J|A|C|S
COMMUNICATIONS

Published on Web 07/08/2003

Correlation between ^{2}H NMR Side-Chain Order Parameters and Sequence Conservation in Globular Proteins

Anthony Mittermaier,[†] Alan R. Davidson,[†,‡] and Lewis E. Kay*[,†,‡,§]

Protein Engineering Network Centres of Excellence and the Departments of Biochemistry, Chemistry, and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8

Received February 25, 2003; E-mail: kay@pound.med.utoronto.ca

NMR measurements of methyl ^{2}H relaxation rates provide site-specific information about the magnitude of side-chain motions in fractionally deuterated proteins. Experiments typically focus on CH_2D moieties and are analyzed to yield an order parameter, S_{axis}, describing the amplitude of ns to ps timescale motions for each methyl group. An S^2_{axis} value of 1 corresponds to complete rigidity and 0 to isotropic averaging of the methyl axis in the molecular frame. The response of S^2_{axis} values to changes in temperature and addition of binding partners has been related to protein stability[1,2] and the affinity and specificity of protein−ligand interactions.[3,4] It is therefore of considerable interest to identify structural correlates with S^2_{axis} values since determinants of side-chain motion likely play an important role in modulating protein function. Although hydrophobic core packing is widely cited as an important consideration, an analysis of S^2_{axis} values for a database of eight proteins did not show strong correlations with either methyl solvent accessibility or packing density.[5] Here we report that S^2_{axis} values for the Fyn SH3 domain, as well as a number of other proteins, show a significantly stronger dependence on residue conservation in sequence alignments of homologous proteins than on measures of solvent exposure calculated from the molecular structures. We suggest that factors restricting the amplitude of side-chain dynamics include evolutionarily conserved structural motifs, as well as, to a small extent, the degree of side-chain burial.

Recently developed experiments for measuring the decay rates of five deuterium spin operators[6] were performed on a sample of the SH3 domain from the Fyn tyrosine kinase. Data were subsequently analyzed to yield S^2_{axis} values. To compare the dynamics of a given residue with the extent to which it is favored in a previously published sequence alignment of SH3 domains,[7] we have defined the degree of preference at any position i to be

$$\Pi_i = \ln(n_{i,X}/N_{i,X}) \qquad (1)$$

where X is the residue occurring at position i in the Fyn sequence, $n_{i,X}$ is the number of sequences in the alignment with residue X at position i, and $N_{i,X}$ is the number of sequences that would be expected to have residue X at position i if the distribution were completely random, i.e., if the probability of residue X occurring at any nongap position in any sequence were equal to the total number of residue X in the alignment divided by the total number of nongap positions in all sequences. Positive preference values therefore indicate an enrichment for residue X at position i, while negative values indicate a deficit, relative to a random distribution.

The intrinsic reorientational freedom of methyl groups increases with increasing separation from the backbone. To allow comparisons between different methyl types, the means (μ_{meth}) and standard deviations (σ_{meth}) of S^2_{axis} values, calculated for each methyl type from a database of eight proteins,[5] were used to compute normalized methyl axis order parameters:

$$S^2_{norm} = (S^2_{axis} - \mu_{meth})/\sigma_{meth} \qquad (2)$$

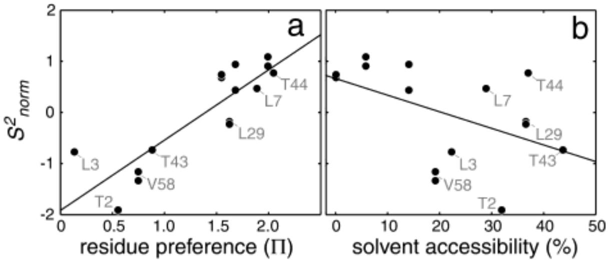

Figure 1. Plots of normalized side-chain methyl axis order parameters (S^2_{norm}, defined by eq 2) versus residue preference (Π, defined by eq 1) (a) and solvent accessibility (b) for the SH3 domain from the Fyn tyrosine kinase.

Values of Π and S^2_{norm} for the Fyn SH3 domain are plotted in Figure 1a. There is a clear tendency for residues that are conserved at their respective positions (high Π) to be less mobile than average (high S^2_{norm}). The correlation coefficient, $r_\Pi = 0.86$, has a statistical significance of $p = 5 \times 10^{-5}$, corresponding to the probability that the observed correlation could be due to chance. In contrast, a comparison of S^2_{norm} and per-residue solvent accessibility values, plotted in Figure 1b, yields a significantly weaker correlation coefficient, $r_{sol} = -0.49$, and statistical significance, $p = 0.07$.

To investigate the generality of this finding, we have examined six additional proteins for which ^{2}H relaxation and sequence alignment data are available. Individual statistical parameters are listed by protein in Table 1. Five of the seven molecules show significant ($p < 0.05$) correlations between Π and S^2_{norm}, and correlations are stronger than for solvent accessibility in all examples except the SAP SH2 domain, which has unusual ligand binding properties compared to other SH2 domains.[8,9] Combining the data to form a single 179-entry sample yields a Pearson linear correlation coefficient for Π and S^2_{norm} values, $r_\Pi = 0.41$, $p_\Pi = 1.5 \times 10^{-8}$, which is much greater than that obtained when S^2_{norm} and solvent accessibility are compared, $r_{sol} = -0.23$, $p_{sol} = 2 \times 10^{-3}$. Fisher z transformations[10] were applied to compare r_Π and r_{sol}, returning a low probability ($p_{diff} = 4\%$) that the true correlation coefficients are equal in magnitude (opposite in sign) and that the apparent difference is due to chance. When data are omitted for several very uncommon residues in the SAP SH2 domain (L25, V37, V40, L46) that are rigid, opposing the overall trend, parameters $r_\Pi = 0.53$, $p_\Pi = 4.6 \times 10^{-14}$, $r_{sol} = -0.22$, $p_{sol} = 3 \times 10^{-3}$, and $p_{diff} = 0.1\%$ are obtained for the combined sample.

The relationship between S^2_{norm} and evolutionary conservation is likely due to the presence of fold-specific structural features that

[†] Department of Biochemistry.
[‡] Department of Medical Genetics.
[§] Department of Chemistry.

10.1021/ja034856q CCC: $25.00 © 2003 American Chemical Society

Table 1. Correlation of Side-Chain Flexibility (S^2_{norm}) with Residue Preference (Π) and Solvent Accessibility

	residue preference		solvent accessibility		
protein[a]	r_Π[b]	p[b]	r_{sol}[c]	p[c]	sample[d]
Fyn SH3 domain	.86	4×10^{-5}	-0.49	7×10^{-2}	15
PLCγ1 SH2 domain	.71	7×10^{-6}	-0.04	8×10^{-1}	31
U1A	.69	3×10^{-5}	-0.46	1×10^{-2}	29
ubiquitin	.65	9×10^{-5}	-0.48	8×10^{-3}	30
Syp SH2 domain	.40	2×10^{-2}	.02	9×10^{-1}	31
drk SH3 domain	.20	4×10^{-1}	-0.05	8×10^{-1}	19
SAP SH2 domain	.09	7×10^{-1}	-0.22	3×10^{-1}	24
SAP SH2 domain[e]	.53	2×10^{-2}	-0.20	4×10^{-1}	18

[a] Deuterium relaxation data have been published for the PLCγ1 C-terminal SH2 domain,[11] U1A protein,[12] ubiquitin,[13] Syp N-terminal SH2 domain,[4] drk N-terminal SH3 domain,[14] and SAP SH2 domain.[15] [b] Linear correlation coefficient (r) and statistical significance (p) between normalized methyl axis order parameters (S^2_{norm}, defined by eq 2), as well as residue preference (Π, defined by eq 1). Previously published alignments of SH3[7] and U1A[16] domains were used. SH2 domain and ubiquitin sequence alignments were obtained from the Pfam protein family database,[17] and individual sequences were weighted according to Henikoff et al.[18] Values of p smaller than 0.05 are considered significant. [c] Linear correlation coefficient (r) and statistical significance (p) between normalized methyl axis order parameters, S^2_{norm}, and solvent accessibility, calculated on a per-residue basis using the program MOLMOL[19] and molecular structures for the Fyn SH3 domain [1SHF[20]], PLCγ1 SH2 domain [2PLE[21]], U1A protein [1FHT[22]], ubiquitin [1UBQ[23]], Syp SH2 domain [1AYD[24]], and SAP SH2 domain [1D1Z[8]], deposited in the Brookhaven Protein Data Bank. The structure of the drk SH3 domain has been solved by Forman-Kay and co-workers and has not been published. [d] Number of methyl groups included in the analysis. Alanine residues have been omitted since these report motions of the backbone. [e] Omitting data for L25, V37, V40, L46.

affect side-chain dynamics, whereby the same interactions that lead to a preference for a particular amino acid type also impose specific side-chain conformational restrictions. The value of r_Π shows that all factors influencing side-chain motions are not reflected in the parameter Π; however, the fact that flexibility depends significantly more on residue preference than on solvent accessibility points to the presence of additional determinants of dynamics whose identification may be facilitated through the use of sequence alignment data. With this in mind, we have examined results for the Fyn SH3 domain in greater detail. In the numbering scheme of Larson et al.,[7] L7, L29, and T44 show degrees of burial similar to those of T2, L3, T43, and V58 and yet are significantly more conserved and less flexible. In the case of T44, this is probably due to a hydrogen bond between the Oγ1 of T44 and the amide proton of residue 46 that is also seen in SH3 domains from the Hck, c-Src, and Lck tyrosine kinases. L7 and L29 present a possible connection between secondary structure and side-chain flexibility. Both positions are exposed to solvent yet show strong preferences for the leucine side-chain. L7 immediately follows the first β-strand, and L29 participates in a classic β-bulge. The conformations at these sites, in which backbone ϕ,ψ angles lie in the right-handed α-helical region of the Ramachandran plot and facilitate sharp bends of extended backbone structure, are conserved in most SH3 domains.[7] The appearance of leucine at position 29 has been linked to a preference for leucine residues at position 1 of classic β-bulges.[25] The basis for this tendency is not known, but the relative rigidity of the L7 and L29 side-chains suggests the presence of interactions in the folded state that restrict their flexibility compared to similarly solvent-exposed residues.

Studies have demonstrated that highly conserved positions in sequence alignments often play specific structural or functional roles.[7] The results presented here show that, in general, such residues are less mobile than average. It is likely, however, that while certain conserved structural or functional motifs involve highly restricted side-chains, others may allow or even require significant conformational freedom. As well, side-chains that are conserved due to specific interactions with binding partners may be mobile when these ligands are not present in the NMR sample. Conversely, uncommon residues may be rigid in cases where they form interactions that are not seen in homologous proteins. Such unusual residues may be identified as large outliers in comparisons of Π and S^2_{norm} values. The association between conserved structural features and flexibility can be more rigorously addressed through analyses of large sets of structural and dynamics information similar to approaches that have identified the sequence preferences of secondary structure motifs. As more NMR side-chain relaxation data are collected, this will become feasible, allowing identification of specific conformational determinants of side-chain dynamics.

Acknowledgment. The authors would like to thank Stefan Larson for supplying U1A sequence data and Dr. Julie Forman-Kay for helpful discussions. This work was supported by a grant from the Canadian Institutes of Health Research. L.E.K. holds a Canada Research Chair in Biochemistry.

Supporting Information Available: Figures showing methyl peak intensity decay curves and relaxation rate consistency relationships, as well as tables of S^2_{axis} values for the Fyn SH3 domain and μ_{meth} and σ_{meth} values for a database of eight proteins (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Lee, A. L.; Wand, J. *Nature* **2001**, *411*, 501−504.
(2) Yang, D.; Mok, Y.-K.; Forman-Kay, J. D.; Farrow, N. A.; Kay, L. E. *J. Mol. Biol.* **1997**, *272*, 790−804.
(3) Lee, A. L.; Kinnear, S. A.; Wand, A. J. *Nat. Struct. Biol.* **2000**, *7*, 72−77.
(4) Kay, L. E.; Muhandiram, D. R.; Wolf, G.; Shoelson, S. E.; Forman-Kay, J. D. *Nature Struct. Biol.* **1998**, *5*, 156−163.
(5) Mittermaier, A.; Kay, L. E.; Forman-Kay, J. D. *J. Biomol. NMR* **1999**, *13*, 181−185.
(6) Millet, O.; Muhandiram, D. R.; Skrynnikov, N.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6439−6448.
(7) Larson, S. M.; Davidson, A. R. *Protein Sci.* **2000**, *9*, 2170−2180.
(8) Poy, F.; Yaffe, M. B.; Sayos, J.; Saxena, K.; Morra, M.; Sumegi, J.; Cantley, L. C.; Terhorst, C.; Eck, M. J. *Mol. Cell* **1999**, *4*, 555−561.
(9) Chan, B.; Lanyi, A.; Song, H. K.; Griesbach, J.; Simarro-Grande, M.; Poy, F.; Howie, D.; Sumegi, J.; Terhorst, C.; Eck, M. J. *Nature Cell Biol.* **2003**.
(10) Zar, J. H. *Biostatistical Analysis*, 2nd ed.; Prentice Hall, Inc.: Englewood Cliffs, NJ, 1984; pp 306−327.
(11) Kay, L. E.; Muhandiram, D. R.; Farrow, N. A.; Aubin, Y.; Forman-Kay, J. D. *Biochemistry* **1996**, *35*, 361−368.
(12) Mittermaier, A.; Varani, L.; Muhandiram, R.; Kay, L. E.; Varani, G. *J. Mol. Biol.* **1999**, *294*, 967−979.
(13) Lee, A. L.; Flynn, P. F.; Wand, J. *J. Am. Chem. Soc.* **1999**, *121*, 2891−2902.
(14) Yang, D.; Kay, L. E. *J. Mol. Biol.* **1996**, *263*, 369−382.
(15) Finerty, P. J. J.; Muhandiram, R.; Forman-Kay, J. D. *J. Mol. Biol.* **2002**, *322*, 605−620.
(16) Larson, S. M.; Ruczinski, I.; Davidson, A. R.; Baker, D.; Plaxco, K. W. *J. Mol. Biol.* **2002**, *316*, 225−233.
(17) Bateman, A.; Birney, E.; Cerruti, L.; Durbin, R.; Etwiller, L.; Eddy, S. R.; Griffiths-Jones, S.; Howe, K. L.; Marshall, M.; Sonnhammer, E. L. L. *Nucleic Acids Res.* **2002**, *30*, 276−280.
(18) Henikoff, S.; Henikoff, J. G. *J. Mol. Biol.* **1994**, *243*, 574−578.
(19) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.
(20) Noble, M. E.; Mussachio, A.; Saraste, M.; Courtneidge, S. A.; Wierenga, R. K. *EMBO J.* **1993**, *12*, 2617−2624.
(21) Pascal, S. M.; Singer, A. U.; Gish, G.; Yamazaki, T.; Shoelson, S. E.; Pawson, T.; Kay, L. E.; Forman-Kay, J. D. *Cell* **1994**, *77*, 461−472.
(22) Avis, J. M.; Allain, F. H.; Howe, P. W.; Varani, G.; Nagai, K.; Neuhaus, D. *J. Mol. Biol.* **1996**, *257*, 398−411.
(23) Vijay-Kumar, S.; Bugg, C. E.; Cook, W. J. *J. Mol. Biol.* **1987**, *194*, 531−544.
(24) Lee, C. H.; Kominos, D.; Jacques, S.; Margolis, B.; Schlessinger, J.; Shoelson, S. E.; Kuriyan, J. *Structure* **1994**, *2*, 423−438.
(25) Chan, A. W. E.; Hutchinson, E. G.; Harris, D.; Thornton, J. M. *Protein Sci.* **1993**, *2*, 1574−1590.

JA034856Q