Journal of Biomolecular NMR (2006) 35: 103–115
DOI 10.1007/s10858-006-9004-2

Article

Prediction of molecular alignment of nucleic acids in aligned media

Bin Wu[a], Michael Petersen[b], Frederic Girard[a], Marco Tessari[a] & Sybren S. Wijmenga[a,*]
[a]*Laboratory of Physical Chemistry-Biophysical Chemistry, Institute of Molecules and Materials, Radboud University Nijmegen, Toernooiveld 1, 6225ED, Nijmegen, The Netherlands;* [b]*Nucleic Acid Center, Department of Chemistry, University of Southern Denmark, 5230, Odense M, Denmark*

Received 24 October 2005; Accepted 14 March 2006

Key words: alignment tensor prediction, electrostatic alignment, NMR, nucleic acids, RNA, residual dipolar couplings

Abstract

We demonstrate – using the data base of all deposited DNA and RNA structures aligned in Pf1-medium and RDC refined – that for nucleic acids in a Pf1-medium the electrostatic alignment tensor can be predicted reliably and accurately via a simple and fast calculation based on the gyration tensor spanned out by the phosphodiester atoms. The rhombicity is well predicted over its full range from 0 to 0.66, while the alignment tensor orientation is predicted correctly for rhombicities up to ca. 0.4, for larger rhombicities it appears to deviate somewhat more than expected based on structural noise and measurement error. This simple analytical approach is based on the Debye–Huckel approximation for the electrostatic interaction potential, valid at distances sufficiently far away from a poly-ionic charged surface, a condition naturally enforced when the charge of alignment medium and solute are of equal sign, as for nucleic acids in a Pf1-phage medium. For the usual salt strengths and nucleic acid sizes, the Debye–Huckel screening length is smaller than the nucleic acid size, but large enough for the collective of Debye–Huckel spheres to encompass the whole molecule. The molecular alignment is then purely electrostatic, but it's functional form is under these conditions similar to that for steric alignment. The proposed analytical expression allows for very fast calculation of the alignment tensor and hence RDCs from the conformation of the nucleic acid molecule. This information provides opportunities for improved structure determination of nucleic acids, including better assessment of dynamics in (multi-domain) nucleic acids and the possibility to incorporate alignment tensor prediction from shape directly into the structure calculation process. The procedures are incorporated into MATLAB scripts, which are available on request.

Introduction

Residual dipolar couplings (RDCs) have come to play an increasingly important role in the study of structure and dynamics of proteins, nucleic acids, and oligosaccharides by NMR (Mollova & Pardi, 2000; Prestegard et al., 2000; Tolman, 2001; Bertini et al., 2002; de Alba & Tjandra, 2002; Bax, 2003; Lipsitz & Tjandra, 2004; Al-Hashimi, 2005; Blackledge, 2005; Latham et al., 2005) – after their initial applications (Tolman et al., 1995; Tjandra & Bax, 1997; Tjandra et al., 1997) – RDCs stem from partial alignment (10^{-3} to 10^{-4}) of the biomolecules. They can be induced by employing an increasing number of different liquid crystalline aligning media (Tjandra & Bax, 1997; Mollova & Pardi, 2000; Blackledge, 2005) with Pf1-phage medium being most popular for nucleic acids (Hansen et al., 1998; Mollova & Pardi, 2000).

*To whom correspondence should be addressed.
E-mail: S.Wijmenga@nmr.ru.nl

Alternatively, for instance for large nucleic acids, the cumulative effect of the large magnetic susceptibilities of the nucleotide bases, can induce sufficient order directly in the magnetic field as shown first by Kung et al. (Kung et al., 1995) and subsequently used to determine structures of e.g. a DNA quadruplex (Al-Hashimi et al., 2001), and a Holliday junction (van Buuren et al., 2004), to derive the stoichiometry in homo-multimeric nucleic-acid complexes (Al-Hashimi et al., 2001), and to study the dynamics of RNA domains (Zhang et al., 2003; Al-Hashimi 2005; Latham et al., 2005). To take full advantage of RDCs knowledge of the molecular alignment tensor (**A**) is important, that is knowledge of its magnitude (A_{ax}), rhombicity (R), and orientation ψ – e.g. defined by the three Euler angles, α, β and γ. The common approach is to extract the five parameters that define **A**, from a combination of RDCs and structure knowledge either via simulated annealing (Clore et al., 1998; Sass et al., 1999; Meiler et al., 2000) or singular-value decomposition (SVD) (Losonczi et al., 1999; Sass et al., 1999). However, to function reliably, this requires a reasonably correct structure and a sufficient number of RDCs ($\gg$5) (Zweckstetter & Bax, 2002). In the absence of structure knowledge, A_{ax} and R can be extracted from the RDCs alone via the (extended) histogram method (Clore et al., 1998; Bryce & Bax, 2004), but this requires that the RDCs sufficiently sample all orientations. Predicted alignment tensors for a given structure have been used to differentiate between monomeric and homodimeric states (Zweckstetter & Bax, 2000), analysis of dynamics in oligosaccharides (Azurmendi & Bush, 2002), refinement of nucleic acid structures (Warren & Moore, 2001), determination of global structure of branched nucleic acids (van Buuren et al., 2004), assessment of domain dynamics in RNA (Zhang et al., 2003), and determination of the relative orientation of protein domains (Bewley & Clore, 2000). The ability to predict dipolar couplings for a given protein structure also provides unique opportunities to classify protein fold families on the basis of unassigned NMR data, potentially increasing data throughput in structural genomics (Valafar & Prestegard, 2003). Moreover, accurate prediction of the alignment tensor from the biomolecular structure paves the way for new global structure optimization of (multi-domain) biomolecules, in which the alignment tensor is continuously calculated for each test structure during the structure derivation process. Such an approach has the advantage, that the number of free parameters is reduced by five, thereby improving the reliability of the structure derivation. In such a protocol the alignment tensor (and RDCs) have to be calculated many times during the structure calculation process, which requires the development of fast (and accurate) prediction methods (Fernandes et al., 2001; Almond & Axelsen, 2002; Azurmendi & Bush, 2002). The Holliday junction structure has been derived from magnetic-field induced RDCs (mRDCs) via such a protocol, because prediction of magnetic field induced alignment is already reliable and fast (van Buuren et al. 2004).

Prediction of the molecular alignment tensor from structure has been demonstrated for steric alignment of proteins, nucleic acids, and carbohydrates – either via simulation or via fast and simple analytical expressions (Zweckstetter & Bax, 2000; Fernandes et al., 2001; Almond & Axelsen, 2002; Azurmendi & Bush, 2002). The advantages of the latter fast prediction over computationally time consuming simulation (Almond & Axelsen, 2002) is the potential of incorporating prediction directly into restrained dynamics calculations to derive structure and to assess (domain) dynamics. Most recently, prediction of mixed steric and electrostatic alignment has been investigated via simulation and experiment for a number of proteins (Ferrarini, 2003; Zweckstetter et al., 2004) and the DNA Dickerson dodecamer (Zweckstetter et al., 2004). For electrostatic alignment, no fast analytical alignment prediction approaches have been proposed and tested so far.

Here, we present simple analytical expressions for predicting electrostatic alignment. We demonstrate the applicability of the analytical expression on nucleic acids structures derived from RDCs measured in the commonly employed Pf1-phage medium. Specifically, we show that for nucleic acids in a Pf1-phage aligning medium the molecular alignment tensor **A** can be predicted reliably and accurately via a simple and fast calculation based on the gyration tensor spanned out by the phosphodiester atoms. Thanks to the recent increase in such nucleic acid RDC-derived structures (see e.g., Mollova & Pardi, 2000) a statistically relevant validation of the proposed prediction method can be carried out.

Materials and methods

To predict the (electrostatic) alignment of nucleic acids in Pf1-medium we employ the following approach. The predicted alignment tensor is derived from the gyration tensor ($\mathbf{R}_g^2$) using the atoms spanned out by the phosphate charges (P and O-atoms of the phosphodiesters) building on the approach of Almond and Axelson (Almond & Axelsen, 2002). The gyration tensor is defined as,

$$R_{g,ij}^2 = N^{-1} \sum_{r=1}^{N} x_i^r x_j^r \tag{1}$$

Here x_i^r $(i, j = x, y, z)$ are the coordinates of the N atoms r (P and O-atoms of the phosphodiesters) in the pdb file, relative to the center of mass. The alignment tensor **A** is then taken to have the same orientation as $\mathbf{R}_g^2$. It's eigenvalues are set equal to the square root of the eigenvalues of $\mathbf{R}_g^2$, denoted as $R_{g,ii}(= (R_{g,ii}^2)^{1/2})$, and it is made traceless,

$$A_{ii} = R_{g,ii} - 1/3 \sum R_{g,ii} \tag{2}$$

The eigenvalues of **A** (in Å) are ordered as usual, $|A_{xx}||A_{yy}|<|A_{zz}|$. The rhombicity $R(= A_{xx-}A_{yy}/A_{ax})$ is then definite positive. The axial component A_a may change sign $(= A_{zz-}1/2(A_{xx} + A_{yy}))$ depending on whether **A** derives from a prolate or oblate probability tensor **P**, defined by Kramer et al. (Kramer et al., 2004) as the probability distribution of the direction of the external field in the molecular fixed frame of reference.

The dipolar couplings between spin-1/2 nuclei P and Q can be calculated from (see e.g., Kramer et al., 2004),

$$D_{PQ} = d_{PQ}[Pf1]F(\vec{v}_{PQ}^{T} \cdot \vec{A} \cdot \vec{v}_{PQ}) \tag{3}$$

Here, d_{PQ} is $-(\mu_0/4\pi)\gamma_P\gamma_Q h/2\pi^2 < r_{PQ}^3 >$ where μ_0 is magnetic permeability of vacuum, γ is the gyromagnetic ratio, h is Planck's constant, and ν_{PQ} is the dimensionless unit vector pointing from spin P to Q, separated by the average distance $< r_{PQ} >$. The phage concentration, [Pf1] (in mg/ml), is introduced because D_{PQ} depends linearly on it, at least in dilute crystalline media (Fernandes et al., 2001). The factor F is a scaling factor (in (ml/mg) $Å^{-1}$) obtained by optimizing the predicted and measured RDC. Finally, one can define the *magnitude of the residual dipolar coupling tensor* D_a (Bax et al., 2001) (here $D_a = d_{PQ}$ [Pf1] F $\mathbf{A_{zz}}/2$).

Note that normalizing **A** (Equation 2) by its trace of its eigenvalues $(= \mathbf{Tr(R_g)})$ yields a dimensionless tensor $\mathbf{A_{sh}}(= \mathbf{A/Tr(R_g)})$ and $\mathbf{A_{sh}} + 1/3\mathbf{E}(= \mathbf{R_g}/\mathrm{Tr}(\mathbf{R_g}))$ equals the shape-dependent part of the probability tensor **P**. The factor ([Pf1] F) scales the probability tensor **P** and depends on a number of factors. It depends on the Pf1-phage concentration, via the integration volume. It also relates to the solute size and Debye–Huckel screening length via the distance dependence of interaction potential. Finally, it includes a scaling factor, which accounts for incomplete Pf1-medium alignment relative to the magnetic field.

Fernandes et al. (Fernandes et al., 2001) have derived simple analytical expressions for the steric alignment of axially symmetric solute particles in a liquid crystalline medium with planar obstruction. Almond and Axelson (Almond & Axelsen, 2002) have expanded on this and shown via simulation that for steric alignment in a liquid crystalline medium with planar obstruction, the alignment tensor is correctly derived from the gyration tensor calculated from all atoms of the solute. For nucleic acids the phosphate charges follow the shape of the molecule. We therefore initially simply postulate that for nucleic acids in Pf1 medium the (electrostatic) alignment tensor can correctly be predicted when derived from the gyration tensor spanned out by the phosphodiester atoms. We then test this by comparison with experimental data. In the theory (see Supplementary material) and results and discussion sections we subsequently show – using the Debye–Huckel approximation to the electrostatic interaction potential – that the electrostatic alignment indeed can be derived from the gyration tensor.

The procedures are incorporated into MATLAB scripts, which are available on request.

Results and discussion

We intend to demonstrate here that for nucleic acids in a Pf1-phage aligning medium the molecular alignment tensor **A** can reliably and accurately be predicted via a simple and fast calculation based on the gyration tensor spanned out by the phosphate charges (P and O-atoms of the phosphodiesters). At present ca. 26 DNA/RNA structures have been derived from RDCs induced in Pf1-phage medium (Table 1). We

restrict our analysis to the data in one medium (the Pf1-phage), because this limits the number of additional factors that can influence the alignment. This number of structures may seem surprisingly small, our analysis set, nevertheless, includes every nucleic acid structure refined with Pf1-phage RDCs deposited at the PDB, subject to availability of RDC data. This set of ca. 26 structures should be large enough to provide sufficient statistics to validate the proposed alignment prediction approach. In addition, the Pf1-phage set includes RNA and DNA molecules of different sizes. It consists of helical stems of different lengths of both A-helix (RNA) and B-helix (DNA) geometries. The helices are straight or bent and/or capped with a hairpin-loop or interrupted via an internal bulge. Also, a pseudoknotted RNA structure, a dimeric GAAA-receptor RNA complex, and a DNA three-way junction is included in the set. Consequently, the shape and the related charge distribution vary considerably within the test set, leading to, for instance, experimental Rhombicities (R_{exp}) that essentially cover the whole range of allowed values, from 0 to 0.66.

To assess the quality of the alignment tensor prediction, we compare for the set of structures the five parameters that define predicted and experimental alignment tensors, i.e. the alignment tensor magnitudes (D_a) and rhombicities (R), and the relative orientation of their principal axes ($\psi_{x,y,z}$). In addition, the predicted and experimental RDCs are compared as an overall indication of the quality of the prediction. If the experimental RDCs and the structure refinement and thus the experimental alignment tensors were error free, the deviations of the predicted parameters from the experimental parameters would directly indicate the quality of the prediction. However, the experimental alignment tensor is not error free and deviations between predicted and experimental values are expected. If they are less or of the same size as the experimental error, one can at least conclude that the deviations do not contradict a perfectly correct prediction. It is therefore important to establish some estimate of the error in the experimental alignment tensor.

The experimental alignment tensors ($\mathbf{A}_{exp}$) are derived from the experimental RDCs via SVD on

Table 1. Validation of prediction of Pf1-mediated alignment for nucleic acids based on gyration tensor spanned out by the phosphodiester atoms[a]

PDB[m]code	Shape[b]	Size (nt)[c]	RDC[d]	$RMSD_{SVD}$[e]	$RMSD_{Pred}$[f]	R_{exp}[g]	R_{pred}[h]	F ($\times 10^{-5}$)[i]	$<\Delta\psi_Z>(^\circ)$ [j]	$<\Delta\psi_{x,y}>(^\circ)$ [k]	R_{cor}[l]
Helical extended											
1naj	DNA	24	18.6	2.9(0)	3.0(1)	0.09(0)	0.12(1)	1.3(1)	2(1)	1(1)	0.93((0)
1fzx	DNA	24	3.9	2.0(0)	5.9(6)	0.01(0)	0.05(2)	3.2(2)	9(1)	35(11)	0.92(1)
1g14	DNA	24	3.7	1.4(0)	5(2)	0.01(0)	0.03(2)	3.2(4)	8(3)	23(14)	0,94(3)
1rvh	DNA	24	3.0	0.8(0)	1.9(8)	0.12(1)	0.09(1)	1.8(1)	4(3)	7(4)	0.97(2)
1rvi	DNA	24	3.3	2.0(0)	3.5(4)	0.06(1)	0.05(1)	1.2(1)	9(1)	7(5)	0.86(2)
1ss7	DNA	32	3.2	1.6(0)	2.7(8)	0.01(0)	0.02(0)	1.4(2)	2(2)	7(5)	0.95(3)
1z31	RNA	32	2.0	6.0(2)	11(1)	0.25(0)	0.09(3)	3.1(4)	4(2)	10(6)	0.91(2)
1xhp	RNA	32	2.4	2.7(1)	4.7(5)	0.13(0)	0.05(1)	1.7(1)	3(1)	18(2)	0.98(1)
1mnx	RNA	42	–	–	–	0.08	0.07(2)	–	–	–	–
1n8x	RNA	36	–	–	–	0.10	0.08(2)	–	–	–	–
1r2p	RNA	34	0.7	2.0(2)	9(2)	0.24(2)	0.05(2)	1.0(2)	12(5)	17(9)	0.75(6)
1nbr	RNA	29	7.3	5.3(3)	7.4(7)	0.15(2)	0.19(5)	1.0(2)	10(4)	17(4)	0.80(4)
1nc0	RNA	20	–	–	–	0.15	0.09(2)	–	–	–	–
1sy4	RNA	24	–	–	–	0.21	0.11(3)	–	–	–	–
1nz1	RNA	24	–	–	–	0.22	0.14(3)	–	–	–	–
1pjy	RNA	22	–	–	–	0.28	0.17(4)	–	–	–	–
1p5n	RNA	34	3.1	2.2(1)	4.3(7)	0.16(1)	0.04(2)	1.0(2)	8(3)	21(12)	0.88(4)
Dimeric complex											
2adt	RNA	86	0.3	1.4(2)	8.9(7)	0.57(1)	0.25(1)	2.3(1)	3(2)	3(2)	0.83(4)
Pseudo knot											
1yg3[n]	RNA	28	3.3	1.2(1)	5.2(4)	0.33(1)	0.59(7)	1.6(2)	23(4)	26(4)	0.63(7)

Table 1. Continued

PDB[m]code	Shape[b]	Size (nt)[c]	RDC[d]	$RMSD_{SVD}$[e]	$RMSD_{Pred}$[f]	R_{exp}[g]	R_{pred}[h]	F ($\times 10^{-5}$)[i]	$<\Delta\psi_Z>$(°) [j]	$<\Delta\psi_{x,y}>$(°) [k]	R_{cor}[l]
1yg3[n]	RNA	25	3.5	1.3(1)	5.5(4)	0.34(1)	0.42(7)	0.8(1)	23(3)	33(7)	0.62(7)
Hooked curved											
1s34	RNA	23	5.1	3.4(1)	7.1(3)	0.45(2)	0.54(9)	0.6(1)	17(3)	24(3)	0.68(5)
1p5m	RNA	55	2.5	2.5(1)	4(1)	0.45(1)	0.32(8)	0.7(2)	9(5)	12(4)	0.90(5)
1p5o	RNA	77	0.8	2.0(1)	9(2)	0.28(2)	0.32(6)	1.0(1)	16(3)	13(5)	0.89(4)
1p5p	RNA	77	0.8	1.9	6.7	0.28	0.31	1.0	14	13	0.95
Three-way junction											
1snj	DNA	36	1.2	1.8(3)	12(1)	0.62(2)	0.53(16)	0.6(2)	32(3)	24(5)	0.57(1)
Near spherical											
1kr8	DNA	7	5.4	0.2(0)	3.7(1)	0.61(3)	0.64(2)	0.5(1)	23(3)	49(6)	0.91(1)
1pqt	DNA	7	5.4	–	–	0.67	0.64(2)	–	–	–	–

[a]Structures refined against Pf1-mediated RDC data and deposited in the pdb bank. Values in bracket are standard deviations (e.g. 3.7(1) means 3.7 ± 0.1). – stand for unavailable experimental RDCs.[b]Shape indication of the molecule. Helical Extended: structures with an extended helical appearance, they include molecules with helical segments, which may contain internal bulge and/or be topped with an apical loop and/or be slightly bent (e.g 1z31, 1xhp, etc). Hooked/Curved: structures that have a strong curved appearance, e.g. 1s34 consists of short 6 bp stem with a folded 11 nt loop, or hockey stick appearance, e.g. 1p5m/o/p the HCV IRES element consist of a helical stem with a helical segment at the top strongly bent away. Near Spherical: DNA stem-loop structures with a very short stem giving a nearly spherical appearance, so that high rhombicity is combined with small anisotropy and errors could relatively easily occur due to slight deviations from the near spherical appearance. The other categories are self-explanatory.[c]Size of molecule in nucleotides (nt).[d]Number of RDCs per residue.[e]$RMSD_{SVD}$ (see text) is the RMSD between the measured RDCs and RDCs calculated via SVD using Pales.[f]$RMSD_{Pred}$ is the RMSD of the measured RDCs and RDCs predicted using gyration tensor based on phosphodiester atoms.[g]R_{exp} is the experimental rhombicity calculated via SVD using Pales (Zweckstetter & Bax, 2000); when RDC data were not available, R_{exp} is taken from the paper.[h]R_{pred} is the predicted rhombicity calculated from the gyration tensor based on phosphodiester atoms.[i]F is the scaling factor that optimizes experimental and predicted RDCs (see text).[j, k] $\Delta\psi_z$ is the angular difference between the principal z-axes of predicted and experimental alignment tensor; the latter are derived via SVD using Pales and calculated as the inverse cosine of the in product of the two vectors; $\Delta\psi_x$ and $\Delta\psi_y$ the corresponding angular deviations of the *x*- and *y*-axes, with $<\Delta\psi_{x,y}>$ equal to 0.5 $(\Delta\psi_x, +\Delta\psi_y)$.[l] R_{cor}, Pearson correlation coefficient between predicted and experimental RDCs.[m]References of the PDB codes are: 1naj (Wu et al., 2003), 1fzx and 1g14 (MacDonald et al., 2001), 1rvh and 1rvi (Stefl et al., 2004), 1ss7 (McAteer et al., 2004), 1z31 (Leeper & Varani, 2005), 1xhp(Sashital et al., 2004; Sashital et al., 2005), 1mnx (Vallurupalli & Moore, 2003), 1n8x (Lawrence et al., 2003), 1r2p (Sigel et al., 2004), 1nbr (McCallum & Pardi, 2003), 1nc0 (Sashital et al., 2003), 1sy4 and 1nz1 (Reiter et al., 2003), 1pjy (Staple & Butcher, 2003), 2adt (Davis et al., 2005), 1yg3 (Cornish et al., 2005), 1s34(Cabello-Villegas et al., 2004), 1p5m, 1p5n, 1p5o and 1p5p (Lukavsky et al., 2003), 1pqt and 1kr8 (Padrta et al., 2002), 1snj (Wu et al., 2004). [n]For 1yg3 two entries are included, one with all residues included and one with three extra-helical residues excluded (G9, A13 and C25). See further discussion in text.

the basis of the known RDC-refined structures. The error in the experimental alignment tensor is therefore determined not only by the accuracy of the measured couplings, but also by errors in the given structure (Zweckstetter & Bax, 2002). These structure errors can be local (structural noise), e.g. errors in sugar pucker, base twist and inclination, as well as overall, i.e. they can represent errors in relative domain orientations, for instance, in the relative orientation of the two parts that constitute bent or kinked helices. As a first indication of such experimental errors we have averaged all parameters defining **A** over the ensembles of deposited structures. The corresponding RMSD (given in parentheses in Table 1 behind relevant values) provides then some indication of the uncertainty in the parameter. Alternatively, the effect of structural uncertainty can be obtained via simulation. Zweckstetter and Bax (Zweckstetter & Bax, 2002) simulated the effect of (local) structural uncertainty by slightly reorienting the RDC vector orientations with a Gaussian cone-shaped distribution. The deviations between the original and final vectors were described by a Gaussian cone-shaped distribution, with a standard deviation σ_{cone}, and a relative probability of $\sin(\beta) \exp(-\beta^2/\sigma^2_{cone})$ for an angle β between the original and modified orientation. A quantitative estimate of the effect of structure noise on D_a, R, and tensor orientation is thus obtained. As a second estimate we therefore employ these Monte-Carlo results presented by Zweckstetter et al. to assess the error in the experimental alignment tensor and from that the significance of the observed deviations between predicted and experimental alignment tensor.

The rhombicity (R) and the alignment tensor orientation are direct indicators of the quality of the prediction of the alignment tensor. The rhombicity is defined as the ratio of $(A_{xx} - A_{yy})$ and A_{zz}. Consequently, aspects such as Pf1-phage order, Pf1-phage concentration, and small-scale uniform dynamics etc., which simply scale the overall degree of alignment, do not affect rhombicity. Thus, the deviations between predicted and experimental values of the rhombicity can directly be related to molecular aspects. The same applies to the alignment tensor orientation. It should be noted however, that the relative error in the rhombicity equals the sum of relative errors in $A_{xx} - A_{yy}$ and A_{zz} and is therefore larger than the error in the axial component, A_a $(= 1.5A_{zz})$. The *magnitudes of the predicted and experimental dipolar coupling tensor*, defined as D_a $(= d_{PQ}$ [Pf1] F $\mathbf{A}_{zz}/2$; see Material and methods) and D_a^{SVD} $(= d_{PQ}\mathbf{D}_{zz}^{SVD}/2$, with A_{zz}^{SVD}, the zz-eigenvalue of the SVD-derived alignment tensor), respectively, are also compared (see Supplementary material, Table S1). As pointed out above, several scaling factors can affect D_a^{SVD} (and D_a) that are difficult to assess accurately, which would make such a comparison less accurate. Therefore, in D_a the optimization factor F was introduced, obtained by minimizing the RMSD between experimental and predicted RDC as explained in Materials and methods. Thus, F incorporates the affect of the scaling factors. Consequently, when D_a and $D_a{}^{SVD}$ are equal, prediction and experiment correspond perfectly, except possibly for simple scaling. A deviation between D_a and $D_a{}^{SVD}$ indicates that prediction and experiment differ by more than just simple scaling. Finally, the overall RDC correspondence is well established by the RMSD ($RMSD_{pred}$) between experimental and predicted RDC. This RMSD should be compared with the RMSD between experimental and SVD-derived RDCs ($RMSD_{SVD}$). The $RMSD_{SVD}$ is the best obtainable correspondence. The residual deviation stems then from RDC measurement error and/or errors in local structure (Zweckstetter & Bax, 2002) as discussed above, as well as error in overall structure and errors due to the neglect of certain types of dynamics (Al-Hashimi, 2005; Blackledge, 2005). For instance, differences in small-scale dynamics of different dipolar vectors may affect their RDCs differently (Blackledge, 2005). In addition, domain dynamics may affect the alignment tensor and thereby affect the RDCs (van Buuren et al., 2004; Al-Hashimi 2005; Blackledge, 2005). Note that uniform small-scale dynamics will simply scale down the RDCs and is accounted for by the scaling factor. Hence, the optimal obtainable experimental RMSD, in the absence of errors in structure and/or dynamics, is the RDC measurement error, which is under normal circumstances about 2 Hz. When only a small number of RDCs is employed, structure refinement is usually underdetermined and the $RMSD_{SVD}$ may become smaller than experimental error in the RDCs. The $RMSD_{pred}$ should per definition be larger than $RMSD_{SVD}$. Finally, we also present the correlation coefficient between predicted and experimental RDCs when applicable.

As can be seen from Figure 1 (and Table 1), the predicted and experimental R correlate quite satisfactory with an overall correlation coefficient of 0.85 and an RMSD (R_{pred}, R_{exp}) of 0.08. The good quality of the prediction is further demonstrated by the good correspondence between predicted and experimental RDCs as shown for four representative examples in Figure 2 (a complete set of correlation plots for all molecules is available in the Supplementary material).

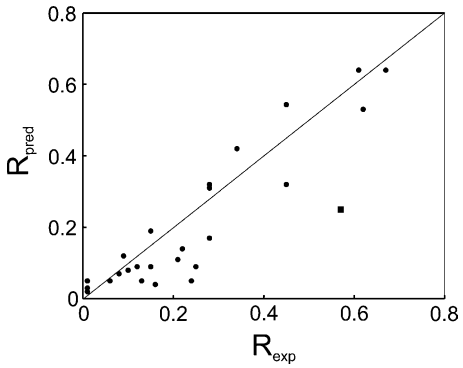

Figure 1. Illustration of the quality of the prediction of the molecular alignment of Pf1-phage aligned RNAs/DNAs, showing predicted rhombicity (R_{pred}) vs. experimental rhombicity (R_{exp}). The data for all 26 Pf1-mediated structures (Table 1) presently deposited in the PDB bank are shown. When experimental RDC data were deposited in the pdb bank, the SVD-derived R_{exp} was used or otherwise as listed in the paper (see also Table 1, column 7). For 2adt R is marked with a different symbol (filled square) to indicate the probable bias in the experimental value (see text). R_{pred} stems from the alignment tensor using the phosphodiester atoms (see also Table 1,column 8).

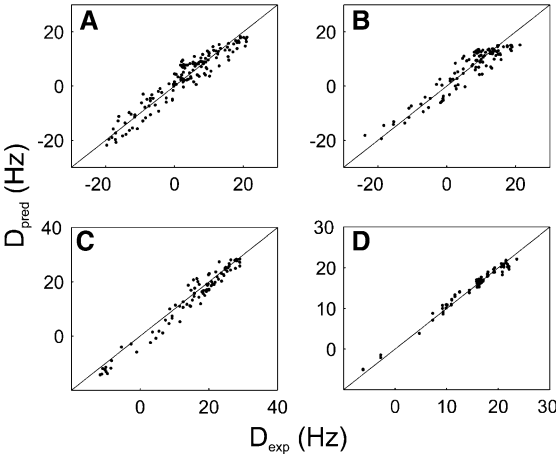

Figure 2. Correlation between experimental and predicated residual dipolar couplings for four representative examples with PDB codes: (A) 1p5m (R_{cor} = 0.95), (B) 1p5n (R_{cor} = 0.94), (C) 1g14 (R_{cor} = 0.98) and (D) 1rvh (R_{cor} = 0.99). Here R_{cor} represents the Pearson correlation coefficient. The literature references to the PDB codes of the corresponding molecules are given in Table 1.

The orientation of the predicted alignment tensor agrees generally quite well with that of the one derived from experimental data via SVD. The *z*-axes are on average quite close ($= 11 \pm 8°$). Within the X/Y plane the deviation is somewhat larger, is $18 \pm 13°$. On average the optimization factor F equals 1.2 ± 0.6 (excluding outliers 1fzx and 1g14). Overall, F is relatively well defined, despite the fact it can be influenced by several experimental factors, e.g. phage orientation, variation in salt strength, temperature, and dynamics. As can be gleaned from Table 1, $RMSD_{pred}$ is on average ($< RMSD_{pred} >$) 5.9 Hz, which is about 2.6 times the error in the experimental RDC data ($< RMSD_{SVD} > = 2.3$ Hz). That $< RMSD_{pred} >$ is larger than $< RMSD_{SVD} >$ is expected even for perfect prediction as discussed above. When comparing individual $RMSD_{pred}$ and $RMSD_{SVD}$ values, it should be kept in mind that the best experimental RMSD, even in the absence of any residual error in structure or dynamics, is the RDC measurement error (ca. 2 Hz). Moreover, during structure refinement, which is underdetermined, in particular with a small number of RDCs, the $RMSD_{SVD}$ can become smaller than this value. Taking 2.9 Hz – the $RMSD_{SVD}$ for 1naj with over 18 RDCs per residue – as a reasonable estimate for the real minimum experimental RMSD, one finds that $RMSD_{pred}$ varies between 1 to 3.0 times this $RMSD_{SVD}$ for all structures (1snj, 4.1, being an exception). In conclusion, the alignment of nucleic acids in Pf1-phage medium appears to be reliably predicted from their structure based on an alignment tensor derived from the radius of gyration spanned out by the phosphate charges.

In the Mechanism Section and in Supplementary material it is shown that nucleic acid alignment in Pf1-medium, although fully electrostatic, is functionally identical to steric alignment for the usual salt conditions. Because 'steric' alignment prediction is implemented in Pales via simulation of all potential solute orientations, a comparison was made with our prediction method. The results of the comparison are presented in Supplementary material (Table S2a and Figures S1a, S2a). The scaled RMSDs ($= RMSD/D_a^{SVD}$) of the two methods are quite similar (Figure S1a). The same holds for the alignment tensor orientations, with the average $\Delta\psi_{x,y,z}$ between the two methods equal to ca. $10 \pm 10°$ with a maximum deviation of ca. 20°. Interestingly, the R calculated via our method yields a somewhat better correlation with experimental (SVD-derived) rhombicity (R_{exp}, Figure S2a). In particular for structures with small R our method appears to give better prediction. It should be pointed out that the 'steric' method implemented in Pales is not very robust with regard to exact position of single atoms (for example rotation of terminal hydroxy groups could change R).

Finally, to more fully verify the correctness of the proposed method, we compared it also with prediction of the electrostatic alignment tensor from structure, implemented in Pales (Zweckstetter et al., 2004). Note that (see also Mechanism Section and in Supplementary material) our proposed method follows from the (approximate) analytical Debye–Huckel solutions of the Poisson–Boltzmann equation for electrostatic potential. In Pales the full Poisson–Boltzmann equation is numerically evaluated to derive the electrostatic potential and subsequent prediction of the 'electrostatic' Pales alignment is done via simulation of all potential solute orientations in this potential. The results of the comparison are presented in Supplementary material (Table S2b and Figures S1b, S2b). The scaled RMSDs ($= RMSD/D_{a\_SVD}$) of the two methods are quite similar (Figure S1b). The R's of the two methods show a very close correlation (Figure S2b). Also the alignment tensor orientations correlate quite closely (Table S2b). Direct

comparison of the two methods yields (Table S2b, last 3 columns) z-axis angular deviations ($|\Delta\psi_z|$) of only $2 \pm 2°$ on average, while the x- and y-axes the angular deviations, $\Delta\psi_{x,y}$, are on average $4 \pm 4°$ and $5 \pm 3°$, respectively; the $\Delta\psi_{x,y}$ data points with $R_{SVD} < 0.02$ (1fzx and 1g14) are excluded in the average because tensor orientation in the X,Y-plane becomes ill defined for very small or zero R. In conclusion, the two methods yield essentially similar 'electrostatic' alignment predictions, which further validates our analytical approach.

It is of interest to examine some specific examples and aspects in more detail. For the Dickerson dodecamer (1naj), which has by far the largest number of RDCs per residue (ca. 18), the predicted alignment is extremely close to the experimental one as evidenced by the different parameters (Table 1). First, the experimental and predicted RDC correlate quite well (correlation coefficient is 0.93). Second, the experimental and predicted Rhombicities are quite close. Third, the $RMSD_{pred}$ and $RMSD_{SVD}$ are essentially equal, and fourth, the F factor is quite close to the average. Most strikingly, the principal axes orientations of the predicted alignment tensor deviate less than 2° (and within 2σ of the ensemble; σ is standard deviation of the ensemble averaged paremeter) from the principal axes orientation of the experimental (SVD) alignment tensor. It is generally accepted that the accuracy of structure determination improves with increasing number of experimental constraints per residue. Thus, given a correct prediction method, the deviation of predicted alignment from the SVD derived tensor should be zero, when a sufficient number of RDCs per residue have been measured. We indeed observe essentially zero deviation for 1naj. This shows that at least for 1naj the prediction method is correct, i.e. precise and accurate within 2° in as far as the orientation is concerned, within 0.03 in as far as the Rhombicity is concerned, and within ca. 1 Hz in as far as the RMSD is concerned. The question that arises now is whether the 1naj precision and accuracy also applies to the whole range of structures.

We consider first some specific examples of large RNA molecules with complex structure. The 30 kDa (86 nt) dimeric complex of the GAAA-receptor RNA is such an example (2adt). The alignment tensor orientation comes out quite correctly ($|\Delta\psi_{z,x/y}| \sim 3°$). The rhombicity, R, is somewhat off (0.57 experimentally vs. 0.25 predicted), but this may well be due to the small number and anisotropic distribution of the RDCs measured. They are mostly from imino resonances and all cluster between 5 Hz and 55 Hz. Even with $R = 2/3$, an isotropic RDC vector distribution would yield values between −55 Hz and 55 Hz and even more negative values for a smaller R. Our experience is that with few RDCs one tends to overestimate R (and underestimate D_a) using SVD with Pales. We indeed find an increase of the R_{SVD} when only aromatic and imino RDCs are used for molecules with small R, e.g. R_{SVD} of 1rvi increases from 0.06 to 0.18, and of 1fzx from 0.01 to 0.08. Thus, the anisotropic distribution of measured RDCs could well explain the discrepancy in R between the gyration tensor method and the SVD.

Another set of examples of large RNAs (55–77 nt) with complex structures, are the entries of the HCV-IRES element, 1p5m, 1p5p, and 1p5o. They all have a hockey stick like shape. Their R's are well predicted (experimental: 0.45/0.28/0.28 vs. predicted: 0.32/0.31/0.32, respectively). Also, the $|\Delta\psi_z|$ are reasonably small (9, 16, and 14°, respectively) and the Pearson correlation coefficients also stay high (>0.9).

A further molecule for special consideration is the pseudoknot from the sugar-cane yellow leaf virus mRNA (1yg3) (Cornish et al., 2005). This molecule contains three extra-helical nucleotides, which are poorly defined in the structure. We found that the exact position of these nucleotides influences the shape of the molecule in a striking manner. When protruding from the core of the molecule, they change the shape of the molecule and the gyration tensor becomes oblate $D_a > 0$ and $R = 0.59$, while if these nucleotides are excluded from the calculation, the gyration tensor is prolate with $D_a < 0$ and $R = 0.43$ (both entries are included in Table 1). Although the correlation between predicted and experimental rhombicity is substantially improved upon exclusion of the extra-helical residues, the orientation of the tensor is hardly improved ($|\Delta\Psi_z| = 23°$). This is in common with other molecules with large rhombicity (see below).

Generally, the residual orientation deviations remain quite small, e.g. $|\Delta\psi_z|$ 15° (excluding, 1snj, 1yg3, and 1kr8) and it seems that the deviations could easily be accommodated by small structural changes or experimental error. With regard to the

$|\Delta\psi_{x,y}|$ values of these structures, it should be noted that a rhombicity smaller than 0.66 scales down the effect of rotations in the X, Y-plane on the calculated RDC compared to rotations away from the z-axis. Consequently, the error in the experimental $\psi_{x,y}$ is expected to be larger than in ψ_z, i.e. $\sigma(\Delta\psi_{x,y}) \approx 2\sigma(\Delta\psi_z)/3R$ as given by Zweckstetter and Bax (Zweckstetter & Bax, 2002). Most of the $|\Delta\psi_{x,y}|$ values follow this relationship and remain relatively small (ca. <20°). Structural noise of up to ca. 5° or measurements error in RDCs of 1.8 Hz could easily explain these angular deviations between predicted and experimental alignment tensor orientations, as can be gleaned from Figure 5 of Zweckstetter and Bax (Zweckstetter & Bax, 2002). For 1snj, 1kr8 (and possibly 1yg3), entries with the largest R, the deviations of the predicted tensor orientation fall just outside the 5°/1.8 Hz range, but still within the 10°/1.8 Hz range. It is potentially possible that due to the higher complexity of molecules with high R structural noise increases, i.e. the chance of conformational deviations becomes larger. Note also that the type of structural noise considered here is local and does not involve global fluctuations. It is important to stress that R is reasonably well predicted for all entries in Table 1, including 1snj, 1kr8, and 1yg3. We further note that these somewhat larger residual orientation deviations for 1snj, 1kr8 and 1yg3 are essentially reproduced with 'steric' and 'electrostatic' Pales and therefore not the result of the simplified procedure. It is tempting to discern a trend, i.e. larger R leads to larger $|\Delta\psi_{z,x/y}|$. However, a correlation plot of different estimates of $|\Delta\psi_z|$ vs. different estimates of R, e.g. of $|\Delta\psi_z|$ of our method and of 'steric' Pales vs. R_{exp} and R_{pred}, shows no discernable correlation (Supplementary material) except that at larger R the angular deviations are larger. Finally, we note that no correlation exists between worse orientation and complexity of the structure, e.g. $|\Delta\psi_z|$ of the dimeric GAAA-receptor (2adt) and of the HCV-IRES element (1p5p/o) is well predicted, but of the pseudoknot (1yg3) it is worse. Taken together the comparison with experimental data strongly suggests that the prediction of the alignment tensor is correct; not only for 1naj, but also for the other structures investigated and that the deviations seen are due to small structural noise (5°/10°) and/or experimental RDC error (1.8 Hz). It appears that for larger R the orientation deviations become somewhat larger. For practical purposes, we conservatively conclude that because the statistical set is relatively small it cannot be excluded that for larger R the predicted tensor orientations may be somewhat off. Taking R, $|\Delta\psi_z|$ and $|\Delta\psi_{x,y}|$ as guidelines, one observes from Table 1, that for $R < \sim 0.4$, our method predicts both R and orientation correctly and for $R > \sim 0.4$, R is right but the orientation may have a somewhat larger deviation.

Mechanism of electrostatic alignment and derivation of analytical expressions

Until now the mechanism of the alignment has not been addressed and below we consider this aspect. The studies of Zweckstetter and Bax (Zweckstetter & Bax, 2001; Zweckstetter et al., 2004) strongly suggest that in the charged Pf1 medium the alignment of a charged biomolecule is at least in part electrostatic for salt strengths below 100 mM. Interestingly, the experimental data and simulations show that for the Dickerson dodecamer the salt-strength dependence of the alignment is weak (Zweckstetter et al., 2004). In the numerical simulation of electrostatic alignment by Zweckstetter et al. (Zweckstetter et al., 2004) and Ferrarini (Ferrarini, 2003), the full Poisson–Boltzman equation was employed to numerically calculate the electrostatic interaction potential between alignment medium and charged solute. The Poisson–Boltzman approach is considered highly reliable, but breaks down (and thus its numerical evaluations) at distances very close to the medium's surface, i.e. at distances of the ion-size (Israelachvili, 1995). It also breaks down when association or binding occurs, which may occur when medium and solute charges are of unequal sign (Israelachvili, 1995). Alternatively, the analytical Debye–Huckel expressions for the electrostatic potential emanating from a negatively charged wall and/or rod exist. They are valid and equal to the full Poisson–Boltzman result at distances sufficiently far away from a poly-ionic charged surface (Brenner & Parsegian, 1974; Manning, 1978; Fixman, 1979; Le Bret & Zimm, 1984; Stroobants et al., 1986; Israelachvili, 1995). This condition is naturally enforced when the charge of alignment medium and solute are of equal sign, because of the mutual repulsion between medium and solute. This

repulsion also circumvents the potential problem of counter-ion association. Consequently, the analytical Debye–Huckel expressions should be valid for nucleic acids aligned in the negatively charged Pf1-medium.

We have employed these well-established analytical Debye–Huckel expressions for the electrostatic potential emanating from a negatively charged wall and/or rod (Brenner & Parsegian, 1974; Manning, 1978; Fixman, 1979; Le Bret & Zimm, 1984; Stroobants et al. 1986; Israelachvili, 1995), and can derive analytical expressions for the interaction potential between charged alignment medium and charged solute and subsequently expressions for the alignment tensor. The details of these derivations are given in Supplementary Material. Here we qualitatively highlight the main aspects of these derivations, which theoretically validate the proposed prediction approach. We also indicate the ranges of validity of the approach and discuss thereby the alignment mechanism. In this way, more insight into the physical parameters determining electrostatic alignment and thus the mechanism is obtained.

It is of value to first briefly summarize the most important parameters that are crucial for the electrostatic alignment as follows from the Debye–Huckel approach we employ. (a) The Debye–Huckel screening length κ^{-1} determines the range of the electrostatic interaction and relates to the salt strength. (b) The size and shape of the molecule. For a simple axially symmetric solute, the length ℓ and axial ratio, p, play an important role ($p = d/\ell$ where d is the diameter). When the solute is of arbitrary shape the gyration tensor, $\mathbf{R_g}^2$, well describes the shape. Its eigenvalues or the square root of the eigenvalues describe the smoothed solute shape. The radius of gyration $R_g(= 1/3\mathrm{Tr}\ \mathrm{R}_{g,ij})$ is then a good parameter for describing the size of the arbitrary shaped solute (it replaces ℓ, which is proportional to R_g). (c) A further important aspect is the charge distribution and its relation to the shape of the molecule. For nucleic acids the charge distribution follows essentially the shape of the molecule. (d) A final aspect is the medium. When the medium is a charged rod, also the rod radius, r_{Pf1}, plays a crucial role. When $\kappa^{-1} > r_{\mathrm{Pf1}}$ the potential emanating from the rod surface is that of a rod (Brenner & Parsegian, 1974), while a smooth transition occurs to a wall-like potential when κ^{-1} decreases and becomes smaller than r_{Pf1} (Israelachvili, 1995). When $\ell < r_{\mathrm{Pf1}}$ the interaction potential and thus the electrostatic alignment is like that of a charged wall, whereas for $\ell > r_{\mathrm{Pf1}}$ rod like conditions apply. In summary, three parameters play a crucial role, namely κ^{-1}, ℓ, and r_{Pf1}. For the phage Pf1 the rod radius r_{pf1} is ca. 34 Å. The usual salt strengths range from ca. 30 to 300 mM, so that $\kappa^{-1} < r_{\mathrm{Pf1}}$ and the potential becomes wall like. For most DNA/RNA molecules considered ℓ ranges up to ca. 100 Å. In other words, $\ell < 2.5\ r_{\mathrm{Pf1}}$ and consequently wall like behavior usually applies in as far as the integration over the volume is concerned.

Figure 3 shows the orientation order parameter $S(r)$ ($\propto A_{\mathrm{ax}}$) of a negatively charged rod-like solute ($R = 0$) as a function of the distance r of the center of mass of the solute to the surface of the negatively charged rod-like aligning medium. Three situations are displayed, viz. $\kappa\ell \approx 5.4$ (Debye–Huckel screening length κ^{-1} is of the order of the solutes' length ℓ), $\kappa\ell \approx 11.2$, and $\kappa\ell \approx 38$ (κ^{-1} much smaller than ℓ). Two opposing effects contribute to $S(r)$. On the one hand, the solute's alignment due to electrostatic interaction is strongest near the medium's surface and decreases exponentially away from it. On the other hand, electrostatic repulsion is strongest near the medium's surface and expels the solute molecules away from it up to $\kappa r \approx 1$, i.e. up to a distance large enough for the electrostatic interaction energy to be of the order of kT. Consequently, although their degree of alignment is large near the surface, they collectively contribute little to $S(r)$, and $S(r)$ is nearly zero near the medium's surface. Further away from the surface, many solute molecules are present, but their degree of alignment becomes increasingly limited. The net effect is that $S(r)$ shows a maximum (Figure 3). Note the analogy of this description with the one used by Fernandes et al. (Fernandes et al., 2001), who derived analytical expression for steric alignment. They divided the volume into a forbidden, restricted, and free region. These regions have similar connotations as the ones described here for the electrostatic interaction, except that the transitions are here less abrupt.

Two crucially different regimes should be considered, viz. κ^{-1} larger than ℓ and κ^{-1} smaller than ℓ. In the limit of κ^{-1} much larger than ℓ, the width at half height of $S(r)$ is proportional to κ^{-1} and the height proportional to $(\kappa\ell)^2$ (Figure 3).

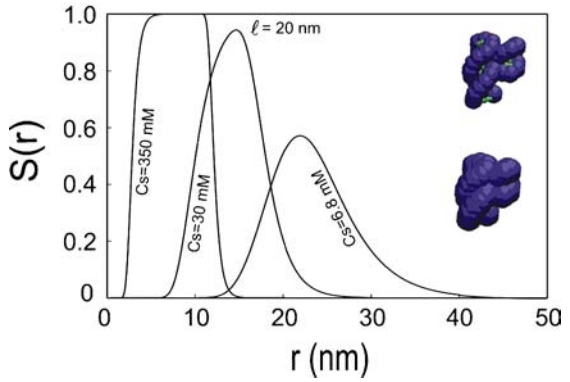

Figure 3. Illustration of the electrostatic alignment mechanism for nucleic acids in a negatively charged medium. S(r) is the orientation order of a negatively charged rod-like molecule of size ℓ of 20 nm (linear charge density 2e per 0.36 nm). S(r) is shown as a function of the distance r away from a negatively charged rod for three salt concentrations, c_s: 350 mM ($\kappa^{-1} = 0.53$ nm, $\kappa\ell = 38$), 30 mM ($\kappa^{-1} = 1.8$ nm, $\kappa\ell = 11.2$); and 6.8 mM ($\kappa^{-1} = 3.7$ nm, $\kappa\ell = 5.4$). To calculate S(r) (see Supporting material) the medium charge density was set equal to that of a Pf1-phage. The electrostatic alignment is due to the repulsive interaction – with the charged Pf1 rod – of the individual Debye–Huckel spheres of radius κ^{-1} located at the phosphate charges. This is shown for two molecules on the right (the Debye–Huckel spheres are indicated in blue). The top-right molecule has a Debye–Huckel screening corresponding to high salt (ca. 350 mM), and the bottom-right molecule corresponding to low salt (ca. 7 mM). One can distinguish, a forbidden region (small r, $r\ \kappa^{-1}$), a restricted region and a free region, as for steric alignment (Fernandes et al., 2001). For high enough salt conditions (c_s between 350 mM and ca. 50 mM), the Debye–Huckel spheres have radii smaller than the overall size of solute (top-right figure) and act like the atoms with a given van der Waals radius in case of steric alignment. Consequently, the electrostatic alignment depends on the distribution of Debye–Huckel spheres in a manner, analogous to the way the steric alignment depends on the distribution of van der Waals spheres. Note that the alignment remains completely electrostatic as long as the collective radii encompass the molecule. Note also that under these high salt conditions S(r) approaches 1 as long $r < \ell$. Consequently, S becomes proportional to ℓ and independent of salt strength. At lower salt (e.g. c_s equal to 6.8 mM) κ^{-1} is much larger than ℓ and this saturation disappears, i.e. S(r) is always smaller than 1. Under these conditions the overall order parameters S from electrostatic alignment becomes salt dependent and proportional to ℓ^2. Note that the corresponding curves for smaller values of ℓ can be obtained by interpolation from the given curves by corresponding scaling by κ^{-1}.

The overall order S is the integral of S(r) over the complete volume with size L_t, where L_t is half the distance between rods, so that $S \propto \kappa^{-1}\ (\kappa\ell)^2/L_t$ in the limit of wall-like behavior. A more extensive derivation is shown in supplementary material, where it is also shown that for an arbitrary-shaped molecule, $S \propto \kappa R_g^2/L_t$. However, under the usual salt conditions (ca. >30 mM), this 'low salt limit' limit is never reached. Instead, κ^{-1} is usually smaller than ℓ and the electrostatic interactions are short range. The individual charges bump into the charged medium's surface and experience strong (electrostatic) hindrance, like the individual surface atoms of the solute in the case of steric hindrance. Consequently, under these conditions the electrostatic alignment behaves essentially as steric alignment, and could be called steric-like alignment. When the charge distribution follows the shape of the molecule, steric and electrostatic alignment are indistinguishable in their functional dependences on the shape of the molecule. The order parameter S(r) is – in this limit – essentially constant within a layer of thickness $\ell/2$ off the surface medium and equal to zero further away (see e.g. left-hand curve in Figure 3). Consequently, the functional form for this steric-like electrostatic alignment is $S \propto \ell/L_t$, as for pure steric alignment. A more extensive derivation is given in supplementary material, where it is also shown that for an arbitrary-shaped molecule, $S \propto R_g/L_t$. Note the absence of salt dependence under these (usual) conditions. The absence of a salt dependence also followed from the simulation of electrostatic alignment on the Dickerson dodecamer by Zweckstetter et al. (Zweckstetter et al., 2004). It is important to further note that for nucleic acids and under the usual (high) salt conditions, κ^{-1} remains large enough for the collection of individual spheres with radius κ^{-1}, to encompass the whole molecule and to prevent true steric interaction (Figure 3). In conclusion, for the usual (high salt) conditions, $\kappa^{-1} > \ell$, the dependence on parameters as size and shape of solute is the same for electrostatic and steric alignment, but the underlying mechanisms are different.

Finally, note that it can readily be shown that the above analysis also applies for solutes of arbitrary shape. Almond and Axelsen (Almond & Axelsen, 2002) have previously shown – considering steric alignment – that when the alignment tensor $\mathbf{A}$ is proportional to size, it has the same orientation as the gyration tensor $\mathbf{R}_g^2$, and its eigenvalues are equal to the square root of those of $\mathbf{R}_g^2$, i.e. $\mathbf{A} \propto (\mathbf{R}_g{}^2)^{1/2} = \mathbf{R}_g$. This proportionality holds thus also true for electrostatic alignment as considered here in the limit when electrostatic interactions are short range.

Concluding remarks

In summary, we have demonstrated using a data base of 26 RDC-based nucleic acid structures that the molecular alignment tensor of Pf1 aligned nucleic acids can accurately be predicted from their structure via a simple and fast calculation based on the gyration tensor spanned out by the phosphate charges. For practical purposes we can conclude the following, for rhombicity $R < \sim 0.4$, our method predicts both R and orientation correctly, while for $R > \sim 0.4$, R is right but the orientation may be somewhat off. We have also shown that for the usual salt strengths and nucleic acid sizes the alignment is completely electrostatic but its functional behavior is like for steric alignment.

The prediction method presented provides opportunities for improved structure calculation of Pf1-aligned (multi-domain) nucleic acids and better assessment of their dynamics. Several (literature) examples of usage of alignment prediction have already been mentioned in the introduction. Here, we highlight and/or elaborate on some (additional) examples. The prediction method could be used to check and thereby validate the refined structures obtained via simulated annealing. The fast prediction method could also be very useful in structure determinations where relatively few RDCs are at hand as it can put R and D_a (and tensor orientation) in the correct range. For instance, an iterative approach could be applied, where R and D_a (and/or alignment tensor orientation), obtained from the gyration tensor, are used in and updated during the simulated annealing structure refinement. Furthermore, it opens the way for new global structure optimization of (multi-domain) nucleic acids, in which the alignment tensor and RDCs are continuously calculated for each test structure during the structure derivation process, similar to the protocol implemented to derive the Holliday junction structure using magnetic-field induced RDCs (van Buuren et al., 2004). As for degenerate solutions, in multi-domain molecules degenerate solutions can be resolved via a combination of prediction of the alignment tensor from structure and geometric constraints (see e.g. (van Buuren et al., 2004)). For assessment of flexibility/dynamics, it is useful for back-calculation of RDCs in molecular dynamics simulations. Most interesting is the possibility to directly simulate the effect of domain dynamics in multi-domain nucleic acids. This ensemble averaging could be incorporated into the structure calculation protocol mentioned above and implemented as for magnetic-field induced RDCs.

Supporting information available:

Table showing comparison of predicted and experimental magnitude of alignment tensor and two Tables and Figures showing comparison of alignment tensor prediction via method presented here with prediction via 'steric' Pales and 'electrostatic' Pales. Complete set of correlation plots of displaying predicted vs. experimental RDCs of RNAs/DNAs in Table 1. Derivation of analytical expressions for the electrostatic alignment presented here based on the Debye–Huckel approximation of the electrostatic potential.

The supporting information is available in electronic format at http://dx.doi.org/10.1007/s10858-006-9008-y.

Acknowledgements

This research was supported by grants from the Dutch Science Foundation (SW) and the Danish National Research Council (MP). We thank Prof. C.W. Hilbers for encouraging comments. We also thank the anonymous reviewers for their constructive suggestions.

References

Al-Hashimi, H.M. (2005) *Chembiochem*, **6**, 1506–1519.
Al-Hashimi, H.M., Majumdar, A., Gorin, A., Kettani, A., Skripkin, E. and Patel, D.J. (2001) *J. Am. Chem. Soc.*, **123**, 633–640.
Al-Hashimi, H.M., Tolman, J.R., Majumdar, A., Gorin, A. and Patel, D.J. (2001) *J. Am. Chem. Soc.*, **123**, 5806–5807.
Almond, M. and Axelsen, J.B. (2002) *J. Am. Chem. Soc.*, **124**, 9986–9987.
Azurmendi, H.F. and Bush, C.A. (2002) *J. Am. Chem. Soc.*, **124**, 2426–1427.
Bax, A. (2003) *Protein Sci.*, **12**, 1–16.
Bax, A., Kontaxis, G. and Tjandra, N. (2001) *Meth. Enzymol.*, **339**, 127–173.
Bertini, L., Luchinat, C. and Parigi, G. (2002) *Prog. Nucl. Magn. Reson. Spectrosc.*, **40**, 249–273.
Bewley, C.A. and Clore, G.M. (2000) *J. Am. Chem. Soc.*, **122**, 6009–6016.
Blackledge, M. (2005) *Prog. Nucl. Magn. Reson. Spectrosc.*, **46**, 23–61.

Brenner, S.L. and Parsegian, A. (1974) *Biophys. J.*, **14**, 327–334.
Bryce, D.L. and Bax, A. (2004) *J. Biomol. NMR*, **28**, 273–287.
Cabello-Villegas, J., Giles, K.E., Soto, A.M., Yu, P., Mougin, A., Beemon, K.L. and Wang, Y.X. (2004) *RNA*, **10**, 1388–1398.
Clore, G.M., Gronenborn, A.M. and Bax, A. (1998) *J. Magn. Reson.*, **133**, 216–221.
Clore, G.M., Gronenborn, A.M. and Tjandra, N. (1998) *J. Magn. Reson.*, **131**, 159–162.
Cornish, P.V., Hennig, M. and Giedroc, D.P. (2005) *Proc. Natl. Acad. Sci. USA*, **102**, 12694–12699.
Davis, J.H., Tonelli, M., Scott, L.G., Jaeger, L., Williamson, J.R. and Butcher, S.E. (2005) *J. Mol. Biol.*, **351**, 371–382.
de Alba, E. and Tjandra, N. (2002) *Prog. Nucl. Magn. Reson. Spectrosc.*, **40**, 175–197.
Fernandes, M.X., Bernado, P., Pons, M. and Torre, J.G.de la (2001) *J. Am. Chem. Soc.*, **123**, 12037–12047.
Ferrarini, A. (2003) *J. Phys. Chem. B*, **107**, 7923–7931.
Fixman, M. (1979) *J. Chem. Phys.*, **70**, 4995–5005.
Hansen, M.R., Mueller, L. and Pardi, A. (1998) *Nat. Struct. Biol.*, **5**, 1065–1074.
Israelachvili, J. (1995) *Intermolecular & Surface Forces* Academic Press, New York.
Kramer, F., Deshmukh, M.V., Kessler, H. and Glaser, S.J. (2004) *Concept Magn. Reson.*, **21**, 10–21.
Kung, H.C., Wang, K.Y., Goljer, I. and Bolton, P.H. (1995) *J. Magn. Reson. Ser. B*, **109**, 323–325.
Latham, M.R., Brown, D.J., McCallum, S.A. and Pardi, A. (2005) *Chembiochem*, **6**, 1492–1505.
Lawrence, D.C., Stover, C.C., Noznitsky, J., Wu, Z.R. and Summers, M.F. (2003) *J. Mol. Biol.*, **326**, 529–542.
Le Bret, M. and Zimm, B. (1984) *Biopolymers*, **23**, 287–312.
Leeper, T.C. and Varani, G. (2005) *RNA*, **11**, 394–403.
Lipsitz, R.S. and Tjandra, N. (2004) *Ann. Rev. Biophys. Biomol. Struct.*, **33**, 387–413.
Losonczi, J.A., Andrec, M., Fischer, M.W.F. and Prestegard, J.H. (1999) *J. Magn. Reson.*, **138**, 334–342.
Lukavsky, P.J., Kim, I., Otto, G.A. and Puglisi, J.D. (2003) *Nat. Struct. Biol.*, **10**, 1033–1038.
MacDonald, D., Herbert, K., Zhang, X.L., Polgruto, T. and Lu, P. (2001) *J. Mol. Biol.*, **306**, 1081–1098.
Manning, G.S. (1978) *Q. Rev. Biophys.*, **11**, 179–246.
McAteer, K., Aceves-Gaona, A., Michalczyk, R., Buchko, G.W., Isern, N.G., Silks, L.A., Miller, J.H. and Kennedy, M.A. (2004) *Biopolymers*, **75**, 497–511.
McCallum, S.A. and Pardi, A. (2003) *J. Mol. Biol.*, **326**, 1037–1050.
Meiler, J., Blomberg, N., Nilges, M. and Griesinger, C. (2000) *J. Biomol. NMR*, **16**, 245–252.
Mollova, E.T. and Pardi, A. (2000) *Curr. Opin. Struc. Biol.*, **10**, 298–302.
Padrta, P., Stefl, R., Kralik, L., Zidek, L. and Sklenar, V. (2002) *J. Biomol. NMR*, **24**, 1–14.
Prestegard, J.H., Al-Hashimi, H.M. and Tolman, J.R. (2000) *Q. Rev. Biophys.*, **33**, 371–424.
Reiter, N.J., Nikstad, L.J., Allmann, A.M., Johnson, R.J. and Butcher, S.E. (2003) *RNA*, **9**, 533–542.
Sashital, D.G., Allmann, A.M., Van Doren, S.R. and Butcher, S.E. (2003) *Biochemistry*, **42**, 1470–1477.
Sashital, D.G., Cornilescu, G. and Butcher, S.E. (2004) *Nat. Struct. Mol. Biol.*, **11**, 1237–1242.
Sashital, D.G., Cornilescu, G. and Butcher, S.E. (2005) *Nat. Struct. Mol. Biol.*, **12**, 99–99.
Sass, J., Cordier, F., Hoffmann, A., Cousin, A., Omichinski, J.G., Lowen, H. and Grzesiek, S. (1999) *J. Am. Chem. Soc.*, **121**, 2047–2055.
Sigel, R.K.O., Sashital, D.G., Abramovitz, D.L., Palmer, A.G., Butcher, S.E. and Pyle, A.M. (2004) *Nat. Struct. Mol. Biol.*, **11**, 187–192.
Staple, D.W. and Butcher, S.E. (2003) *Nucl. Acids Res.*, **31**, 4326–4331.
Stefl, R., Wu, H.H., Ravindranathan, S., Sklenar, V. and Feigon, J. (2004) *Proc. Natl. Acad. Sci. USA*, **101**, 1177–1182.
Stroobants, A., Lekkerkerker, H.N.W. and Odijk, T. (1986) *Macromolecules*, **19**, 2232–2238.
Tjandra, N. and Bax, A. (1997) *Science*, **278**, 1111–1114.
Tjandra, N., Omichinski, J.G., Gronenborn, A.M., Clore, G.M. and Bax, A. (1997) *Nat. Struct. Biol.*, **4**, 732–738.
Tolman, J.R. (2001) *Curr. Opin. Struc. Biol.*, **11**, 532–539.
Tolman, J.R., Flanagan, J.M., Kennedy, M.A. and Prestegard, J.H. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 9279–9283.
Valafar, H. and Prestegard, J.H. (2003) *Bioinformatics*, **19**, 1549–1555.
Vallurupalli, P. and Moore, P.B. (2003) *J. Mol. Biol.*, **325**, 843–856.
van Buuren, B.N.M., Schleucher, A., Wittmann, V., Griesinger, C., Schwalbe, H. and Wijmenga, S.S. (2004) *Angew. Chem. Int. Ed.*, **43**, 187–192.
Warren, J.J. and Moore, P.B. (2001) *J. Biomol. NMR*, **20**, 311–323.
Wu, B., Girard, F., van Buuren, B., Schleucher, J., Tessari, M. and Wijmenga, S. (2004) *Nucl. Acids Res.*, **32**, 3228–3239.
Wu, Z.G., Delaglio, F., Tjandra, N., Zhurkin, V.B. and Bax, A. (2003) *J. Biomol. NMR*, **26**, 297–315.
Zhang, Q., Throolin, R., Pitt, S.W., Serganov, A. and Al-Hashimi, H.M. (2003) *J. Am. Chem. Soc.*, **125**, 10530–10531.
Zweckstetter, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 3791–3792.
Zweckstetter, M. and Bax, A. (2001) *J. Biomol. NMR*, **20**, 365–377.
Zweckstetter, M. and Bax, A. (2002) *J. Biomol. NMR*, **23**, 127–137.
Zweckstetter, M., Hummer, G. and Bax, A. (2004) *Biophys. J.*, **86**, 3444–3460.