Supporting Information for

p53 searches on DNA by rotation-uncoupled sliding at C-terminal tails and restricted hopping of core domains

Tsuyoshi Terakawa[1], Hiroo Kenzaki[1], and Shoji Takada[1, 2]

SI text

Protein model for folded domains

To model folded domains of p53 (the core and TET domains), we used an Atomic-Interaction-based Coarse-Grained (AICG) model developed by Li *et al*. (1) based on Go-like model (2). In this model, one CG particle represents an amino acid locating the particle at C_α atom positions. The model is defined by the potential energy function,

$$V_{AICG} = V_{bond}^{AICG} + V_{angle}^{AICG} + V_{dihedral}^{AICG} + V_{contact}^{AICG} + V_{excluded} + V_{ele}.$$

V_{bond}^{AICG} is the potential energy function for bond stretching,

$$V_{bond}^{AICG} = \sum_I k_b (r^I - r_0^I)^2,$$

where k_b is a constant, r^I is the length of the *I*-th virtual bond that connect *I*-th and *I*+1-th C_α atoms, and r_0^I is the length in the native structure. V_{angle}^{AICG} is the potential energy function for virtual bond angle bending,

$$V_{angle}^{AICG} = \sum_I k_a^{X_I} (\theta^I - \theta_0^I)^2,$$

where $k_a^{X_I}$ is a constant, θ^I is the *I*-th virtual bond angle defined by *I*-th, *I*+1-th, and *I*+2-th C_α atoms, and θ_0^I is the angle in the native structure. X_I represents that *I*+1-th C_α atom is assigned to a structure category *X* by DSSP (3). Here, *X* stands for glycine (*G*), helix (*H*), β-strand (*E*), turn (*T*), or random coil (*C*). $V_{dihedral}^{AICG}$ is the potential energy function for dihedral angle twisting,

$$V_{dihedral}^{AICG} = \sum_I \left\{ \varepsilon_\phi^{X_I} [1 - \cos(\phi^I - \phi_0^I)] + \varepsilon_\phi^{X_I} [1 - \cos 3(\phi^I - \phi_0^I)]/2 \right\},$$

where $\varepsilon_{\phi}^{X_I}$ is a constant, ϕ^I is the I-th virtual dihedral angle defined by I-th, I+1-th, I+2-th and I+3-th C_α atoms, and ϕ_0^I is the dihedral angle in the native structure. $V_{contact}^{AICG}$ is the potential energy function for native contact interactions,

$$V_{contact}^{AICG} = \sum_{I<J-3}^{native} \varepsilon^{IJ} \left[5 \left(\frac{r_0^{IJ}}{r^{IJ}} \right)^{12} - 6 \left(\frac{r_0^{IJ}}{r^{IJ}} \right)^{10} \right],$$

where ε^{IJ} is a constant, r^{IJ} is the distance between I-th and J-th C_α atoms, and r_0^{IJ} is the distance in the native structure. The summation is taken only for the natively contacting residue pairs. The residue pairs are considered to be in contact if they are separated by at least four residues in sequence and the distance between the closest heavy atoms of them is within 6.5 Å in the native structure. The interaction strength ε^{IJ} is written as $\varepsilon^{IJ} = \varepsilon_{nloc} W^{IJ}$ where ε_{nloc} is a constant and W^{IJ} is calculated by the linear regression procedure described in (1). $V_{excluded}$ is the potential energy function for excluded volume effect,

$$V_{excluded} = \sum_{I<J-3}^{non-native} \varepsilon_{ex} (C/r^{IJ})^{12},$$

where ε_{ex} and $C = 6.0$ are constants. The summation is taken only for the natively non-contacting amino acid pairs. V_{ele} is the potential energy function for electrostatic interactions derived from Debye-Huckel theory,

$$V_{ele} = \sum_{I<J} \frac{q_I q_J}{4\pi\varepsilon_0 \varepsilon_k r^{IJ}} e^{-r^{IJ}/\kappa_D},$$

where q_I is charge, ε_0 is a electric constant, $\varepsilon_k = 78.0$ is a dielectric constant. κ_D is so-called Debye length,

$$\kappa_D = \left(\frac{\varepsilon_0 \varepsilon_k k_B T}{2 N_A e^2 I} \right)^{0.5},$$

where k_B is the Boltzmann constant, T is the temperature, N_A is the Avogadro number, e is the elementary electric charge, and I is the ionic strength. The ionic strength is defined as $I = 0.5\sum z_i^2 c_i$ where $z_i = q_i/e$ and the c_i is the molar density per m^3. All the Lys, Arg, and His residues have +1 charges, and all the Asp and Glu residues have -1 charges. The constants used in each potential energy function ($k_b = 109.94$, $k_a^G = 13.40$, $k_a^H = 40.03$, $k_a^E = 17.32$, $k_a^T = 19.35$, $k_a^C = 11.70$, $\varepsilon_\phi^G = 0.29$, $\varepsilon_\phi^H = 1.76$, $\varepsilon_\phi^E = 1.32$, $\varepsilon_\phi^T = 0.82$, $\varepsilon_\phi^C = 0.81$ and $\varepsilon_{nloc} = 0.37$) were set to the value obtained in (1) by the fluctuation matching procedure. We used this model in the core domain (residues 91-289) and the TET domain (residues 326-356) where the reference structures were obtained from PDB code 2XWR (4) and 1AIE (5), respectively. The inter-chain contact interactions ($V_{contact}^{AICG}$) were only imposed in TET and, but not included in the core domain.

Protein model for disordered regions

For disordered regions of p53 (NTD, linker, and CTD), we used a model containing the statistical potentials for virtual bond angles and dihedral angles. The statistical potentials were constructed from generic loop structures in PDB. In our previous work, the model moderately reproduced the profiles of small angle X-ray scattering and NMR residual dipolar coupling of the intrinsically disordered p53 NTD (6). The potential energy function of this model is defined by

$$V_{stat} = V_{bond}^{stat} + V_{angle}^{stat} + V_{dihedral}^{stat} + V_{excluded} + V_{ele}.$$

V_{bond}^{stat} is the potential energy function for bond stretching,

$$V_{bond}^{stat} = k_b(r^I - b)^2,$$

where $b = 3.8$ is constants and r^I is the length of the *I*-th virtual bond that connect *I*-th and *I*+1-th C_α atoms.

In order to construct the potential energy functions for virtual bond angle (V_{angle}^{stat}) and virtual dihedral angle ($V_{dihedral}^{stat}$), we first constructed a generic set of probability distributions from the dataset of 13598 protein structures in PDB (http:// http://www.rcsb.org/). These structures have mutual sequence identity less than 30%. For each of the proteins, using DSSP (3) for assigning the secondary structure, we extracted

four consecutive loop residues (residues which are not assigned to helix or strand). The virtual bond angles θ were classified by the amino acid type of the central residue. For every central residue types, we obtained histograms with the bin size of 10 degrees. Thus, totally, 20 probability distributions $P(\theta)$ were constructed. The virtual dihedral angles ϕ were classified by the central pair of amino acid types. For every pairs we obtained histograms with the bin size of 10 degrees. Thus, totally 400 probability distributions $P(\phi)$ were obtained. By comparing these 400 probability distributions, we found that the distributions are similar when the central pairs of amino acid are R_1-Gly, R_2-Pro, Ala-R_3, Arg-R_3, Asn-R_3, Asp-R_3, Cys-R_3, Gln-R_3, Glu-R_3, Gly-R_3, His-R_3, Ile-R_3, Leu-R_3, Lys-R_3, Met-R_3, Phe-R_3, Pro-R_3, Ser-R_3, Thr-R_3, Trp-R_3, Tyr-R_3, or Val-R_3 where R_1 represents all amino acids, R_2 represents all amino acids except Gly, and R_3 represents all amino acids except Gly and Pro (Fig. S3). The distribution of Gly-Pro is not similar with any other distributions. Then, we averaged the similar distributions and re-constructed totally 23 probability distribution $P(\phi)$. Then, we inverted these probability distributions in ordered to obtain the potential energy function for a virtual bond angle (Fig. S4),

$$V_{angle}^{stat} = -k_B T \ln \frac{P(\theta)}{\sin\theta}$$

and for a virtual dihedral angle (Fig. S3),

$$V_{dihedral}^{stat} = -k_B T \ln P(\phi),$$

where k_B is the Boltzmann constant and $T = 300.0$ is temperature. When we calculated a force for a virtual bond angle in simulation, we interpolated these tabulated values with the spline interpolation to obtain a continuous potential energy function. When we calculated a force for a virtual dihedral angle in simulation, to satisfy the periodic boundary condition of ϕ, we fit the tabulated data by the truncated Fourier series as

$$f(\phi) = \sum_{m}^{3} k_m \sin(m\phi) + \sum_{n}^{3} k_n \cos(n\phi) + C,$$

where k_m, k_n, and C are Fourier coefficients.

The interactions of virtual bond angle defined by *I*-th, *I*+1-th, and *I*+2-th C_α atoms where *I* is 1-89 (NTD), 289-324 (linker region), and 356-391 (CTD) were calculated by using

V_{angle}^{stat}. The interactions of virtual dihedral angle defined by *I*-th, *I*+1-th, *I*+2-th and *I*+3-th C_α atoms where I is 1-89 (NTD), 288-324 (linker region), and 355-390 (CTD) were calculated by using $V_{dihedral}^{stat}$. $V_{excluded}$ and V_{ele} are same as those described in "Protein model for folded domains".

DNA model

For DNA model, we used 3SPN.1 model developed by de Pablo group (7). This model captured the ionic strength dependency of melting temperature, the persistence length and the heat capacity profiles of dsDNA. In this model, CG particles represent the three chemical moieties comprising a nucleotide (sugar, phosphate, and nitrogenous base). The 3SPN.1 model is defined by the potential energy function,

$$V_{DNA} = V_{bond}^{DNA} + V_{angle}^{DNA} + V_{dihedral}^{DNA} + V_{stuck}^{DNA} + V_{base}^{DNA} + V_{excluded}^{DNA} + V_{solv}^{DNA} + V_{ele}.$$

V_{bond}^{DNA} is the potential energy function for bond stretching,

$$V_{bond}^{DNA} = \sum_I k_1^{DNA}(r^I - r_0^I)^2 + k_2^{DNA}(r^I - r_0^I)^4,$$

where $k_1^{DNA} = 0.1839$ and $k_2^{DNA} = 183.9$ are constants, r^I is the *I*-th virtual bond length, and r_0^I is the length in the native structure. V_{angle}^{DNA} is the potential energy function for bond angle bending,

$$V_{angle}^{DNA} = \sum_I k_a^{DNA}(\theta^I - \theta_0^I)^2,$$

where $k_a^{DNA} = 128.73$ is constant, θ^I is the *I*-th virtual bond angle, and θ_0^I is the bond angle in the native structure. $V_{dihedral}^{DNA}$ is the potential energy function for dihedral angle twisting,

$$V_{dihedral}^{DNA} = \sum_I k_\phi^{DNA}[1 - \cos(\phi^I - \phi_0^I)],$$

where $k_\phi^{DNA} = 5.1492$ is constant, ϕ^I is the *I*-th virtual dihedral angle, and ϕ_0^I is the dihedral angle in the native structure. V_{stuck}^{DNA} is the potential energy function for base stacking,

$$V_{stuck}^{DNA} = \sum_{I<J} 4\varepsilon \left[\left(\frac{\sigma_{IJ}^{DNA}}{r^{IJ}} \right)^{12} - \left(\frac{\sigma_{IJ}^{DNA}}{r^{IJ}} \right)^{6} \right],$$

where $\varepsilon = 0.1839$ is constant, σ_{IJ}^{DNA} is pair-dependent constant and r^{IJ} is the distance between *I*-th and *J*-th beads. The summation is taken only for the bead pairs that are in a single strand of DNA and the distance between them is within 9.0 Å in the native structure. V_{base}^{DNA} is the potential energy function for base pairing,

$$V_{base}^{DNA} = \sum_{I<J} 4\varepsilon_{bIJ}^{DNA} \left[5 \left(\frac{\sigma_{bI}^{DNA}}{r^{IJ}} \right)^{12} - 6 \left(\frac{\sigma_{bI}^{DNA}}{r^{IJ}} \right)^{10} \right],$$

where ε_{bIJ}^{DNA} and σ_{bI}^{DNA} are pair-dependent constant. ε_{IJ}^{DNA} is 0.3678 for AT pair and 0.4656 for CG pair. σ_{bI}^{DNA} is 2.9002 for AT pair, and 2.8694 for CG pair. The summation is taken only for all complementary base pairs that do not participate in V_{stuck}^{DNA}. A complementary base pair is considered to be hydrogen-bonded when the separation between bases is $r^{IJ} < \left(\sigma_{IJ}^{DNA} + 2.0\ \text{Å}\right)$ in the native structure. $V_{exclude}^{DNA}$ is the potential energy function for excluded volume effect,

$$V_{excluded}^{DNA} = \sum_{I<J} \begin{cases} 4\varepsilon \left[\left(\frac{\sigma_0^{DNA}}{r^{IJ}} \right)^{12} - \left(\frac{\sigma_0^{DNA}}{r^{IJ}} \right)^{6} \right] + \varepsilon \left(r^{IJ} < r_{coff} \right) \\ 0 \left(r^{IJ} \geq r_{coff} \right) \end{cases}$$

where $\sigma_0^{DNA} = 2^{-1/6} r_{coff}$ and r_{coff} is constant. r_{coff} is 1.00 for mismatch pairs and 6.86 otherwise. The summation is taken only for natively non-contacting pairs. V_{solv}^{DNA} is the potential energy function for solvation energy,

$$V_{solv}^{DNA} = \sum_{I<J} \mathcal{E}_s^{DNA} \left[1 - e^{-\alpha \left(r^{IJ} - r_s^{DNA} \right)} \right]^2,$$

where $\mathcal{E}_s^{DNA}$ is an ion concentration dependent constant, $\alpha = 5.333$ and $r_s^{DNA} = 13.38$ are constant. The summation is taken only for inter-strand sugar bead pairs. V_{ele} is the same as that described in "Protein model for folded domains".

Inter chain interaction

The potential energy function for the inter-protein-chain and the protein-DNA interaction is

defined by

$$V_{inter-chain} = V_{excluded} + V_{ele}.$$

where $V_{excluded}$ and V_{ele} are same as those described in "Protein model for folded domains". This potential is used in all the cases other than TET. In the case of TET, we considered the inter-chain native contact interaction in addition to them.

Reference

1. Li W, Wolynes P, Takada S (2011) Frustration, specific sequence dependence, and nonlinearity in large-amplitude fluctuations of allosteric proteins. *Proc Natl Acad Sci USA* 108:3504-9.

2. Clementi C, Nymeyer H, Onuchic N (2000) Topological and energetic factors: Whet Determines the Structural Details of the Transition State Ensemble and En-route Intermediates for protein Folding An investigation for Small Globular Proteins. *Journal of Molecular Biology* 298:937-953.

3. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577-637.

4. Natan E et al. (2011) Interaction of the p53 DNA-Binding Domain with Its N-Terminal Extension Modulates the Stability of the p53 Tetramer. *J Mol Biol* 409:358-68.

5. Mittl PRE, Chène P, Grütter MG (1998) Crystallization and structure solution of p53 (residues 326–356) by molecular replacement using an NMR model as template. *Acta Crystallogr D Biol Crystallogr* 54:86-9.

6. Terakawa T, Takada S (2011) Multiscale Ensemble Modeling of Intrinsically Disordered Proteins: p53 N-Terminal Domain. *Biophys J* 101:1450-8.

7. Sambriski EJ, Schwartz DC, de Pablo JJ (2009) A mesoscale model of DNA and its renaturation. *Biophys J* 96:1675-90.

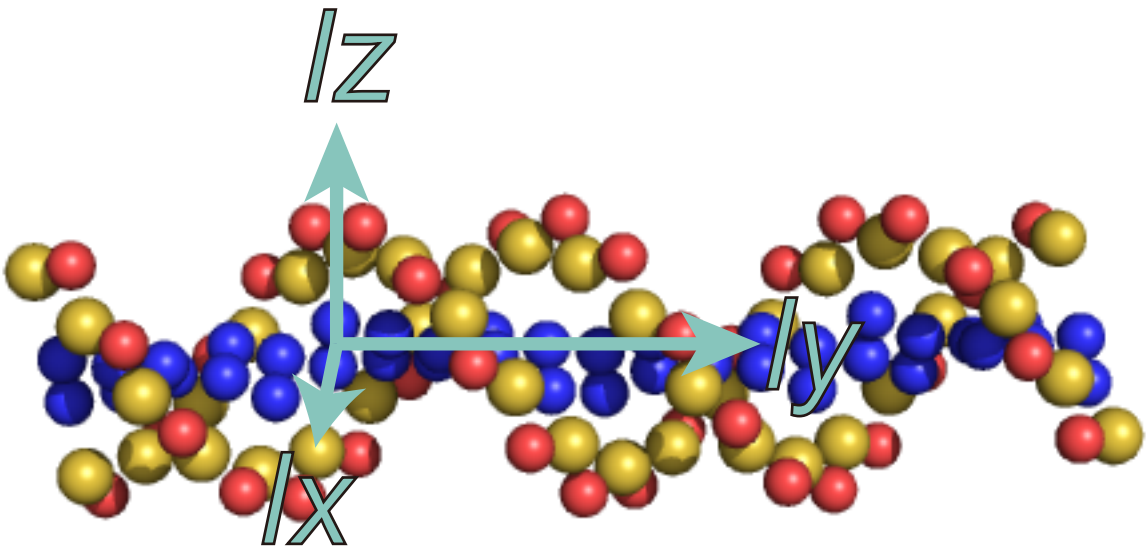

Figure S1: The definition of local coordinate axis used for the analysis of the protein rotation around DNA. The yellow beads, the red beads, and the blue beads represent coarse-grained sugar, phosphate, and base beads, respectively.

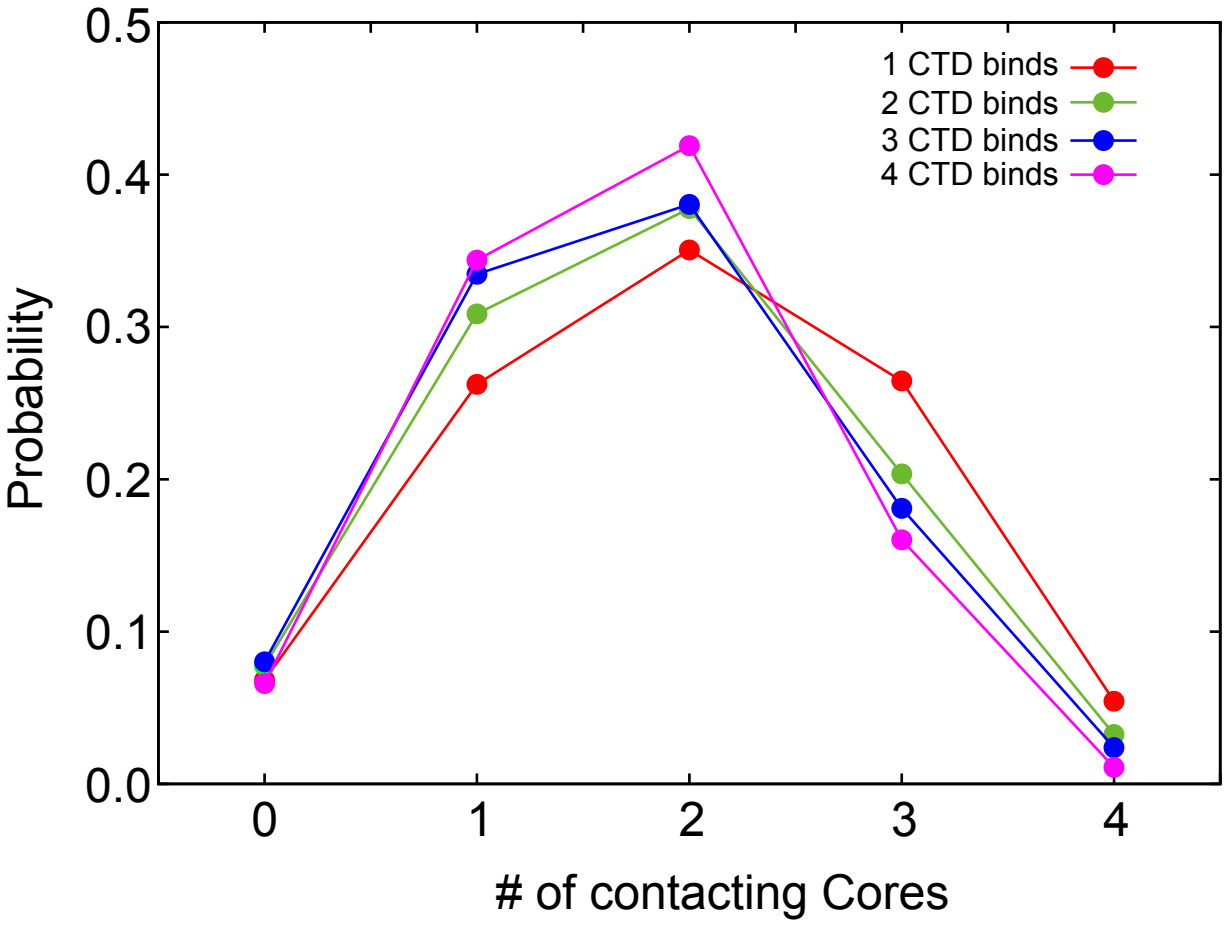

Figure S2: The probability distribution of the number of contacting Cores in the ensemble where n CTD binds (n = 1, 2, 3, 4) The analysis is done using the simulation in 200 mM ion.

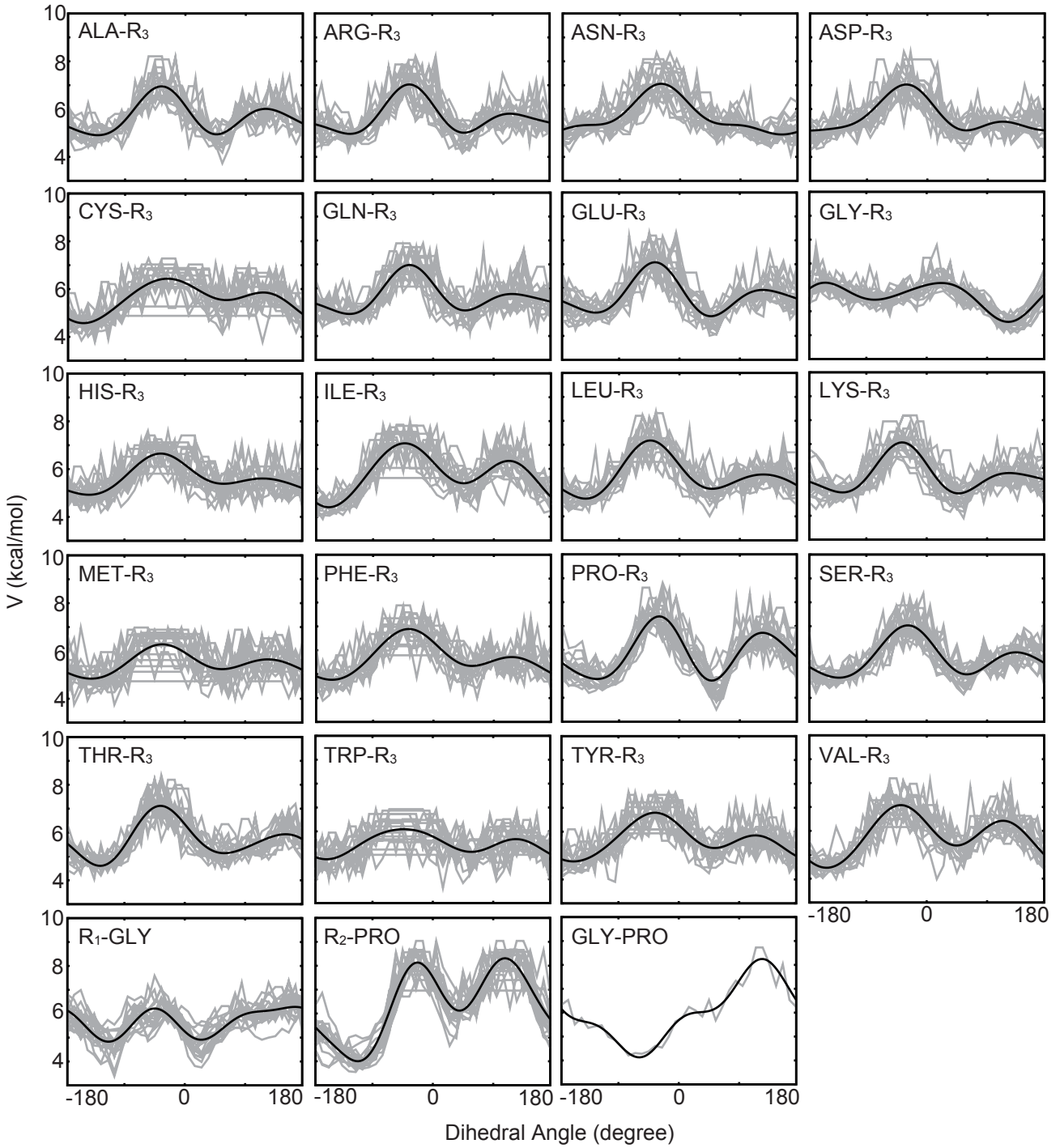

Figure S3: The potential energy functions of virtual dihedral angle. R_1 represents all amino acids, R_2 represents all amino acids except Gly, and R_3 represents all amino acids except Gly and Pro. The potential energy functions for each specific R are depicted with the grey line. We average these potential energy functions, fit them with the truncated Fourier series, and depict them with black lines.

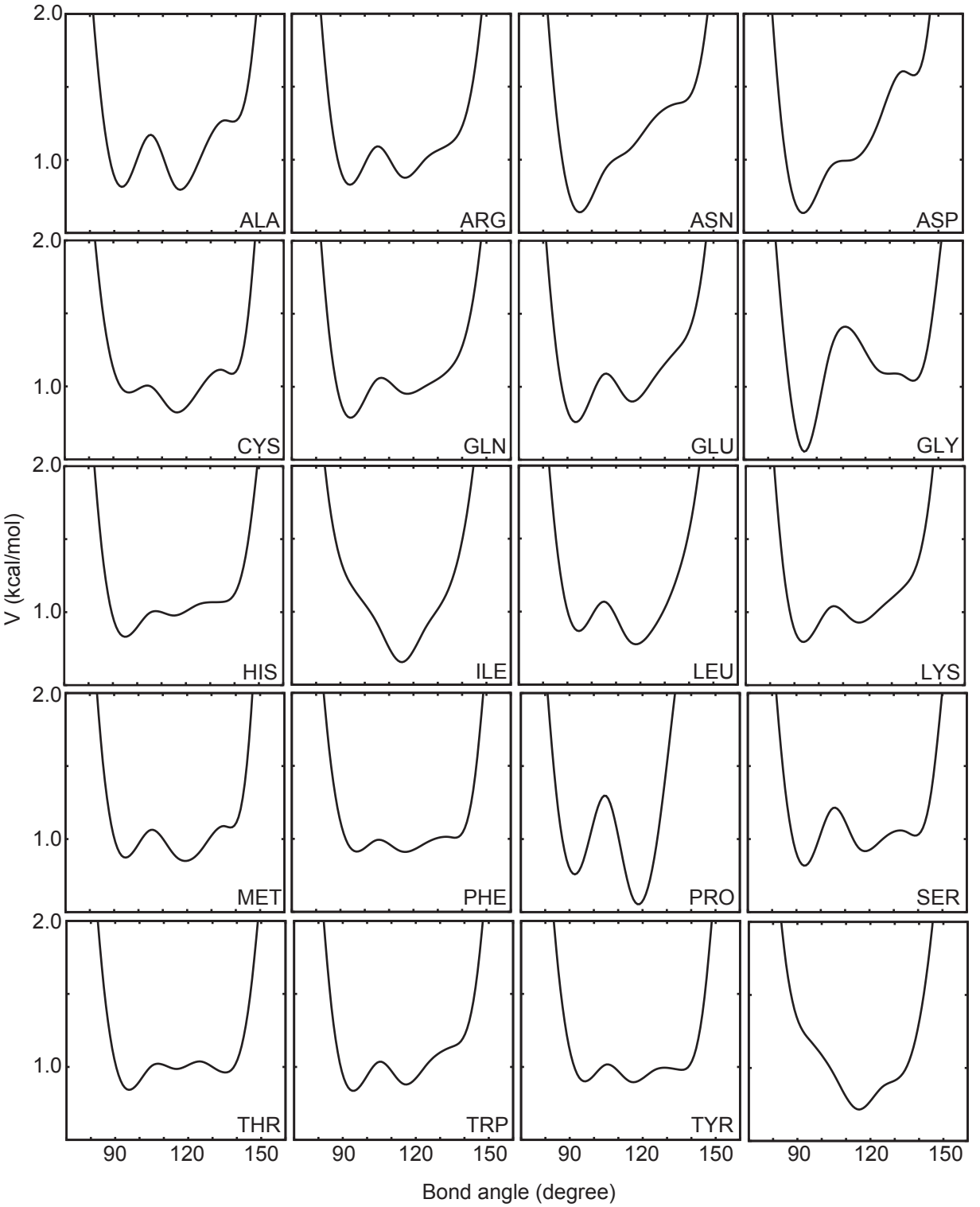

Figure S4: The potential energy functions of virtual bond angle.