J|A|C|S COMMUNICATIONS

Published on Web 04/17/2009

Conservation of Side-Chain Dynamics Within a Protein Family

Anthony B. Law,[†] Ernesto J. Fuentes,[‡] and Andrew L. Lee*[,†,§]

Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, Department of Biochemistry, University of Iowa, Iowa City, Iowa 52242, and Division of Medicinal Chemistry and Natural Products, School of Pharmacy, UNC-Chapel Hill, Chapel Hill, North Carolina 27599

Received December 24, 2008; E-mail: drewlee@unc.edu

Protein dynamics are important for ligand binding, enzyme catalysis, conformational change, and protein folding. NMR relaxation experiments now allow for quantitative characterization of the dynamics of both backbone and side-chain bonds, as reflected in the Lipari-Szabo order parameter, S^2.[1,2] Order parameters from a multitude of proteins have shown, on average, that backbone atoms are primarily rigid, whereas side chains are heterogeneously mobile on the ps-ns time scale.[3] The heterogeneity in methyl side-chain dynamics (SCD) is not yet well understood, as evidenced by the challenge in predicting SCD parameters, even though most methyl side chains are packed into the hydrophobic core. Yet, the variability in SCD is likely required for diversity of protein function. Thus, specific patterns of flexibility may be associated with specific functions.

Because NMR studies of SCD have been relatively few, comparisons of SCD in structurally or functionally similar proteins have been limited. Here we test the idea that proteins with similar function (and tertiary fold) have similar dynamics. We enlarge the SCD pool by characterizing dynamics for three PDZ (PSD-95/Discs Large/Zonula Occludens-1) domains and show that SCD in these domains are more similar than naively expected. These findings suggest that dynamic similarity is driven in large part by fold and/or function, as opposed to local structure.

Side-chain motion on the ps-ns time scale is captured by the methyl symmetry axis order parameter, S^2_{axis}, corresponding to the angular restriction of the C−CH_3 bond[4] and referred to simply as S^2. Values range from 0 to 1, where 1 indicates complete rigidity in the molecular frame. ^{2}H methyl relaxation was used,[5] as previously described,[6] to obtain S^2 values from three PDZ domains: the third PDZ domain from PSD-95, which, because it has three PDZ domains is denoted PSD95 (3/3); the PDZ domain from Erbin, i.e., Erbin (1/1); and hPTP1e (2/5) which was characterized previously.[6] The function of PDZ domains is to bind C-terminal tail sequences of target proteins. The three PDZ domains were structurally aligned to define the sites for comparing S^2 (using pdb structures 1bfe, 2h3l, and an unpublished crystal structure of hPTP1e (2/5)). Based on these alignments, the sequence identity was ~30% for each of the three PDZ pairs. We compared S^2 to obtain absolute differences in SCD, as opposed to use of S^2_{norm}.[7] The pairwise comparisons are shown in Figure 1A. The rules used for comparing nonidentical side chains are given in the figure legend. The correlation coefficient for the SCD comparison of PSD95 (3/3) and Erbin (1/1) is 0.72. Because this type of comparison is between two proteins of the same family, we designate this correlation coefficient as r_{fam}. For PSD95 (3/3) vs hPTP1e (2/5) and Erbin (1/1) vs hPTP1e (2/5), r_{fam} is 0.67 and 0.56, respectively (Table 1). The average r_{fam} is 0.65 for all comparisons, indicating very similar SCD among these PDZ domains.

One potential explanation for these relatively high correlations in SCD is sequence similarity. Because different methyl types have characteristic distributions of S^2, simply having the same methyl types at aligned positions may contribute to the correlation. To test this, we compared S^2_{exp} values with corresponding values expected based purely on methyl type and designated the correlation coefficient r_{seq}.

High values of r_{fam} are not driven by sequence identity. The dependence of SCD similarity on amino acid sequence was assessed via calculation of $\langle r_{seq} \rangle$ from simulated sets of S^2 based on known distributions for the six methyl-containing residues.[7] Specifically, for

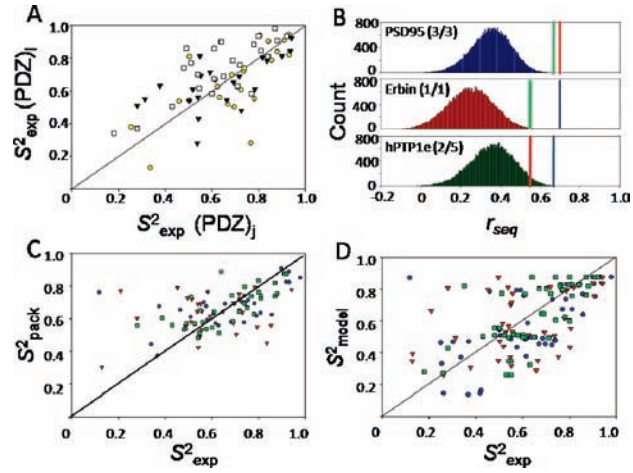

Figure 1. Pairwise correlations of experimental SCD and sequence/packing based predictions of SCD for three PDZ domains. (A) Inter-PDZ correlation of S^2_{exp} of PDZ$_i$ with S^2_{exp} of PDZ$_j$: PSD-95 (3/3) vs Erbin (yellow circles); Erbin vs hPTP1e (2/5) (black triangles); hPTP1e (2/5) vs PSD-95 (3/3) (white squares). For comparison of S^2 between different methyl types, simple and consistent rules were used: (1) singly occurring methyl groups (Ala, Thr, Met) are compared directly to one another and to the average of the two Val-γ or two Leu-δ groups; (2) Ile-γ methyls are compared to averages of Val groups and to Ala and Thr methyls; (3) Ile-δ methyls are compared to averages of Leu groups and to Met methyls; and (4) Val and Leu methyls are compared to the averages (if both are available) of each other. (B) Histograms of r_{seq} for PSD-95 (3/3) (blue), Erbin (red), and hPTP1e (2/5) (green). Vertical lines show the corresponding r_{fam} values, based on comparison to the other two PDZ domains (indicated by color). (C) Correlation of S^2_{pack} with S^2_{exp}; the correlation coefficient is given by r_{pack}. Color coding (blue circles, red triangles, green squares) is same as that in B. (D) Correlation of S^2_{model} with S^2_{exp}; the correlation coefficient is given by r_{comb}. Color coding is same as that in B.

Table 1. Correlations of SCD in PDZ Domains

	inter	intra (expt vs predicted)			
	r_{fam}	$\langle r_{seq} \rangle$	r_{pack}	r_{comb}	methyl identity
PSD95 (3/3)	0.72[a]	0.35 (0.10)	0.53	0.62	59%[a]
Erbin (1/1)	0.56[b]	0.25 (0.12)	0.27	0.47	45%[b]
hPTP1e (2/5)	0.67[c]	0.36 (0.10)	0.73	0.67	77%[c]
average:	0.65	0.32	0.51	0.59	60%

[a] PSD95 (3/3) vs Erbin (1/1). [b] Erbin (1/1) vs hPTP1e (2/5). [c] hPTP1e (2/5) vs PSD95 (3/3).

[†] Department of Biochemistry and Biophysics, UNC−Chapel Hill.
[‡] University of Iowa.
[§] School of Pharmacy, UNC−Chapel Hill.

10.1021/ja809915a CCC: $40.75

each of the three PDZ domains, S^2_{exp} values were replaced with 10,000 S^2 values drawn randomly from Gaussians parametrized according to methyl type as in the calculation of S^2_{norm}.[7] S^2 values <0 or >1 made up less than 2% of the data and were discarded. Because the distributions are independent of fold, the simulated order parameters are "structureless" in that they only reflect the methyl type. For each PDZ domain, 10,000 r_{seq} values were obtained from correlation of S^2_{exp} values with the corresponding S^2 values from each simulated PDZ (Figure 1B). In the case of PSD95 (3/3), the average value of the distribution, $\langle r_{seq} \rangle$, is 0.35. The low values of $\langle r_{seq} \rangle$ for each of the domains indicate that methyl type alone contributes relatively little to the observed SCD (Table 1). Importantly, because $\langle r_{seq} \rangle$ for each PDZ domain falls far below the corresponding r_{fam} values, this also indicates that SCD *similarity* in these PDZ domains is not explained by sequence conservation. Consistent with this, SCD similarity calculated only from identical residues is slightly lower than when calculated from nonidentical residues (average r_{fam} values of 0.6 and 0.7, respectively). Thus, based on Figure 1A and 1B, we reach our primary conclusion that SCD is a conserved property among these PDZ domains (and by extension, the PDZ family) and that ps-ns side-chain motion is highly organized and dependent upon overall protein fold and/or function.

Structural factors were considered next. The high similarity in SCD may have contributions from local and nonlocal structural factors. In general, studies have failed to show a good correlation of SCD (S^2_{exp}) with structural parameters such as local packing density, accessible surface area, depth of burial, or B-factors;[3,8] even MD simulations typically do not accurately capture side-chain order,[9] although specific cases of high correlation ($r \sim 0.8$) have been reported.[10−12] However, a reasonable correlation between the occluded surface percent (OSP) measure of packing density[13] and S^2_{norm} was recently reported.[14] For each PDZ domain, we sought to distinguish local from nonlocal contributions to SCD. We considered r_{local}, a correlation coefficient between S^2_{exp} and ideal S^2 values that would result only from local packing. Values of r_{local} therefore indicate the extent to which SCD is dominated by local structure. We also sought to assess the basis of *conservation* in SCD. In principle, the value of r_{local} represents the maximal contribution of local factors to SCD similarity between nonidentical proteins (or here, PDZ domains), where the maximal value will only be attained in the unlikely scenario in which all corresponding side chains from the two proteins have identical local structural environments. Thus, realistically, if r_{fam} is equivalent or greater than r_{local}, the similarity in PDZ domain SCD is driven at least in part by nonlocal factors.

To test this assertion, we calculated two estimates for r_{local}: (1) r_{pack}, in which S^2_{exp} is correlated with S^2 predicted from local structural packing density (using OSP), and (2) r_{comb}, in which S^2_{exp} is correlated with S^2 predicted from a model that combines local structure and sequence information. For the first method, S^2 was predicted from local packing using the linear correlation of OSP with S^2_{norm} ($r = 0.41$ for a diverse set of 9 proteins, data not shown): $S^2_{pack} = \mu_{meth} + \sigma_{meth}(3.60 \cdot \mathrm{OSP} - 1.15)$, where μ_{meth} and σ_{meth} are the mean and standard deviation of the known S^2 distribution for the relevant methyl type, and the constants are obtained from the regression of S^2_{norm} vs OSP. We define r_{pack} as the correlation of S^2_{exp} with the predicted S^2_{pack} (Figure 1C). On average, r_{pack} is 0.50, significantly lower than r_{fam} (Table 1). For the second method, a model that combines methyl type and local packing information[15] was used to predict S^2 (referred to as S^2_{model}). r_{comb} is defined as the correlation between S^2_{exp} and S^2_{model} (Figure 1D) and has an average value of 0.59, also lower than r_{fam} (Table 1). It is noted that some values of S^2_{model} exhibit tight clustering about known methyl-specific averages, which has the effect of narrowing the range of S^2_{model} values and increasing r_{comb} slightly. It is also noted that r_{pack} and r_{comb} are independent of the crystal structure resolution, since the lowest correlations are for the 1.0 Å Erbin (1/1) structure. Also, use of a higher resolution PSD95 (3/3) structure (1TQ3) changes r_{pack} and r_{comb} by <2%. Thus, on average, the inter-PDZ correlation r_{fam} exceeds both intra-PDZ measures of r_{local} (r_{pack} and r_{comb}). We conclude that local structural factors, with or without sequence information, do not fully account for the similarity in PDZ SCD.

Collectively, these data indicate that the SCD of a different PDZ domain is a much stronger predictor of SCD than predictions from sequence and at least as strong a predictor (and in most cases, better) than what is expected from local structure. Not only are the SCD simply conserved, but they also appear to be dependent on nonlocal factors. This is consistent with the idea of conserved long-range or correlated motions that are supported by the larger PDZ architecture. Such motions may also support the common PDZ function of C-terminus binding, further suggesting that SCD, and fast dynamics in general, are under evolutionary pressure. We note that correlated motions have been shown to exist in hPTP1e (2/5).[16]

In summary, NMR methyl relaxation data on three PDZ domains have provided experimental evidence for organized, conserved, ps-ns side-chain dynamics. We note that this conclusion is distinct from the correlation previously reported between sequence conservation and side-chain rigidity.[7] In addition, these conserved motions likely reflect long-range effects present in PDZ domains. Consistent with our findings here, prediction of S^2 based on ensembles of X-ray structures of "high-sequence similarity" (>90% sequence identity, or bound to different ligands) shows strong correlation with S^2_{exp} (average $r = 0.66$).[17] S^2_{axis} has also recently been predicted from structure-based modeling of rotamer states, with average $r = 0.57$.[18]

Finally, we note that while the dynamics of these homologous PDZ proteins are quite similar, variance in SCD remains. Our future studies will be directed at the location of variance and its bearing on specificity differences in PDZ function.

Acknowledgment. This work was funded by NSF Grant MCB-0344354 (A.L.L.), a diversity supplement to NIH grant GM066009 (A.L.L.), an NSF minority fellowship DBI-0314399 (E.J.F.), and NIH MSTP Grant T32 GM008719.

References

(1) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4546–4559.
(2) Jarymowycz, V. A.; Stone, M. J. *Chem. Rev.* **2006**, *106*, 1624–71.
(3) Igumenova, T. I.; Frederick, K. K.; Wand, A. J. *Chem. Rev.* **2006**, *106*, 1672–99.
(4) Lipari, G.; Szabo, A. *J. Am. Chem. Soc.* **1982**, *104*, 4559–4570.
(5) Muhandiram, D. R.; Yamazaki, T.; Sykes, B. D.; Kay, L. E. *J. Am. Chem. Soc.* **1995**, *117*, 11536–11544.
(6) Fuentes, E. J.; Der, C. J.; Lee, A. L. *J. Mol. Biol.* **2004**, *335*, 1105–15.
(7) Mittermaier, A.; Davidson, A. R.; Kay, L. E. *J. Am. Chem. Soc.* **2003**, *125*, 9004–5.
(8) Mittermaier, A.; Kay, L. E.; Forman-Kay, J. D. *J. Biomol. NMR* **1999**, *13*, 181–185.
(9) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature (London)* **2005**, *433*, 128–32.
(10) Hu, H.; Clarkson, M. W.; Hermans, J.; Lee, A. L. *Biochemistry* **2003**, *42*, 13856–68.
(11) Showalter, S. A.; Johnson, E.; Rance, M.; Bruschweiler, R. *J. Am. Chem. Soc.* **2007**, *129*, 14146–7.
(12) Best, R. B.; Clarke, J.; Karplus, M. *J. Mol. Biol.* **2005**, *349*, 185–203.
(13) Pattabiraman, N.; Ward, K. B.; Fleming, P. J. *J. Mol. Recognit.* **1995**, *8*, 334–44.
(14) Okamura, H.; Makino, K.; Nishimura, Y. *J. Mol. Biol.* **2007**, *367*, 1093–117.
(15) Ming, D.; Bruschweiler, R. *J. Biomol. NMR* **2004**, *29*, 363–8.
(16) Dhulesia, A.; Gsponer, J.; Vendruscolo, M. *J. Am. Chem. Soc.* **2008**, *130*, 8931–9.
(17) Best, R. B.; Lindorff-Larsen, K.; DePristo, M. A.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10901–6.
(18) Friedland, G. D.; Linares, A. J.; Smith, C. A.; Kortemme, T. *J. Mol. Biol.* **2008**, *380*, 757–74.

JA809915A