Ms # JA062715T /Revised

Supporting Information

An efficient computational method for predicting rotational diffusion tensors of globular proteins using an ellipsoid representation

Yaroslav E Ryabov, Charles Geraghty, Amitabh Varshney, and David Fushman

List of the 841 PDB files used in this study:

4znf, 4tgf, 4fgf, 5znf, 1zwf, 1xif, 1rcf, 1phf, 1prv, 1psf, 1psv, 1ptf, 1qnf, 1whf, 1thv, 1tif, 1tiv, 1tyv, 1kpf, 1hmf, 1hnf, 1haf, 1hcv, 1hc6, 1hev, 1iyv, 1nif, 1nef, 1oav, 1lrv, 1mof, 1mbf, 1bif, 1bkf, 1ba6, 1bc6, 1brf, 1cnv, 1cof, 1ctf, 1aj6, 1ak6, 1amf, 1anf, 1anv, 1aaf, 1abv, 1acf, 1ad6, 1az6, 1apf, 1arf, 1arv, 1a06, 1a26, 1gof, 1gcf, 1div, 1dmf, 1ddf, 1dff, 1drv, 1eaf, 1egf, 1erv, 2stv, 2hgf, 2omf, 2liv, 1zqw, 1ytw, 1rlw, 1rmg, 1sig, 1pog, 1vig, 1vjw, 1thg, 1tig, 1tpg, 1tsg, 1uag, 1udg, 1utg, 1jkw, 1jdw, 1jsg, 1hbg, 1huw, 1iow, 1ncg, 1new, 1ldg, 1bbg, 1ah7, 1aiw, 1acw, 1af7, 1ag7, 1aww, 1fkw, 1fow, 1edg, 1exg, 1erg, 1erw, 2eng, 6mht, 1zfd, 1xat, 1rkd, 1rmd, 1ret, 1ryt, 1rpt, 1smd, 1szt, 1std, 1pht, 1pot, 1pft, 1ppt, 1pud, 1put, 1vid, 1vnd, 1vsd, 1vvd, 1wit, 1wkt, 1tit, 1tnt, 1tbd, 1tud, 1tvt, 1uxd, 1jdd, 1jud, 1kid, 1krt, 1kst, 1hkt, 1hmt, 1hcd, 1ift, 1igd, 1nct, 1np4, 1ojt, 1opd, 1lit, 1lbd, 1lst, 1mit, 1mat, 1mbd, 1med, 1mrt, 1mut, 1bm4, 1bdd, 1bed, 1bfd, 1bg4, 1brt, 1bw4, 1chd, 1cid, 1cpt, 1ctt, 1agt, 1ag4, 1aqt, 1art, 1atd, 1av4, 1awd, 1a9t, 1a5t, 1fid, 1fit, 1fnd, 1fct, 1fxd, 1gnd, 1gpt, 1gw4, 1dmd, 1dad, 1ddt, 1dst, 1eit, 1elt, 1ecd, 1edt, 1eft, 1erd, 1eut, 2pgd, 2ptd, 2tct, 2knt, 2had, 2hft, 2lbd, 2mrt, 2cmd, 2crd, 2abd, 2end, 3ukd, 3lzt, 3bct, 5eau, 1rie, 1rtu, 1sfe, 1stu, 1pne, 1pou, 1pbe, 1pce, 1vie, 1vpu, 1tle, 1tfe, 1une, 1uae, 1kte, 1hoe, 1neu, 1nre, 1lbu, 1mwe, 1ble, 1blu, 1bme, 1cc5, 1cfe, 1c25, 1aie, 1ak5, 1alu, 1ac5, 1af5, 1awe, 1a8e, 1fre, 1fsu, 1gne, 1dpe, 2uce, 2leu, 2cae, 3pte, 8dfr, 4i1b, 4mt2, 5icb, 1xnb, 1xer, 1yer, 1yub, 1rcb, 1sp2, 1svb, 1svr, 1phr, 1pi2, 1pmr, 1pdr, 1prb, 1prr, 1ps2, 1vib, 1wab, 1wer, 1tib, 1tfr, 1trb, 1udb, 1u9b, 1jer, 1knb, 1ksr, 1hlb, 1hnr, 1har, 1iab, 1nar, 1neb, 1ner, 1ngr, 1nxb, 1ntr, 1obr, 1oyb, 1opr, 1ovb, 1lab, 1ldr, 1leb, 1lfb, 1mxb, 1bhb, 1bmb, 1bnb, 1bor, 1bdb, 1bf2, 1byb, 1cdb, 1cfb, 1cfr, 1c52, 1air, 1ai2, 1aj2, 1akr, 1am2, 1aab, 1ab2, 1adr, 1ad2, 1ag2, 1aqb, 1aqr, 1arb, 1a32, 1a62, 1fmb, 1far, 1fbr, 1fdr, 1fsb, 1gcb, 1gpr, 1gur, 1dhr, 1dmr, 1dar, 1dyr, 1egr, 1eur, 2rn2, 2reb, 2sob, 2pnb, 2por, 2pf2, 2lhb, 2mbr, 2adr, 2apr, 2a0b, 2fmr, 2fcr, 2fxb, 2gar, 2gpr, 2dkb, 2dtb, 2dtr, 3seb, 3vub, 3kar, 3cyr, 3dfr, 1ycc, 1ygs, 1rhs, 1ris, 1rec, 1rfs, 1rgs, 1rss, 1sis, 1sqc, 1plc, 1pls, 1pmc, 1poc, 1pdc, 1pfc, 1pgs, 1pyc, 1puc, 1vjs, 1vnc, 1vcc, 1tis, 1tn3, 1tfs, 1tys, 1jpc, 1kjs, 1hcc, 1hfc, 1ifc, 1igs, 1nls, 1ncs, 1ois, 1onc, 1opc, 1orc, 1lis, 1mjc, 1mmc, 1mdc, 1msc, 1bds, 1bec, 1bgc, 1bp3, 1chc, 1clc, 1cbs, 1cxc, 1aj3, 1alc, 1al3, 1am3, 1ans, 1aac, 1aa3, 1ads, 1ass, 1avc, 1a1s, 1a3c, 1fas, 1fds, 1fyc, 1fts, 1fus, 1ghc, 1gks, 1gpc, 1dhs, 1dlc, 1dec, 1drs, 1ehs, 1esc, 2sns, 2sn3, 2sas, 2pkc, 2plc, 2wbc, 2hts, 2cy3, 2ctc, 3tss, 3b5c, 3c2c, 3grs, 8abp, 4ptp, 1rip, 1rgp, 1sap, 1sbp, 1sfp, 1php, 1pkp, 1qdp, 1qyp, 1vtp, 1tap, 1ulp, 1hip, 1hyp, 1htp, 1ihp, 1inp, 1nhp, 1nfp, 1opp, 1mpp, 1mrp, 1mup, 1bhp, 1bmp, 1bnp, 1bdp, 1bgp, 1br0, 1akp, 1ak0, 1amp, 1aop, 1acp, 1ac0, 1aep, 1afp, 1ap0, 1at0, 1aw0, 1a8p, 1a0p, 1flp, 1fgp, 1gvp, 1g3p, 1dtp, 1enp, 1erp, 2prp, 2hsp, 2mlp, 2bsp, 2cbp, 2cyp, 4hb1, 5p21, 6taa, 6fd1, 7rsa, 1zaq, 1yua, 1rp1, 1sh1, 1sra, 1plq, 1paa, 1pba, 1pca, 1pda, 1pea, 1ptq, 1wba, 1wea, 1tca, 1uba, 1ubq, 1hma, 1ha1, 1iba, 1ica, 1noa, 1oxa, 1osa, 1lla, 1lba, 1lxa, 1mla, 1mba, 1mb1, 1mfa, 1bia, 1ba1, 1bea, 1be1, 1bp1, 1bv1, 1cpq, 1ah1, 1ak1, 1alq, 1aoa, 1aba, 1apa, 1apq, 1aua, 1a8q, 1a91, 1a6q, 1fna, 1fua, 1gca, 1gsa, 1ema, 2pia, 2ts1, 2u1a, 2baa, 2cba, 2fha, 2gf1, 2gsq, 3cla, 8acn, 6rxn, 1zin, 1ztn, 1ppn, 1vin, 1ten, 1tpn, 1jon, 1han, 1hxn, 1nin, 1nfn, 1ntn, 1omn, 1mfn, 1myn, 1btn, 1csn, 1ctn, 1akn, 1adn, 1axn, 1gln, 1gen, 1din, 1emn, 2fdn, 2ebn, 3prn, 3ctn, 1zfo, 1zto, 1xjo, 1roo, 1rpo, 1rro, 1sro, 1pdo, 1pfo, 1who, 1klo, 1ido, 1lfo, 1boo, 1bco, 1bdo, 1beo, 1bso, 1coo, 1ceo, 1cgo, 1cpo, 1aho, 1ako, 1azo, 1a6o, 1dro, 5nul, 6cel, 1xbl, 1sxl, 1vil, 1tml, 1tal, 1tul, 1kal, 1hcl, 1iml, 1nkl, 1lml, 1lcl, 1ldl, 1lxl, 1lvl, 1mil, 1mml, 1mdl, 1bol, 1bal, 1btl, 1cjl, 1cvl, 1ail, 1aol, 1ayl, 1fbl, 1gal, 1dsl, 1eal, 1ecl, 1esl, 119l, 153l, 2sil, 2sxl, 2ptl, 2abl, 2ezl, 2erl, 3nul, 1xsm, 1rom, 1slm, 1psm, 1iam, 1ncm, 1lam, 1mzm, 1bam, 1cem, 1ahm, 1aim, 1amm, 1fdm, 1ezm, 2hvm, 2gdm, 1skz, 1ucz, 1hcz, 1idz, 1nsj, 1maz,

1mrj, 1brz, 1buz, 1czj, 1ctj, 1ajj, 1akz, 1abz, 1adz, 1afj, 1axj, 1ayj, 1apj, 1awj, 1a3z, 1a7j, 1fsz, 1ftz, 1grj, 4pfk, 1phk, 1pbk, 1tlk, 1tsk, 1urk, 1idk, 1npk, 1mek, 1msk, 1bak, 1bgk, 1cuk, 1aak, 1auk, 1a3k, 1fmk, 2sak, 2ilk, 2abk, 2ack, 3lck, 1pnh, 1pch, 1pex, 1vhh, 1vgh, 1toh, 1tdx, 1uch, 1kuh, 1hfh, 1hp8, 1ixh, 1nox, 1ndh, 1mgx, 1bnh, 1bnx, 1cd8, 1cex, 1cfh, 1cyx, 1csh, 1amx, 1an8, 1adx, 1af8, 1axh, 1ayx, 1ap8, 1ash, 1a1x, 1a68, 1ffh, 1gox, 1dtx, 2pth, 2hfh, 2ayh, 2ech, 2ezh, 3il8, 3cox, 1ra9, 1rsy, 1r69, 1pii, 1pmi, 1pmy, 1pty, 1qli, 1vii, 1v39, 1whi, 1tmy, 1tfi, 1try, 1uby, 1jli, 1hfi, 1hpi, 1hqi, 1opy, 1lki, 1lci, 1lsi, 1mai, 1msi, 1bo9, 1bb9, 1ciy, 1cby, 1cdy, 1cpy, 1ah9, 1aly, 1amy, 1aoy, 1af9, 1agi, 1aty, 1a0i, 1gky, 1gai, 1doi, 1dxy, 1eny, 1ery, 2phy, 2pii, 2tgi, 2acy, 2dri, 3chy.

Supporting Table 1. Comparison of the predictions of the inertia-equivalent ellipsoid model[1] with the experimental data for several proteins (set A). f is the scaling factor required to match the experimental and calculated τ_c values: $\tau_c^{exp} = f^3 \tau_c^{calc}$.

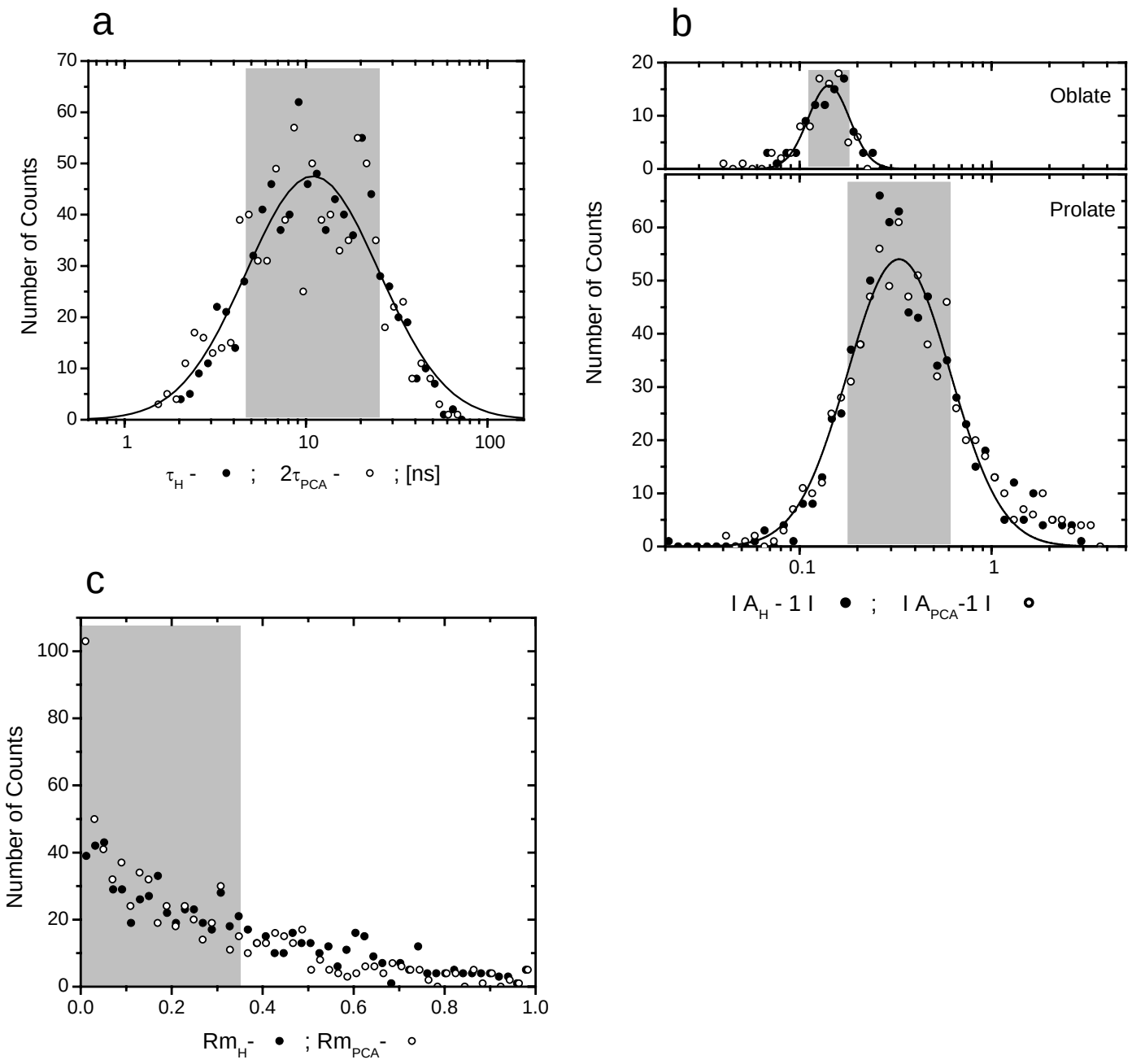
			Inertia-equivalent Ellipsoid		
		Experim.	f	$f = 2.6$	
Protein	PDB	τ_c [ns]	matching	τ_c Calc. [ns]	% Diff.
Malate synthase G	1y8b	55.3	2.514	61.18	11
Human serum albumin	1a06	41.0	2.255	62.85	53
Maltose binding protein	1ezp	28.6	2.370	37.77	32
β-lactoglobulin A (dimer)	1bsy	23.2	2.459	27.44	18
Δ^5-3 ketosteroid isomerase	1buq	18.0	2.474	20.90	16
Leukemia inh. Factor	1lki	14.9	2.710	13.16	-12
Trypsin	2blv	14.8	2.693	13.32	-10
Yellow fluorescent protein	2yfp	14.8	2.546	15.76	6
Green fluorescent protein	1w7s	14.2	2.505	15.87	12
Carbonic anhydrase	2cab	14.0	2.425	17.26	23
HIV-1 protease	1bvg	13.0	2.592	13.12	1
Savinase	1svn	12.4	2.489	14.13	14
Interleukin-1β	6i1b	12.4	2.737	10.63	-14
Ribonuclease H	2rn2	11.7	2.618	11.47	-2
Cytochrome C2	1c2n	10.4	2.891	7.57	-27
β-lactoglobulin A (mono)	1bsy	9.7	2.526	10.59	9
Apomyoglobin	1bvc	9.5	2.438	11.52	21
Lysozyme	1hwa	8.3	2.578	8.52	3
Barstar C40/83A	1bta	7.4	2.976	4.94	-33
Eglin c	1egl	6.2	2.646	5.88	-5
Cytochrome b_5	1wdb	6.1	2.772	5.04	-17
Calbindin-D9k+Ca^{2+}	2bca	5.1	2.719	4.46	-13
Ubiquitin	1ubq	5.0	2.759	4.18	-16
Calbindin-D9k	1clb	4.9	2.631	4.73	-3
BPTI	1pit	4.4	2.784	3.58	-19
Protein G	1igd	3.7	2.434	4.51	22
Xfin-zinc finger DBD	1znf	2.4	3.151	1.35	-44
Mean absolute value			2.6 (0.2)		17 %

Supporting Table 2. Comparison of the predictions of the FAST-HYDRONMR program[2] (version fast-hydronmr7c2lnx.exe) with the experimental data for proteins of set A.

Protein	PDB	Experim.	FAST-HydroNMR AER=3.2 Å	
		τ_c [ns]	τ_c Calc. [ns]	% Diff.
Malate synthase G	1y8b	55.3	68.2	23
Human serum albumin	1a06	41.0	66.1	61
Maltose binding protein	1ezp	28.6	43.0	50
β-lactoglobulin A (dimer)	1bsy	23.2	27.6	19
Δ^5-3 ketosteroid isomerase	1buq	18.0	22.4	24
Leukemia inh. Factor	1lki	14.9	13.1	-12
Trypsin	2blv	14.8	13.4	-9
Yellow fluorescent protein	2yfp	14.8	14.6	-1
Green fluorescent protein	1w7s	14.2	16.0	13
Carbonic anhydrase	2cab	14.0	17.1	22
HIV-1 protease	1bvg	13.0	14.1	8
Savinase	1svn	12.4	13.6	10
Interleukin-1β	6i1b	12.4	12.3	-1
Ribonucleaase H	2rn2	11.7	12.2	4
Cytochrome C2	1c2n	10.4	7.44	-28
β-lactoglobulin A (mono)	1bsy	9.7	11.5	19
Apomyoglobin	1bvc	9.5	11.0	16
Lysozyme	1hwa	8.3	9.6	16
Barstar C40/83A	1bta	7.4	5.8	-22
Eglin c	1egl	6.2	6.0	-3
Cytochrome b_s	1wdb	6.1	6.4	5
Calbindin-D9k+Ca^{2+}	2bca	5.1	4.8	-6
Ubiquitin	1ubq	5.0	4.6	-8
Calbindin-D9k	1clb	4.9	4.8	-2
BPTI	1pit	4.4	4.5	2
Protein G	1igd	3.7	4.6	24
Xfin-zinc finger DBD	1znf	2.4	1.7	-29
Mean absolute value				16 %

Supporting Table 3. Comparison of the PCA-based predictions for the overall correlation time τ_c for the 12 NMR-derived structures from protein set A (Table 1) with and without hydrogen atoms.

Protein	PDB code	Exp.	HLT = 2.8 Å				% difference between τ_{calc} with and without hydrogens
			With hydrogens		Without hydrogens		
		τ_c [ns]	τ_{calc} [ns]	Diff. from exp. %	τ_{calc} [ns]	Diff. from exp. %	
Malate synthase G	1y8b	55.3	45.7	-17	45.1	-18	-2
Maltose binding protein	1ezp	28.6	28.2	-1	27.2	-5	-4
Δ^5-3 ketosteroid isomerase	1buq	18.0	16.7	-7	16.4	-9	-2
HIV-1 protease	1bvg	13.0	12.9	-1	12.4	-5	-5
Interleukin-1β	6i1b	12.4	11.1	-10	10.8	-13	-3
Cytochrome C2	1c2n	10.4	7.7	-26	7.5	-28	-4
Lysozyme	1hwa	8.3	9.2	11	8.9	7	-3
Barstar C40/83A	1bta	7.4	6.0	-19	5.9	-20	-2
Calbindin-D9k+Ca^{2+}	2bca	5.1	5.0	-2	5.1	0	2
Calbindin-D9k	1clb	4.9	5.4	10	5.3	8	-2
BPTI	1pit	4.4	4.6	5	4.4	0	-4
Xfin-zinc finger DBD	1znf	2.4	2.2	-8	2.1	-12	-5
Mean absolute value				10 %		10 %	3.2 %

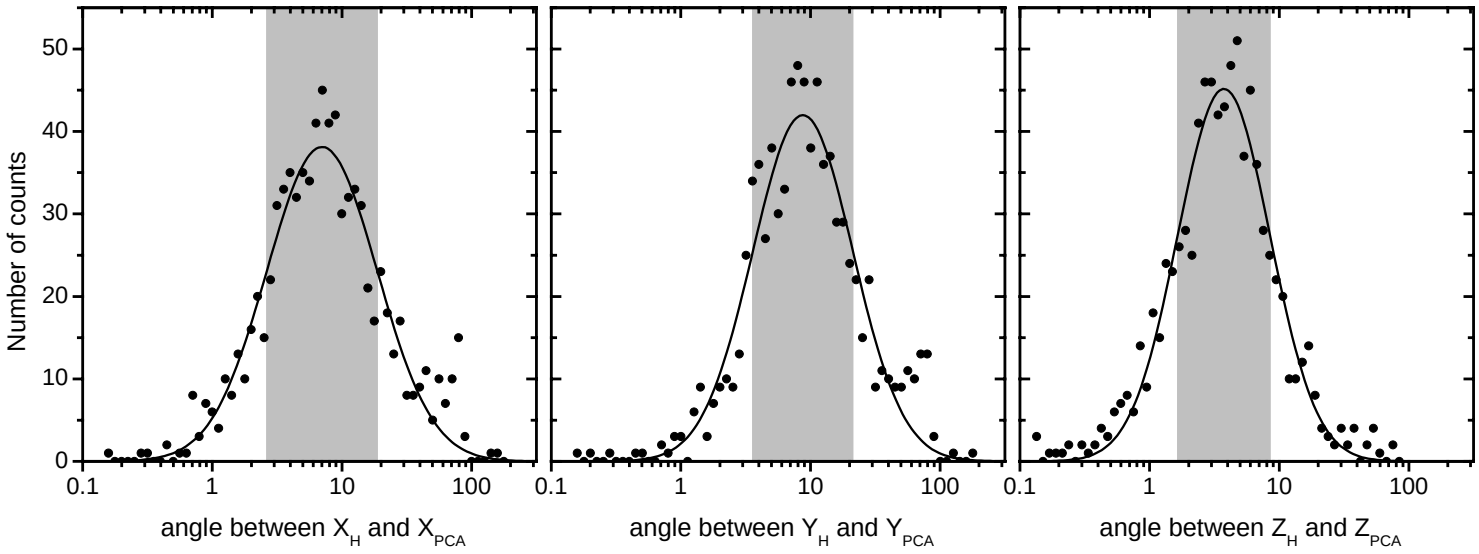


Supporting Figure 1. Statistical properties of the rotational diffusion tensors for a representative set of 841 monomeric proteins (set B). These data were calculated using HYDRONMR with AER=3.2 Å (solid symbols) and the PCA based method for HLT =0.0 Å (open symbols), also indicated by the subscripts "H" and "PCA", respectively.

(a) The distribution of the overall rotational correlation times, τ_H and $2\tau_{PCA}$. Both sets of the correlation time values obey the same log-normal law, Eq.S1 (below), represented by a solid line. The maxima of these distributions correspond to 10.9±0.4 ns; the other parameters are presented in Supporting Table 4. The shaded area indicates the confidence interval, from $-\sigma$ to σ around the maximum of the log-normal distribution, which contains 68 % of all data; in this case from 4.7 to 25.4 ns. This plot was obtained by distributing the data among bins of constant width of $d\underline{\ln}x = 0.05$ (where $x=\tau_H$ or $2\tau_{PCA}$).

(b) Statistical distributions of the anisotropies (Eq.9, main text) of the oblate (top panel) and prolate diffusion tensors. Both $|A_H - 1|$ and $|A_{PCA} - 1|$ obey the same log-normal law (Eq.S1), represented by a solid line. Parameters of these distributions are presented in Supporting Table 4. The maxima of these distributions correspond to $A_0 \approx 0.858 \pm 0.003$ for oblate and $A_0 \approx 1.330 \pm 0.006$ for prolate tensors. The shaded areas correspond to the confidence intervals from $-\sigma$ to σ, containing 68 % of all data; these intervals cover anisotropies from 0.819 to 0.889 for oblate and from 1.18 to 1.61 for prolate diffusion tensors, respectively. This plot was obtained by distributing the data among bins of constant width of $d\underline{\ln}x = 0.05$ (where $x = |A_H - 1|$ or $|A_{PCA} - 1|$).

(c) Statistical distributions of the rhombicities *Rm* (Eq.9) of the diffusion tensors. The shaded area corresponds to the range $0 < Rm < 0.35$ containing 68 % of the data.



Supporting Figure 2. Distribution of the tilt angles between the principal axes for the diffusion tensors predicted for the 841 proteins of set B using HYDRONMR with AER=3.2 Å (subscript "H") and the PCA-based method with HLT=0.0 Å (subscript "PCA"). All angles are in degrees. All data obey log-normal distributions (Eq.S1) shown by solid lines in all panels. Parameters of these distributions are presented in Supporting Table 4. The maxima of these distributions are at $X_0 = 7.0^o \pm 0.3^o$, $Y_0 = 8.7^o \pm 0.4^o$, and $Z_0 = 3.7^o \pm 0.1^o$. The shaded areas in all panels indicate confidence intervals from $-\sigma$ to σ, containing 68 % of all data: these intervals are from 2.6^o to 19^o for the angles between the x-axes, from 3.6^o to 21^o between the y-axes, and from 1.6^o to 8.5^o for the angles between the z-axes. These plots were obtained by distributing the data among bins of constant width of $d\underline{\ln}x = 0.05$ (where x are the corresponding angles). The log-normal distribution for the tilt angles between PAFs of the tensors is surprising, as these angles presumably reflect differences between the methods plus random errors in the eigenvectors.

The data presented in Supporting Figures 1 and 2 where analyzed using a log-normal distribution

$$dn = \frac{n_0}{\sigma\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\left(\frac{\ln x - \ln x_0}{\sigma} \right)^2 \right\} d\ln x \qquad \text{(S1)}$$

Supporting Table 4. Parameters of the log-normal distribution of the overall tumbling time (τ_H, τ_{PCA}), the anisotropy (A_H, A_{PCA}) of the rotational diffusion tensors calculated for the representative set of 841 proteins (set B), and the tilt angles (in degrees) between the corresponding principal axes of the diffusion tensors calculated using HYDRONMR with AER=3.2 Å (subscript "H") and the PCA-based method with HLT=0.0 Å (subscript "PCA") for the set of 841 proteins (set B). The actual data and the corresponding log-normal distribution functions are shown in the Supporting Figures 1 and 2.

x	n_0	σ	x_0
τ_H, $2\,\tau_{PCA}$ [ns]	43.7±1.7	0.84±0.035	10.9±0.4 [ns]
1-A_H, 1-A_{PCA} (oblate)	4.2±0.3	0.24±0.023	0.142±0.003
A_H-1, A_{PCA}-1 (prolate)	36.1±0.9	0.61±0.012	0.330±0.006
angle between X_H and X_{PCA}	40.6±1.2	0.98±0.035	7.0±0.3^o
angle between Y_H and Y_{PCA}	40.6±1.3	0.89±0.035	8.7±0.4^o
angle between Z_H and Z_{PCA}	40.4±1.0	0.82±0.023	3.7±0.1^o

References

(1) Taylor, W. R.; Thornton, J. M.; Turnell, W. G. *Journal of Molecular Graphics* **1983**, *1*, 30-38.
(2) Ortega, A.; Garcia de la Torre, J. *J Am Chem Soc* **2005**, *127*, 12764-12765.