Residual dipolar couplings: are multiple independent alignments always possible?

Victoria A. Higman · Jonathan Boyd · Lorna J. Smith · Christina Redfield

Abstract RDCs for the 14 kDa protein hen egg-white lysozyme (HEWL) have been measured in eight different alignment media. The elongated shape and strongly positively charged surface of HEWL appear to limit the protein to four main alignment orientations. Furthermore, low levels of alignment and the protein’s interaction with some alignment media increases the experimental error. Together with heterogeneity across the alignment media arising from constraints on temperature, pH and ionic strength for some alignment media, these data are suitable for structure refinement, but not the extraction of dynamic parameters. For an analysis of protein dynamics the data must be obtained with very low errors in at least three or five independent alignment media (depending on the method used) and so far, such data have only been reported for three small 6–8 kDa proteins with identical folds: ubiquitin, GB1 and GB3. Our results suggest that HEWL is likely to be representative of many other medium to large sized proteins commonly studied by solution NMR. Comparisons with over 60 high-resolution crystal structures of HEWL reveal that the highest resolution structures are not necessarily always the best models for the protein structure in solution.

Keywords Residual dipolar couplings · Hen egg-white lysozyme · Dynamics · Protein alignment · Ubiquitin

Introduction

Since residual dipolar couplings (RDCs) were first introduced as a means for studying protein structure, several studies have suggested methods for extracting information on protein dynamics from RDCs (Meiler et al. 2001; Tolman 2002; Deschamps et al. 2005; Bouvignies et al. 2006; Yao et al. 2008; Salmon et al. 2009). These methods require data from at least three to five independent alignment media with very low errors in the RDCs and any structural models used. Although it is possible to change the alignment of a protein by changing the alignment medium (Ramirez and Bax 1998), obtaining such data is a challenge (Ruan and Tolman 2005; Tolman 2009) and has so far only been reported for three very similar proteins: ubiquitin (Peti et al. 2002; Lakomek et al. 2006), GB3 (Ulmer et al. 2003; Yao et al. 2008) and GB1 (Ruan et al. 2008).

RDCs measured for hen egg-white lysozyme (HEWL) in three alignment media have previously been used for structure refinement (Schwalbe et al. 2001) and to characterize side-chain orientations (Higman et al. 2004). Here we report RDCs for HEWL in a further five alignment media and examine the limits on obtaining a sufficient number of independent alignments with low errors. We also examine errors arising from structural models. HEWL is a good model system because it is a medium-sized (14 kDa) protein containing both α-helical and β-sheet secondary structure. In addition, it has been studied extensively as a model protein by NMR (Buck et al. 1995; Schwalbe et al. 2001), X-ray diffraction (Vaney et al. 1996;...
Walsh et al. 1998) and computational methods (Smith et al. 1995; Buck and Karplus 1999; Soares et al. 2004).

Materials and methods

Sample preparation

All HEWL sample concentrations were approximately 1.4 mM. Details of the sample preparation and data collection for HEWL in media 1–3 are described elsewhere (Boyd and Redfield 1999; Higman et al. 2004).

 Ether bicelles doped with La$^{3+}$ (medium 4)

A sample containing 3.8% w/v bicelles was prepared from a 10% stock solution of D13OPC, DHOPC, CTAB, 1,2-dimyristoyl-sn-glycero-3-phosphoethanolamine-N-diethylenetriamine-pentaacetic acid (DMPE-DTPA) and LaCl$_3$ in a 3:1:0.4:0.07:0.06 ratio. The pH of the solution was adjusted to 6.1 and the sample contained 5% D$_2$O.

 Cetylpyridinium bromide/NaBr/n-hexanol (medium 5)

Procedures for using CpBr/NaBr/n-hexanol mixtures as an alignment medium were taken from Barrientos et al. (2000). The final sample contained 5% (w/v) CpBr/n-hexanol (1:1.33 (w/w)), 25 mM NaBr, 10 mM acetate buffer (pH 3.8) and 5% D$_2$O.

 C12E6/n-hexanol (medium 6)

Liquid crystals from n-alkyl-poly(ethylene glycols) and n-alkyl alcohols were prepared as described by Rückert and Otting (2000). To make a 600 µl sample, 12 µl of n-hexanol were added to 300 µl of a 10% C12E6 stock solution. The sample tube was shaken. When the two compounds had mixed sufficiently the solution was inspected for a bluish tinge indicative of the formation of the liquid crystalline phase. The protein was dissolved in 300 µl 90% H$_2$O/10% D$_2$O (pH 3.8) and added to the C12E6/n-hexanol mixture. The sample tube was shaken and again inspected for the bluish tinge. Finally the sample was briefly centrifuged to remove bubbles prior to pipetting it into the NMR sample tube.

 Strained polyacrylamide gels (medium 7)

Polyacrylamide gels were made as described by Sass et al. (2000). 6, 8, 9 and 12% polyacrylamide gels were cast in 3.5 mm inner diameter tubing. Gels were washed, cut to a length of 3 cm and dried in an oven at 37°C overnight. Samples were made by placing a dried gel in a Shigemi NMR tube and then adding 475 µl of protein solution at pH 3.8 containing 10% D$_2$O. The plunger was used to compress the gel to about 2.1 cm.

 Pf1 phage (medium 8)

HEWL was dissolved in 0.1 M potassium phosphate buffer at pH 7.0 (90% H$_2$O/10% D$_2$O) followed by the addition of Pf1 bacteriophage from a 50 mg/ml stock solution (ASLA Biotech, Riga). This resulted in precipitation of the protein around the phage. NaCl was added from a 5 M stock solution whereupon the precipitate dissolved after shaking. The sample was then briefly centrifuged to remove bubbles. The final sample used for measurement of RDCs contained 10 mg/ml phage and 500 mM NaCl.

NMR

All NMR experiments were performed on home-built NMR spectrometers (fitted with magnets manufactured by Oxford Instruments, Eynsham, Oxfordshire) located in the Department of Biochemistry, University of Oxford. Experiments were carried out on spectrometers with 1H operating frequencies of 500, 600 and 750 MHz. The NMR data were processed and spectra were analysed using Felix 2.3 (Accelrys).

The 1H-15N residual dipolar couplings were measured from the 1J$_{NH}$ splitting appearing in the 15N dimension of an HSQC experiment which incorporated a S$^3$E pulse sequence element (Meissner et al. 1997) or were recorded using the IPAP procedure (Ottiger et al. 1998). Measurements were made in an isotropic solution as well as in the anisotropic media. The temperatures and parameters used for the different samples are given in the Electronic supplementary material. The data were zero-filled once in the direct dimension and a Gaussian window function was applied. In the indirect dimension the data were zero-filled to 8192 points and a 90° shifted sine-bell window function was applied. Peak picking was carried out using an in-house FORTRAN77 program. Peaks are identified as local maxima within a defined search region with intensity above a threshold value. The peak positions (chemical shifts) in the two dimensions are interpolated using a three-point quadratic fit.

Comparison of RDCs with X-ray structures

X-ray diffraction coordinates for 56 high resolution structures (2 Å or better) of HEWL were obtained from the Brookhaven Protein Data Bank (Berman et al. 2000). The monoclinic structures contain two molecules per unit cell and were therefore divided into A and B structures. The 1IEE and 3LZT structures contained alternative
conformations for parts of the backbone, thus two coordinate sets, designated $a$ and $b$, were created to include these alternate conformations. This resulted in a total of 68 sets of co-ordinates. Hydrogen atoms were added to these co-ordinates using the HBUILD feature in XPLOR (Brünger 1992) followed by energy minimisation (Boyd and Redfield 1998).

The principal components ($A_{xx}$, $A_{yy}$ and $A_{zz}$) and orientation ($\phi$, $\theta$ and $\psi$) of the molecular alignment tensor for each X-ray structure were fitted to minimize the $\chi^2$ between experimental and calculated residual dipolar couplings using an in-house program written in FORTRAN77. Residues with heteronuclear $^1$H–$^{15}$N NOE values of less than 0.7 were excluded from the analysis (residues 85, 100-104 and 129) (Buck et al. 1995). Residues 47 and 49 as well as 69–73 were also omitted; these regions are involved in crystal contacts in several structures. In addition, residues 88 and 118, which gave consistently high $\chi^2$ values, were omitted from the analysis.

The fitted values of the principal components and orientation of the alignment tensor were then used to calculate the residual dipolar couplings expected for the backbone $\mathrm{N–HN}$ groups. The $\mathrm{N–HN}$ bond length in the amide groups was assumed to be 1.02 Å.

$Q$-values were calculated during the process of fitting the structure to the experimental RDCs using the equation from Cornilescu et al. (1998):

$$Q = \left[ \sum_{i=1}^{N} \left( \frac{D_{i}^{\text{obs}} - D_{i}^{\text{calc}}}{N} \right)^2 \right]^{1/2} / D_{\text{rms}}$$

(2.2)

where $D_{i}^{\text{obs}}$ and $D_{i}^{\text{calc}}$ are the experimentally measured and predicted dipolar couplings, respectively. $D_{\text{rms}}$ is the root mean square value of all measured RDCs and the sum is over the N residues for which RDC measurements have been made.

Programs were written in FORTRAN77 to perform singular value decomposition of the RDC data matrix (Tolman 2002) and to calculate the angular RMSD of the $\mathrm{N–HN}$ vectors in crystal structures.

Results and discussion

$^{15}$N–$^1$H$^2$ RDC data have so far been published for HEWL in three alignment media. Two of these are positively charged, $7.5\%$ ester bicelles doped with CTAB (medium 1) (Boyd and Redfield 1999) and $5\%$ ether bicelles doped with CTAB (medium 2) (Higman et al. 2004), the third is either uncharged or very mildly negatively charged, $5\%$ ester bicelles (medium 3) (Boyd and Redfield 1999). Data were acquired in a further five media: $3.8\%$ ether bicelles doped with CTAB, the metal chelator DMPE-DTPA and La$^{3+}$ (positively charged, medium 4), CpBr/hexanol/NaBr (positively charged, medium 5), C12E6/hexanol (uncharged, medium 6), compressed $6\%$ polyacrylamide gel (uncharged, medium 7) and $10 \text{ mg/ml Pf1 bacteriophage}$ (negatively charged, medium 8). The alignment tensors were fitted using the 193L HEWL crystal structure (Vaney et al. 1996), the fitting parameters are given in Table 1. Axis system 1 in Fig. 1a shows the relative orientations of the eight alignment tensors alongside the 193L crystal structure rotated into each of these co-ordinate frames.

The protein adopts four main alignments (Fig. 1a, axis system 2) dominated by its shape and charge. The elongated shape of HEWL ($D_{||}/D_\perp = 1.3$) is the overriding factor in aligning the protein in both uncharged and positively charged media (media 1–7) and accounts for three of the four main alignments (Fig. 1a, grey, pale blue and pale green in axis system 2). Interestingly, the steric alignment induced by the compressed polyacrylamide gel differs significantly from that observed in the other uncharged and positively charged media, and is not simply related by a factor of $-\frac{1}{2}$ (Fig. 1a dark blue/pale green). Although the alignments shown in grey and pale blue in Fig. 1a appear distinct, they are still highly correlated (Table 2) and may not be strictly independent. Alignment in a strongly negatively charged medium, such as Pf1 phage, provides the fourth main alignment (Fig. 1a, purple/black). However, due to the highly positively charged surface of HEWL (pI $\approx 11$), the protein strongly interacts with the negatively charged phage and a salt concentration of 500 mM was required in order to obtain even a medium quality spectrum of HEWL (Figure S1). This is undesirable, since the additional salt may influence the structure and/or dynamics of the protein. This highlights a further problem when trying to modulate the alignment tensor with

| Medium | \(D_{||}\) | R$^2$ | $\beta^a$ | $\phi^a$ | $\psi^a$ | Q-value |
|--------|----------|------|--------|--------|--------|--------|
| Medium 1 | 12.09 | 0.15 | 136.5 | 156.0 | 53.4 | 0.1902 |
| Medium 2 | 11.38 | 0.18 | 132.8 | 160.3 | 91.1 | 0.1490 |
| Medium 3 | 16.03 | 0.31 | 126.5 | 135.2 | 89.0 | 0.1512 |
| Medium 4 | 14.04 | 0.48 | 121.5 | 162.9 | 75.2 | 0.1751 |
| Medium 5 | 6.88 | 0.13 | 143.1 | 169.0 | 48.0 | 0.2069 |
| Medium 6 | 10.84 | 0.39 | 122.4 | 158.8 | 93.3 | 0.1615 |
| Medium 7 | 4.86 | 0.39 | 54.8 | 112.0 | 112.7 | 0.2116 |
| Medium 8 | $-9.55$ | 0.17 | 40.0 | 127.7 | 57.0 | 0.2273 |

* See text and supplementary material for composition of different media

b Magnitude of the alignment tensor
c Rhombicity
d Euler angles
different media: certain media place significant constraints on the physical conditions (temperature, pH, ionic strength). SECONDA analysis (Hus and Brüscheville 2002) of the HEWL data across all eight media or any sub-group of seven media (Fig. 2) gives rise to a ratio between the 5th and 6th eigenvalue, $q$, of 1.1–3.4. Furthermore, the ratio between the 1st and 5th eigenvalues, an indicator for the signal to noise level, is significantly greater than 1 in all cases. Thus the error level across the different media is not sufficient for information about dynamics to be extracted.

Interestingly, the ten alignments of ubiquitin measured by Peti et al. (2002) and reported by Hus et al. (2003), show a larger number of different alignments with $p = 11.5$ (Fig. 1b). Ubiquitin appears to be more sensitive to small differences between the alignment media and the subtle interplay between steric and electrostatic forces. This is most likely because ubiquitin is a more spherical...
### Table 2  Correlation coefficients between different RDC datasets measured for HEWL

| Medium 1 | Medium 2 | Medium 3 | Medium 4 | Medium 5 | Medium 6 | Medium 7 | Medium 8 |
|----------|----------|----------|----------|----------|----------|----------|----------|
| Medium 1 | 0.97     | 0.91     | 0.81     | 0.94     | 0.85     | -0.61    | 0.46     |
| Medium 2 | 0.95     | 0.95     | 0.92     | 0.98     | 0.73     | 0.66     | -0.67    |
| Medium 3 | 0.92     | 0.78     | 0.94     | 0.73     | 0.78     | -0.44    | 0.84     |
| Medium 4 | 0.73     | 0.87     | 0.94     | 0.73     | 0.84     | -0.44    | 0.71     |
| Medium 5 | 0.66     | 0.70     | 0.66     | 0.57     | 0.70     | -0.52    |
| Medium 6 | 0.67     | 0.70     | 0.30     | 0.70     | 0.70     | -0.85    |
| Medium 7 | -0.85    | -0.85    | -0.85    | -0.85    | -0.85    | -0.85    | -0.85    |

Fig. 2 SECONDA analysis (Hus and Brüschweiler 2002; Hus et al. 2003) for the eight sets of RDCs measured for HEWL. RDCs were normalised prior to the analysis as described by Lakomek et al. (2008) The eigenvalues are plotted in decreasing size and the ratio between the 5th and 6th eigenvalue, \( \rho \), is indicated. The ratio between the 1st and 5th eigenvalues is provided in parentheses. The analysis included all eight data sets (a) or all data sets excluding medium 1 (b), medium 2 (c), medium 3 (d), medium 4 (e), medium 5 (f), medium 6 (g), medium 7 (h) or medium 8 (i).
protein (D$_\parallel$/D$_\perp$ = 1.17) (Tjandra et al. 1995) with a more varied surface charge distribution than HEWL (Fig. 3).

Further limitations to the use of RDCs to extract dynamic information come from the experimental errors, since insufficient accounting for errors can lead to misinterpretation of the RDC data. Experimental errors for ubiquitin and GB3 have been reported as lying between 0.1 and 0.5 Hz (Tolman et al. 1997; Peti et al. 2002; Clore and Schwieters 2004; Yao and Bax 2007) which is within the range of errors required for accurate extraction of dynamic parameters from RDCs (Yao et al. 2008). Although experimental errors for HEWL are within this range, the percentage error increases and Q-values clearly increase when the magnitude of the alignment tensor decreases (e.g. media 5 and 7, Table S1). In addition, the experimental error will increase in cases where the spectral quality is reduced, either due to increased protein size and longer $\tau_c$ values or due to excessive interactions between the protein and the alignment medium that adversely affect the spectral quality, e.g. positively charged HEWL with negatively charged Pf1 phage, or HEWL in 9–12% strained polyacrylamide gel (Figures S1, S2, Table S1 in the supplementary material) (Sass et al. 2000). Such factors will make it difficult to obtain the low errors required for the extraction of dynamic information from RDCs.

Finally, the quality of X-ray structures places a limit on the ability to extract dynamic information for those methods that use an explicit model (Bernado and Blackledge 2004a, b). It has been shown that both structural noise within a crystal structure (Zweckstetter and Bax 2002) and the placement of hydrogen atoms on crystal structures (Ulmer et al. 2003; Higman 2004) can affect the fit between the structure and a set of RDCs. A very large number of X-ray structures is available for HEWL (Higman et al. 2004). Interestingly, we find that the correlation between crystal structure resolution and Q-value is rather low ($R$ = 0.4, Fig. 4). All four crystal forms in which HEWL can be crystallized (tetragonal, triclinic, orthorhombic and monoclinic) have representative structures with both low and high Q-values and no crystal form represents a significantly better model for the solution structure of HEWL than the others. Particularly high Q-values, however, are observed for some monoclinic structures. The structural noise in the form of the angular RMSD of the NH vector (usually estimated to be on the order of $5^\circ$) (Bernado and Blackledge 2004a; Ruan et al. 2008) was determined for the 20 highest resolution structures (0.92-1.64 Å) and found to be 4.5° in secondary structure, but 7.6° when including loops. This rise is not surprising, since the 20 crystal structures represent several crystal forms. The largest variations between these are seen in the loops which are often involved in crystal contacts. This highlights the fact that protein crystal structures cannot be assumed to represent the average solution structure observed by NMR accurately. It also provides an explanation for the observation made by Bernado et al. that the

---

**Fig. 3** Surface representations of a HEWL (pdb code 193L (Vaney et al. 1996)) and b ubiquitin (pdb code 1UBQ (Vijay-Kumar et al. 1987), excluding the three flexible C-terminal residues) illustrating the overall shape of the proteins and their surface electrostatic potential (positive in blue, negative in red)

**Fig. 4** Plot of crystal structure resolution vs. average Q-value across all eight RDC data sets for HEWL. Panel B shows an expansion of part of panel A. The crystal form is indicated as tetragonal (red square), monoclinic (black circle), orthorhombic (blue triangle) or triclinic (gold diamond). A correlation coefficient ($R$) of only 0.4 is obtained between the average Q-values and X-ray structure resolution.
A possible strategy (Yao and Bax 2007), but may be less effective for use of positively charged media (Latham et al. 2008). While negative charge dominated the alignment and prevented the elongated structures as models, not only does structural noise at a level of 5–8° have to be considered, but also the accuracy with which the model reflects the solution structure probed by NMR. Although it has been possible to obtain data in four independent alignment media for HEWL, the homogeneity across these media and/or the errors for this medium-sized protein are only sufficient to allow the extraction of structural, and not dynamic, parameters from the RDCs.

Conclusions

We have shown firstly, using HEWL as an example, that the physical attributes of a protein may limit the number of independent alignments which it can adopt. Ubiquitin, GB1 and GB3 may be uniquely suited to undergoing modulations in their alignment tensors in alignment media with only small differences in their steric and electrostatic properties, while the limitations we have identified with HEWL could be typical of many larger or more complex proteins. Indeed, in a study on the alignment of an RNA construct, it was found that it was very difficult to modulate the alignment tensor, since the elongated structure and high negative charge dominated the alignment and prevented the use of positively charged media (Latham et al. 2008). Lanthanide binding tags (Wöhnert et al. 2003), which can provide alternative modes of alignment, may provide the best solution to this problem, but involve additional molecular biology steps. Conservative mutagenesis as a means to alignment tensor modulation provides a further possibility (Yao and Bax 2007), but may be less effective with proteins which are highly charged or significantly elongated. Secondly, the size of the protein and the properties of the alignment media used will have a significant impact on the experimental errors and thus on the ability to extract dynamic information. Finally, when using X-ray structures as models, not only does structural noise at a level of 5–8° have to be considered, but also the accuracy with which the model reflects the solution structure probed by NMR. Although it has been possible to obtain data in four independent alignment media for HEWL, the homogeneity across these media and/or the errors for this medium-sized protein are only sufficient to allow the extraction of structural, and not dynamic, parameters from the RDCs.

Acknowledgments

V. A. H. was funded by the BBSRC and St. Peter’s College, Oxford. C. R. was funded by the Wellcome Trust (Grant number 079440).

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

Barrientos LG, Dolan C, Gronenborn AM (2000) Characterization of surfactant liquid crystal phases suitable for molecular alignment and measurement of dipolar couplings. J Biomol NMR 16: 329–337
Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242
Bernadó P, Blackledge M (2004a) Anisotropic small amplitude peptide plane dynamics in proteins from residual dipolar couplings. J Am Chem Soc 126:4907–4920
Bernadó P, Blackledge M (2004b) Local dynamic amplitude on the protein backbone from dipolar couplings: toward the elucidation of slower motions in biomolecules. J Am Chem Soc 126: 7760–7761
Bouvignies G, Markwick P, Brüschweiler R, Blackledge M (2006) Simultaneous determination of protein backbone structure and dynamics from residual dipolar couplings. J Am Chem Soc 128:15100–15101
Boyd J, Redfield C (1998) Defining the orientation of the 15N shielding tensor using 15N NMR relaxation data for a protein in solution. J Am Chem Soc 120:9692–9693
Boyd J, Redfield C (1999) Characterization of 15N chemical shift anisotropy from orientation-dependent changes to 15N chemical shifts in dilute bicelle solutions. J Am Chem Soc 121:7441–7442
Brünger AT (1992) XPLOR version 3.1: a system for X-ray crystallography and NMR. Yale University Press, New Haven, CT
Buck M, Karplus M (1999) Internal and overall peptide group motion in proteins: molecular dynamics simulations for lysozyme compared with results from X-ray and NMR spectroscopy. J Am Chem Soc 121:9645–9658
Buck M, Boyd J, Redfield C, MacKenzie DA, Jeenes DJ, Archer DB, Dobson CM (1995) Structural determinants of protein dynamics: analysis of 15N NMR relaxation measurements for main-chain and side-chain nuclei of hen egg white lysozyme. Biochemistry 34:4041–4055
Clore GM, Schwiters CD (2004) Amplitudes of protein backbone dynamics and correlated motions in a small alpha/beta protein: correspondence of dipolar coupling and heteronuclear relaxation measurements. Biochemistry 43:10678–10691
Cornilescu G, Marquardt JL, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837
Deschamps M, Campbell ID, Boyd J (2005) Residual dipolar couplings and some specific models for motional averaging. J Magn Res 172:118–132
Higman VA (2004) The use of bicelles and other ordered media to study protein structure and dynamics. DPhil Thesis, University of Oxford
Higman VA, Boyd J, Smith LJ, Redfield C (2004) Asparagine and glutamine side-chain conformations in solution and crystal: a comparison for hen egg-white lysozyme using residual dipolar couplings. J Biomol NMR 30:327–346
Hus J-C, Brüschweiler R (2002) Principal component method for assessing structural heterogeneity across multiple alignment media. J Biomol NMR 24:123–132
...