Supporting Information

On the Optimization of the Protein Backbone Dihedral Angles by Means of Hamiltonian Reweighting

Christian Margreitter and Chris Oostenbrink*

Institute for molecular modeling and simulation, University of natural resources and life sciences, Muthgasse 18, 1190 Vienna, Austria

E-mail: chris.oostenbrink@boku.ac.at

List of Tables

S1 Non-bonded parameters used for blocking for the Ac-X-NHMe simulations . 2

S2 List of the experimental target values 3

S3 Secondary structure basin definitions 4

S4 Dynamical features of 54A7 and suggested combinations 5

S5 Parameter combinations for the individually optimized amino acids 6

List of Figures

S1 Validation of the accuracy of Hamiltonian reweighting 7

S2 Example of scatter plot showing the J-value and propensity deviations for combination #5623 8

S3 Ramachandran plot prediction by Hamiltonian reweighting 9

S4 Correlation plot of experimental J-values with calculated values 10

Table S1: Parameters used for blocking for the Ac-X-NHMe simulations. Integer atom codes and mass codes according to the GROMOS 54A7 force field.[1]

	Acetyl-		
	Integer Atom Code (IAC)	Masscode	Charge
CM	16	5	0.00
C	12	12	0.45
O	1	16	-0.45
	N-Methyl-		
	Integer Atom Code (IAC)	Masscode	Charge
C	12	12	0.45
O	1	16	-0.45
N	7	14	-0.31
H	21	1	0.31
CT	15	5	0.00

Table S2: List of the target values extracted from the studies of Avbelj et al.[2] and Grdadolnik et al.[3] The propensities have been retrieved from Raman spectroscopy data calibrated by the J-values of the former study.

	target values			
	J-value	α	β	P_{II}
ALA	6.06	0.11	0.60	0.29
GLY	5.85	0.66[a]	0.12[a]	0.22[a]
ARG	6.85	0.07	0.54	0.39
ASN	7.45	0.02	0.58	0.40
ASP	6.93	0.05	0.46	0.49
CYS	7.31	0.03	0.54	0.43
GLU	6.63	0.05	0.36	0.59
GLN	7.14	0.08	0.48	0.44
HIS	7.89[b]	0.04[b]	0.58[b]	0.38[b]
ILE	7.33	0.02	0.52	0.46
LEU	6.88	0.10	0.35	0.55
LYS	6.83	0.04	0.41	0.55
MET	7.02	0.03	0.47	0.50
PHE	7.18	0.06	0.49	0.45
SER	7.02	0.04	0.47	0.49
THR	7.35	0.03	0.58	0.39
TRP	6.91	0.02	0.44	0.54
TYR	7.13	0.07	0.47	0.46
VAL	7.30	0.02	0.51	0.47

[a] Although the special distribution of the dihedral angles in the case of glycine may lead to a great over-emphasis of the defined three basins, we apply the same renormalization procedure as in the other cases (see text for justification).

[b] The values of histidine have not been taken into account in the ranking protocol.

Table S3: Listing of the dihedral angle areas used for classification. This table holds the basins,[4] which were used for the classification with DISICL[4,5] for all compounds investigated. The original values were reported by Hollingsworth et al.[6] A structure with a pair of ϕ/ψ backbone dihedrals, that falls within one of these regions is assigned to belong to the secondary structure element given in the first column. All remaining structures are considered to be unclassified. The experimental comparison was based on helical_R, β and P_II.

	definitions			
type	ϕ_1 [°]	ϕ_2 [°]	ψ_1 [°]	ψ_2 [°]
helical_R	-95	-40	-70	-32
	-107	-40	-32	-12
	-165	-95	-70	-32
	-107	-40	-12	8
	-150	-67	8	40
	-150	-107	-32	8
β	-135	-100	95	150
	-175	-135	95	136
	-180	-135	136	180
	-135	-105	150	180
	-180	-135	-180.1	-160
helical_L	38	140	-25	75
P_II	-100	-30	95	180
	-100	-60	-180.1	-150

Table S4: Comparison of dynamical parameters of different combinations. On the left of each column, the number of visits per basin is reported, while on the right the average residence time per basin is given in picoseconds. Some combinations in our suggested set increase the torsional energy barriers (especially in the ϕ-angle, see figure 5), which in turn leads to higher residence times. Still, a sufficient number of transitions between the basins is observed. All values are reported for trajectories of 100 ns length.

system	parameters	#α	<time> [ps]	#β	<time> [ps]	#P_{II}	<time> [ps]
GLY	54A7	737	2.0	15158	1.6	3385	1.8
GLY	#81883	7116	3.1	3650	1.6	1845	1.6
ALA	54A7	681	16.3	16233	2.8	15183	2.1
ALA	#12572	914	11.3	5614	4.5	4448	13.7
ALA	#12572 (334 K)	1369	9.1	7466	3.5	6089	9.3
ALA	#12572 (354 K)	1839	8.0	8502	3.2	6948	7.6
CYS	54A7	732	24.5	14684	3.1	12952	2.1
CYS	#5623	560	6.8	9962	4.9	7673	5.6
VAL	54A7	368	33.6	16183	2.1	17247	2.5
VAL	#86516	113	2.9	9712	5.4	7238	5.9

Table S5: Parameter combinations for the individually optimized amino acids.

	backbone angle energy potentials						description
	ϕ			ψ			
	K [$\frac{kJ}{mol}$]	shift [°]	mult.	K [$\frac{kJ}{mol}$]	shift [°]	mult.	
#67656	5.0	0.0	3	3.0	0.0	2	arginine
	5.0	180.0	6	5.0	0.0	2	
#96795	5.0	0.0	3	5.0	180.0	2	asparagine
	5.0	180.0	2	5.0	180.0	2	
#31541	3.0	180.0	2	1.0	180.0	2	aspartate
	5.0	0.0	3	3.0	180.0	3	
#50989	3.0	180.0	2	1.0	180.0	3	cysteine
	5.0	0.0	3	5.0	180.0	3	
#83856	5.0	0.0	3	3.0	180.0	2	glutamine
	5.0	180.0	6	5.0	180.0	6	
#87099	5.0	0.0	3	5.0	0.0	2	glutamate
	5.0	180.0	6	5.0	0.0	2	
#89861	3.0	180.0	2	5.0	0.0	2	histidine
	5.0	0.0	3	5.0	0.0	6	
#9315	5.0	0.0	3	1.0	0.0	6	leucine
	5.0	180.0	2	1.0	180.0	3	
#80612	5.0	0.0	3	3.0	180.0	3	lysine
	5.0	180.0	6	5.0	0.0	2	
#28751	5.0	0.0	3	1.0	180.0	6	methionine
	5.0	180.0	2	3.0	0.0	1	
#25538	5.0	0.0	3	1.0	180.0	3	phenylalanine
	5.0	180.0	6	3.0	0.0	2	
#67652	5.0	0.0	3	3.0	0.0	1	serine
	5.0	180.0	6	5.0	0.0	2	
#93571	5.0	0.0	3	5.0	0.0	2	tryptophane
	5.0	180.0	6	5.0	180.0	3	
#6094	5.0	0.0	3	1.0	0.0	1	tyrosine
	5.0	180.0	6	1.0	180.0	6	
#28191	3.0	0.0	3	1.0	180.0	6	isoleucine
	5.0	180.0	6	3.0	0.0	1	
#92991	3.0	0.0	3	5.0	0.0	2	threonine
	5.0	180.0	6	5.0	180.0	3	
#24957	3.0	0.0	3	1.0	180.0	3	valine
	5.0	180.0	6	3.0	0.0	1	

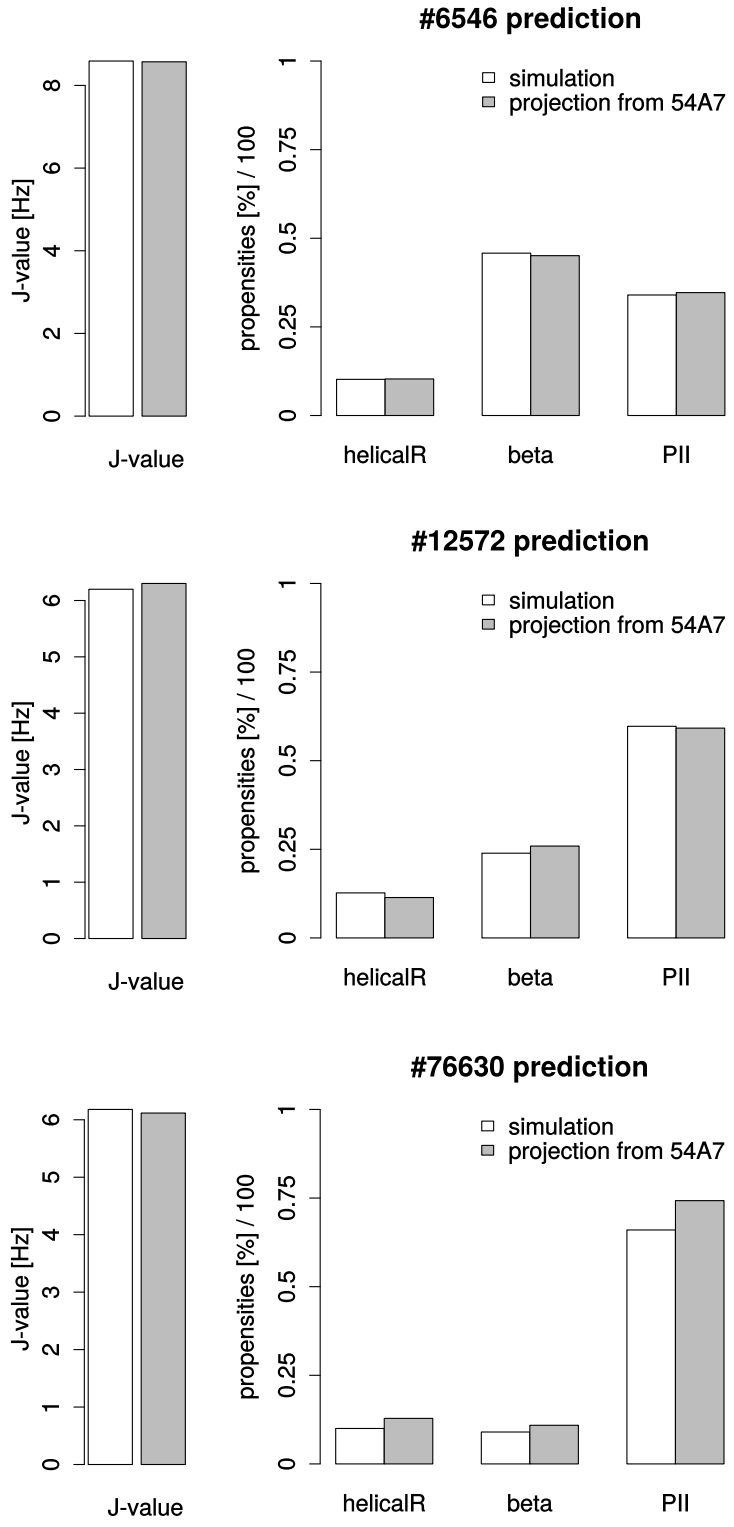

Figure S1: Shows the accuracy of the projections based on the 54A7 trajectories for alanine for three different sets of backbone dihedral potentials. Note, that the projection is quite accurate even when different trends (e.g. propensities of #76630 versus those of #6546) are observed.

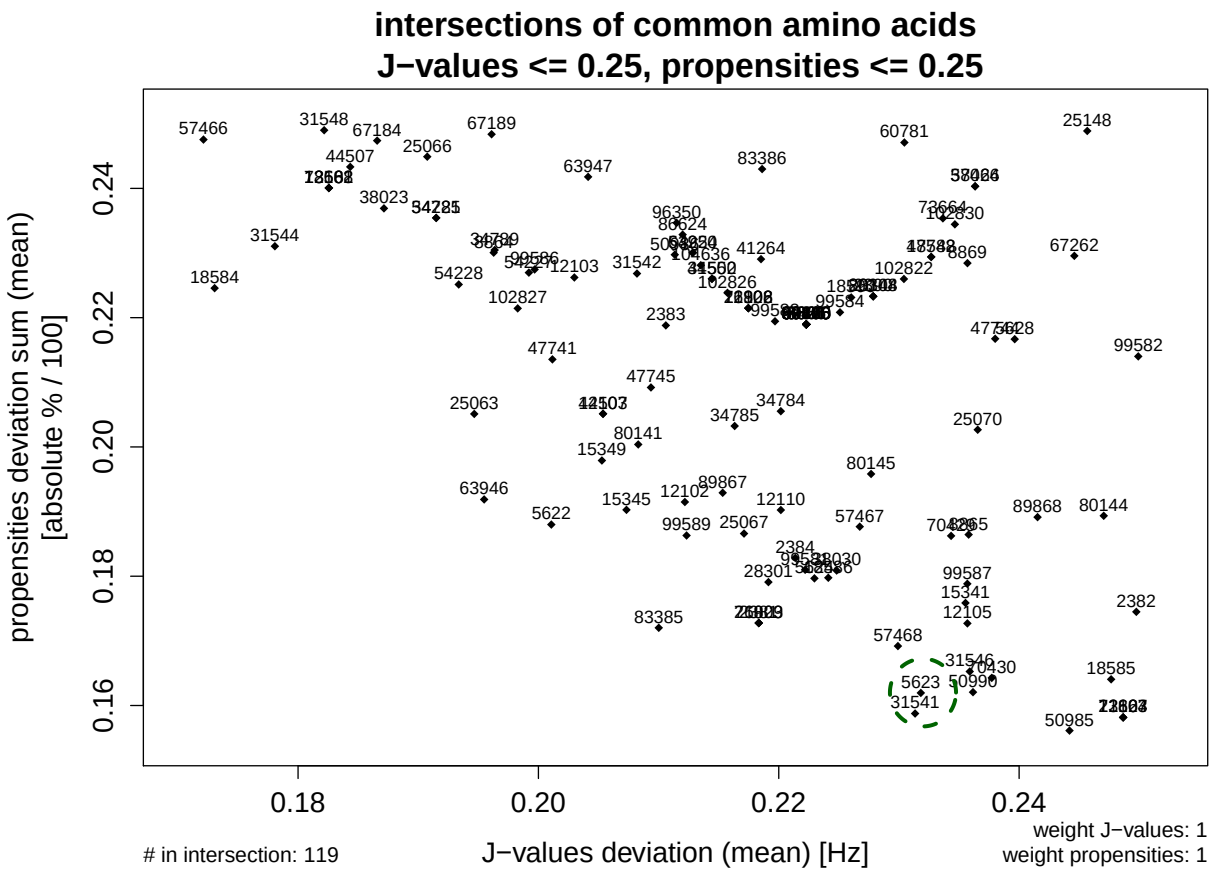

Figure S2: Example of scatter plot showing the deviation in J-value reproduction (x-axis) versus the summed up deviation in propensities (y-axis) for a selection of combinations from our screening set. In principle all combinations in this scatter plot could lead to reasonable agreement with experimental data. The combination #5623, which has been proposed for the description of the "common" amino acids (see main manuscript), is marked with a green dashed circle.

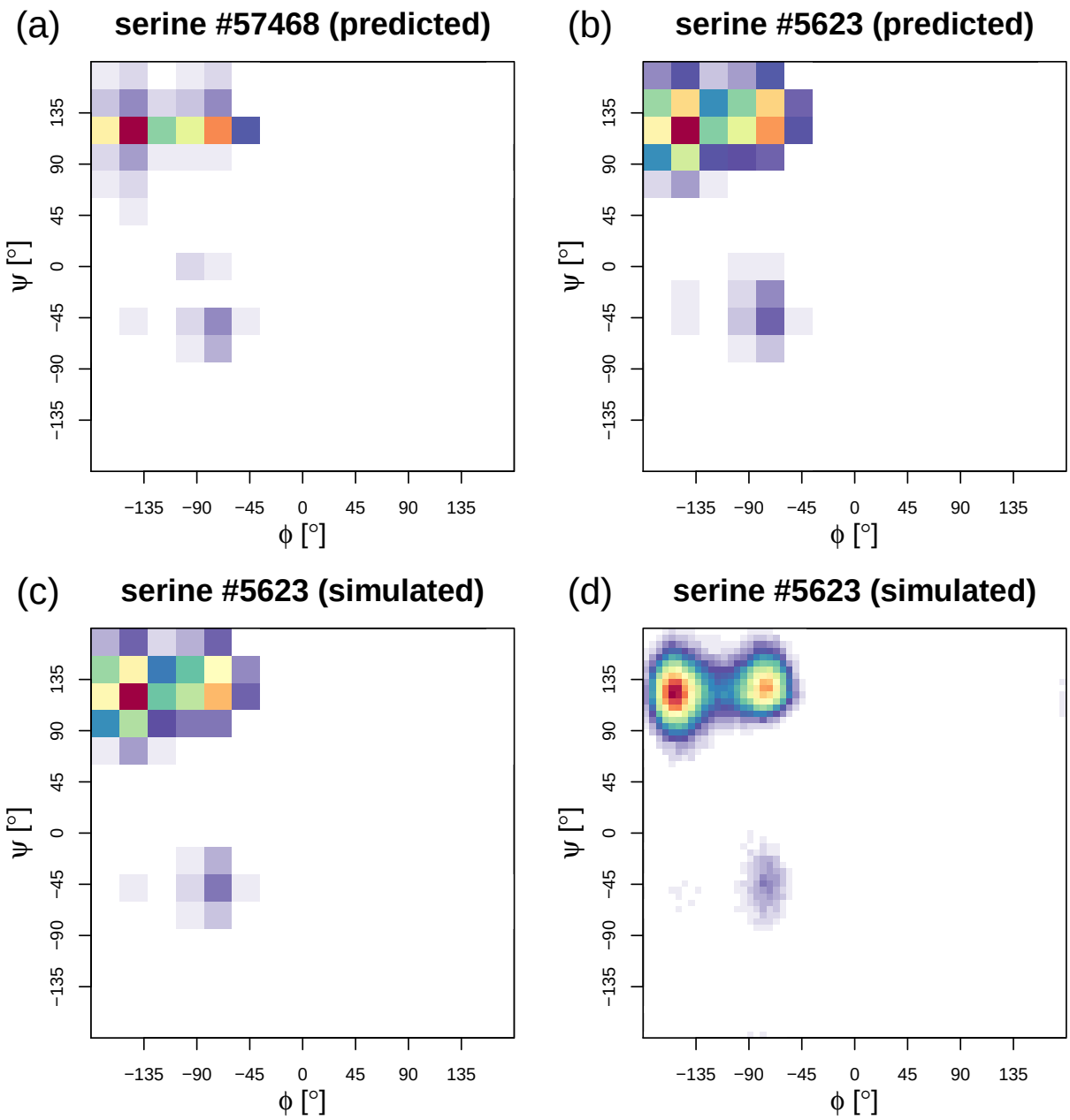

Figure S3: Prediction accuracy for the 15x15 bin Ramachandran plots on the example of serine. For combination #57468, which produced very good results in terms of matching the experimental results, a very narrow distribution within the basins is observed (a). Combination #5623 performs only slightly worse in the agreement with the experimental data, but leads to a wider sampling (b). The reliability of the prediction is shown by comparison to actual simulation data in panel (c). In (d), the simulated result with a higher resolution is shown.

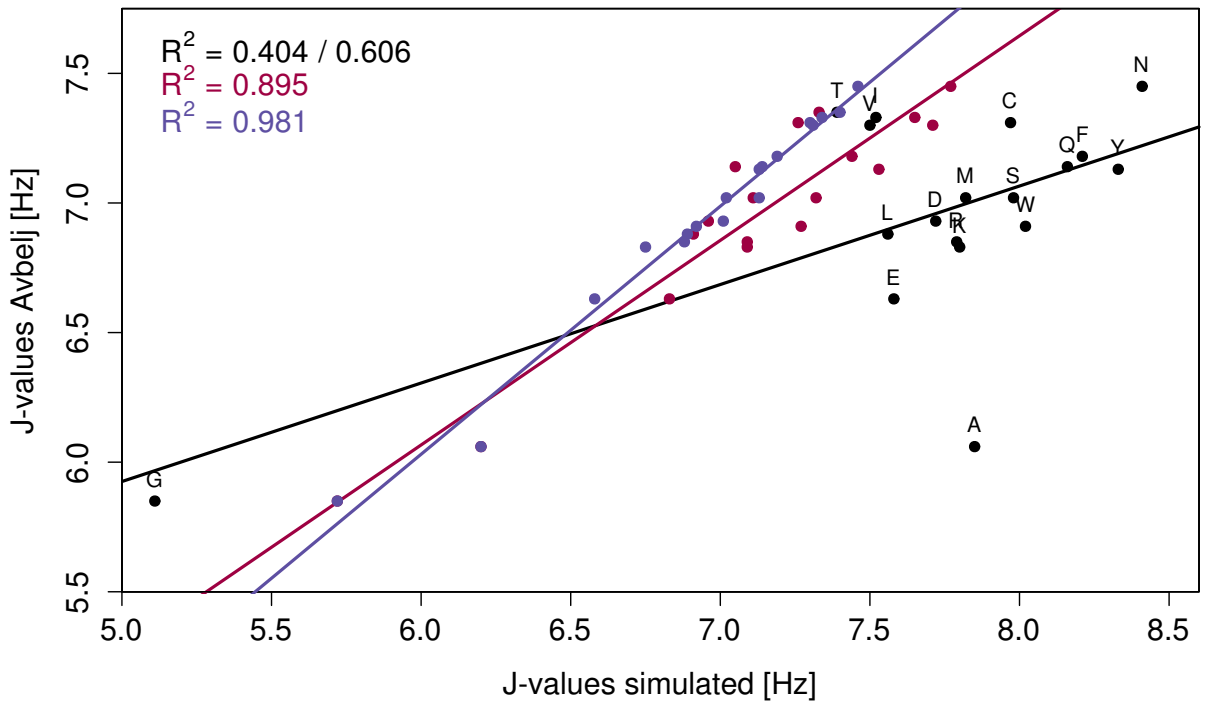

Figure S4: Correlation plot of experimental J-values with the calculated ones, based on 54A7 (black), our suggested set (red) and the individually optimized (blue) parameters. Since alanine is a strong outlier for 54A7, we also report the correlation coefficient for the remaining residues only which leads to a slight improvement (see second value). The slopes are 0.380, 0.397, 0.789 and 0.957, and the intercepts are 4.025, 3.952, 1.331 and 0.289 for 54A7, 54A7 without alanine, our suggested set and the individually optimized set, respectively. The markers for Ala and Gly are positioned identically in the proposed and the individually optimized sets.

References

(1) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.

(2) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. Intrinsic Backbone Preferences Are fully Present in Blocked Amino Acids. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 1272–1277.

(3) Grdadolnik, J.; Mohacek-Grosev, V.; Baldwin, R. L.; Avbelj, F. Populations of the three major Backbone Conformations in 19 Amino Acid Dipeptides. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 1794–1798.

(4) Nagy, G.; Oostenbrink, C. Dihedral-Based Segment Identification and Classification of Biopolymers I: Proteins. *J. Chem. Inf. Model.* **2014**, *54*, 266–277.

(5) Nagy, G.; Oostenbrink, C. Dihedral-Based Segment Identification and Classification of Biopolymers II: Polynucleotides. *J. Chem. Inf. Model.* **2014**, *54*, 278–288.

(6) Hollingsworth, S. A.; Lewis, M. C.; Berkholz, D. S.; Wong, W.-K.; Karplus, P. A. (ϕ,ψ)-Motifs: a purely Conformation-Based, Fine-Grained Enumeration of Protein Parts at the Two-Residue Level. *J. Mol. Biol.* **2012**, *416*, 78–93.