Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supplementary Information

HyRes: a coarse-grained model for multi-scale enhanced sampling of disordered protein conformations

Xiaorong Liu[1] and Jianhan Chen[1,2]*

[1]Department of Chemistry and [2]Department of Biochemistry and Molecular Biology,

University of Massachusetts Amherst, Amherst, MA 01003, USA

Table S1. Mapping of side chain atoms from the CHARMM 22 atomistic model in HyRes. A single CB bead is used in all the other residues (except Gly).

Residue	CG bead	Atom names in CHARM 22
Lys	CB	CB HB1 HB2 CG HG1 HG2 CD HD1 HD2
	CC	CE HE1 HE2 NZ HZ1 HZ2 HZ3
Arg	CB	CB HB1 HB2 CG HG1 HG2 CD HD1 HD2
	CC	NE HE CZ NH1 HH11 HH12 NH2 HH21 HH22
His	CB	CB HB1 HB2 CG
	CC	CD2 HD2 NE2
	CD	ND1 HD1 CE1 HE1
Phe	CB	CB HB1 HB2 CG CD1 HD1
	CC	CD2 HD2 CE2 HE2
	CD	CE1 HE1 CZ HZ
Tyr	CB	CB HB1 HB2 CG CD1 HD1
	CC	CD2 HD2 CE2 HE2
	CD	CE1 HE1 CZ OH HH
Trp	CB	CB HB1 HB2 CG
	CC	CD1 HD1 NE1 HE1
	CD	CD2 CE2
	CE	CZ2 HZ2 CH2 HH2
	CF	CE3 HE3 CZ3 HZ3

Table S2. Sequences of model peptides used in this work**.** All peptides were capped with an acetyl group at N-terminus and N-methyl amide at C-terminus.

Peptide	**κ**	**Sequence**
Gly_{10}		GGGGG GGGGG
$(AAQAA)_3$		AAQAA AAQAA AAQAA
KID		TDSQK RREIL SRRPS YRKIL NDLSS DAP
ACTR		EGQSD ERALL DQLHT LLSNT DATGL EEIDR ALGIP ELVNQ GQALE PK
NCBD		PNRSI SPSAL QDLLR TLKSP SSPQQ QQQVL NILKS NPQLM AAFIK QRTAK YVANQ PGMQ
RS		GAMGP SYGRS RSRSR SRSRS RSRS
GB1m3		KKWTY NPATG KFTVQ E
$(EK)_{25}$	0.0009	EKEKE KEKEK EKEKE KEKEK EKEKE KEKEK EKEKE KEKEK EKEKE KEKEK
	0.0025	EEEKK KEEEK KKEEE KKKEE EKKKE EEKKK EEEKK KEEEK KKEEE KKKEK
	0.0139	KEKKK EKKEE KKEEK EKEKE KEEKK KEEKE KEKEK KKEEK EKEEK KEEEE
	0.0140	KEKEK KEEKE KKEEE KKEKE KEKKK EEKKK EEKEE KKEEK KKEEK EEEKE
	0.0245	KEKEE KEKKK EEEEK EKKKK EEKEK EKEKE EKKEE KKKKE EKEEK EKEKE
	0.0273	EEEKK EKKEE KEEKK EKKEK EEEKK KEKEE KKEEE KKKEK EEEEK KKKEK
	0.0450	EEEEK KKKEE EEKKK KEEEE KKKKE EEEKK KKEEE EKKKK EEEEK KKKEK
	0.0450	KKKKE EEEKK KKEEE EKKKK EEEEK KKKEE EEKKK KEEEE KKKKE EEEKE
	0.0624	EEKKE EEKEK EKEEE EEKKE KKEKK EKKKE EKEKE KKKEK KKKEK EEEKE
	0.0834	EKKKK KKEEK KKEEE EEKKK EEEKK KEKKE EKEKE EKEKK EKKEE KEEEE
	0.0841	EKEKK KKKEE EKKEK EEEEK EEEEK KKKKE KEEEK EEKKE EKEKK KEEKK
	0.0864	EKKEE EEEEK EKKEE EEKEK EKKEK EEKEK KEKKK EKKEE EKEKK KKEKK
	0.0951	KEKKK EKEKK EKKKE EEKKK EEEKE KKKEE KKEKK EKKEE EEEEE KEEKE
	0.1311	EKKEK EEKEE EEKKK KKEEK EKKEK KKKEK KKKKE EEEEE KEEKE KEKEE
	0.1354	KKEKK EKKKE KKEKK EEEKE KEKKE KKKKE KEKKE EEEEE EEKEE KKEEE

	0.1458	EKEKE EKKKE EKKKK EKKEK EEKKE KEKEK KEEEE EEEEE KEKKE KKKKE
	0.1643	EKEKK KKKKE KEKKK KEKEK KEKKE KEEEK EEKEK EKKEE KKEEE EEEEE
	0.1677	KEEKK EEEEE EEKEE KKKKK EKKKE KKEEE KKKEE KKKEE EEEEK KKKEK
	0.1941	EEEEE KKKKK EEEEE KKKKK EEEEE KKKKK EEEEE KKKKK EEEEE KKKKK
	0.2721	EEKEE EEEEK EEEKE EKKEE EKEKK EKKEK EEKKE KKKKK KKKKK KKEEE
	0.2737	EEEEE EEEEK EKKKK KEKEE KKKKK KEKKE KKKKE KKEEE EEEKE EEKKK
	0.3218	KEEEE KEEKE EKKKK EKEEK EKKKK KKKKK KKKEK KEEEE EEEEK EKEEE
	0.3545	EEEEE KEEEE EEEEE EEKEE KEKKK KKKEK KKKKK KEKEK KKKEK KEEKK
	0.4456	EEEEK EEEEE KEEEE EEEEE EEEKK KEEKK KKKEK KKKKK KEKKK KKKKK
	0.5283	EEEEE EEEEE EKEEE EKEEK EEKEK KKKKK KKKKK KKKKK KKEEK KEEKE
	0.6101	KEEEE EEEKE EKEEE EEEEE EKEEE EKEEK KKKKK KKKKK KKKKK KKKKE
	0.6729	KKEKK KEKKE EEEEE EEEEE EEEEE EEEEK EEKKK KKKKK KKKKK KKEKK
	0.7666	EKKKK KKKKK KKKKK KKKKK KKEEE EEEEE EEEEE EEEEE KKEEE EEKEK
	0.8764	KEEEE KEEEE EEEEE EEEEE EEEEE EEKKK KKKKK KKKKK KKKKK KKKKK
	1.0000	EEEEE EEEEE EEEEE EEEEE EEEEE KKKKK KKKKK KKKKK KKKKK KKKKK

Table S3. Parameters of U_{dihedral} for side chain χ (N_i-CA_i-CB_i-CC_i).

Residues	k_χ(kcal/mol)	n	δ
Lys	0.3	3	0
Lys	0.5	1	75
Arg	0.3	3	0
Arg	0.5	1	75
His	0.6	3	0
His	0.4	1	75
His	0.1	1	50
Phe	0.8	3	180
Phe	0.3	1	100
Phe	0.1	1	50
Tyr	0.9	3	180
Tyr	0.3	1	90
Tyr	0.1	1	50
Trp	0.8	3	0
Trp	0.8	1	85

Table S4. Parameters of U_{dihedral} applied to dihedral ψ' (CB_i-CA_i-C_i-O_i)

Residues	$k_{\psi'}$(kcal/mol)	n	δ
Asp	0.3	1	240
Asn	0.3	1	240
Thr	0.3	1	240
Cys	0.3	1	240
Val	0.3	1	240

Table S2. vdW interaction energies (in kcal/mol) for Gly_{10} in representative compact and extended states derived from two independent sets of 10-ns explicit solvent simulations. The compact and extended states are mimicked by restraining the peptide end-to-end distances to 7 and 20 Å, respectively.

Simulation	Energy term	U_{cmp}	U_{ext}	U_{cmp} - U_{ext}
1	$U_{vdw}^{intra-pept}$	-1.78	0.33	-2.11
	U_{vdw}^{inter}	-46.69	-49.84	3.15
	$U_{vdw}^{intra-solv}$	8358.30	8359.74	-1.44
	U_{vdw}^{tot}	8309.82	8310.22	-0.40
2	$U_{vdw}^{intra-pep}$	-0.78	-0.42	-0.36
	U_{vdw}^{inter}	-47.07	-48.91	1.85
	$U_{vdw}^{intra-solv}$	8357.60	8359.23	-1.62
	U_{vdw}^{tot}	8309.76	8309.90	-0.14

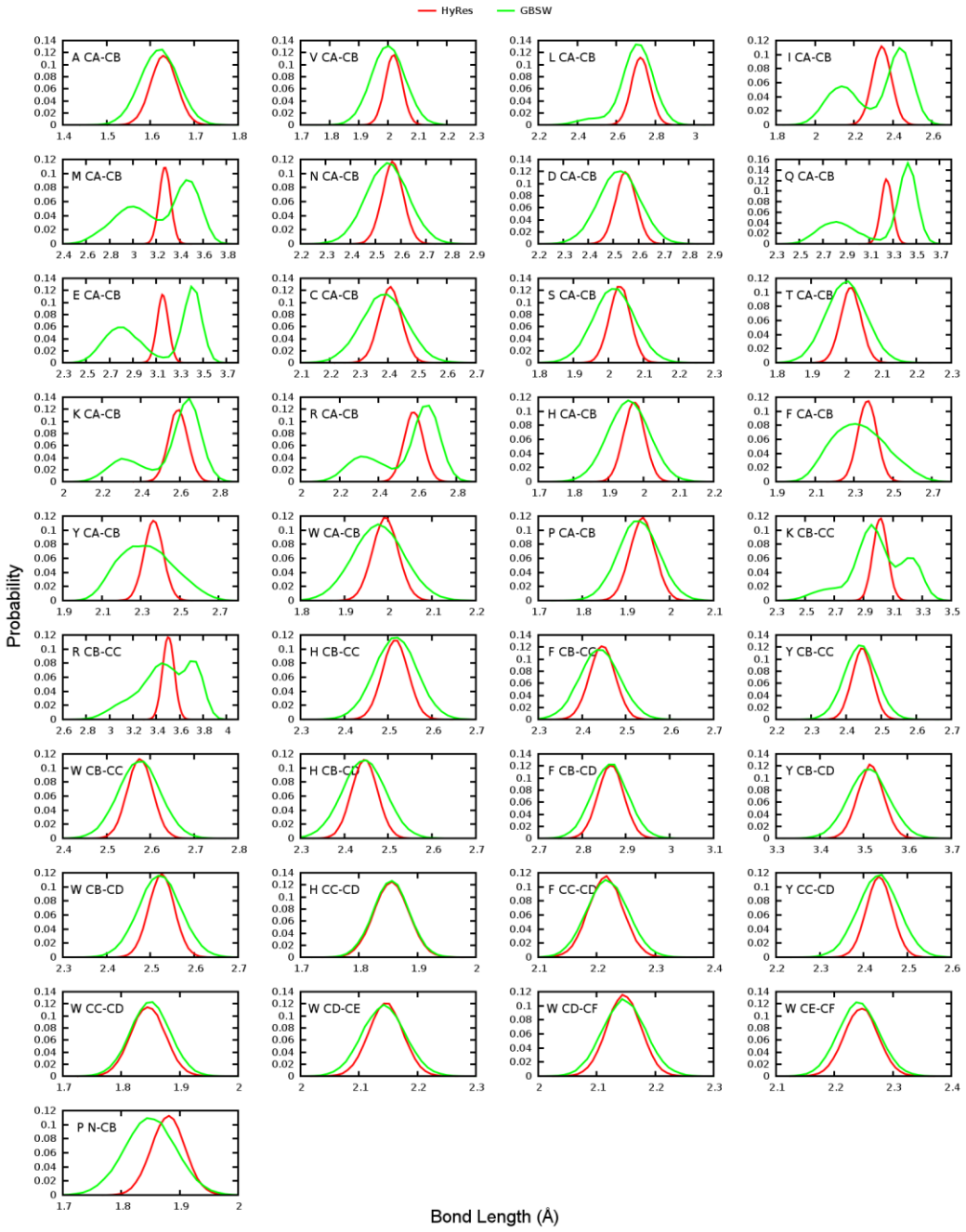

Figure S1. Distributions of side chain virtual bond lengths from MD simulations of dipeptides using the GBSW and HyRes models.

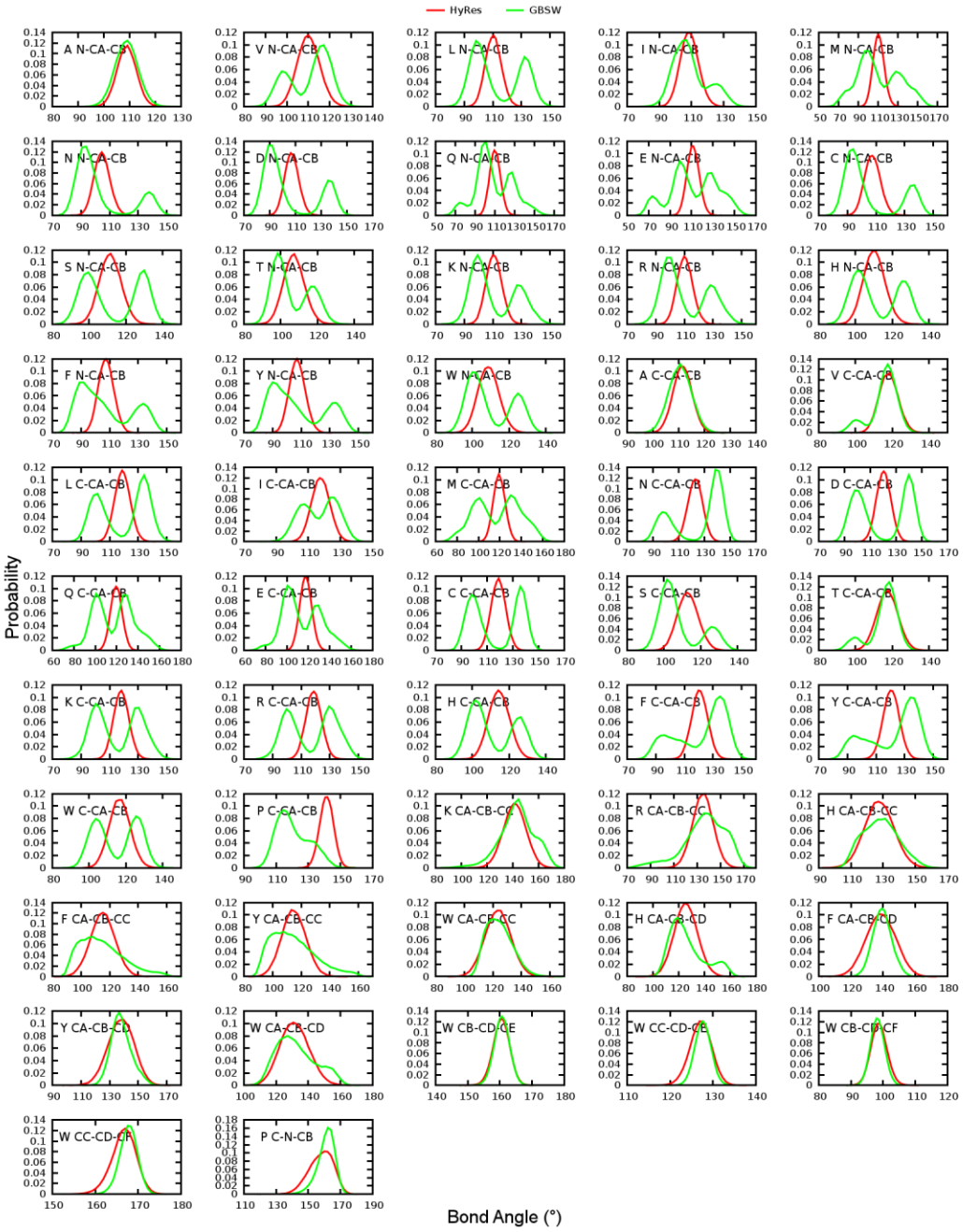

Figure S2. Distributions of side chain virtual bond angles from MD simulations of dipeptides using the GBSW and HyRes models.

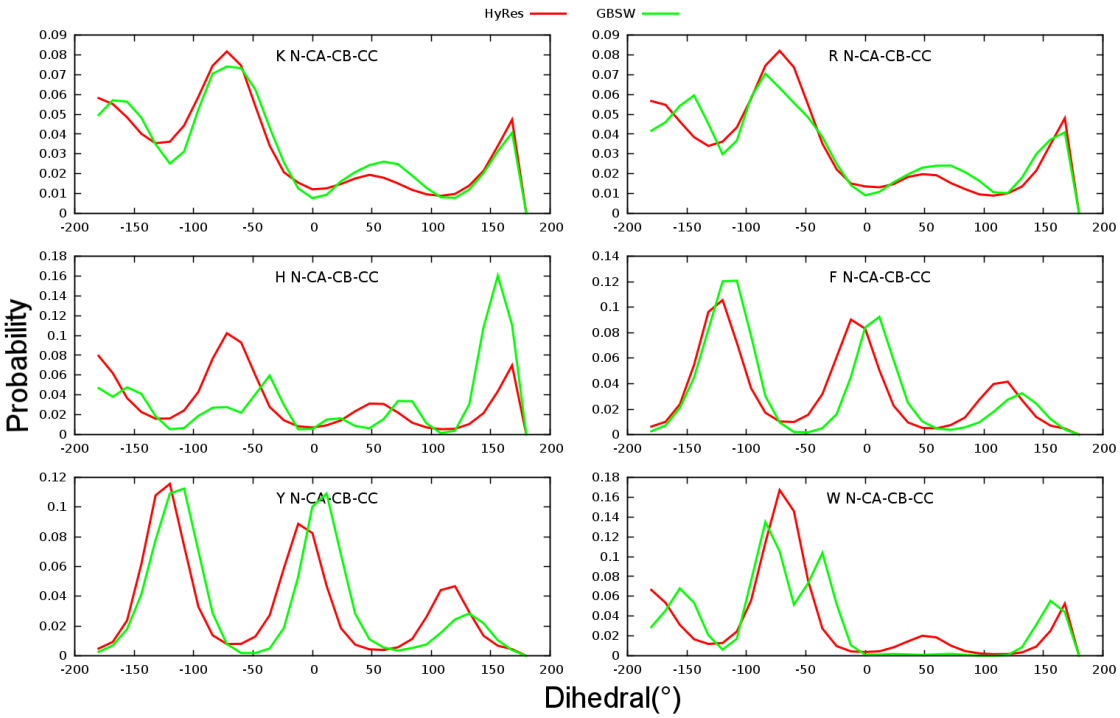

Figure S3. Probability distributions of dihedral χ in dipeptides obtained from CG and GBSW atomistic simulations.

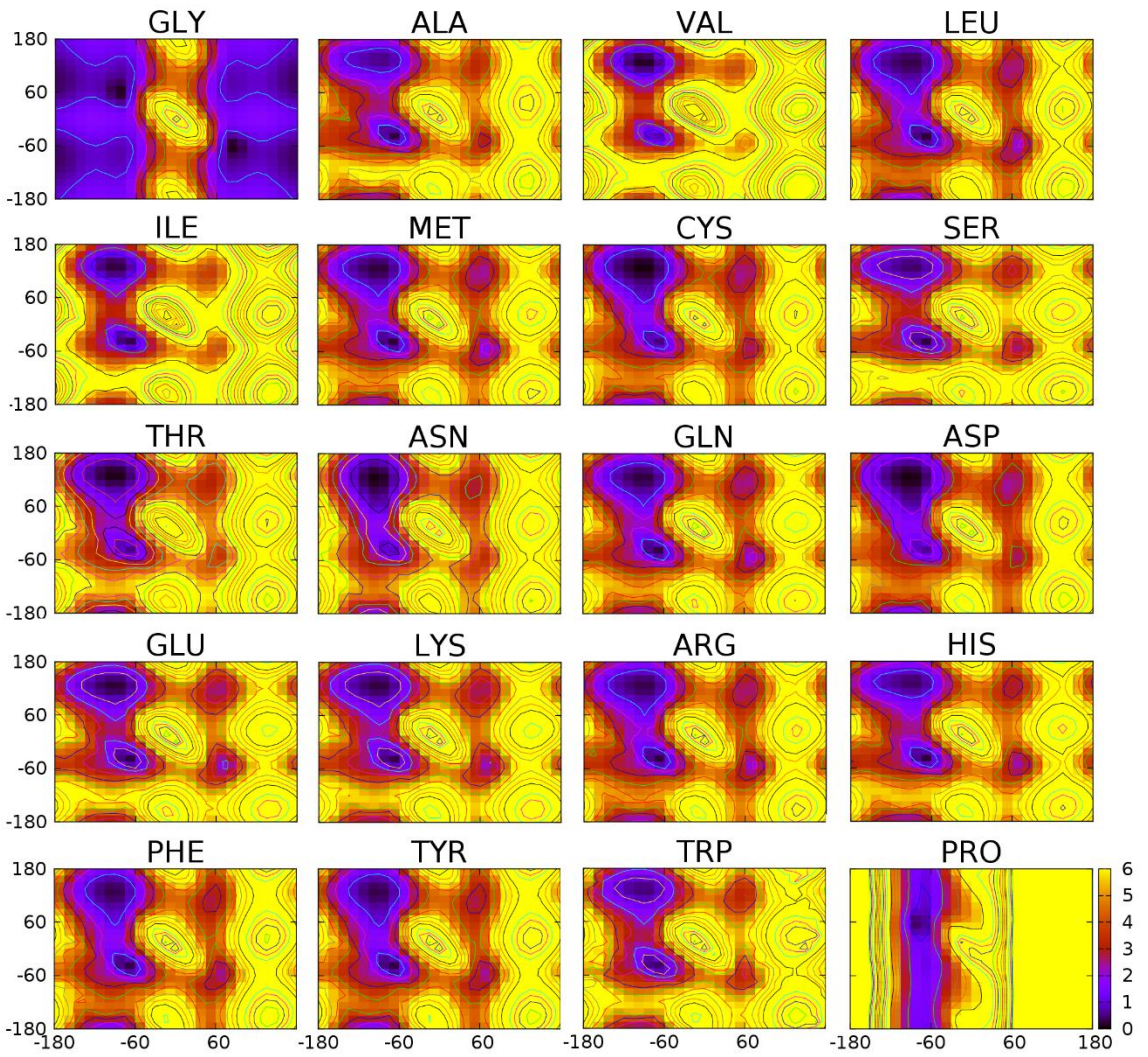

Figure S4. Backbone ϕ/ψ adiabatic energy surfaces (in kcal/mol) of all 20 dipeptides in the HyRes model. The surface was calculated by energy minimization with ϕ and ψ restrained at specified values. The surface was shifted such that the minimum value for each system was zero.

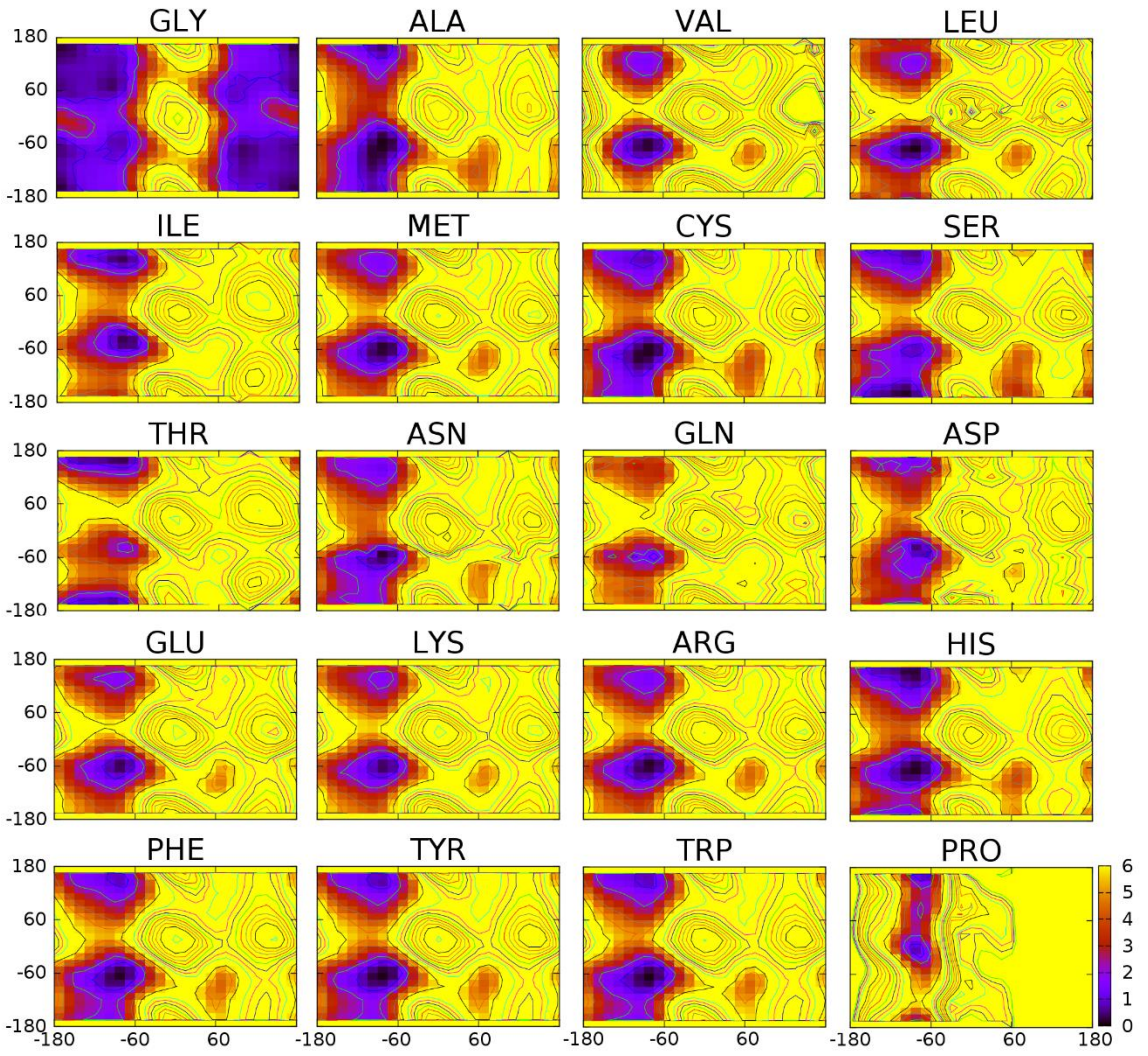

Figure S5. Backbone ϕ/ψ adiabatic energy surfaces (in kcal/mol) of all 20 dipeptides in the GBSW implicit solvent model. The surface was calculated by energy minimization with ϕ and ψ restrained at specified values. The surface was shifted such that the minimum value for each system was zero.

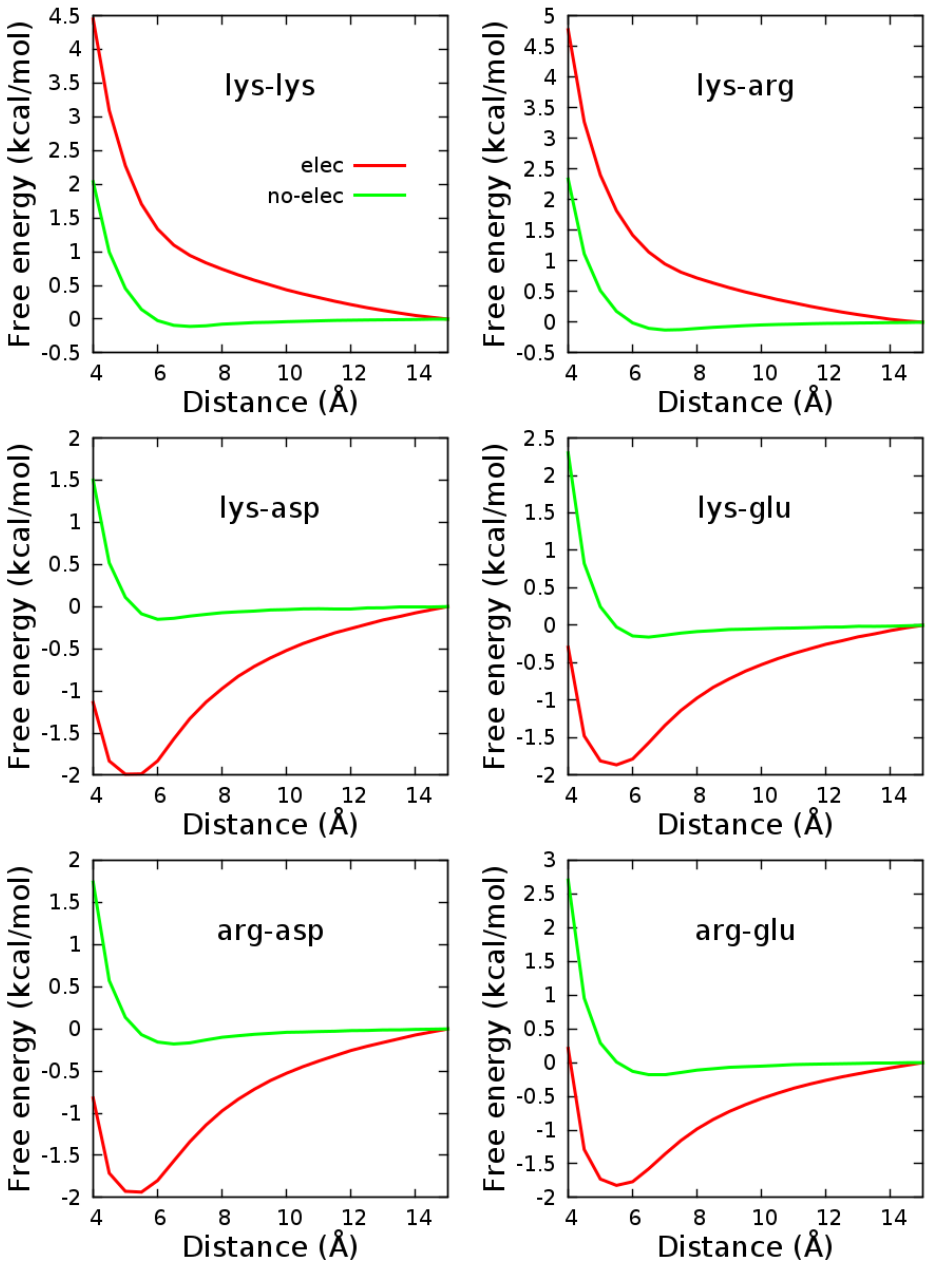

Figure S6. Free energy profiles as a function of separation distance between the center of mass of charged amino acid side chain analogs with (red traces) and without (green traces) electrostatic interactions.

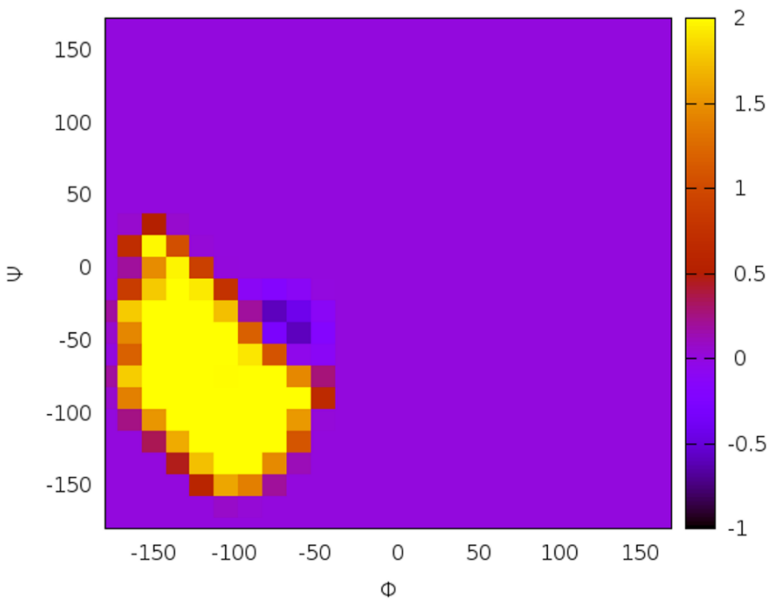

Figure S7. Backbone φ/ψ CMAP cross-term (in kcal/mol) in the HyRes model, which includes a small energy basin to stabilize α-helixes and a energy barrier to suppress the sampling of π-helixes.

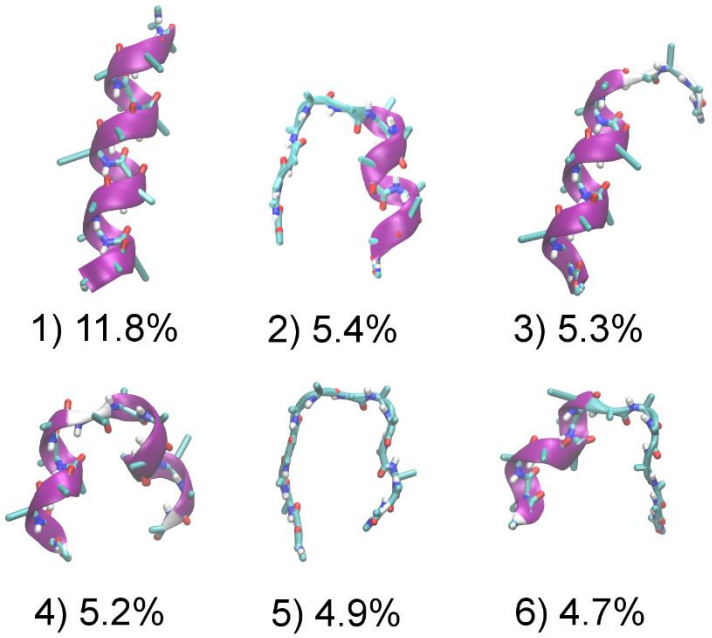

Figure S8. Centroids and populations of six largest clusters for $(AAQAA)_3$ structure ensemble in the control simulation using HyRes model. All heavy atoms were used to compute RMSD between structures, and a fixed radius of 4 Å was used to define clusters.

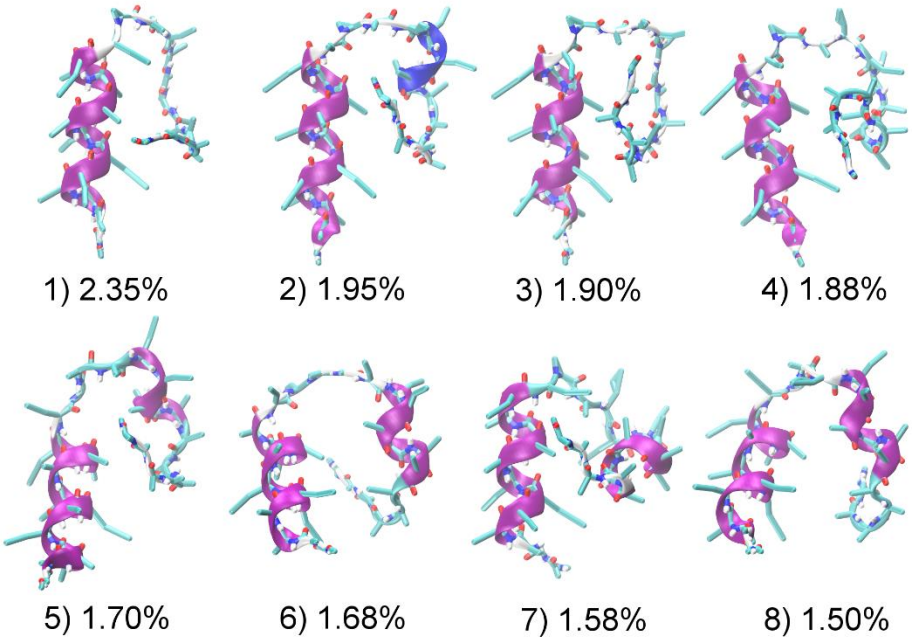

Figure S9. Centroids and populations of eight largest clusters for KID structure ensemble in the control simulation using HyRes model. CA and CB atoms were used to compute RMSD between structures, and a fixed radius of 4 Å was used to define clusters.

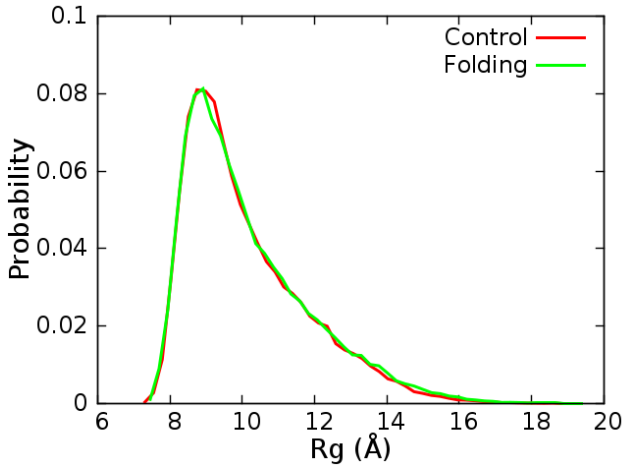

Figure S10. Probability distribution of R_g of RS peptide obtained from HyRes simulations at 300 K.

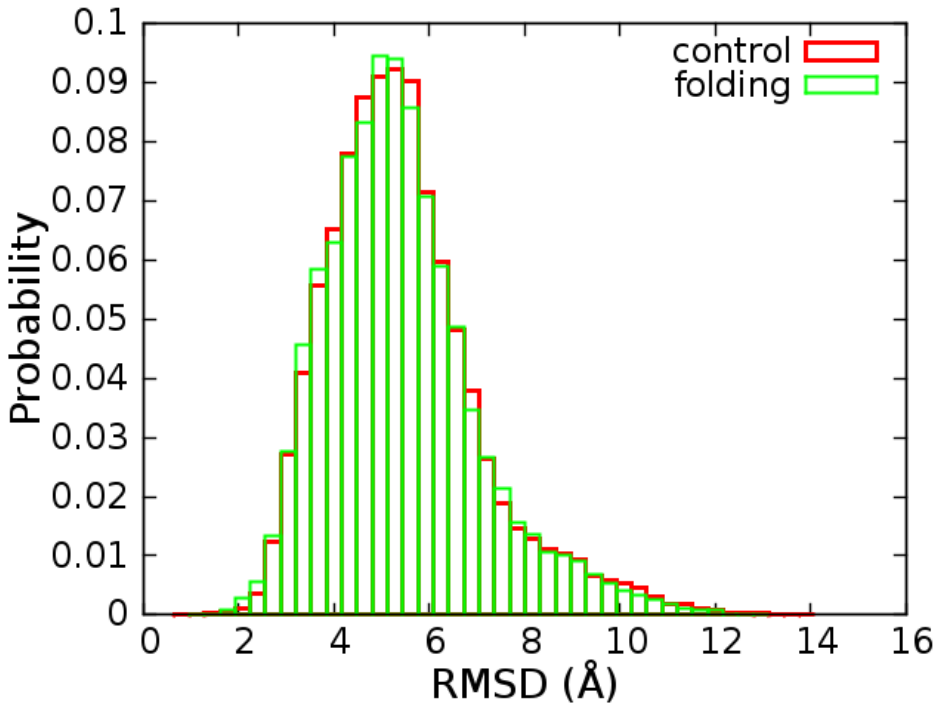

Figure S11. Probability distribution of backbone RMSD of GB1m3 peptide with respect to the folded state from HyRes simulations at 300 K.

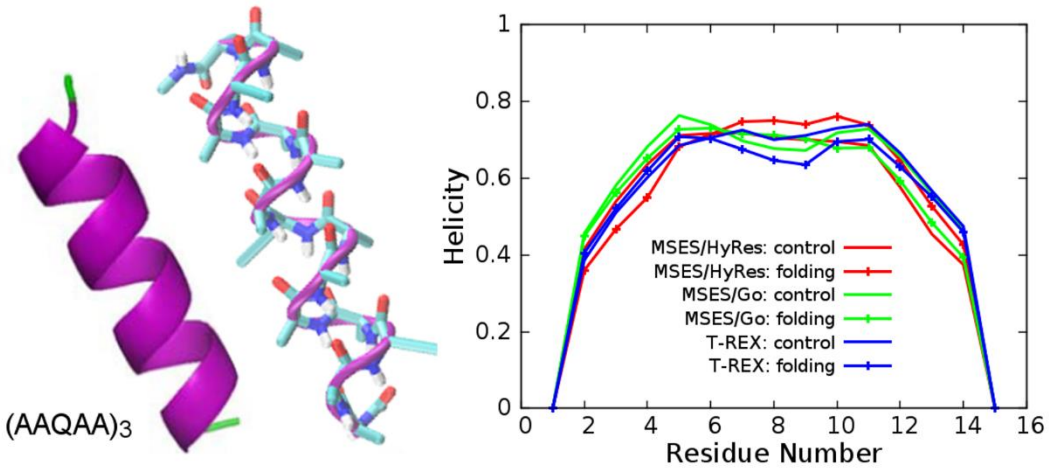

Figure S12. (Left) AT and HyRes representation of folded $(AAQAA)_3$. The AT model is shown in cartoon and HyRes model in Licorice representation, with the backbone highlighted in purple. (Right) Residual helicity profiles of $(AAQAA)_3$ at 270 K obtained from MSES/HyRes, MSES/Gō, and T-REX simulations (see **Table 1** of the main text). Note that only the second half of each trajectory was used in these calculations.