iso-types can be infinitely expanded by continuously varying the partial occupancies as illustrated for H_2O/Na^+ hybrids in Table 2 and H_2O/K^+ hybrids in Table 3.

One might naively assume that differences in shape, arising from differences in ionic/atomic radii could differentiate various species. The molecular radius of H_2O is greater than the ionic radius of Mg^{2+}, suggesting that the electron density from a Mg^{2+} ion should be of greater amplitude at the peak center and of less breadth than that of a water molecule. However, the effective dispersion of the electrons about the atom center, which is described in the structure factor equation as the rms amplitude of the atomic displacement, varies with

thermal fluctuations, positional disorder, and data quality. The contributions of those effects generally obscure differences in ionic/atomic radii. This effect is illustrated in Fig. 1.

8
Compensating Parameters

The problem extends beyond scattering iso-types and partial occupancies. Models with different electron densities can fit the same data equally well. Changes in one parameter are compensated in the fit by changes in other parameters. For example, decreasing the number of electrons of a K^+ ion (achieved by decreasing the occupancy or by changing the atom type from K^+ to Na^+ or H_2O) is compensated wholly or partially by decreasing the thermal factor. This compensation is related to determinacy (ratio of data to parameters). Therefore, hydration/counter-ion occupancies are generally not refined. Instead occupancies are fixed at a default value of 1.00 (the number of significant figures is absurd). This default value is arbitrary. In many cases switching an occupancy from 1.00 to 0.80 would not affect map quality or refinement statistics. Adjustments in thermal factors would eat up the difference.

9
Coordination Fingerprints

As explained above, the hydration/counter-ion region of nucleic acids is especially difficult to characterize unambiguously. Yet there is some hope provided by coordination fingerprints. The octahedral geometry and the ligand-to-metal distances of a well-ordered Mg^{2+} do allow reliable identification. It is conceivable that Na^+ could be distinguished from H_2O by a coordination fingerprint. The ideal distance from Na^+ to ligand is less than that from H_2O to ligand (Brown [2] has provided useful surveys of coordination geometries). However, in the macromolecular experiment, the experimentalist must interpret subtle differences in geometries because the observed distances are generally occupancy-weighted averages (Tables 2 and 3). How does one interpret nominally short contacts between water molecules? Do they arise from partial Na^+ occupancy or simply from coordinate error? The database contains very few examples of Na^+ ions with unambiguous coordination geometry, many with unreasonable geometry, and none with reasonable estimates of atom-type error. Many if not most of the Na^+ ions in the NDB have coordination geometries consistent with those of water molecules.

10
Revaluation of Published Structures

Considering the limitations imposed by mixed and partial occupancies, scattering iso-types, compensating parameters, and occupancy-weighted geometries it is simply impossible to unambiguously extract the correct identities of the contributing species from the data (in the absence of anomalous scattering data). The hydration/counter-ion region cannot be reliably characterized by standard macromolecular x-ray methods. We believe there is a realistic possibility that many or even most 'water molecules' and ions are mis-assigned and/or incorrectly described in the NDB. In the following sections we provide some examples where the mis-assignment of atom-types has been experimentally revealed.

11
Mis-assigned Peaks

11.1
Z-DNA

A very high resolution structure (1.0 Å resolution) of the spermine form of Z-DNA [3] was determined by Martin Egli, Loren Williams, and Qi Gao, as postdoctoral researchers, in the laboratory of Alexander Rich. Water molecules were assigned to sum and difference electron density peaks, including a series of regularly spaced peaks in the minor groove, which formed a 'spine of hydration'. The water molecules were well-behaved, with reasonable geometry and interactions, realistic thermal factors, good electron density, etc. The final published structure (NDB entry) contains water molecules within the 'spine of hydration', which are assigned atom types of O (hydrogen atoms are implicit).

After that structure had been completed and published, evolution of cryocrystallographic techniques allowed the same group to collect the first ever data set from a flash-frozen crystal of DNA [4], consisting of the same spermine-form Z-DNA. Lowering the temperature dampens molecular motions, increasing the information content of the diffraction data. The data obtained from a flash-frozen crystal indicated that the minor groove contained a spermine molecule. Lowering the temperature converted the isolated peaks in the minor groove that had been assigned to a spine of hydration into a tube of electron density. The spine of hydration morphed into a spermine molecule. At room temperature the methylene groups of the spermine molecule were thermally disordered, and so not visible in the electron density maps. The amino groups of the spermine formed hydrogen bonds to the floor of the minor groove and were ordered at room temperature, and appeared as spheres of electron density. The peaks of electron density in the minor groove had been mis-assigned during the original refinement of the room temperature structure. They were not water

molecules but were the amino groups of a spermine molecule. Our retrospective assessment is that our original conclusion, that peaks of electron density indicated a spine of hydration, was an over-interpretation of the data. However, even the revised structure, with occupancies of 1.00 for the spermine molecule, etc., is also an over-interpretation.

11.2 B-DNA #1

By 1998 there were 68 isomorphous members of the CGCXAATTYGCG (X=G or A, Y=C or T) dodecamer family in the NDB. Collectively those structures contain thousands of water molecules, and no monovalent cations, divalent cations, or polyamines.[1] We proposed an alternative interpretation of the data [5–7] in which the grooves of the B-DNA in those crystals are decorated with various types of cations. That proposal is now accepted, based in part on substitution experiments with scattering cations that give distinctive (anomalous) scattering information [8–10] (also see work by Egli and coworkers [11]). It is now clear that much of the hydration/counter-ion region contains partial and mixed-occupancy water molecules and cations. The results of anomalous experiments indicate that the A-tract minor groove contains monovalent cations. Increasing data quality has revealed that the major groove is associated with hydrated magnesium ions and spermine molecules (Fig. 1). Our retrospective assessment is that the original models, which for example assumed that peaks of electron density indicated a well-ordered spine of hydration [12, 13], were over-interpretations of the data. The predominant monovalent cation in the crystallization solutions was Na^+ (a scattering iso-type with water). Models with Na^+/H_2O partial occupancy hybrids would therefore fit the data equally as well as models with water only. There was never any experimental basis for setting the atom-types as pure water or their occupancies as 1.00.

11.3 B-DNA #2

In 1998 we identified a Mg^{2+} ion in the major groove of CGCGAATTCGCG [7]. The ion was well-behaved in the refinement, with excellent electron density, thermal factors, geometry, etc. The location of the Mg^{2+} was confirmed by other investigators [14, 15]. It is a 'consensus' cation, observed in structures obtained from a variety of crystallization conditions and DNA modification. The major groove Mg^{2+} has been assigned a structural role by several investigators, and is thought to contribute to the famous 'dodecamer bend'.

[1] A spermine molecule identified in an early dodecamer structure has been revised, and is now considered to be a hydrated magnesium ion.

In 2001 we discovered with an anomalous scattering experiment that the major groove Mg^{2+} is not fully occupied [9]. A monovalent signal is subtle in the sum and difference maps but is unambiguous in the anomalous map. The monovalent site is displaced somewhat from the divalent site, although the proximity is such that occupancy of either site by a cation would preclude occupancy of the other (the sum of the occupancies cannot exceed 1.00). Since the Mg^{2+} is only partially occupied, the role of the Mg^{2+} in contributing to the dodecamer bend is unclear.

Our retrospective assessment is that our original major groove Mg^{2+} ion model was an over-interpretation of the data. The good thermal factor/good geometry criterion was not sufficient to ascribe an occupancy of 1.00 to the Mg^{2+}.

11.4 DNA–Drug Complexes

In X-ray structures of several early DNA–anthracycline complexes, well-defined and fully occupied Na^+ ions mediate interactions between the intercalated chromophore and the DNA [16, 17]. In considering the reliability of those Na^+ assignments the following factors require consideration:

1. The Na^+ atom-type assignments are based on geometric considerations. It was assumed that six ligands surrounding an electron density peak is definitively an indicator of Na^+ ion occupancy.
2. After initial refinement, the geometry of the ligands surrounding the Na^+ peak was restrained, to 'ideal' octahedral Na^+ geometry.
3. The definition of ideal Na^+ geometry has evolved over time. Six-coordinate octahedral geometry is no longer considered a reliable indicator of sodium. As noted by Jeffery [18], water molecules engaging in bifurcated hydrogen bonding can be six-coordinate.
4. Structures of some DNA–anthracycline complexes lack localized cations [19, 20]. The Na^+ ions in the original structure have not proved to be fully reproducible.
5. Recent anomalous experiments show that there are additional monovalent cation sites in the DNA–anthracycline complexes [21]. In the anomalous experiments the original Na^+ sites are not as highly occupied as other sites.

Our retrospective assessment is that localized cations and electrostatic forces are indeed important in structure, thermodynamics, and sequence specificity of DNA–ligand complexes. Favorable interactions of adriamycin and cations with the sequence-specific electrostatic landscape of DNA may be universal characteristics of DNA–small molecule interactions and may be useful in sequence-specific ligand design. However, the certitude and the details of the original Na^+ descriptions are in error.

12 Summary

Water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites in the hydration/counter-ion regions of nucleic acids. In general, crystallographic models of hydration/counter-ion regions are biased simplifications. In assessing the reliability of crystallographic models one must consider many factors:

1. With scattering iso-types one can construct many different models with indistinguishable electron density maps.
2. Even models with different electron density maps can give similar fits of model to data.
3. The geometries of the coordinating ligands of common species (Mg^{2+} excluded) in the hydration/counter-ion region of DNA are very similar to each other, and are effectively identical within the limitations of macromolecular crystallography.
4. The good thermal factor/good geometry criterion is not sufficient to ascribe an occupancy of 1.00.
5. Published structures and the database that contains them do not provide realistic representations of uncertainty.
6. The atom type assignments and occupancies in the hydration/counter region are generally subjective.
7. In many cases electron density peaks should be assigned to wild-card atoms (10 electrons, identity unknown).
8. Anomalous scattering experiments provide important information that allows one to characterize the hydration/counter-ion region with greater accuracy than was previously possible.

References

1. Berman HM, Olson WK, Beveridge DL, Westbrook J, Gelbin A, Demeny T, Hsieh S-H, Srinivasan AR, Schneider B (1992) The nucleic acid database. a comprehensive relational database of three-dimensional structures of nucleic acids. Biophys J 63:751–759
2. Brown ID (1988) What factors determine cation coordination numbers. Acta Crystallogr B44:545–553
3. Egli M, Williams LD, Gao Q, Rich A (1991) Structure of the pure-spermine form of Z-DNA (magnesium free) at 1 Å resolution. Biochemistry 30:11388–11402
4. Bancroft D, Williams LD, Rich A, Egli M (1994) The low temperature crystal structure of the pure-spermine forms of Z-DNA reveals binding of a spermine molecule in the minor groove. Biochemistry 33:1073–1086
5. Shui X, Sines C, McFail-Isom L, VanDerveer D, Williams LD (1998) Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. Biochemistry 37:16877–16887
6. McFail-Isom L, Sines C, Williams LD (1999) DNA structure: cations in charge? Curr Opin Struct Biol 9:298–304

7. Shui X, McFail-Isom L, Hu GG, Williams LD (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. Biochemistry 37:8341–8355
8. Woods K, McFail-Isom L, Sines CC, Howerton SB, Stephens RK, Williams LD (2000) Monovalent cations sequester within the a-tract minor groove of [D(CGCGAATTCGCG)]2. J Am Chem Soc 122:1546–1547
9. Howerton SB, Sines CC, VanDerveer D, Williams LD (2001) Locating monovalent cations in the grooves of B-DNA. Biochemistry 40:10023–10031
10. Tereshko V, Minasov G, Egli M (1999) A "hydrat-ion" spine in a B-DNA minor groove. J Am Chem Soc 121:3590–3595
11. Tereshko V, Wilds CJ, Minasov G, Prakash TP, Maier MA, Howard A, Wawrzak Z, Manoharan M, Egli M (2001) Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. Nucleic Acids Res 29:1208–1215
12. Drew HR, Dickerson RE (1981) Structure of a B-DNA dodecamer. Iii. Geometry of hydration. J Mol Biol 151:535–556
13. Kopka ML, Fratini AV, Drew HR, Dickerson RE (1983) Ordered water structure around a B-DNA dodecamer. A quantitative study. J Mol Biol 163:129–146
14. Chiu TK, Kaczor-Grzeskowiak M, Dickerson RE (1999) Absence of minor groove monovalent cations in the crosslinked dodecamer CGCGAATTCGCG. J Mol Biol 292:589–608
15. Minasov G, Tereshko V, Egli M (1999) Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. J Mol Biol 291:83–99
16. Wang AH, Ughetto G, Quigley GJ and Rich A (1987) Interactions between an anthracycline antibiotic and DNA: molecular structure of daunomycin complexed to D(Cpgpt-papcpg) at 1.2 Å resolution. Biochemistry 26:1152–1163
17. Frederick CA, Williams LD, Ughetto G, van der Marel GA, van Boom JH, Rich A, Wang AH-J (1990) Structural comparison of anti-cancer drug-DNA complexes: adriamycin and daunomycin. Biochemistry 29:2538–2549.
18. Jeffrey GA (1997) An introduction to hydrogen bonding. Oxford University Press, New York
19. Williams LD, Egli M, Ughetto G, van der Marel GA, van Boom JH, Quigley GJ, Wang AH-J, Rich A, Frederick CA (1990) Structure of 11-deoxydaunomycin bound to DNA containing a phosphorothioate. J Mol Biol 215:313–320
20. Lipscomb LA, Peek ME, Zhou FX, Bertrand JA, VanDerveer D, Williams LD (1994) Water ring structure at DNA interfaces: hydration and dynamics of DNA–anthracycline complexes. Biochemistry 33:3649–3659
21. Howerton SB, Nagpal A, Williams LD (2003) Surprising roles for electrostatic interactions in DNA–ligand complexes. Biopolymers 69:87–99