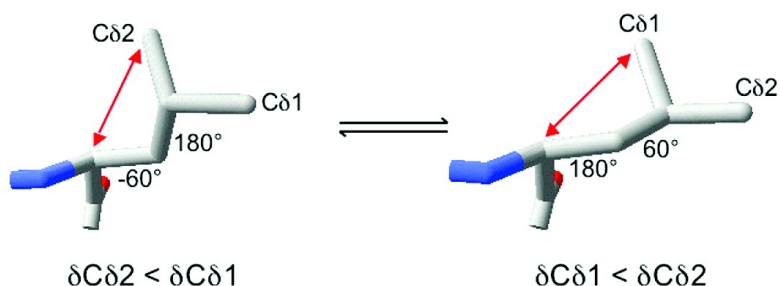


## Dependence of Amino Acid Side Chain C Shifts on Dihedral Angle: Application to Conformational Analysis

Robert E. London, Brett D. Wingad, and Geoffrey A. Mueller

*J. Am. Chem. Soc.*, **2008**, 130 (33), 11097-11105 • DOI: 10.1021/ja802729t • Publication Date (Web): 25 July 2008

Downloaded from <http://pubs.acs.org> on February 8, 2009



### More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 1 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

[View the Full Text HTML](#)

## Dependence of Amino Acid Side Chain $^{13}\text{C}$ Shifts on Dihedral Angle: Application to Conformational Analysis

Robert E. London,\* Brett D. Wingad, and Geoffrey A. Mueller

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, 111 T.W. Alexander Drive, MD MR-01, Research Triangle Park, North Carolina 27709

Received April 14, 2008; E-mail: london@niehs.nih.gov

**Abstract:** Chemical shift data from the BiomagResDataBank and conformational data derived from the protein data bank have been correlated in order to explore the conformational dependence of side chain  $^{13}\text{C}$  resonance shifts. Consistent with predictions based on steric compression, upfield shifts for  $\text{C}_\gamma$  resonances of Thr, Val, Ile, Leu, Met, Arg, Lys, Glu, and Gln residues correlate with both the number of heavy atom (nonproton)  $\gamma$ -substituents and with *gauche* conformational orientations of  $\gamma$ -substituents. The  $^{13}\text{C}$  shift/conformation correlations are most apparent for  $\text{C}_\gamma$  carbons but also can be observed at positions further from the backbone. Intraresidue steric conflict leads to a correlation between upfield-shifted side chain  $^{13}\text{C}$  resonances and statistically lower probabilities in surveys of protein side chain conformation. Illustrative applications to the DNA pol  $\lambda$  lyase domain and to dihydrofolate reductase are discussed. In the latter case,  $^{13}\text{C}$  shift analysis indicates that the conformation of the remote residue V119 on the  $\beta\text{F}$ - $\beta\text{G}$  loop is correlated with the redox state of the bound pyridine nucleotide cofactor, providing one basis for discrimination between substrate and product. It is anticipated that  $^{13}\text{C}$  shift data for protein sidechains can provide a useful basis for the analysis of conformational changes even in large, deuterated proteins. Additionally, the large dependence of the leucine methyl shift difference,  $\delta\text{C}\delta 1 - \delta\text{C}\delta 2$ , on both  $\chi 1$  and  $\chi 2$  is sufficient to allow this parameter to be used as a restraint in structure calculations if stereospecific assignment data are available.

### 1. Introduction

Although nuclear Overhauser effect interactions provide the primary basis for determining solution structures by NMR spectroscopy, structural information also has been derived from other parameters. For example, the use of protein chemical shift information as a structural tool has steadily increased. Chemical shift indexing (CSI)<sup>1,2</sup> and TALOS<sup>3</sup> analysis of backbone shifts have been demonstrated to provide reliable information on protein secondary structure. Furthermore, forcefields have been developed to refine NMR structures against  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$ , and  $^1\text{H}$  shifts.<sup>4,5</sup> Most impressively, the inclusion of chemical shift-based restraints in molecular force field calculations leads to significant improvements in protein structure prediction.<sup>6</sup> However, extending shift information to specific side chain conformations has proven to be more difficult.

Conversely, predicting shifts from structure also has been widely studied in order to better understand the physical basis of the chemical shift. Sidechain proton shifts are predicted by several convenient web-based programs which utilize density

functional calculations<sup>7</sup> or are based on combinations of empirical correlations, classical formulas, and quantum mechanical determinations.<sup>8</sup> Extensive quantum mechanical calculations of protein  $^{13}\text{C}$  shifts have been performed by Oldfield and co-workers.<sup>9,10</sup> The program SPARTA predicts backbone chemical shifts by matching  $\phi$ ,  $\psi$ , and  $\chi 1$  angles and sequence similarity of three residue groups with a database of high resolution X-ray structures for which the backbone chemical shifts have been assigned.<sup>11</sup> Although these studies focus primarily on backbone shifts, some examples of side chain shifts have also been presented.<sup>10,12,13</sup> Structure-based predictions for the  $^1\text{H}$  shifts of side chain nuclei have provided useful structural insights,<sup>14</sup> but the  $^{13}\text{C}$  shift predictions for side chain nuclei have generally proven to be less useful. Neither of the currently available web-based prediction programs SHIFTS<sup>7</sup> or SHIFTX<sup>8</sup>

(7) Xu, X. P.; Case, D. A. *J. Biomol. NMR* **2001**, *21*, 321–33.

(8) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. *J. Biomol. NMR* **2003**, *26*, 215–40.

(9) Havlin, R. H.; Laws, D. D.; Bitter, H. M. L.; Sanders, L. K.; Sun, H. H.; Grimley, J. S.; Wemmer, D. E.; Pines, A.; Oldfield, E. *J. Am. Chem. Soc.* **2001**, *123*, 10362–10369.

(10) Havlin, R. H.; Le, H. B.; Laws, D. D.; deDios, A. C.; Oldfield, E. *J. Am. Chem. Soc.* **1997**, *119*, 11951–11958.

(11) Shen, Y.; Bax, A. *J. Biomol. NMR* **2007**, *38*, 289–302.

(12) Pearson, J. G.; Le, H. B.; Sanders, L. K.; Godbout, N.; Havlin, R. H.; Oldfield, E. *J. Am. Chem. Soc.* **1997**, *119*, 11941–11950.

(13) Sun, H. H.; Sanders, L. K.; Oldfield, E. *J. Am. Chem. Soc.* **2002**, *124*, 5486–5495.

(14) Kirby, T. W.; DeRose, E. F.; London, R. E. *FASEB J.* **2005**, *19*, A308–A308.

(15) Neal, S.; Berjanskii, M.; Zhang, H.; Wishart, D. S. *Magn. Reson. Chem.* **2006**, *44* (S1), S158–67.

(1) Wishart, D. S.; Sykes, B. D. *J. Biomol. NMR* **1994**, *4*, 171–80.

(2) Wishart, D. S.; Sykes, B. D.; Richards, F. M. *Biochemistry* **1992**, *31*, 1647–51.

(3) Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999**, *13*, 289–302.

(4) Kuszewski, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1995**, *107*, 293–297.

(5) Kuszewski, J.; Qin, J.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson., Ser. B* **1995**, *106*, 92–96.

(6) Cavalli, A.; Salvatella, X.; Dobson, C. M.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 9615–9620.

provide  $^{13}\text{C}$  chemical shift values for residue sidechains, presumably because the correlation coefficients relating experimental to predicted data were substantially poorer than those obtained for  $^1\text{H}$  shift data or for the  $^{13}\text{C}$  and  $^{15}\text{N}$  backbone data, and since side chain shifts beyond  $C\beta$  have been considered not to differ substantially from random coil values.<sup>8</sup> The programs SHIFTOR<sup>15</sup> and PREDITOR<sup>16</sup> predict side chain  $\chi_1$  values from data that include  $^{13}\text{C}\gamma$  shifts, so that the observed correlation may be derived in part from a conformational dependence of  $C\gamma$ , but any such relationship remains implicit in the analysis.

Early studies of  $^{13}\text{C}$  shifts in aliphatic hydrocarbons by Grant and co-workers demonstrated a significant “ $\gamma$ -substituent effect”, according to which substituents positioned three bonds from the observed nucleus contribute to the  $^{13}\text{C}$  shift based on their number<sup>17</sup> and subtended dihedral angle.<sup>18,19</sup> The effect is believed to arise primarily from steric crowding leading to carbon–hydrogen bond polarization;<sup>18,20</sup> however it subsequently has become clear that additional factors are also involved.<sup>21,22</sup> Tonelli interpreted carbonyl and  $C\beta$  shifts in several peptides based on  $\gamma$ -effect contributions.<sup>23,24</sup> Pearson et al. have calculated a significant dihedral angle dependence for the  $C\gamma_1$  and  $C\gamma_2$   $^{13}\text{C}$  shifts in studies of *N*-formylvaline.<sup>12</sup> These results support the conclusion that, despite some controversy over the physical basis for the  $\gamma$ -effect,<sup>22</sup> it can be extrapolated to studies of protein side chain conformation. In order to further explore this possibility, we have investigated the extent to which correlations between side chain dihedral angles ( $\chi$  values) correlate with experimental  $^{13}\text{C}$  shift data for a set of well-characterized proteins. We also have evaluated the parameters  $\Delta C\gamma_{12} = \delta C\gamma_1 - \delta C\gamma_2$  and  $\Delta C\delta_{12} = \delta C\delta_1 - \delta C\delta_2$  for valine and leucine, respectively, since many of the shift contributions arising from other factors might be expected to be similar for these pairs of methyl carbons, enhancing the predictive value of the shift difference.

## 2. Methods

In order to evaluate potential correlations between  $^{13}\text{C}$  shift and conformation, shift data were obtained from the Biomag Res data bank (BMRB)<sup>25</sup> and correlated with structural data from the protein data bank (PDB).<sup>26</sup> After considering various sample subsets, we combined the sample sets of structures and shifts that previously had been selected for determination of the relationships for TALOS<sup>3</sup> and SHIFTX.<sup>8</sup> Both studies attempted to compile data that featured reliable shift assignments and a corresponding high-resolution crystal structure. Importantly, none of the data is reported to have paramagnetic atoms present, which can lead to significant contact and pseudocontact shifts.

The BMRB contains NMR chemical shift information and other parameters derived from NMR studies. The information needed was

extracted and collated using Perl scripts that parse through the BMRB-STAR files. The shift data were corrected to remove any referencing biases introduced when the original measurements were taken using the recommendations of Wang et al.<sup>27</sup> For the  $^{13}\text{C}$  shifts of the prochiral methyl groups of valine and leucine, the data were filtered by ambiguity code = 1 (as defined in the BMRB), to include only residues which had been stereospecifically assigned. We note that according to the papers describing the structural analyses, in most cases these assignments have been based on the analysis of NOE data. For comparison we also tabulated a list of five proteins for which the methyl stereo assignments were made using the method of Neri et al.<sup>28</sup> The name, pdb, and bmrB identification of these proteins are as follows: malate synthase G (1D8C, 5471), maltose binding protein (1omp, 4354), putidaredoxin (1OQR, 2278), phage 434 repressor (1R69, 195), and P22 c2 repressor (1ADR, 486).

The structural information required for calculating the bond angles can be found at the Protein Data Bank. We used the structural visualization programs MOLMOL<sup>29</sup> and PyMol (DeLano Scientific, LLC 2008) to calculate the bond angles. We note that, for isoleucine residues, the program MOLMOL gave discrepancies in the calculated angles depending on the atomic nomenclature used. After manually visualizing the structures and rotameric conformations for a subset of our assembled database, we found PyMol did not have the same problem and utilized this software for the isoleucine  $\chi_2$  calculations.

The amino acid type, shifts, and angles were correlated to analyze for trends again using Perl scripting. The results were saved in formats that could easily be read into a spreadsheet. We created templates, readable by Excel or Open Office Charts, that would read the shift and angle data and then automatically separate, analyze, and graph them while still including editable fields for filtering data based on other criteria (vide infra). This makes the process easy and quickly repeatable. The spreadsheet calculates statistical quantities including mean and standard deviation, performs *t* tests to verify significance in the results, allows for visual inspection, and reports the quantity of interest, the relative shift of the more sterically compressed conformation. Typical output is presented as Supporting Information. The *t* test compared the mean shift of two rotamers to determine if they were derived from significantly different populations. In most cases there was better than 99% confidence that the mean shifts were different, although when the shift difference was much less than 1, the confidence could be as low as 50%. We find that, generally, the significance attributed to the *t* test correlates with the calculated shift differences presented, as might be expected.

In our final analysis, we have applied the following filters in order to eliminate poorer quality structures. Structural analyses have shown that, for the amino acid sidechains in proteins, the conformations cluster around the relatively stable *gauche* and *trans* conformers.<sup>30</sup> Structures with angles not falling into these ranges are generally characterized by higher *B*-factors. For this reason, we ultimately utilized only data falling within a  $40^\circ$  range of these minima, i.e. with  $\chi$  angles equal to  $\pm 60^\circ \pm 20^\circ$  and  $180^\circ \pm 20^\circ$ . In addition, we only use X-ray data which diffracted to  $2.0 \text{ \AA}$  or better and *B*-factors less than 40.

Similarly the shift data needed to be examined for errors.  $^{13}\text{C}$  shift data falling greater than 2 standard deviations from the mean were eliminated from the analysis, in order to account for anomalies or incorrect assignments. For example, we found some lysine  $C\gamma$  shifts of  $>49$  ppm and methionine  $C\epsilon$  shifts  $>30$  ppm which are

- (16) Berjanskii, M. V.; Neal, S.; Wishart, D. S. *Nucleic Acids Res.* **2006**, *34*, W63–9.  
 (17) Grant, D. M.; Paul, E. G. *J. Am. Chem. Soc.* **1964**, *86*, 2984.  
 (18) Cheney, B. V.; Grant, D. M. *J. Am. Chem. Soc.* **1967**, *89*, 5319.  
 (19) Saito, H. *Magn. Reson. Chem.* **1986**, *24*, 835–852.  
 (20) Grant, D. M.; Cheney, B. V. *J. Am. Chem. Soc.* **1967**, *89*, 5315.  
 (21) Jiao, D.; Barfield, M.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 10883–10887.  
 (22) Seidl, P. R.; Leal, K. Z.; Costa, V. E. U.; Mollmann, M. E. S. *Magn. Reson. Chem.* **1998**, *36*, 261–266.  
 (23) Tonelli, A. E. *J. Am. Chem. Soc.* **1980**, *102*, 7635–7637.  
 (24) Tonelli, A. E. *Biopolymers* **1984**, *23*, 819–829.  
 (25) Doreleijers, J. F.; Mading, S.; Maziuk, D.; Sojourner, K.; Yin, L.; Zhu, J.; Markley, J. L.; Ulrich, E. L. *J. Biomol. NMR* **2003**, *26*, 139–146.  
 (26) Berman, H.; Henrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980–980.

- (27) Wang, L. Y.; Eghbalian, H. R.; Bahrami, A.; Markley, J. L. *J. Biomol. NMR* **2005**, *32*, 13–22.  
 (28) Neri, D.; Szyperski, T.; Otting, G.; Senn, H.; Wuthrich, K. *Biochemistry* **1989**, *28*, 7510–7516.  
 (29) Koradi, R.; Billeter, M.; Wuthrich, K. *J. Mol. Graphics* **1996**, *14*, 51–5.  
 (30) MacArthur, M. W.; Thornton, J. M. *Acta Crystallogr., Sect. D* **1999**, *55*, 994–1004.

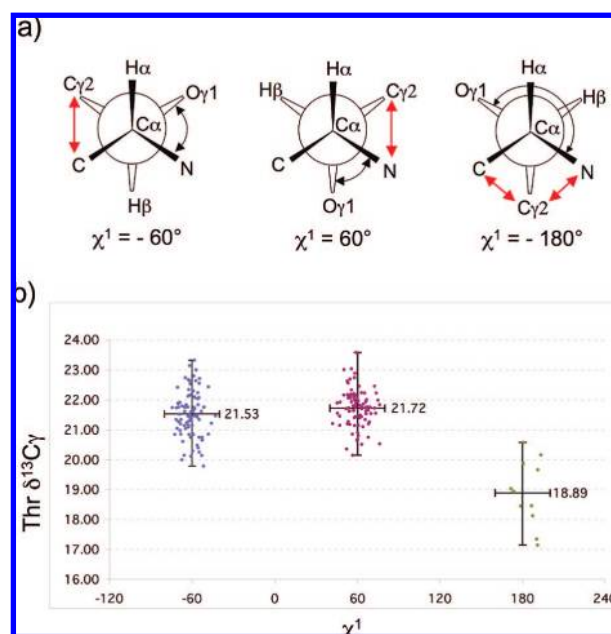
clearly in error. Statistically, eliminating data 2 standard deviations from the mean will filter out 4.4% of the data. In the Supporting Information, we have included graphs and statistics showing the data used in this study, before and after the filtering procedures described above.

Statistical preferences for side chain conformations were taken from tables compiled by Penel et al.<sup>31</sup> and are based on a set of 298 protein structures corresponding to 78 978 amino acids. The protein selection was based on a survey of the October 1997 pdb data set to select protein structures solved at a resolution higher than 2.0 Å, with an *R*-factor lower than 20%, and having less than 25% identity.<sup>32</sup>

### 3. Results and Discussion

In the studies described here, the analysis was restricted to aliphatic residues leucine, isoleucine, valine, and methionine and polar residues threonine, lysine, arginine, glutamate, and glutamine. Glycine, alanine, serine, and cysteine lack  $\text{C}\gamma$  carbons and so cannot show an analogous  $\gamma$ -effect; the side chain conformation of proline is constrained, although dynamic,<sup>33,34</sup> while residues aspartate, asparagine, phenylalanine, tyrosine, histidine, and tryptophan have quaternary  $\text{C}\gamma$  nuclei which can be subject to different shift contributions; for example, the aspartate  $\text{C}\gamma$  shift varies with protonation state.<sup>35</sup> The  $\text{C}\delta$  nuclei of the aromatic residues above are, however, expected to exhibit a conformational dependence on  $\chi_1$  and  $\chi_2$ , and this will be explored elsewhere.

In correlating  $^{13}\text{C}$  shift data with side chain conformational data derived from the protein data bank, it is important to note the significant differences between the conditions of crystallographic and NMR studies which ultimately limit the analysis. The most critical differences are temperature: generally  $\sim 100^\circ\text{K}$  for crystal structures vs  $\sim 300^\circ\text{K}$  for NMR solution studies, precipitants that cocrystallize and lattice contact interactions in the crystalline state, and the effects of buffer interactions and molecular aggregation in solution. The temperature difference is generally the most important factor. The higher temperatures required for solution state studies generally correspond to conformational mixtures, while the lower temperature of the crystallographic studies selects for the lowest energy conformer for most residues. The effects of temperature are expected to be most significant for surface residues, as discussed in greater detail for the analysis of lysine residues presented below. Further, this limitation is much more significant for the analysis of side chain conformation than for backbone conformation, which has been the object of most earlier studies.<sup>2,7,8,12,36</sup> Lattice contacts also can result in a significant conformational bias, although it is expected that analysis of a large data set will reduce most of this bias since the effects of such interactions are presumably random. The use of high concentrations of precipitants in crystallographic studies may systematically perturb the conformations of some residues; however since these occur in only a few, well-defined positions in the crystal, they probably are of limited significance. Nevertheless, motional and conformational properties can be significantly impacted by crystal packing, and breathing motions in solution can relax not



**Figure 1.** (a) Three stable rotameric conformations of threonine. The significant steric interactions are indicated by red arrows. (b) Threonine  $^{13}\text{C}\gamma$  shifts corresponding to the three stable rotameric states. Shift data were included for  $\pm 20^\circ$  ranges.

only interconversion barriers but also severe, longer range steric constraints that destabilize or even eliminate particular conformations in the crystal. Examples include Val1 in the cyclic peptide antamanide<sup>37</sup> and Val21 in the  $\beta 1$  immunoglobulin binding domain of protein G (GB1).<sup>38</sup>

**3.1.  $\chi_1$  Correlations.** Of the residues evaluated, threonine, with its single  $\gamma$ -methyl group, provides the most ideal system for evaluation of the potential significance of the  $\gamma$ -effect in protein side chains. Figure 1a summarizes the  $\chi_1$  definitions for the three stable conformers of threonine;  $\chi_1 = 180^\circ$  corresponds to the conformer in which  $\text{C}\gamma$  is *gauche* to both N and  $\text{C}'$ , while, for  $\chi_1 = \pm 60^\circ$ , one of the *gauche* substituents is a proton. Hence, threonine  $\text{C}\gamma$  is predicted to experience a relative upfield  $^{13}\text{C}$  shift for residues with  $\chi_1$  values near  $180^\circ$ . Figure 1b summarizes the data for this residue obtained from the analysis of the protein test sample. In generating this figure, only shift data within a  $40^\circ$  range, i.e., corresponding to dihedral angles of  $\pm 60^\circ \pm 20^\circ$ , and  $180^\circ \pm 20^\circ$ , were used. In addition, shift values  $> 2$  standard deviations from the mean were also deleted, in order to eliminate extreme outliers from the analysis. Based on the data shown in Figure 1, it is immediately apparent that the  $\text{C}\gamma$  resonance corresponding to  $\chi_1 = 180^\circ$  is shifted upfield. This effect can be expressed quantitatively by comparing the  $^{13}\text{C}$  shift for the conformation(s) with two *gauche* heavy atoms to the shift for conformation(s) with a single *gauche* interaction:

$$\text{Thr } \text{C}\gamma: \delta_{180} - \frac{1}{2}(\delta_{60} + \delta_{-60}) = -2.74 \text{ ppm}$$

where  $\delta_{\chi_1}$  corresponds to the mean shift corresponding to conformers with dihedral angle  $\chi_1 \pm 20^\circ$ . As is also apparent

(31) Penel, S.; Hughes, E.; Doig, A. J. *J. Mol. Biol.* **1999**, *287*, 127–143.

(32) Hobohm, U.; Sander, C. *Protein Sci.* **1994**, *3*, 522–524.

(33) London, R. E. *J. Am. Chem. Soc.* **1978**, *100*, 2678–2685.

(34) Schmidt, J. M.; Brusweiler, R.; Ernst, R. R.; Dunbrack, R. L.; Joseph, D.; Karplus, M. *J. Am. Chem. Soc.* **1993**, *115*, 8747–8756.

(35) London, R. E. *J. Magn. Reson.* **1980**, *38*, 173–177.

(36) Villegas, M. E.; Vila, J. A.; Scheraga, H. A. *J. Biomol. NMR* **2007**, *37*, 137–146.

(37) Straus, S. K.; Bremi, T.; Ernst, R. R. *J. Biomol. NMR* **1997**, *10*, 119–128.

(38) Franks, W. T.; Zhou, D. H.; Wylie, B. J.; Money, B. G.; Graesser, D. T.; Frericks, H. L.; Sahota, G.; Rienstra, C. M. *J. Am. Chem. Soc.* **2005**, *127*, 12291–12305.

**Table 1.**  $\chi_1$  Conformer  $^{13}\text{C}\gamma$  Shift Perturbations and Statistical Probabilities

residue	carbon	$\chi_1$ value corresponding to maximum steric conflict <sup>a</sup>	$\delta_{\chi_1} - \frac{1}{2}(\delta_{\chi_1'} + \delta_{\chi_1''})^a$ ppm	probability of rare $\chi_1$ conformer <sup>b</sup>
Thr	C $\gamma$ 2	180° ( <i>t</i> )	-2.74	8%
Ile	C $\gamma$ 1	60°	-2.02	13%
	C $\gamma$ 2	180°	-1.35	10%
Val <sup>c</sup>	C $\gamma$ 1	60°	-1.03	8%
	C $\gamma$ 2	-60°	-1.50	20%
Arg	C $\gamma$	60° ( $g^-$ )	-0.78	9%
Gln	C $\gamma$	60° ( $g^-$ )	-0.54	8%
Glu	C $\gamma$	60° ( $g^-$ )	-0.59	11%
Leu	C $\gamma$	60° ( $g^-$ )	-0.13	2%
Lys	C $\gamma$	60° ( $g^-$ )	-1.01	8%
Met	C $\gamma$	60° ( $g^-$ )	-0.34	9%

<sup>a</sup> The  $\chi_1$  value selected corresponds to the conformer in which both the CO and NH groups have a *gauche* orientation relative to the observed C $\gamma$ , while in orientations  $\chi_1'$  and  $\chi_1''$  only one of these groups has a *gauche* orientation relative to C $\gamma$ . <sup>b</sup> Conformer statistics based on Penel et al.;<sup>31</sup> see also <http://pbil.univ-lyon1.fr/members/penel/manchester/preferences.html>. <sup>c</sup> Val has a different convention for rotamers:  $g^-$  is -60°, *t* is 60°, and  $g^+$  is 180°.

from Figure 1b, the conformer in which both N and C' adopt a *gauche* orientation relative to C $\gamma$  tends to be the least energetically favorable and hence is statistically the least populated, so that there are typically fewer examples in the data set corresponding to this conformer (Table 1). Considering the data prior to shift filtering, an observed  $^{13}\text{C}$  shift for Thr C $\gamma$  that falls below the mean of 18.9 ppm corresponding to  $\chi_1 = 180^\circ$  serves as a correct indicator of this conformation with 64% accuracy. However, based on the raw data, this cutoff misses about half of the conformers that have  $\delta^{13}\text{C}\gamma$  greater than the mean value and also misidentifies the conformation in five cases.

All of the aliphatic C $\gamma$  nuclei in the data set showed an upfield bias in the more sterically crowded conformer. The isoleucine methylene C $\gamma$  also showed one of the larger shift perturbations:

$$\text{Ile C}\gamma 1: \delta_{60} - \frac{1}{2}(\delta_{180} + \delta_{-60}) = -2.02 \text{ ppm}$$

$$\text{Ile C}\gamma 2: \delta_{180} - \frac{1}{2}(\delta_{60} + \delta_{-60}) = -1.35 \text{ ppm}$$

For isoleucine,  $\chi_1 = 60^\circ$ , corresponding to the  $g^-$  conformer, places C $\gamma$ 1 in a *gauche* orientation relative to N and C'.

Interpretation of the shifts for the valine C $\gamma$ 1 and C $\gamma$ 2 methyl groups is based on stereospecific assignments and yields the following:

$$\text{Val C}\gamma 1: \delta_{60} - \frac{1}{2}(\delta_{180} + \delta_{-60}) = -1.03 \text{ ppm}$$

$$\text{Val C}\gamma 2: \delta_{-60} - \frac{1}{2}(\delta_{60} + \delta_{180}) = -1.50 \text{ ppm}$$

It is important to emphasize that our data are derived from only those assignments that are coded in the BMRB database as unique and, therefore, stereospecifically assigned. In general, the upfield shift observed for valine is expected to be somewhat smaller than that for threonine or isoleucine due to the greater symmetry of the side chain and hence the increased likelihood of thermal averaging of the conformations. For both isoleucine and valine, the significant differences observed between the upfield shifts of the two C $\gamma$  nuclei can be qualitatively understood in terms of thermal averaging of the different rotameric states, which is most significant for the least stable

**Table 2.** Dependence of Mean  $^{13}\text{C}$  Shift Differences on  $\chi_1$  or  $\chi_2$ 

I: Valine $\chi_1$	$\delta\text{C}\gamma 1$	$\delta\text{C}\gamma 2$	$\Delta\text{C}\gamma 12$	
			all data	Neri-selected
-60°	22.05	20.1	1.95	1.31 ± 2.8
+60°	20.87	21.23	-0.36	0.05 ± 1.7
180°	21.74	21.97	-0.26	-1.01 ± 1.5

II: Leucine $\chi_2$	$\delta\text{C}\delta 1$	$\delta\text{C}\delta 2$	$\Delta\text{C}\delta 12$	
			all data	Neri-selected
-60° <sup>aa</sup>	24.9	26.2	-1.3	0.093
+60°	24.45	25.79	-1.34	-2.3 ± 1.6
180°	25.17	23.84	1.33	3.0 ± 0.5

<sup>a</sup> The leucine database included only two examples corresponding to  $\chi_2 = -60^\circ$  and one example for the Neri-selected group that was stereospecifically assigned using isotopic labeling.<sup>28</sup>

rotamer. This effect is discussed in greater detail for the case of leucine considered below.

Assuming that some of the other factors influencing the  $^{13}\text{C}$  shift will be similar or partially averaged for the two valine C $\gamma$  nuclei, we can consider the shift difference  $\Delta\text{C}\gamma 12 = \delta\text{C}\gamma 1 - \delta\text{C}\gamma 2$  as a potentially useful parameter for the analysis of valine side chain conformation. Correlation of mean shift values with  $\chi_1$  indicates that  $\Delta\text{C}\gamma 12 = 1.95$  for  $\chi_1 \approx -60^\circ$  but is near 0 (no correlation) for  $\chi_1 = 60^\circ$  and  $180^\circ$  (Table 2).

Since statistically  $\chi_1 \approx 60^\circ$  corresponds to only 8% of the valine rotamers, small values for  $\Delta\text{C}\gamma 12$  are more likely to correspond to the  $\chi_1 = 180^\circ$  ( $g^+$ ) conformer. Interestingly, an even larger  $\Delta\text{C}\gamma 12$ ,  $\chi_1$  correlation is observed for malate synthase G, the largest protein analyzed in detail; of the 41 assigned valine residues, 39 could be correlated with dihedral angles in the crystal structure, 1D8C.<sup>39,40</sup> The parameter  $\Delta\text{C}\gamma 12 = 2.86, -0.10, \text{ and } -1.11$  for  $\chi_1 \approx -60^\circ, +60^\circ, \text{ and } 180^\circ$ , respectively. Based on this effect, we also tabulated the values for a set of five proteins (see methods) for which the methyl groups were stereospecifically assigned based on isotopic labeling.<sup>28</sup> These results extend the correlations with  $\chi_1 \approx +60^\circ$  still close to 0, but  $\chi_1 \approx 180^\circ$  corresponding to a more negative shift difference of -1 ppm (Table 2). The correlation apparent in Table 2 is observed for the Val1 rotamers in the cyclic peptide antamanide, which can be individually observed in the solid state.<sup>37</sup> NOE data were used to assign the resonances to two rotameric forms: (Rotamer I)  $\chi_1 = -60^\circ$ ;  $\delta\text{C}\gamma 1, \delta\text{C}\gamma 2 = 20.5, 16.1$  and (Rotamer II)  $\chi_1 = 180^\circ$ ;  $\delta\text{C}\gamma 1, \delta\text{C}\gamma 2 = 19.5, 17.1$ . Thus, the transition from  $g^-$  to  $g^+$  conformations corresponds to  $\Delta\text{C}\gamma 1 = -1.0$  ppm and  $\Delta\text{C}\gamma 2 = +1.0$  ppm.

The data for aspartate and asparagine were very limited. Since the side chain C $\gamma$  shift of Asp and Asn residues rarely contains any information for structure determination, it is not usually reported or assigned. More importantly, the carbonyl C $\gamma$  carbons fall outside the aliphatic hydrocarbon originally evaluated by Grant and Cheney, so that there would be no predictions for these residues.<sup>20</sup> The limited data available indicate either no significant shift or even a downfield shift for the most sterically crowded  $\chi_1 = 60^\circ$  conformer. In addition, the shift for the aspartate C $\gamma$  will be dependent on the titration state of the residue, which will vary significantly, depending on the individual pK values and the pH of the study.<sup>35</sup>

(39) Tugarinov, V.; Kay, L. E. *J. Am. Chem. Soc.* **2004**, *126*, 9827–9836.  
(40) Howard, B. R.; Endrizzi, J. A.; Remington, S. J. *Biochemistry* **2000**, *39*, 3156–3168.

**Table 3.**  $\chi_2$  and  $\chi_3$  Shift Correlations

residue	carbon	steric shift calculation <sup>a</sup>	value	probability of $\chi_2$ conformer <sup>b</sup>		
				<i>g</i> <sup>-</sup>	<i>g</i> <sup>+</sup>	<i>t</i>
$\chi_2$ correlations						
Ile	C $\delta$ 1	$\delta_{-60} - \frac{1}{2}(\delta_{60} + \delta_{180})$	-2.73	10	15	75
Leu	C $\delta$ 1	$\delta_{60} - \delta_{180}$	-0.72	43	5	52
	C $\delta$ 2	$\delta_{180} - \delta_{60}$	-1.95	43	5	52
Arg	C $\delta$	$\frac{1}{2}(\delta_{-60} + \delta_{60}) - \delta_{180}$	-0.36	9	12	79
Lys	C $\delta$	$\frac{1}{2}(\delta_{-60} + \delta_{60}) - \delta_{180}$	-0.38	12	16	72
$\chi_3$ correlations						
Met	C $\epsilon$	$\frac{1}{2}(\delta_{60} + \delta_{-60}) - \delta_{180}$	-1.78			
Lys	C $\epsilon$	$\frac{1}{2}(\delta_{60} + \delta_{-60}) - \delta_{180}$	-0.18			

<sup>a</sup> The shift calculations subtract the shift values for less sterically crowded conformation(s) from the values for the more sterically crowded conformation(s). In most cases, this corresponds to two conformers with a single *gauche* heavy atom, compared with a conformer with no *gauche* heavy atoms. For the case of leucine, only two conformations are significantly populated. <sup>b</sup>  $\chi_2$  conformations were determined from the Penel et al.<sup>31</sup> Although the *g*<sup>+</sup> conformation for leucine is predicted to experience the greatest steric crowding and hence the greatest upfield shift, there is insufficient data corresponding to this conformation to allow this determination.

The results for the remaining  $\chi_1$ , C $\gamma$  shift correlations of Arg, Glu, Gln, Leu, Lys, and Met are summarized in Table 1. For the group of residues shown in Table 1, there are apparently no exceptions to the prediction that the conformation characterized by the most significant steric conflict shows lowest statistical probability and the greatest upfield C $\gamma$  shift. In a few cases, notably for leucine and methionine, there was so little data that corresponded to the sterically disfavored conformer that any numerical value for the upfield  $\gamma$ -shift should be regarded as only a qualitative indicator. A more serious limitation on the data results from the assumptions of single rotamer occupancy for all side chains, including those on the protein surface. This is particularly the case for charged residues, since most of these are located on the surface of the protein.

**3.2.  $\chi_2/\chi_3$  Correlations.** Conformational- $^{13}\text{C}$  shift correlations for  $\chi_2$  are summarized in Table 3. For the case of leucine, as for several other residues, the shift pattern changes in that, for each  $\delta$  methyl group, two possible conformations contain a single, *gauche* heavy atom, while in the third conformation both *gauche* atoms are hydrogen. Comprehensive analyses of the protein data bank indicate that leucine side chains exist predominantly in only two rotameric states:  $\chi_1 = -60^\circ$ ,  $\chi_2 = 180^\circ$ , and  $\chi_1 = 180^\circ$ ,  $\chi_2 = +60^\circ$ .<sup>30,31</sup> Consistent with this expectation, in our analysis of stereospecifically assigned leucines there are only two examples where the structure indicates  $\chi_2 = -60^\circ$ . Further, solid-state NMR results indicate that there can be rapid interconversion between the two significantly populated states,<sup>41,42</sup> although the motion may be more complex than described by this simple two-site jump model.<sup>43</sup> Macarthur and Thornton concluded that the X-ray structures are generally not able to distinguish one conformation from the other, particularly when low resolution results in ambiguity.<sup>30</sup> Dunbrack and Karplus had earlier noticed similar problems in the interpretation of electron density maps for leucine residues.<sup>44</sup> We therefore have analyzed the leucine data

based on the assumption of only two significant populated side chain conformational states, giving the following:

$$\text{Leu C}\delta 1: \delta_{\chi_2=60} - \delta_{\chi_2=180} = -0.72 \text{ ppm}$$

$$\text{Leu C}\delta 2: \delta_{\chi_2=180} - \delta_{\chi_2=60} = -1.95 \text{ ppm}$$

where the above calculation is based on the fact that, for C $\delta$ 1, the  $\chi_2 = 60^\circ$  conformation exhibits greater steric crowding, while, for C $\delta$ 2, the  $\chi_2 = 180^\circ$  conformation shows more crowding, and the above correlations assume that  $\chi_2$  can adopt only two different values.

It is interesting to note that, in the case of leucine, the apparent steric shift differences for the two C $\delta$  carbon resonances differ by more than a factor of 2. The most probable explanation for this large variation relates to the fact noted above that, in solution at  $\sim 300^\circ\text{K}$ , we are not looking at pure conformations. More specifically, we can predict this result based on the assumption that the effect of the higher temperature is to go from a protein structure corresponding to a global energy minimum to a structure in which additional conformational states are populated based on the local minima that characterize the individual residues. In the following discussion, we assume that the local stability of the rotamer conformations is correlated with the observed rotamer probabilities given by Penel et al.<sup>31</sup> For the case of leucine, this implies that the *trans* conformation corresponding to  $\chi_2 = 180^\circ$  is more stable (Table 3) and hence more likely than the *gauche* conformation to persist in solution at higher temperature. Alternatively, a residue observed to adopt the  $\chi_2 = 60^\circ$  conformation in the crystal will have a greater tendency to equilibrate with the  $\chi_2 = 180^\circ$  rotamer in solution. The averaging of the shifts between states can be modeled quantitatively by amending the relation given above to the following:

$$\text{Leu C}\delta 1: [p_1\delta_{\chi_2=60} + (1-p_1)\delta_{\chi_2=180}] - [p_1\delta_{\chi_2=180} + (1-p_1)\delta_{\chi_2=60}]$$

$$\text{Leu C}\delta 2: [p_2\delta_{\chi_2=180} + (1-p_2)\delta_{\chi_2=60}] - [p_2\delta_{\chi_2=60} + (1-p_2)\delta_{\chi_2=180}]$$

The above expressions for the thermally averaged results become equivalent to the nonthermally averaged result as  $p_1, p_2 \rightarrow 1$  and are identically equal to 0 if there is complete thermal averaging ( $p_1 = 0.5$ ;  $p_2 = 0.5$ ). In the above expressions,  $p_1$  is the probability that a nominal  $\chi_2 = 60^\circ$  conformation identified in the crystal remains in this rotameric state in solution at high temperature, and  $p_2$  is the probability that a nominal  $\chi_2 = 180^\circ$  conformation identified in the crystal remains in the *trans* conformation at high temperature. Given that statistically the  $\chi_2 = 180^\circ$  conformation is more stable,  $p_2 > p_1$ , consistent with a larger upfield shift for Leu C $\delta$ 2. As an illustrative calculation, if we assume that the upfield shift experienced by a leucine methyl carbon in a *gauche* conformation in the absence of rotameric averaging,  $\Delta\text{C}$ , is the same for both C $\delta$  methyl carbons, the above relations simplify to

$$\text{Leu C}\delta 1: \Delta\text{C}(2p_1 - 1)$$

$$\text{Leu C}\delta 2: \Delta\text{C}(2p_2 - 1)$$

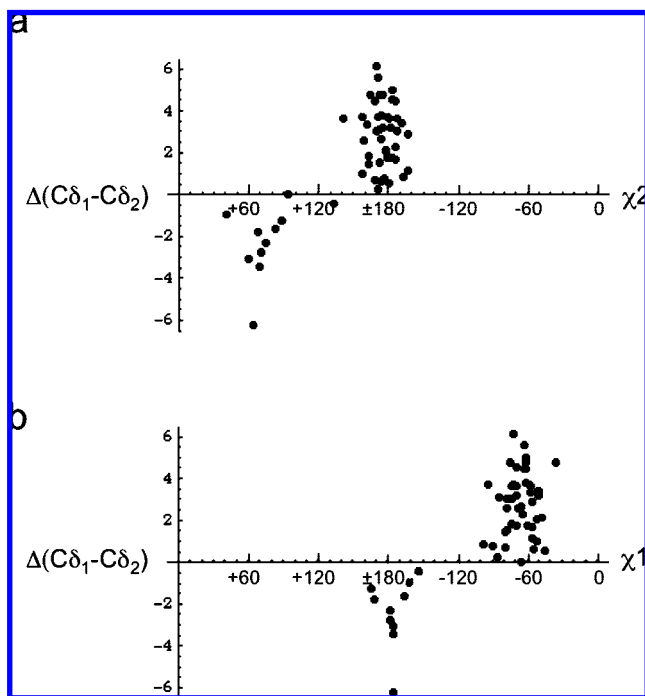
Setting  $\Delta\text{C} = -4$  ppm,  $p_1 = 0.59$ , and  $p_2 = 0.75$  gives a Leu C $\delta$ 1 shift of  $-0.72$  ppm and a Leu C $\delta$ 2 shift of  $-2.0$  ppm, similar to the observed values given above. In addition to predicting that the actual  $\gamma$ -effect shift contribution is significantly larger than the observed values, this model is consistent with NMR data showing interconversion of the two stable

(41) Batchelder, L. S.; Sullivan, C. E.; Jelinski, L. W.; Torchia, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 386–389.

(42) Colnago, L. A.; Valentine, K. G.; Opella, S. J. *Biochemistry* **1987**, *26*, 847–854.

(43) Yang, D. W.; Mittermaier, A.; Mok, Y. K.; Kay, L. E. *J. Mol. Biol.* **1998**, *276*, 939–954.

(44) Dunbrack, R. L.; Karplus, M. *J. Mol. Biol.* **1993**, *230*, 543–574.



**Figure 2.** Analysis of leucine methyl shifts in malate synthase G. (A) Variation of  $\Delta C\delta_{12}$  with  $\chi_2$  and (b) correlation with  $\chi_1$ . As is apparent from the figure, most of the observed conformations correspond to  $\chi_1 = -60^\circ$ ,  $\chi_2 = 180^\circ$ , with a smaller proportion corresponding to  $\chi_1 = 180^\circ$ ,  $\chi_2 = 60^\circ$ . Data were analyzed for 54 leucine residues for which both NMR shift data<sup>39</sup> and crystallographic  $\chi_1/\chi_2$  data<sup>40</sup> were available.

leucine conformations.<sup>41,42</sup> Qualitatively similar effects also are expected to be present for the  $C\gamma$  nuclei of valine and isoleucine, although, in these cases, the effects result from mixing three conformational states. In general, the carbon for which the state with the greatest number of *gauche* interactions is the least stable statistically will show the most significant reduction in shift due to averaging with more stable conformers. For valine, this corresponds to  $C\gamma_1$  ( $\chi_1 = 60^\circ$ ), and for isoleucine this corresponds to  $C\gamma_2$  ( $\chi_1 = 180^\circ$ ).

Returning to the crystallographic comparisons for leucine, we note that the very strong correlation between  $\chi_1$  and  $\chi_2$  for this residue implies that the  $C\delta$  shifts should also be dependent on  $\chi_1$ .<sup>31</sup> Correlating the mean  $C\delta$  shifts with the  $\chi_1$  dihedral angles yields

$$\text{Leu } C\delta_1: \delta_{\chi_1=180} - \delta_{\chi_1=60} = -1.06 \text{ ppm}$$

$$\text{Leu } C\delta_2: \delta_{\chi_1=60} - \delta_{\chi_1=180} = -2.17 \text{ ppm}$$

As expected, the larger  $\gamma$ -effect shift correlation for  $C\delta_2$  remains even when the analysis is based on  $\chi_1$ .

As in the case of valine discussed above, we can consider the parameter  $\Delta C\delta_{12} = \delta C\delta_1 - \delta C\delta_2$  as a potentially useful indicator of the leucine side chain conformation, if stereo assignments are known. The strong conformational coupling of the two dihedral angles results in pronounced correlations between  $\Delta C\delta_{12}$  and both  $\chi_2$  and  $\chi_1$ . Plots of  $\Delta C\delta_{12}$  for the leucine residues in malate synthase G as a function of both  $\chi_2$  and  $\chi_1$  are shown in Figure 2.<sup>39</sup> For the  $\chi_2$  dependence, the parameter  $\Delta C\delta_{12} = 1.33$  and  $-1.34$  ppm for  $\chi_2 \approx 180^\circ$  and  $+60^\circ$ , respectively (Table 2). For the complete data set, there were only two examples with  $\chi_2 = -60^\circ$ , consistent with the above discussion. An even more pronounced dependence was observed for the set of five proteins stereospecifically assigned

by isotopic labeling (Table 2). In this case, the sensitivity of the  $\Delta C\delta_{12}$  parameter to  $\chi_1$  or  $\chi_2$  exceeds the standard deviation of the value. The limited accessibility of conformational space and the large dependence of  $\Delta C\delta_{12}$  on dihedral angles support the use of this parameter for the analysis of conformational changes in leucine residues. It is likely that many of the leucine residues exhibiting larger positive or negative  $\Delta C\delta_{12}$  values correspond to purer conformational states, while residues exhibiting values closer to 0 are more likely to correspond to nearly equal rotamer mixtures.

For methionine, lysine, and arginine, the  $C\delta$  is *gauche* to  $C\alpha$  for  $\chi_2 = \pm 60^\circ$  and *gauche* to two hydrogen atoms for  $\chi_2 = 180^\circ$ . Hence, the effect of steric crowding should be evidenced as a negative value for  $\frac{1}{2}(\delta_{-60} + \delta_{60}) - \delta_{180}$ . Since, in methionine, the  $\delta$  position is occupied by a sulfur atom, we only obtain data for lysine and arginine  $C\delta$  atoms (Table 3), both of which show modest upfield shifts.

$$\text{Lys } C\delta: \frac{1}{2}(\delta_{60} + \delta_{-60}) - \delta_{180} = -0.38 \text{ ppm}$$

$$\text{Arg } C\delta: \frac{1}{2}(\delta_{60} + \delta_{-60}) - \delta_{180} = -0.36 \text{ ppm}$$

As discussed above, the small shift differences here undoubtedly relate to the fact that these two charged residues are located predominantly on the surface and, in contrast with the low temperature crystallographic studies, the NMR studies correspond to significant conformational mixtures. For a few cases in which these residues are involved in salt bridges which constrain the conformations, as in the example of DNA polymerase  $\lambda$  discussed below, a larger effect can be observed. These steric interactions are also expected to produce upfield shifts for  $C\alpha$ . However, the  $^{13}C\alpha$  shifts tend to be dominated by the backbone angle dependences that have been extensively studied, so we have not explored these correlations further.

In the calculation of shift correlations related to Met  $\chi_3$ , we noted a significant variability dependent on whether the mean or median shift value was used. This was traced to three extreme shift values for Met  $C\epsilon$ : Met1,  $\delta^{13}C = 26.99$  ppm (Influenza A virus NS1 protein, pdb code 1AIL BMRB, code 4317; Met60,  $\delta^{13}C = 37.53$  and Met 136,  $\delta^{13}C = 32.53$  ppm, human fibroblast collagenase, pdb code 1HFC, BMRB code 4064). Additionally, shift data for plastocyanin (pdb code 1PLC, BMRB code 4019) was found to correspond frequently to outliers, as in the case of Met57 ( $\delta^{13}C = 13.8$  ppm). The high values above were considered to represent assignment problems, while the low value for plastocyanin was considered to indicate a difference in referencing, and so these values were dropped from the analysis.

The  $\gamma$ -shift perturbation observed for methionine is surprisingly large, particularly in view of the fact that the order parameters for methionine  $C\epsilon$  are typically very low.<sup>45</sup> Hence, this value may reflect the paucity of shift data available for this carbon, which is not readily assigned using standard techniques. Interestingly, the  $N\epsilon$  shift of arginine showed a large  $\gamma$ -shift effect of  $-0.87$  ppm, although the original Grant–Cheney analysis did not extend to noncarbon atoms and there were only 24 data points available for analysis.<sup>20</sup>

**3.3. Applications.** It is anticipated that the  $^{13}C$  shift,  $\chi_1$  correlations described here are most useful for the analysis of

(45) Skrynnikov, N. R.; Millet, O.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 6449–6460.

protein conformational changes. Several illustrative examples are given below.

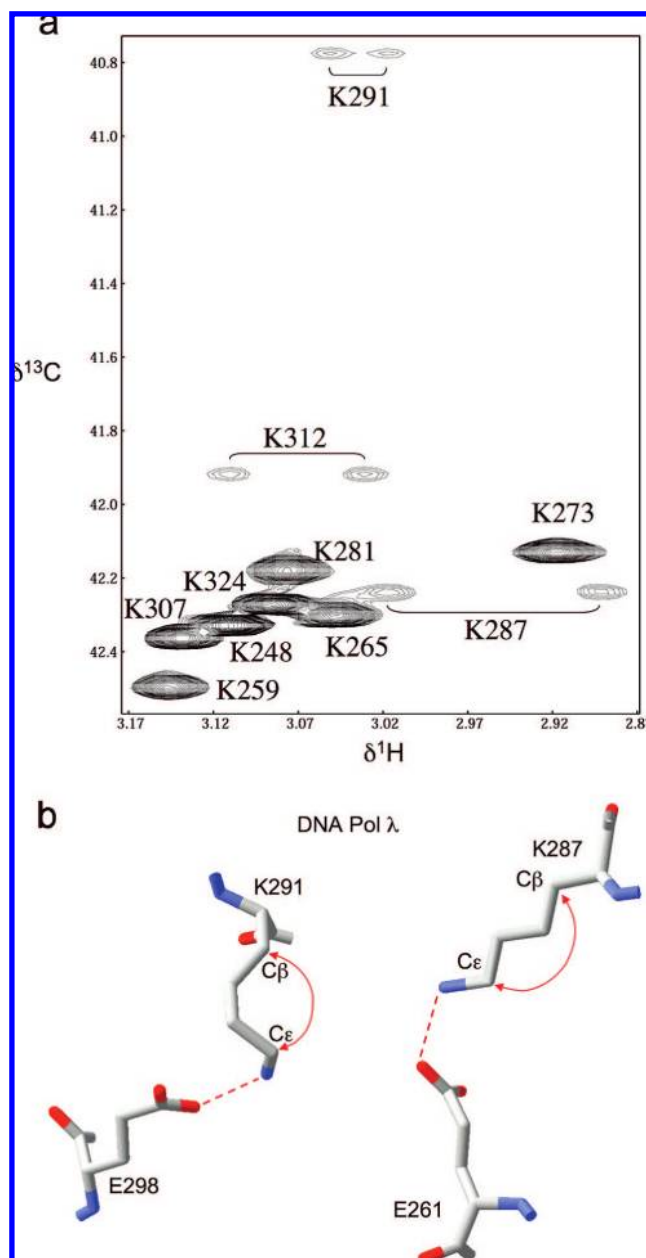
**3.3.1. DNA Polymerase  $\lambda$  Lyase Domain.** In general, the polymethylene chain of lysine represents the most tractable structure for revealing correlations between  $^{13}\text{C}$  shift values and molecular conformation in proteins. However, the statistical data for lysine are expected to be weak for two reasons. First the lower temperature ( $\sim 100$  vs  $300$  K) of most X-ray data will result in populating the lowest energy state to a much greater extent than is achieved in the NMR experiment. As an estimate of this effect, the *gauche/trans* energy difference for a polymethylene chain is approximately  $0.5$  kcal/mol.<sup>46</sup> Using the following relations

$$\frac{p_g}{p_t} = e^{-\Delta E/RT}$$

$$2p_g + p_t = 1$$

yields an 86% *trans* conformation probability at  $100$  K compared with a 54% *trans* probability at  $300$  K. Second, lysines are commonly located on the surface where there is generally a lack of conformational restraint in solution while in a crystal there may be conformational bias due to lattice contacts. Therefore the analysis of lysine shift and structural data is expected to be problematic. Nevertheless, useful examples of the conformational dependence of lysyl  $^{13}\text{C}$  shifts exist in the literature. Studies of  $[\epsilon\text{-}^{13}\text{C}]$  lysine-labeled lyase domain from DNA polymerase  $\lambda$  indicate that the  $^{13}\text{C}$  shifts for 9 of the 10 lysine residues in the domain cluster near  $\delta^{13}\text{C} = 42.2$  ppm, while the shift for K291 is  $40.8$  ppm (Figure 3a).<sup>47</sup> Although the  $^{13}\text{C}$  shift is also sensitive to the titration state of the  $\epsilon$ -amino group, deprotonation leads to a downfield rather than an upfield shift for  $\text{C}\epsilon$ , and in addition, the  $\text{pK}$  of this residue was found to be typical for surface-exposed lysine residues, so that the  $-1.4$  ppm shift difference could not be explained on this basis. Instead, an analysis of the crystal structure of Pol  $\lambda$  (pdb code: 1RZT) indicates a  $\chi_3 = -77.2$  in the  $g^+$  conformation as a result of a salt bridge with E298.<sup>48</sup> The other salt-bridged lysine residue, K287, is held in an all *trans* configuration (Figure 3b). Presumably, the other surface exposed lysine residues, which include all but Lys312, exist as *gauche/trans* mixtures, with the ratio determined mostly by the *gauche/trans* energy difference and other inter-residue steric effects. Although there were no useful inter-residue NOE cross peaks to help position K291 in this study, the notable upfield shift for  $\text{C}\epsilon$  indicates that the side chain conformation is constrained to have a *gauche* conformation for  $\chi_3$ .<sup>49</sup>

**3.3.2. Dihydrofolate Reductase.** Dihydrofolate reductase, a target of both chemotherapeutic and bacterial antifolate drugs, is one of the most extensively studied enzymes. One of the interesting conformational aspects of DHFR activity concerns the interaction of the "Met20 loop" located at the active site and the  $\beta\text{F}$ - $\beta\text{G}$  loop connecting  $\beta$  strands F and G. This second loop comes in contact with the Met20 loop and appears to influence ligand binding. NMR relaxation studies and mutational analysis have supported the conclusion that the  $\beta\text{F}$ - $\beta\text{G}$  loop



**Figure 3.** (a)  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectrum of  $[\epsilon\text{-}^{13}\text{C}]$ lysine polymerase  $\lambda$  lyase domain (residues 242–327) obtained on a Unity INOVA 800. Three of the residues, K287, K291, and K312, show inequivalent  $\epsilon$  protons under the conditions of the study. Other sample conditions were as follows: 100 mM KCl, 2 mM deuterated DTT, 0.1%  $\text{NaN}_3$  (w/v), 90%  $\text{H}_2\text{O}$  and 10%  $^2\text{H}_2\text{O}$  (v/v), and  $50 \mu\text{M}$  DSS, used as a chemical shift standard.<sup>47</sup> (b) Structural data showing salt bridge interactions for K287–E261 and K291–E298, taken from structure 1RZT.<sup>48</sup>

plays a significant catalytic role by influencing the conformational distribution of the Met20 loop which in turn modulates ligand affinity and relative orientation.<sup>50,51</sup> For example, deletion of residue G121 in the  $\beta\text{F}$ - $\beta\text{G}$  loop reduced the binding affinity for NADPH nearly 3-fold more than the affinity for  $\text{NADP}^+$ . Osborne et al. recently have determined the side chain  $^{13}\text{C}$  shifts for many of the residues of *E. coli* DHFR in two ternary complexes: DHFR•Folate• $\text{NADP}^+$  (complex I), corresponding to a closed conformation of the Met20 loop, and DHFR•

(46) Scherer, J. R.; Snyder, R. G. *J. Chem. Phys.* **1980**, *72*, 5798–5808.

(47) Gao, G. H.; DeRose, E. F.; Kirby, T. W.; London, R. E. *Biochemistry* **2006**, *45*, 1785–1794.

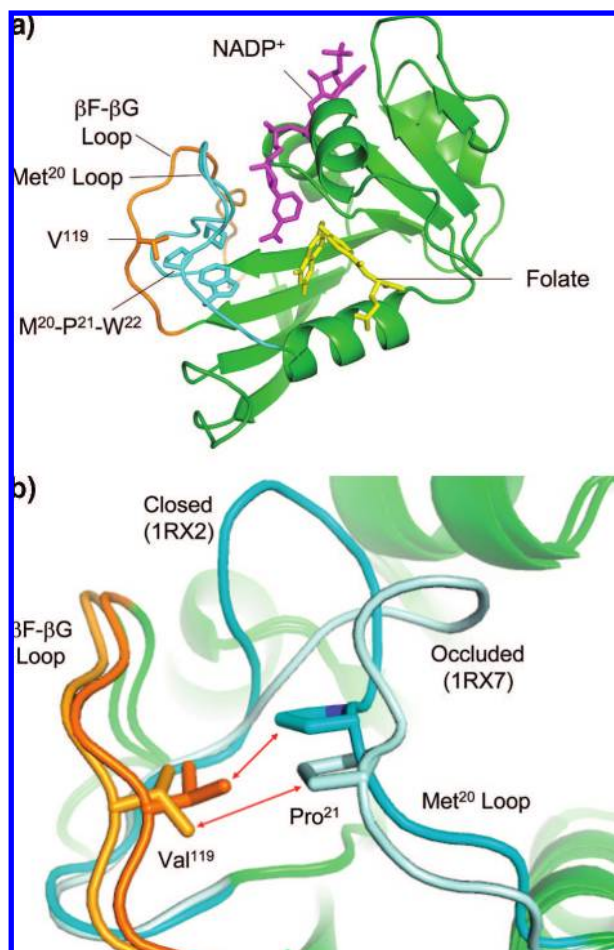
(48) Garcia-Diaz, M.; Bebenek, K.; Krahn, J. M.; Blanco, L.; Kunkel, T. A.; Pedersen, L. C. *Mol. Cell* **2004**, *13*, 561–572.

(49) DeRose, E. F.; Kirby, T. W.; Mueller, G. A.; Bebenek, K.; Garcia-Diaz, M.; Blanco, L.; Kunkel, T. A.; London, R. E. *Biochemistry* **2003**, *42*, 9564–9574.

(50) Miller, G. P.; Benkovic, S. J. *Biochemistry* **1998**, *37*, 6327–6335.

(51) Miller, G. P.; Benkovic, S. J. *Biochemistry* **1998**, *37*, 6336–6342.





**Figure 4.** Structure of *E. coli* dihydrofolate reductase showing the Met20 and  $\beta$ F- $\beta$ G loops. (a) Structure 1RA2<sup>53</sup> corresponding to the DHFR•Folate•NADP<sup>+</sup> complex. The Met20 loop is indicated in cyan, and the  $\beta$ F- $\beta$ G loop, in orange. NADP<sup>+</sup> is shown in magenta, and folate, in yellow. Sidechains for V119 on the  $\beta$ F- $\beta$ G loop and several residues in the Met20 loop are also explicitly indicated. (b) Superposition of active site loops in structures 1RX2 (corresponding to a ternary DHFR•Folate•NADP<sup>+</sup> complex in which the Met20 loop adopts a closed conformation) and 1RX7 (corresponding to a binary DHFR•Folate complex with the Met20 loop in an occluded conformation). Both structures show residues 9–24 of the Met 20 loop and 117–122 of the  $\beta$ F- $\beta$ G loop, as well as the sidechains for P21 and V119. The proximity of the P21-V119 sidechains is significantly greater in the closed conformation of the Met20 loop, which is presumably the basis for the observed shift perturbation and corresponding conformational change of V119.

Folate•DHNADPH (complex II), corresponding to an occluded conformation of the Met20 loop (Figure 4).<sup>52</sup> DHNADPH corresponds to 5,6-dihydroNADPH, a stabilized, reduced co-factor analogue. Evaluation of side chain methyl <sup>13</sup>C shifts indicates that the only valine residue assigned in both structures for which the C $\gamma$  values show a significant variation is V119, located in the  $\beta$ F- $\beta$ G loop. The C $\gamma$ 1, C $\gamma$ 2 shifts for Val119 are 21.59, 18.64 and 20.89, 20.23, for complexes I and II, respectively. Although the methyl groups were not stereospecifically assigned, the presumed C $\gamma$ 2 resonance undergoes a significant  $-1.59$  ppm upfield shift in complex I relative to complex II, and the parameter  $\Delta$ C $\gamma$ <sub>12</sub> decreases from  $\Delta$ C $\gamma$ <sub>12</sub><sup>I</sup> = 2.95 to  $\Delta$ C $\gamma$ <sub>12</sub><sup>II</sup> = 0.66. Based on the data analysis presented above, the shifts correspond to an increased probability for the *g*<sup>-</sup>,  $\chi$ 1 =  $-60^\circ$  conformer in complex I, with complex II

corresponding to  $\chi$ 1 values of  $180^\circ$  or  $60^\circ$ . Since the  $\chi$ 1 =  $+60^\circ$  conformation for valine is relatively rare (8%), the V119 conformational switch is probably from  $\chi$ 1 =  $-60^\circ$  to  $180^\circ$  in going from the complex with NADP<sup>+</sup> to DHNADPH and from a closed to an occluded conformation of the Met20 loop. As can be seen from a comparison of the two loop conformations in Figure 4b, the Val119 side chain comes in close contact with Pro21 on the Met20 loop in the closed conformation, while Pro21 moves away from Val119 when the Met20 loop adopts the occluded conformation. Evaluation of the Val119 conformation for all of the crystal structures described by Sawaya and Kraut<sup>53</sup> indicates that in most of the structures observed in the *P*<sub>21</sub>*2*<sub>1</sub>*2*<sub>1</sub><sup>B</sup> space group, which was considered the most reliable basis for interpreting the relationship of ligand binding to conformation due to the lower number of lattice contacts for the flexible loops, Val119 has an eclipsed conformation. This result probably indicates a dynamic averaging process. Nevertheless, for a few of the closed conformations (1RX1, 1RX9) Val119 adopts a *g*<sup>-</sup>,  $\chi$ 1 =  $-60^\circ$  conformation, and for most of the observed occluded conformations, which correspond to the *P*<sub>61</sub> and *P*<sub>65</sub> space groups, Val119  $\chi$ 1  $\approx 180^\circ$ , consistent with the conclusions of the <sup>13</sup>C shift analysis. A similar conformational conclusion also can be derived from analysis of coupling constant data.<sup>54</sup>

#### 4. Conclusions

Correlation of the <sup>13</sup>C shifts for protein side chains with the  $\chi$ <sub>*i*</sub> values derived from the protein data bank indicates a readily observed  $\gamma$ -effect, according to which an upfield shift is observed for the most sterically hindered conformation. It is likely that these correlations significantly understate the magnitude of the effect, since at 300 K most side chains, particularly those on the protein surface, populate multiple conformational states, so that the measured values do not correspond to the pure conformational states indicated in the pdb. Although the details of the mechanism remain unclear, the steric origin of the shifts is consistent with the strong correlation between lower (upfield) C $\gamma$  shifts and lower statistical conformational probability.<sup>31</sup> As noted above, qualitatively similar predictions of  $\gamma$ -effect shifts in *N*-formylvaline have been determined by Pearson et al.<sup>12</sup> Interestingly, in these calculations the shielding is always greatest near  $\chi$ 1 =  $\pm 60^\circ$  and falls off significantly at  $\chi$ 1 =  $0^\circ$ , a result which is not readily explained by simple steric arguments. The side chain shift data also have some absolute predictive values for the analysis of  $\chi$ 1 values, and the more extensive shift data included in the PREDITOR torsion angle prediction program yield  $\chi$ 1 values that agree with the crystallographic results in 64% of the cases, for predictions based only on chemical shift data.<sup>16</sup> Density functional studies of Villegas et al. also have indicated that  $\delta^{13}\text{C}\alpha$  and  $\delta^{13}\text{C}\beta$  show a significant dependence on  $\chi$ 1 and a smaller but significant dependence on  $\chi$ <sub>*i*</sub>>1 for 11 of 19 residues.<sup>36</sup> The most significant side chain <sup>13</sup>C shift/conformation correlation observed in the present study corresponds to the  $\Delta$ C $\delta$ <sub>12</sub> parameter for leucine. Although not all of the proteins gave as well separated data as malate synthase G (Figure 2), assignment of the leucine side chain conformation to one of the two stable alternatives based on the sign of the  $\Delta$ C $\delta$ <sub>12</sub> parameter was consistent with crystallographic data >92% of the time for the five protein

(52) Osborne, M. J.; Venkitakrishnan, R. P.; Dyson, H. J.; Wright, P. E. *Protein Sci.* **2003**, *12*, 2230–2238.

(53) Sawaya, M. R.; Kraut, J. *Biochemistry* **1997**, *36*, 586–603.

(54) Schnell, J. R.; Dyson, H. J.; Wright, P. E. *Biochemistry* **2004**, *43*, 374–383.

sample set stereospecifically assigned by isotopic labeling, and it is likely that values close to 0 often correspond to nearly equal concentrations of interconverting conformers. For this data set, most of the discrepancies correspond to  $\chi_2$  values outside of the usual ranges. The corresponding  $\Delta C\gamma_{12}$  parameter for valine is less reliable; omitting the rare *t* conformer as well as a few noncanonical conformers, the  $g^-/g^+$  conformers are correctly predicted in 80% of the cases. Other residues showing a modest correlation included Ile  $\delta C\gamma_1 < 26.0$  ppm implying a  $g^+$  or  $g^-$  conformer, Ile  $\delta C\delta < 11.6$  ppm corresponding to  $\chi_2 = -60^\circ$ , and Thr  $\delta C\gamma < 19.6$  ppm corresponding to  $\chi_1 = 180^\circ$ . In most of these cases, there is a significant number of both false negatives, i.e. similar  $\chi_1$  values that fail to satisfy the shift cutoff, and false positives, other conformations satisfying the shift cutoff criteria. Further, as noted above, most side chains do not adopt a single conformation under typical solution state conditions.

These correlations should be of general use for the interpretation of conformational changes in proteins, particularly in cases for which NOE and/or coupling data are not available. The dramatic upfield shift observed for the  $C_\epsilon$  of K291 in DNA polymerase  $\lambda$  lyase domain is readily interpreted on the basis of the *gauche* orientational constraint resulting from the salt bridge, while there is no apparent alternative explanation for such an unusual shift. In this case, no inter-residue NOE constraints could be observed that would indicate the residue conformation and the involvement in a salt bridge. Hence, for this example, the  $^{13}\text{C}$  shift gives useful structural information not readily obtained from NOE restraints. Evaluation of available data for *E. coli* dihydrofolate reductase indicated that Val119 on the  $\beta\text{F}-\beta\text{G}$  loop that is not in direct contact with the pyridine nucleotide cofactor is subject to subtle conformational influences that depend on the redox state of the bound pyridine nucleotide.

The statistically more stable  $180^\circ$  conformation is correlated with binding of the reduced cofactor analogue, DHNADPH.

Given the current popularity of introducing "selective methyl reprotonation" in large, deuterated proteins,<sup>55</sup> and the fact that some of the clearest correlations are obtained for threonine and the aliphatic residues, we believe that such shift perturbations provide a useful basis for analyzing conformational changes in these systems. Further, such conformationally related shift data can be derived independently based on relaxation measurements,<sup>56</sup> so that it may be possible to associate dynamic behavior with specific excited-state conformational changes. Further, the large correlation obtained in the case of leucine  $\Delta C\delta_{12}$  with  $\chi_1$  and  $\chi_2$  is sufficient to allow this parameter to be used as a restraint in structure calculations, if stereospecific assignment data are available.

**Acknowledgment.** This research was supported by the Intramural Research Program of the NIH, National Institute of Environmental Health Sciences. The authors wish to thank Dr. Ron Venters and Dr. Eugene DeRose for a critical reading of the manuscript.

**Supporting Information Available:** Raw and filtered shift data as a function of  $\chi_i$  and accompanying Newman projections are provided as Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

JA802729T

(55) Goto, N. K.; Gardner, K. H.; Mueller, G. A.; Willis, R. C.; Kay, L. E. *J. Biomol. NMR* **1999**, *13*, 369–374.

(56) Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Am. Chem. Soc.* **2002**, *124*, 12352–12360.