

**Smooth statistical torsion angle potential derived from a large  
conformational database via adaptive kernel density estimation  
improves the quality of NMR protein structures**

Guillermo A. Bermejo, G. Marius Clore, and Charles D. Schwieters

**Supporting Information**

**Table SI.** Probability Density Expressions Extracted from the Torsion Angle Database

<b>Residue Type</b>	<b>Residues in Database <sup>a</sup></b>	<b>Probability Density Function <sup>b</sup></b>	<b>Statistical Approximation <sup>c</sup></b>	<b>Energy Term Applied To: <sup>d</sup></b>
Gly	All glycines (93,113)	$p(\phi, \psi)$	None	All glycines
cis-Pro	All cis-prolines (2,343)	$p(\phi, \psi, \chi_2)$	$\chi_2$ determines other ring torsion angles	All cis-prolines
trans-Pro	All trans-prolines (46,686)	$p(\phi, \psi, \chi_2)$	$\chi_2$ determines other ring torsion angles	All trans-prolines
prePro <sub>Ala</sub>	All non-glycine, non-proline, pre-proline residues (39,093)	$p(\phi, \psi)$	None	pre-proline alanines
prePro <sub>Thr</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1)$	None	pre-proline threonines
prePro <sub>Val</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1)$	None	pre-proline valines
prePro <sub>Ser</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1)$	None	pre-proline serines
prePro <sub>Cys</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1)$	None	pre-proline cysteines
prePro <sub>Asp</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1) \times$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline aspartates
	All aspartates (57,985)	$p(\chi_2)$		
prePro <sub>Asn</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1) \times$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline asparagines
	All asparagines (44,158)	$p(\chi_2)$		
prePro <sub>Ile</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1) \times$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline isoleucines
	All isoleucines (59,385)	$p(\chi_2)$		
prePro <sub>Leu</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1) \times$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline leucines
	All leucines (95,250)	$p(\chi_2)$		
prePro <sub>His</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)	$p(\phi, \psi, \chi_1) \times$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline histidines
	All histidines (24,061)	$p(\chi_2)$		

**Table SI. (Continued.)**

Residue Type	Residues in Database <sup>a</sup>	Probability Density Function <sup>b</sup>	Statistical Approximation <sup>c</sup>	Energy Term Applied To: <sup>d</sup>
prePro <sub>Tyr</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All tryptophans (15,671)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2)$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline tryptophans
prePro <sub>Tyr</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All tyrosines (38,484)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2)$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline tyrosines
prePro <sub>Phe</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All phenylalanines (45,572)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2)$	$\chi_2 \perp \phi, \psi, \chi_1$	pre-proline phenylalanines
prePro <sub>Met</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All methionines (14,319)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2, \chi_3)$	$\chi_2, \chi_3 \perp \phi, \psi, \chi_1$	pre-proline methionines
prePro <sub>Glu</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All glutamates (46,950)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2, \chi_3)$	$\chi_2, \chi_3 \perp \phi, \psi, \chi_1$	pre-proline glutamates
prePro <sub>Gln</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All glutamines (31,349)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2, \chi_3)$	$\chi_2, \chi_3 \perp \phi, \psi, \chi_1$	pre-proline glutamines
prePro <sub>Arg</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All arginines (38,701)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2, \chi_3, \chi_4)$	$\chi_2, \chi_3, \chi_4 \perp \phi, \psi, \chi_1$	pre-proline arginines
prePro <sub>Lys</sub>	All non-glycine, non-proline, non-alanine, pre-proline residues (35,471)  All lysines (36,834)	$p(\phi, \psi, \chi_1) \times$  $p(\chi_2, \chi_3, \chi_4)$	$\chi_2, \chi_3, \chi_4 \perp \phi, \psi, \chi_1$	pre-proline lysines

**Table SI. (Continued.)**

Residue Type	Residues in Database <sup>a</sup>	Probability Density Function <sup>b</sup>	Statistical Approximation <sup>c</sup>	Energy Term Applied To: <sup>d</sup>
Ala	All non-pre-proline alanines (98,680)	$p(\phi, \psi)$	None	non-pre-proline alanines
Thr	All non-pre-proline threonines (58,953)	$p(\phi, \psi, \chi_1)$	None	non-pre-proline threonines
Val	All non-pre-proline valines (75,992)	$p(\phi, \psi, \chi_1)$	None	non-pre-proline valines
Ser	All non-pre-proline serines (58,156)	$p(\phi, \psi, \chi_1)$	None	non-pre-proline serines
Cys	All non-pre-proline cysteines (9,577)	$p(\phi, \psi, \chi_1)$	None	non-pre-proline cysteines
Asp	All non-pre-proline aspartates (54,955)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline aspartates
Asn	All non-pre-proline asparagines (41,623)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline asparagines
Ile	All non-pre-proline isoleucines (56,534)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline isoleucines
Leu	All non-pre-proline leucines (90,123)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline leucines
His	All non-pre-proline histidines (22,619)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline histidines
Trp	All non-pre-proline tryptophans (15,110)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline tryptophans
Tyr	All non-pre-proline tyrosines (36,671)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline tyrosines
Phe	All non-pre-proline phenylalanines (43,564)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$	non-pre-proline phenylalanines
Met	All non-pre-proline methionines (13,727)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2, \chi_3)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$ $\chi_3 \perp \phi, \psi$ given $\chi_1, \chi_2$	non-pre-proline methionines
Glu	All non-pre-proline glutamates (45,332)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2, \chi_3)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$ $\chi_3 \perp \phi, \psi$ given $\chi_1, \chi_2$	non-pre-proline glutamates
Gln	All non-pre-proline glutamines (30,031)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2, \chi_3)}{p(\chi_1)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$ $\chi_3 \perp \phi, \psi$ given $\chi_1, \chi_2$	non-pre-proline glutamines
Arg	All non-pre-proline arginines (37,112)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2, \chi_3)P(\chi_2, \chi_3, \chi_4)}{p(\chi_1)P(\chi_2, \chi_3)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$ $\chi_3 \perp \phi, \psi$ given $\chi_1, \chi_2$ $\chi_4 \perp \phi, \psi$ given $\chi_2, \chi_3$	non-pre-proline arginines
Lys	All non-pre-proline lysines (35,281)	$\frac{p(\phi, \psi, \chi_1)P(\chi_1, \chi_2, \chi_3)P(\chi_2, \chi_3, \chi_4)}{p(\chi_1)P(\chi_2, \chi_3)}$	$\chi_2 \perp \phi, \psi$ given $\chi_1$ $\chi_3 \perp \phi, \psi$ given $\chi_1, \chi_2$ $\chi_4 \perp \phi, \psi$ given $\chi_2, \chi_3$	non-pre-proline lysines

<sup>a</sup> Residues in the torsion angle database used to estimate the probability density function (the total number of instances is indicated in parenthesis).

<sup>b</sup> Expression used to represent the full joint probability density function. For simplicity, conditionals are omitted (for residue type Gly,  $p(\phi, \psi)$  stands for  $p(\phi, \psi|\text{Gly})$ , etc.). For certain prePro residue subtypes (e.g., prePro<sub>Asp</sub>) the different probability density components arise from different residue populations in the database.

<sup>c</sup> Approximation used to break down the full probability density function into components of lower dimensionality ( $\perp$  means “independent of”).

<sup>d</sup> Residues to which the associated statistical potential is applied during structure calculations.