Published in final edited form as:

Phys Chem Chem Phys. 2017 December 13; 19(48): 32421–32432. doi:10.1039/c7cp06736d.

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

Xiaorong Liu^a and Jianhan Chen^{a,b}

^aDepartment of Chemistry, University of Massachusetts Amherst, Amherst, MA 01003, USA

^bDepartment of Biochemistry and Molecular Biology, University of Massachusetts Amherst, Amherst, MA 01003, USA

Abstract

Efficient coarse-grained (CG) models can be coupled with atomistic force fields to accelerate the sampling of atomistic energy landscapes in the multi-scale enhanced sampling (MSES) framework. This approach may be particularly suitable for generating atomistic conformational ensembles of intrinsically disordered proteins (IDPs). While MSES is relatively robust to inherent CG artifacts, achieving optimal sampling efficiency requires CG modeling to generate local and long-range fluctuations that are largely consistent with those at the atomistic level. Here, we describe a new hybrid resolution CG model (HyRes) for MSES simulations of disordered protein states, which is specifically designed to provide semi-quantitative secondary structure propensities together with a qualitative description of long-range nonspecific interactions. The HyRes model contains an atomistic description of the backbone with intermediate resolution side chains. The secondary structure propensities are tuned by adjusting the backbone hydrogen-bonding strength and ϕ/ψ torsion profile. The sizes and covalent geometries of side chains are parameterized to reproduce distributions derived from atomistic simulations. Lennard-Jones parameters for sidechain beads are assigned to reproduce statistical potentials derived from protein structural database, and then globally parameterized with nonspecific electrostatic interactions to reproduce free energy profiles of pair wise interactions and key conformational properties of model peptides. Application of HyRes to MSES simulations of small IDPs suggests that it is capable of driving faster structural transitions at the atomistic level and increasing the convergence rate compared to $\text{C}\alpha$ -only G β -like models previously utilized. With further optimization, we believe that the new CG model could greatly improve the efficiency of MSES simulations of larger and more complex IDPs frequently involved in cellular signaling and regulation.

Graphical Abstract

Correspondence to: Jianhan Chen.

[†]Electronic Supplementary Information (ESI) available: five tables and 12 figures. See DOI:

Conflicts of interest

There are no conflicts to declare.

Introduction

Intrinsically disordered proteins (IDPs) remain fully or partially disordered in the unbound state under physiological conditions^{1–4}. Compared with globular proteins, IDPs are enriched in polar and charged residues, and deficient in hydrophobic residues⁵. They often play significant roles in cellular signalling and regulation^{1, 6–10}, and are frequently associated with human diseases including cancer and neurodegenerative diseases^{7, 11–14}. There is thus a great need to understand the physical basis of IDP functions in molecular recognition and cellular signaling. The dynamic nature of unbound IDPs, however, requires their properties to be described by heterogeneous ensembles that are not amenable to traditional experimental techniques geared towards structured proteins. Averaged properties measured on disordered protein states are generally insufficient to uniquely define the underlying ensembles via various restrained structure calculation or selection selection protocols^{4, 15–18}. Reliable physics-based atomistic simulations thus have an important role to play.

We recently adapted the multi-scale enhanced sampling (MSES) approach of Kidera and coworkers¹⁹ to couple G²-like coarse-grained (CG) models for enhanced sampling of protein conformations in atomistic (AT) force fields^{20, 21}. In brief, both AT and CG representations of the system are simulated simultaneously, but they don't interact directly with each other except through a MSES coupling potential (U_{MSES}). Potential energy of the hybrid system is shown in Eq. 1,

$$U_{\text{hybrid}}(r_{\text{AT}}, r_{\text{CG}}, \lambda) = U_{\text{AT}}(r_{\text{AT}}) + U_{\text{CG}}(r_{\text{CG}}) + \lambda U_{\text{MSES}}(r_{\text{AT}}, r_{\text{CG}}) \quad (1)$$

where U_{AT} and U_{CG} are AT and CG potentials, respectively. U_{MSES} is designed to restrict the structural divergence between AT and CG models along selected degrees of freedom. A proper U_{MSES} is expected to leverage rapid CG fluctuation to accelerate structural transitions at the atomistic level, and the scaling factor λ determines how closely the AT and CG configurations could track each other. Bias introduced by MSES coupling is removed using Hamiltonian replica exchange (H-REX), which communicates coupled conformational dynamics to the limit of $\lambda=0$ where AT and CG models are completely independent and proper ensembles could be recovered at both resolutions. Our MSES approach also incorporates temperature replica exchange (T-REX) to further enhance the sampling efficiency^{20, 21}. Application to small β -hairpins and helical peptides supports that MSES could dramatically accelerate the sampling of peptide conformational transitions. Compared

with other multiscale sampling methods such as direct coordinate swapping in resolution exchange^{22–24}, coupling AT and CG models through U_{MSES} provides several significant advantages. It offers flexibility in controlling energetic impact of diverged conformations at AT and CG levels, which improves exchange efficiency and provides superior scalability to large systems. Also, it prevents the CG model from dictating the conformational dynamics and makes MSES more robust against inherent CG artifacts.

A key bottleneck in sampling large-scale conformational rearrangement, either for folded proteins or IDPs, is that these transitions are often entropically limited^{20–22}. Kinetics of such transition is temperature insensitive, and the popular temperature-based REX protocols becomes ineffective. In this regard, MSES could provide unique advantages, because CG modeling reduces the conformational space and may be particularly effective in driving conformational search at atomistic level. However, the efficiency of MSES depends critically on the ability of CG modeling to generate local and long-range structural fluctuations that are largely consistent with those expected at the atomistic level. Otherwise, the CG model would attempt to drive atomistic transitions through unfavorable pathways, which will hinder the convergence of MSES simulations. Previous studies have shown that coupling the AT model to a simple C α -based G β -like model could significantly increase atomistic structural transitions in simulations of both β -hairpin²⁰ and small IDPs²¹. However, these studies also suggest that the efficiency of MSES for simulating IDPs is limited by the existence of inconsistent structural fluctuations between C α -based CG and atomistic models²¹. This problem is anticipated to become more severe for larger and more complex IDPs. There is thus a need to explore or develop alternative CG models for more efficient MSES simulations of IDPs.

CG modeling is an area of intensive research and numerous CG protein models have been described in the literature. Some CG force fields treat solvent molecules implicitly, for example, the united-residue (UNRES) model^{25–30}, PRIME (protein intermediate resolution model)^{31–34}, Bereau and Deserno's CG model³⁵, OPEP (optimized potential for efficient structure prediction) force field^{36–40} and PRIMO force field⁴¹. Others include solvent molecules explicitly, such as MARTINI^{42, 43} and PACE⁴⁴. We refer the reader to references^{45–48} for more complete reviews. We note that virtually all existing CG models are designed to work independently (i.e., as alternatives to AT models), which presents formidable challenges and requires stringent optimization and/or more sophisticated effective potential energy functions. We have evaluated several CG protein models and the results suggest that none of them is capable of provide a balanced description of disordered protein ensembles⁴⁹. Here, we describe a new hybrid resolution CG model (HyRes) that has been specifically designed for MSES simulations of IDPs. A guiding hypothesis is that, the CG model should be able to capture essential conformational features of IDPs to achieve best possible efficiency of MSES simulations. In particular, most regulatory IDPs are known to contain some level of residual secondary structures (mostly helical) and transient long-range contacts with functional implications^{3, 15, 50, 51}. Importantly, the new model does not have to be highly accurate to be effective in MSES simulations, because the quality of structure ensembles derived from MSES simulations is ultimately determined by the AT model. In addition, HyRes should be designed with simple potential energy function to minimize the computational overhead in MSES simulations. To this end, the new model

contains a united-atom representation of the backbone to semi-quantitatively describe secondary structure propensities. Side chains are represented at intermediate resolution, to provide a qualitative description of long-range nonspecific interactions. The model is parameterized based on atomistic and CG simulations of model compounds and peptides as well as statistical potentials derived from protein structure database. The new model is then validated using a set of model IDPs and its efficacy for MSES simulations evaluated using short IDPs.

Models and Methods

Peptide model with hybrid resolutions

In the HyRes model, the peptide backbone is represented at the united-atom level and side chains at an intermediate resolution. The goal is to allow semi-quantitative description of secondary structure propensities together with a qualitative description of geometries and pair-wise interactions between side chains. Detailed description of peptide backbone, particularly the hydrogen bonding geometry, has been shown to be critical for realistic modeling of peptide conformational equilibrium⁵². Figure 1 summarizes the representation of all 20 amino acids in HyRes. There is no side chain bead in Gly, while there are two side chain beads in Lys and Arg, three side chain beads in His, Phe, and Tyr, five side chain beads in Trp, and one side chain bead in the other amino acids. Detailed mapping of side chain atoms from the atomistic model^{53–55} to CG beads are summarized in Table S1.

Effective potential energy function

The total interaction energy U is shown in Eq. 2. It consists of eight terms, including bond (U_{bond}), angle (U_{angle}), dihedral (U_{dihedral}), improper dihedral (U_{improper}), backbone ϕ/ψ torsion cross-term (U_{CMAP}), Lennard-Jones interactions (U_{LJ}), hydrogen bonding (U_{Hbond}) and electrostatic potentials (U_{elec}),

$$U = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{CMAP}} + U_{\text{Hbond}} + U_{\text{LJ}} + U_{\text{elec}} \quad (2)$$

$$U_{\text{bond}} = \sum_{\text{bonds}} k_b (b - b_0)^2 \quad (3)$$

$$U_{\text{angle}} = \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \quad (4)$$

$$U_{\text{dihedral}} = \sum_{\text{dihedrals}} k_\chi [1 + \cos(n\chi - \delta)] \quad (5)$$

$$U_{\text{improper}} = \sum_{\text{improper}} k_{\psi} (\psi - \psi_0)^2 \quad (6)$$

$$U_{\text{Hbond}} = \sum_{\text{Hbonds}} \varepsilon_{\text{HB}} \left[5 \left(\frac{\sigma}{r} \right)^{12} - 6 \left(\frac{\sigma}{r} \right)^{10} \right] \cos^4 \theta_{\text{AHD}} \quad (7)$$

$$U_{\text{CMAP}} = \sum_{\substack{\text{non-Gly, non-Pro} \\ \text{residues}}} U_{\text{CMAP}}(\varphi, \psi) \quad (8)$$

The bond (U_{bond}) and angle (U_{angle}) interactions are described as harmonic potentials (Eq. 3 and 4). U_{angle} is found between all adjacent bonds, except those constituting a triangle (such as for the His side chain) and the $\text{C}_{i-1}\text{-N}_i\text{-CB}_i$ angle for Pro. Bond and angle parameters for the peptide backbone are obtained directly from the CHARMM19 united-atom force field⁵⁶. The equilibrium values and force constants of other bonds and angles are optimized to reproduce distributions obtained from atomistic simulations of dipeptides in the GBSW implicit solvent⁵⁷ based on the mapping scheme (Table S1). The force constants are restricted to be between 200 to 600 kcal/mol/Å² for bonds and to 20–90 kcal/mol/rad² for angles during the optimization.

Dihedral potentials (U_{dihedral} shown in Eq. 5) are applied to peptide ϕ ($\text{C}_{i-1}\text{-N}_i\text{-CA}_i\text{-C}_i$), ω ($\text{CA}_{i-1}\text{-C}_{i-1}\text{-N}_i\text{-CA}_i$), χ ($\text{N}_i\text{-CA}_i\text{-CB}_i\text{-CC}_i$, if present) and ψ ($\text{CB}_i\text{-CA}_i\text{-C}_i\text{-O}_i$, only for a few selected residues, see below). Parameters related to ϕ are obtained directly from the CHARMM 19 force field⁵⁶. The peptide bond strongly favors the *trans*-conformation for all non-proline amino acids. Therefore, for ω , n is set to 2 and δ to 180° in Eq. 5, and a 9.00 kcal/mol energy barrier is used here to separate *cis*- and *trans*- conformations. The barrier reflects a slight increase from the 8.2 kcal/mol value in CHARMM19, such that only *trans*-conformation is sampled in CG simulations of dipeptides. However, for ω in $\text{X}_{i-1}\text{-Pro}_i$, both *cis*- and *trans*-conformations are accessible. To allow efficient sampling of both *cis*- and *trans*-conformations, a smaller energy barrier (5.40 kcal/mol) is applied to ω in this case, which is based on the previous report that the barrier of *trans*- to *cis*- rotation for $\text{X}_{i-1}\text{-Pro}_i$ is ~65% of that for $\text{X}_{i-1}\text{-nonPro}_i$ ⁵⁸. U_{dihedral} related to χ is applied to residues Lys, Arg, His, Phe, Tyr and Trp, and parameters were obtained by fitting the distribution of χ from atomistic simulations of dipeptides.

Improper potentials (U_{improper}) are required to maintain the chirality of $\text{C}\alpha$ atoms in non-Gly amino acids (i.e., on $\text{CA}_i\text{-N}_i\text{-C}_i\text{-CB}_i$, where CA_i is the central atom), planarity of peptide bonds (i.e., on $\text{N}_i\text{-C}_{i-1}\text{-CA}_i\text{-H}_i$ for non-Pro amino acids and $\text{C}_i\text{-CA}_i\text{-N}_{i+1}\text{-O}_i$ for all amino acids), and planarity of the Trp side chain ring. Although representing multiple atoms in the side chain by one bead could impact the geometry of $\text{C}\alpha$ -connected atoms/beads to different

degrees for different amino acids, we treat the chirality of C α atoms uniformly (i.e., 35.3°) for all non-Gly amino acids in the HyRes model.

The secondary structure propensities are mainly determined by the backbone hydrogen bonding interactions (U_{Hbond}) and ϕ/ψ torsion cross-term (via the CMAP term in CHARMM⁵⁹) (U_{CMAP}). U_{Hbond} is applied on peptide backbone N, H and O atoms, and depends on both the N-O distance and O-H-N angle (r and θ_{AHD} in Eq. 7, respectively). σ in Eq. 7 is set to the ideal hydrogen bond donor-acceptor distance of 2.9 Å. In the current model, the same CMAP is applied to all non-Gly, non-Pro residues. The main role of the CMAP term is to fine tune secondary structure preferences together with backbone hydrogen-bonding interactions. Specifically, stabilizing or dis-stabilizing potentials may be added to ϕ/ψ regions corresponding to certain secondary structures, such as α -helix, π -helix, parallel and anti-parallel β strand, and poly-proline II (PPII) structures.

Nonspecific van der Waals (vdW) interactions between atoms separated by at least 3 virtual bonds are described by Lennard-Jones (LJ) potential (U_{LJ}) (Eq. 9), where $r_{\text{min}}^i/2$ and ϵ_i are the vdW radius and interaction strength of bead i . LJ parameters of the backbone atoms were first obtained from the CHARMM 19 force field, and then the interaction strengths were scaled down to account for lack of solvent screening of protein-protein dispersion interactions (see **Results and Discussions** for details). Initial vdW radii of side chain beads are assigned by matching the total volume occupied by corresponding atoms in the AT model, and ϵ_i values assigned based on Miyazawa and Jernigan's (MJ) statistical potentials⁶⁰. These parameters are then optimized based on free energy profiles of pair-wise interactions and model peptide simulations (see **Results and Discussions** for details).

$$U_{\text{LJ}} = \sum_{i,j} \epsilon_{i,j} \left[\left(\frac{r_{i,j}^{\text{min}}}{r_{i,j}} \right)^{12} - 2 \left(\frac{r_{i,j}^{\text{min}}}{r_{i,j}} \right)^6 \right]$$

$$\text{with } r_{i,j}^{\text{min}} = \frac{r_i^{\text{min}} + r_j^{\text{min}}}{2}, \text{ and } \epsilon_{i,j} = \sqrt{\epsilon_i \epsilon_j} \quad (9)$$

There is no charge on either backbone atoms or side chain CG beads, except for those corresponding to charged groups in Lys, Arg, Asp, and Glu. Nonspecific electrostatic interactions are described using the Debye-Hückel-type potential,

$$U_{\text{elec}} = \sum_{\text{chrg. pairs}} \frac{q_i q_j}{4\pi\epsilon_r\epsilon_0 r_{i,j}} e^{-\frac{r_{i,j}}{D}} \quad (10)$$

where q_i is the charge of CG bead i , $r_{i,j}$ is the inter-charge distance, and ϵ_0 is the permittivity of vacuum. D is the Debye screening length. At 300 K, D (in Å) depends ionic strength (I , in M) according to $\sqrt{9.48/I}$. Note that similar potentials have been widely used in various CG models and shown to provide an adequate description of long-range electrostatic

interactions, even though accurate modeling of electrostatic solvation in heterogeneous protein environments requires more detailed and computationally more intensive implicit solvent theories. In the present work, I is set to 0.01 M, to roughly account for the low salt conditions in many NMR experiments. ϵ_r is the effective dielectric constant. The final value of ϵ_r , 20, was chosen in the current model based on simulations of several highly charged model peptides (see **Results and Discussions** for details).

AT and CG simulation protocols

The HyRes model has been implemented in CHARMM^{61, 62}. Multiple sets of AT and CG simulations were performed in explicit or implicit solvent for backbone and side chain analogs, dipeptides, and various model peptides. The protocols of these simulations are summarized below. Unless otherwise noted, Langevin dynamics was always used with a friction coefficient of 0.2 ps⁻¹ and a time step of 2 fs. SHAKE algorithm⁶³ was used to constrain the length of all bonds involving hydrogen atoms.

Dipeptide simulations—Explicit solvent simulations of the Ala dipeptide (Ace-Ala-Nme) were performed using both AMBER ff99SB-ILDN⁶⁴ and CHARMM36m⁶⁵ force fields, as implemented in GROMACS 2016.1⁶⁶. 586 TIP3P water molecules⁶⁷ were used to solvate the dipeptide. Simulations were performed in the NPT ensemble at 298 K and 1 atm pressure under periodic boundary conditions. The PME method⁶⁸ was used to treat long-range electrostatic interactions. Short-range nonbonded interactions were truncated at 1.0 nm in the AMBER ff99SB-ILDN simulation, and 1.2 nm in CHARMM36m simulation. Each simulation lasted for 200 ns. In addition, all 20 dipeptides were simulated at 300 K for 300 ns in the GBSW implicit solvent, to derive the distributions of side chain conformations for parameterization of bonding parameters as well as for initial assignment of vdW radii of side chain beads.

Free energy of side chain interactions—Potentials of mean force (PMFs) for pairwise interactions between hydrophobic side chains were calculated by umbrella sampling simulations at 300 K in both the CHARMM36 explicit solvent and GBSW implicit solvent⁵⁷ atomistic force fields, and compared to the results from the HyRes model. In atomistic simulations, harmonic potentials with a force constant of 5.0 kcal/mol/Å² were imposed on the center of mass distance between the pair of side chain analogs. 17 windows were used, spanning from 4 to 12 Å at a spacing of 0.5 Å. The simulation time at each window was 5 and 8 ns for explicit and implicit solvent calculations, respectively. In the CG simulation, the force constant of harmonic restraining potentials was 1.0 kcal/mol/Å². A total of 12 windows were used, spanning the reaction coordinate from 4 to 15 Å at a spacing of 1 Å. Each window was sampled for 200 ns. The weighted histogram analysis method (WHAM)⁶⁹ was then utilized to calculate the free energy profiles along the center of mass separation distance.

Model peptide simulations—Multiple model peptides were used in parameterization and validation of HyRes, including Gly₁₀, (AAQAA)₃, 30 variants of the (EK)₂₅ peptide⁷⁰, the 24-residue arginine/serine (RS) peptide, the 28-residue kinase inducible transactivation domain of transcription factor CREB (KID), the activation domain of the activator for

thyroid hormone and retinoid receptors (ACTR, residues 1040–1086), nuclear coactivator binding domain of the CREB binding protein (NCBD, residues 2059–2117), and a β -hairpin GB1m3⁷¹. The sequences of all model peptides are summarized in Table S2.

Conformational ensembles of relatively shorter and simpler Gly₁₀, (AAQAA)₃, RS peptide, KID, and GB1m3 peptides in HyRes were calculated using direct MD simulations at 300 K. For each peptide, two independent simulations were performed, one starting from the folded or helical structure (control run) and the other starting from fully extended structure (folding run). The folded configuration of KID was obtained from RCSB Protein Data Bank (PDB) entry 1KDX⁷². Comparison of results from folding and control runs demonstrate that convergence can be achieved with 200 ns MD.

Conformational ensembles of ACTR, NCBD and (EK)₂₅ peptides in HyRes were calculated using T-REX simulations, performed using CHARMM in combination with the MMTSB toolset⁷³. For (EK)₂₅ peptides, T-REX simulations were initiated from fully extended structures. For ACTR and NCBD, both control and folding runs were performed for convergence analysis. The folded conformations in control runs of ACTR and NCBD were obtained from their complex structure (PDB: 1KBH⁷⁴). Each simulation involved eight replicas with temperatures exponentially spaced between 300 and 500 K, and exchanges between neighboring replicas were attempted every 2 ps. The simulations lasted for 200 ns per replica. Averaged exchange acceptance ratio was ~20% for all three peptides. Conformations sampled at the lowest temperature were collected for analysis.

For Gly₁₀, four additional 10 ns explicit solvent simulations were performed at 300 K under NPT conditions using the CHARMM36m force field to estimate solvent screening of dispersion interactions. The peptide was solvated by 7240 TIP3P water molecules in a 60 Å-cubic box. In these simulations, the peptide end-to-end distance was restrained to either 7 or 20 Å to represent the compact or extended conformation states. We also carried out Monte Carlo (MC) simulation of free Gly₁₀ using ABSINTH model⁷⁵ implemented in CAMPARI package to calculate the conformational properties. ABSINTH has been highly successful in describing the IDP sequence composition-structural preference relationship^{75–78}, and thus provides a good reference for examining the conformational properties of Gly₁₀. The MC simulation follows the protocol as defined in “file abs3.2_opls.prm” in CAMPARI. The peptide was put in a droplet of 80 Å radius and simulated at 300 K under NVT conditions. The simulation lasted 41,000,000 MC steps, with the first 1,000,000 steps excluded in the analysis. A 200-ns simulation of Gly₁₀ was also performed in the GBSW implicit solvent for comparison.

MSES simulations

In MSES simulations of (AAQAA)₃, AT and CG models were coupled using U_{MSES} imposed on (i, i+4) C α atoms that form native helical contacts. U_{MSES} with soft asymptote was used to reduce energy penalty for significantly different conformations between AT and CG models and avoid exchange bottleneck between neighboring replicas²¹,

$$\begin{aligned}
 U_{\text{MSES}}(r_{\text{AT}}, r_{\text{CG}}) &= \sum_i 0.5k_i (d_i^{\text{AT}} - d_i^{\text{CG}})^2, \quad \text{if } |\Delta d_i| \leq d_s \\
 &= \sum_i A + \frac{B}{(d_i^{\text{AT}} - d_i^{\text{CG}})^s} + f_{\text{max}} (d_i^{\text{AT}} - d_i^{\text{CG}}), \quad \text{if } |\Delta d_i| > d_s \\
 &\text{with } \Delta d_i = d_i^{\text{AT}} - d_i^{\text{CG}}
 \end{aligned} \tag{11}$$

where d_i^{AT} and d_i^{CG} are Ca-Ca distances for residue pair i in atomistic and CG models, respectively. d_s is the distance threshold where U_{MSES} smoothly switches from quadratic to the soft asymptote. s determines how quickly the maximal force, f_{max} , is approached at large distance difference $|\Delta d_i|$. A and B are identified by requiring both MSES coupling energy and force to be continuous at $|\Delta d_i| = d_s$. The H-REX protocol previously optimized for disordered proteins²¹ is used in this work. For comparison, three independent sets of simulations were performed using T-REX in GBSW implicit solvent and MSES with the GBSW model coupled to the Ca-only G-like model (MSES/G) or the HyRes model (MSES/HyRes). For each protocol, both control and folding simulations were initiated from helical and extended structures, respectively, for convergence analysis. All simulations were performed with 16 replicas and temperatures ranging from 270 K to 500 K. Other simulation details are listed in Table 1.

Results and Discussions

Side chain bonded parameter optimization

Packing of peptide side chains has long been recognized as a key factor in determining protein structure (and folding)⁷⁹. A flexible CG protein model thus needs to provide a reasonably realistic description of side chain geometries. Accordingly, the equilibrium bond and angle parameters (as well as vdW radii) of the HyRes model are assigned to maximally reproduce atomistic distributions according the mapping scheme (Table S1). The corresponding bond and angle force constants are assigned to reproduce the root-mean-squared (RMS) fluctuations from atomistic simulations. As summarized in Figure 2, the average values of all virtual bond lengths and angles observed in the atomistic GBSW model are well reproduced in HyRes. Importantly, the CG model is also able to reproduce the distributions of most virtual bonds and angles, except those displaying multi-modal distributions (Figure S1 and S2). These are the cases where the side chain samples multiple rotamers at the atomic resolution and reflect an inherent consequence of coarse-graining. For these bonds and angles, the associated standard deviation derived from atomistic simulations are larger than those in the HyRes model (Figure 2). Once the bond and angle parameters are assigned, side chain χ ($\text{N}_i\text{-CA}_i\text{-CB}_i\text{-CC}_i$) dihedral potentials are parameterized by fitting atomistic distributions to cosine expansions for residues Lys, Arg, His, Phe, Tyr and Trp. The resulting parameters are summarized in Table S3. As shown in Figure S3, all atomistic distributions along χ can be well reproduced by HyRes. Additional dihedral potentials are imposed on ψ' ($\text{CB}_i\text{-CA}_i\text{-C}_i\text{-O}_i$), to finely tune the helical propensities of individual amino acids (independent of the backbone ϕ/ψ CMAP cross-term). Since Asp, Asn, Thr, Cys and Val have been suggested to have lower helical propensities than other non-Gly, non-Pro

residues^{80, 81}, weak U_{dihedral} along ψ' is applied to these residues to capture this feature (see Table S4).

Scaling of backbone vdW interactions

The original CHARMM19 vdW parameters for the backbone over-estimate the effective strengths of nonspecific dispersion interactions in HyRes due to a lack of solvent screening, and this leads to severe over compaction in peptide simulations. To address this limitation, we uniformly scale down the strengths of backbone vdW interactions (ϵ_1 in Eq. 9). Determination of the appropriate scaling factor is guided by explicit solvent simulations of Gly₁₀, a highly flexible peptide that does not have any side chains and is thus particularly sensitive to the balance of water-mediated backbone-backbone interactions. Two sets of explicit solvent simulations of Gly₁₀ were performed on representative “compact” and “extended” states to estimate the ratio between net contribution of vdW interactions and the intra-peptide component alone:

$$f = \frac{U_{\text{vdw}}^{\text{tot}}(\text{cmp}) - U_{\text{vdw}}^{\text{tot}}(\text{ext})}{U_{\text{vdw}}^{\text{intra}}(\text{cmp}) - U_{\text{vdw}}^{\text{intra}}(\text{ext})} \quad (12)$$

The results show that the net contribution of vdW interactions to the stabilization of compact states is about 20–40% of the intra-peptide component alone (see Table S5 for details). That is, solvent screening reduces the energetic contribution of nonspecific vdW interactions to conformational collapse by 60–80%. Accordingly, vdW interactions of all backbone atoms (ϵ_1) are reduced to 30% of the original CHARMM19 values to approximately account for solvent screening in HyRes. The scaling significantly reduces over-compaction. The conformational properties of Gly₁₀ calculated using HyRes agrees well with results from various atomistic simulations in implicit and explicit solvents (Table 2 and Figure 3). We note that GBSW is known to generate overly compact ensembles (e.g., compared to ABSINTH, blue vs green traces in Figure 3). HyRes appears to display a small bias towards compaction, which is likely due to lack of more sophisticated implicit treatment of solvent. Benchmark simulations suggest that the compaction bias is smaller for more complex and longer peptides (see below), and the current choice of scaling appears appropriate.

Side chain non-bonded parameters

The vdW radii of all amino acid side chain beads were first assigned based on the average volume occupied by corresponding atoms in GBSW implicit solvent simulations of dipeptides, as $R = \sqrt[3]{3V/4\pi}$. The initial values of vdW radii were fine-tuned by examining the backbone ϕ/ψ adiabatic energy surfaces of dipeptides and PMFs of side chain pair-wise interactions. The backbone ϕ/ψ energy surface is particularly sensitive to the covalent geometry and vdW radii. Other non-bonded parameters such as hydrogen bonding interactions and vdW interaction strengths have relatively weaker effects. It was found that special 1–4 vdW radii were needed for most amino acid side chain beads to reproduce results from GBSW implicit solvent, e.g., comparing Figures S4 and S5. The final assignment of vdW radii is summarized in Table 3. MD simulations of Ala dipeptide were

also performed to compare HyRes with AMBER ff99SB-ILDN⁶⁴ and CHARMM36m⁶⁵ explicit solvent force fields. The results are shown in Figure 4. Although details in free energy landscapes are different among all three models, probabilities of visiting various secondary structure regions (e.g., α helical vs extended states) appear similar. Note that inherent secondary structure propensities of HyRes have also been finely tuned using the ϕ/ψ CMAP and ψ' torsion terms based on model peptide simulations. This will be described in detail in the following sections.

To determine vdW interaction strengths of side chain beads, we followed a deconvolution strategy previously described by Bereau and Deserno³⁵ to reproduce the 20×20 MJ matrix using 20 parameters (ϵ'_i) based on the Lorentz-Berthelot (LB) mixing rule (Eq. 9). The final values of 20 ϵ'_i were listed in Table 4. As shown in Figure 5, the pair-wise vdW interaction strengths calculated using 20 ϵ'_i is in excellent agreement with the original MJ matrix, with the overall RMSD of 0.25 RT. Only 11 out of the 210 pair-wise interactions deviate by greater than 0.5 RT (Table 5). Deviation of Cys-Cys pair-wise interaction from MJ potential may be related to disulfide bond formation in proteins. Among these 11 pairs of residues, nine pairs are between charged residues or between charged and hydrophobic residues. Interactions between oppositely charged residues were underestimated by this deconvolution method, while overestimated for residues with same charges (Table 5). When explicit electrostatic interactions are taken into account between charged residues, the actual strengths of interactions involving charged residues agree better with the MJ matrix (Figure S6). Importantly, the nonspecific attractive interactions between oppositely charged residues in HyRes are ~ 2 kcal/mol (Figure S6), which is very similar to results obtained from atomistic simulations in explicit solvent^{44, 83}.

The deconvoluted parameters (ϵ'_i) for 20 amino acids are then rescaled based on the PMFs of pair-wise interactions between side chain analogs using atomistic results as a reference. During the rescaling process, the minimum and maximum values of ϵ'_i (for Lys and Leu, respectively) are adjusted while the relative scale $(\epsilon'_i - \min[\epsilon'_i]) / (\max[\epsilon'_i] - \min[\epsilon'_i])$ of each residue is maintained. Optimal agreement with atomistic PMFs were found with $\min[\epsilon'_i] = 0.1$ kcal/mol and $\max[\epsilon'_i] = 0.9$ kcal/mol. For residues with multiple CG beads in the side chain, such as Lys, Arg, His, Phe and Trp (see Figure 1), the scaled ϵ'_i needs to be further partitioned to the constituent CG beads based on the mapping scheme (Table S1). The final parameters are summarized in Table 3. As illustrated in Figure 6, the resulting PMFs of side chain interactions in HyRes are in excellent agreement with atomistic results. Note that HyRes tends to shift the minima of PMFs towards larger values, which is a direct consequence of using CG beads to represent groups of atoms with anisotropic structures. An implication is that HyRes likely remains limited in accurate description of side chain packing. However, the main objective of developing HyRes is to provide a better CG modeling for MSES simulation of atomistic ensembles. A highly precise of long-range interactions at the CG level is not expected to be critical.

Tuning of electrostatic interactions

The Debye–Hückel-type potential has proven highly successful in description nonspecific protein interactions⁸⁴. A main parameter subject to tuning in Eq. 10 is the effective dielectric constant ϵ_r , which directly scales the strengths of electrostatic interactions. The optimal value for a given CG model should be chosen to provide a proper balance among various competing short- and long-range interactions. We evaluate such balance by examining the conformational properties of a set of 30 variants of (EK)₂₅ peptide, which were originally designed by Das and Pappu⁷⁰ to study how charge patterning governs the conformational preferences of IDPs. A parameter, κ , ranging from 0 to 1, has been defined to quantify the charge patterning, with low values for well-mixed sequences and high values for segregation of oppositely charged residues in the sequence. As shown Figure 7, the ensemble averaged radius of gyration (R_g) of (EK)₂₅ peptide decreases as κ increases, a trend that is well reproduced by HyRes with $\epsilon_r = 20$. This suggests that the overall balance between electrostatic and other interactions is reasonable. Nonetheless, HyRes tends to generate conformational ensembles more compact than ABSINTH for some sequences with intermediate κ (e.g., between 0.1 and 0.4). This again likely reflects inherent limitations associated with CG representations and greatly simplified energy function employed (particularly, treatment of solvation).

Balancing backbone hydrogen-bonding and torsion potentials

Once all vdW parameters were assigned, the strength of backbone hydrogen bonding interactions was tuned together with the ϕ/ψ CMAP term to recapitulate the secondary structure propensities of model peptides. In this work, we mainly focused on helix-coil transitions and used (AAQAA)₃ as a primary model peptide in backbone tuning. (AAQAA)₃ is partially helical that is about 20% folded at 300 K⁸⁵; it has been frequently used in atomistic and CG force field optimization^{21, 57, 86}. The backbone parameters optimized based on (AAQAA)₃ were then fine-tuned using KID, a slightly more complex IDP that folds into two helices upon binding to specific target KIX⁷². NMR secondary chemical shifts and atomistic simulations have shown that N-terminal helical segment is about 50–60% folded, while the second helical segment is only 10–15% folded^{87, 88}. The final CMAP is shown in Figure S7, and hydrogen bond interaction strength (ϵ_{HB} in Eq. 7) is set to 1.5 kcal/mol. Figure 8 depicts the residual helicity profiles of (AAQAA)₃ and KID calculated from the HyRes model, with representative structures shown in Figure S8 and S9, respectively. The results are highly consistent with both NMR and previous atomistic simulations, suggesting that the HyRes model is able to capture the sequence dependence of helical propensity quite well. This property is expected to be highly beneficial for driving faster atomistic sampling in MSES simulations of IDPs.

HyRes simulations of nontrivial IDPs

The HyRes model was used to study the RS peptide, a highly charged, disordered peptide that has been extensively characterized by NMR⁸⁹, small angle X-ray scattering (SAXS)⁹⁰, and various atomistic simulations^{65, 90, 91}. RS peptide in HyRes simulations showed averaged R_g of 10.1 Å at 300 K (Figure S10), which agreed well with SAXS measurements (12.62 Å)⁹⁰ by considering CG bead radii of 2–3 Å (Table 3). The optimized HyRes model

was further validated using two longer IDPs with more complex conformational properties, ACTR and NCBD. These two IDPs can synergistically fold into a stable helical complex upon binding⁹². In the unbound state, ACTR has a very low level of residual helices, while NCBD contains has molten globule characteristics with folded-like helical contents^{93, 94}. Key conformational properties obtained from T-REX simulations in HyRes are summarized in Figure 9. The results show that HyRes yields residual helices in the correct regions revealed in previous studies⁵⁰. It also successfully predicts higher helical contents in NCBD than ACTR, even though the model over-estimates the helical content of ACTR while under-estimating that of NCBD. The calculated R_g distribution of ACTR also agrees well with the result from ABSINTH simulations. However, the overall dimension of ACTR seems to be under-estimated compared to SAXS measurements, which yielded an average $R_g \sim 26.3$ Å at 278 K for a longer 71-residue construct⁹⁵ (vs. the 47-residue segment simulated here). Over-estimation of the helical content is likely a key contributing factor here. For the larger peptide NCBD (59 residues), HyRes generates largely globular conformational ensembles with an averaged R_g of ~ 11.4 Å, which is in reasonable agreement with the value of ~ 15.2 Å derived from SAXS (after taking into account CG bead radii of 2–3 Å in HyRes, see Table 3)⁹⁶. For a β -hairpin peptide GB1m3, which is $\sim 86\%$ folded at 298 K based on NMR measurement⁷¹, the current HyRes model could sample the folded state, but underestimates its stability (Figure S11). Taken together, these benchmark simulations support the notion that HyRes has achieved the main design objective, which is to provide a semi-quantitative description of secondary structures and a qualitative description of the long-range conformational properties. We note that the model could be further improved by including better treatment of hydrogen bonds and solvation.

Efficiency of HyRes in driving helix-coil transitions in MSES simulations

Previous work has shown that MSES approach is able to accelerate sampling of protein conformations in the rough, complex atomistic energy landscapes, even when the atomistic model is only coupled with a simple C α -based G β -like model^{20, 21}. MSES simulations of disorder proteins should benefit further from CG models that can generate both local and long-range structural fluctuations consistent with those observed in atomistic models. The performance of the newly developed HyRes model in MSES simulations is examined using the model peptide (AAQAA)₃. As shown in Figure S12, atomistic helicity profiles obtained from T-REX and MSES simulations, coupled to either the C α -only G β -like model or HyRes, are essentially identical as expected. This validates that MSES simulations are able to generate proper canonical ensemble free of bias due to MSES coupling. The efficacy of using HyRes to accelerate atomistic conformational transitions in MSES simulations is then assessed by computing the reversible helix/coil transition rate for each residue in (AAQAA)₃ at the atomistic level. As summarized in Figure 10, the atomistic helix/coil transition rate observed in MSES/G β simulation is ~ 2 fold of that in atomistic T-REX simulations. Using the HyRes model in MSES simulations further accelerates atomistic structural transitions to ~ 3 fold (red traces in Figure 10). Although the exact values of helix-coil transition rate depend on a lot of factors, such as capping groups, peptide length⁹⁷, concentration, temperature⁹⁸, hydrodynamics^{99, 100}, etc., our result demonstrates that the HyRes model could drive faster conformational transitions in atomistic energy landscapes. Faster helix-coil transitions translate well into faster convergence rate in averaged conformational

properties such as residue helicity profiles. As shown in Figure 11, the residual helicity profiles of (AAQAA)₃ obtained from MSES/HyRes simulations converge rapidly in both control and folding runs. Especially in the folding runs, MSES/HyRes simulation became converged faster than both MSES/G and T-REX simulations.

Conclusions

Understanding the structural basis of IDPs recognition and regulation is critical in unveiling the mechanisms of IDP functions and associated diseases. Previous studies have demonstrated that the MSES scheme is highly suitable for characterizing conformational properties of IDPs. One of the factors that limits the MSES sampling efficiency is the ability of CG models to generate both local and long-range structural fluctuations that are largely consistent with AT models. Ca-only G-like models have proven to be inadequate. Here, we describe a new HyRes model designed especially for MSES simulations of IDPs. The model treats solvent molecules implicitly. The peptide backbone is at atomistic resolution, which allows for semi-quantitative description of secondary structure propensities of IDPs. Side chains are represented at an intermediate resolution with one or more CG beads, whose parameters were carefully tuned provide a qualitative description of transient peptide-peptide interactions and long-range conformational properties. Benchmark simulations of several IDPs with various level of complexities demonstrate that the HyRes model has achieved all design objectives. Initial application to simple model IDPs suggests HyRes can drive faster atomistic transitions in MSES and leads to faster convergence in atomistic structure ensembles. With further optimization of this CG model using more peptides containing β strands, the HyRes model could be extended to study IDPs with more complex secondary structures. We also note that achieving the best sampling efficiency with coupling to HyRes will require further optimization of the MSES coupling scheme and H/T-REX protocol. These developments are ongoing, which can be anticipated to further improve the efficiency of MSES for studies of larger and more complex IDPs frequently involved in cellular signaling and regulation.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work is supported by the National Institutes of Health (GM114300). The computing was performed at Massachusetts Green High Performance Computing Center (MGHPCC).

References

1. Wright PE, Dyson HJ. Nat Rev Mol Cell Bio. 2015; 16:18–29. [PubMed: 25531225]
2. Uversky VN, Oldfield CJ, Dunker AK. J Mol Recognit. 2005; 18:343–384. [PubMed: 16094605]
3. Dyson HJ, Wright PE. Nat Rev Mol Cell Bio. 2005; 6:197–208. [PubMed: 15738986]
4. Chen JH. Archives of biochemistry and biophysics. 2012; 524:123–131. [PubMed: 22579883]
5. Romero P, Obradovic Z, Li XH, Garner EC, Brown CJ, Dunker AK. Proteins-Structure Function and Genetics. 2001; 42:38–48.
6. Mitrea DM, Kriwacki RW. Febs Lett. 2013; 587:1081–1088. [PubMed: 23454209]

7. Babu MM, van der Lee R, de Groot NS, Gsponer J. *Curr Opin Struct Biol.* 2011; 21:432–440. [PubMed: 21514144]
8. Smock RG, Gierasch LM. *Science.* 2009; 324:198–203. [PubMed: 19359576]
9. Boehr DD, Nussinov R, Wright PE. *Nature chemical biology.* 2009; 5:789–796. [PubMed: 19841628]
10. Gsponer J, Futschik ME, Teichmann SA, Babu MM. *Science.* 2008; 322:1365–1368. [PubMed: 19039133]
11. Lu HC, Chung SS, Fornili A, Fraternali F. *Frontiers in molecular biosciences.* 2015; 2:47. [PubMed: 26322316]
12. Vacic V, Markwick PR, Oldfield CJ, Zhao X, Haynes C, Uversky VN, Iakoucheva LM. *PLoS Comput Biol.* 2012; 8:e1002709. [PubMed: 23055912]
13. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK. *J Mol Biol.* 2002; 323:573–584. [PubMed: 12381310]
14. Uversky VN, Dave V, Iakoucheva LM, Malaney P, Metallo SJ, Pathak RR, Joerger AC. *Chemical reviews.* 2014; 114:6844–6879. [PubMed: 24830552]
15. Click TH, Ganguly D, Chen J. *Int J Mol Sci.* 2010; 11:5292–5309. [PubMed: 21614208]
16. Fisher CK, Stultz CM. *Curr Opin Struct Biol.* 2011; 21:426–431. [PubMed: 21530234]
17. Kruschel D, Zagrovic B. *Molecular BioSystems.* 2009; 5:1606–1616. [PubMed: 20023721]
18. Schuler B, Eaton WA. *Curr Opin Struct Biol.* 2008; 18:16–26. [PubMed: 18221865]
19. Moritsugu K, Terada T, Kidera A. *J Chem Phys.* 2010; 133:224105. [PubMed: 21171681]
20. Zhang WH, Chen JH. *J Chem Theory Comput.* 2014; 10:918–923. [PubMed: 26580171]
21. Lee KH, Chen J. *Journal of computational chemistry.* 2015
22. Lwin TZ, Luo R. *The Journal of chemical physics.* 2005; 123:194904. [PubMed: 16321110]
23. Liu P, Shi Q, Lyman E, Voth GA. *J Chem Phys.* 2008; 129:114103. [PubMed: 19044946]
24. Lyman E, Zuckerman DM. *J Chem Theory Comput.* 2006; 2:656–666. [PubMed: 26626672]
25. Liwo A, Khalili M, Czaplewski C, Kalinowski S, Oldziej S, Wachucik K, Scheraga HA. *J Phys Chem B.* 2007; 111:260–285. [PubMed: 17201450]
26. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *Journal of computational chemistry.* 1997; 18:849–873.
27. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. *Journal of computational chemistry.* 1997; 18:874–887.
28. Oldziej S, Czaplewski C, Liwo A, Chinchio M, Nancias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kazmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Gibson KD, Scheraga HA. *Proc Natl Acad Sci U S A.* 2005; 102:7547–7552. [PubMed: 15894609]
29. Liwo A, Khalili M, Scheraga HA. *Proc Natl Acad Sci U S A.* 2005; 102:2362–2367. [PubMed: 15677316]
30. Maisuradze GG, Senet P, Czaplewski C, Liwo A, Scheraga HA. *J Phys Chem A.* 2010; 114:4471–4485. [PubMed: 20166738]
31. Smith AV, Hall CK. *Proteins-Structure Function and Genetics.* 2001; 44:344–360.
32. Cheon M, Chang I, Hall CK. *Proteins.* 2010; 78:2950–2960. [PubMed: 20740494]
33. Nguyen HD, Hall CK. *Proc Natl Acad Sci U S A.* 2004; 101:16180–16185. [PubMed: 15534217]
34. Nguyen HD, Hall CK. *Journal of the American Chemical Society.* 2006; 128:1890–1901. [PubMed: 16464090]
35. Bereau T, Deserno M. *J Chem Phys.* 2009; 130.
36. Maupetit J, Tuffery P, Derreumaux P. *Proteins.* 2007; 69:394–408. [PubMed: 17600832]
37. Derreumaux P. *J Chem Phys.* 1999; 111:2301–2310.
38. Chebaro Y, Derreumaux P. *Proteins: Structure, Function, and Bioinformatics.* 2009; 75:442–452.
39. Chebaro Y, Pasquali S, Derreumaux P. *J Phys Chem B.* 2012; 116:8741–8752. [PubMed: 22742737]
40. Kynast P, Derreumaux P, Strodel B. *Bmc Biophys.* 2016; 9.

41. Kar P, Gopal SM, Cheng YM, Predeus A, Feig M. *J Chem Theory Comput.* 2013; 9:3769–3788. [PubMed: 23997693]
42. Monticelli L, Kandasamy SK, Periole X, Larson RG, Tieleman DP, Marrink SJ. *J Chem Theory Comput.* 2008; 4:819–834. [PubMed: 26621095]
43. de Jong DH, Singh G, Bennett WFD, Arnarez C, Wassenaar TA, Schafer LV, Periole X, Tieleman DP, Marrink SJ. *J Chem Theory Comput.* 2013; 9:687–697. [PubMed: 26589065]
44. Han W, Schulten K. *J Chem Theory Comput.* 2012; 8:4413–4424. [PubMed: 23204949]
45. Kmiecik S, Gront D, Kolinski M, Wieteska L, Dawid AE, Kolinski A. *Chemical reviews.* 2016; 116:7898–7936. [PubMed: 27333362]
46. Ingolfsson HI, Lopez CA, Uusitalo JJ, de Jong DH, Gopal SM, Periole X, Marrink SJ. *Wiley interdisciplinary reviews Computational molecular science.* 2014; 4:225–248. [PubMed: 25309628]
47. Marrink SJ, Tieleman DP. *Chemical Society reviews.* 2013; 42:6801–6822. [PubMed: 23708257]
48. Sterpone F, Melchionna S, Tuffery P, Pasquali S, Mousseau N, Cragnolini T, Chebaro Y, St-Pierre JF, Kalimeri M, Barducci A, Laurin Y, Tek A, Baaden M, Nguyen PH, Derreumaux P. *Chemical Society reviews.* 2014; 43:4871–4893. [PubMed: 24759934]
49. Jia ZG, Chen JH. *Journal of computational chemistry.* 2016; 37:1725–1733. [PubMed: 27130454]
50. Zhang W, Ganguly D, Chen J. *Plos Computational Biology.* 2012; 8:e1002353. [PubMed: 22253588]
51. Parigi G, Rezaei-Ghaleh N, Giachetti A, Becker S, Fernandez C, Blackledge M, Griesinger C, Zweckstetter M, Luchinat C. *J Am Chem Soc.* 2014; 136:16201–16209. [PubMed: 25331250]
52. Jia Z, Chen J. *J Comput Chem.* 2016; 37:1725–1733. [PubMed: 27130454]
53. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M. *J Phys Chem B.* 1998; 102:3586–3616. [PubMed: 24889800]
54. MacKerell AD, Feig M, Brooks CL. *J Am Chem Soc.* 2004; 126:698–699. [PubMed: 14733527]
55. MacKerell AD, Feig M, Brooks CL. *Journal of computational chemistry.* 2004; 25:1400–1415. [PubMed: 15185334]
56. Neria E, Fischer S, Karplus M. *J Chem Phys.* 1996; 105:1902–1921.
57. Chen JH, Im WP, Brooks CL. *Journal of the American Chemical Society.* 2006; 128:3728–3736. [PubMed: 16536547]
58. Pierson NA, Chen LX, Russell DH, Clemmer DE. *Journal of the American Chemical Society.* 2013; 135:3186–3192. [PubMed: 23373819]
59. Mackerell AD, Feig M, Brooks CL. *Journal of computational chemistry.* 2004; 25:1400–1415. [PubMed: 15185334]
60. Miyazawa S, Jernigan RL. *J Mol Biol.* 1996; 256:623–644. [PubMed: 8604144]
61. Brooks BR, Brucoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. *Journal of computational chemistry.* 1983; 4:187–217.
62. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, Won Y, Archontis G, Bartels C, Boresch S, Caflisch A, Caves L, Cui Q, Dinner AR, Feig M, Fischer S, Gao J, Hodoseck M, Im W, Kuczera K, Lazaridis T, Ma J, Ovchinnikov V, Paci E, Pastor RW, Post CB, Pu JZ, Schaefer M, Tidor B, Venable RM, Woodcock HL, Wu X, Yang W, York DM, Karplus M. *Journal of computational chemistry.* 2009; 30:1545–1614. [PubMed: 19444816]
63. Ryckaert JP, Ciccotti G, Berendsen HJC. *J Comput Phys.* 1977; 23:327–341.
64. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. *Proteins.* 2010; 78:1950–1958. [PubMed: 20408171]
65. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, de Groot BL, Grubmuller H, MacKerell AD Jr. *Nature methods.* 2017; 14:71–73. [PubMed: 27819658]
66. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E. *SoftwareX.* 2015; 1:19–25.

67. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. *J Chem Phys.* 1983; 79:926–935.
68. Darden T, York D, Pedersen L. *The Journal of chemical physics.* 1993; 98:10089–10092.
69. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM. *Journal of computational chemistry.* 1992; 13:1011–1021.
70. Das RK, Pappu RV. *Proc Natl Acad Sci U S A.* 2013; 110:13392–13397. [PubMed: 23901099]
71. Fesinmeyer RM, Hudson FM, Andersen NH. *Journal of the American Chemical Society.* 2004; 126:7238–7243. [PubMed: 15186161]
72. Radhakrishnan I, PerezAlvarado GC, Parker D, Dyson HJ, Montminy MR, Wright PE. *Cell.* 1997; 91:741–752. [PubMed: 9413984]
73. Feig M, Karanicolas J, Brooks CL. *J Mol Graph Model.* 2004; 22:377–395. [PubMed: 15099834]
74. Demarest SJ, Martinez-Yamout M, Chung J, Chen H, Xu W, Dyson HJ, Evans RM, Wright PE. *Nature.* 2002; 415:549–553. [PubMed: 11823864]
75. Vitalis A, Pappu RV. *Journal of computational chemistry.* 2009; 30:673–699. [PubMed: 18506808]
76. Mao AH, Crick SL, Vitalis A, Chicoine CL, Pappu RV. *Proc Natl Acad Sci U S A.* 2010; 107:8183–8188. [PubMed: 20404210]
77. Das RK, Pappu RV. *Proc Natl Acad Sci U S A.* 2013; 110:13392–13397. [PubMed: 23901099]
78. Das RK, Ruff KM, Pappu RV. *Curr Opin Struct Biol.* 2015; 32:102–112. [PubMed: 25863585]
79. Clementi C, Garcia AE, Onuchic JN. *J Mol Biol.* 2003; 326:933–954. [PubMed: 12581651]
80. Chakrabartty A, Kortemme T, Baldwin RL. *Protein Science.* 1994; 3:843–852. [PubMed: 8061613]
81. Pace CN, Scholtz JM. *Biophys J.* 1998; 75:422–427. [PubMed: 9649402]
82. Drake JA, Pettitt BM. *J Comput Chem.* 2015; 36:1275–1285. [PubMed: 25952623]
83. Chen JH. *J Chem Theory Comput.* 2010; 6:2790–2803. [PubMed: 26616080]
84. Kim YC, Hummer G. *J Mol Biol.* 2008; 375:1416–1433. [PubMed: 18083189]
85. Shalongo W, Dugad L, Stellwagen E. *Journal of the American Chemical Society.* 1994; 116:8288–8293.
86. Best RB. *Curr Opin Struc Biol.* 2012; 22:52–61.
87. Radhakrishnan I, Perez-Alvarado GC, Dyson HJ, Wright PE. *FEBS Lett.* 1998; 430:317–322. [PubMed: 9688563]
88. Ganguly D, Chen J. *J Am Chem Soc.* 2009; 131:5214–5223. [PubMed: 19278259]
89. Xiang S, Gapsys V, Kim HY, Bessonov S, Hsiao HH, Mohlmann S, Klaukien V, Ficner R, Becker S, Urlaub H, Luhrmann R, de Groot B, Zweckstetter M. *Structure.* 2013; 21:2162–2174. [PubMed: 24183573]
90. Rauscher S, Gapsys V, Gajda MJ, Zweckstetter M, de Groot BL, Grubmuller H. *J Chem Theory Comput.* 2015; 11:5513–5524. [PubMed: 26574339]
91. Lee KH, Chen JH. *Journal of computational chemistry.* 2017; 38:1332–1341. [PubMed: 28397268]
92. Demarest SJ, Martinez-Yamout M, Chung J, Chen HW, Xu W, Dyson HJ, Evans RM, Wright PE. *Nature.* 2002; 415:549–553. [PubMed: 11823864]
93. Demarest SJ, Deeckongkit S, Dyson HJ, Evans RM, Wright PE. *Protein Sci.* 2004; 13:203–210. [PubMed: 14691235]
94. Ebert MO, Bae SH, Dyson HJ, Wright PE. *Biochemistry.* 2008; 47:1299–1308. [PubMed: 18177052]
95. Kjaergaard M, Nørholm AB, Hendus-Altenburger R, Pedersen SF, Poulsen FM, Kragelund BB. *Protein Science.* 2010; 19:1555–1564. [PubMed: 20556825]
96. Kjaergaard M, Teilum K, Poulsen FM. *Proc Natl Acad Sci U S A.* 2010; 107:12535–12540. [PubMed: 20616042]
97. Wang T, Zhu Y, Getahun Z, Du D, Huang CY, Degradó WF, Gai F. *J Phys Chem B.* 2004; 108.
98. Gornall JL, Terentjev EM. *Physical review E, Statistical, nonlinear, and soft matter physics.* 2008; 77:031908.
99. Sterpone F, Derreumaux P, Melchionna S. *J Chem Theory Comput.* 2015; 11:1843–1853. [PubMed: 26574390]

100. Chiricotto M, Melchionna S, Derreumaux P, Sterpone F. J Chem Phys. 2016:145.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

HyRes representation of all 20 natural amino acids. The CG side chain beads are shown in VDW representation to proportion, with the atomistic model shown in ball-and-stick to illustrate the mapping scheme (summarized in Table S1). Representative backbone and side chain beads in the HyRes model are labeled in red.

Figure 2. Averaged values and standard deviations of bond lengths (top) and angles (bottom) for all 20 amino acids, derived from 300 ns of GBSW and 100 ns HyRes simulations of dipeptides at 300 K.

Figure 3. Probability distribution of the end-to-end distance, R_g and SASA of Gly₁₀ obtained from HyRes, GBSW and ABSINTH simulations at 300 K.

Figure 4. Backbone ϕ/ψ free energy profiles of alanine dipeptide obtained from 200 ns of MD simulations using (A) HyRes, (B) AMBER ff99SB-ILDN, and (C) CHARMM36m.

Figure 5.

$\varepsilon'_{i,j}$ parameters from original MJ statistical potentials (upper left) and in HyRes model by calculating $\varepsilon'_{i,j}$ from ε'_i using the LB mixing rule (lower right).

Figure 6. Free energy profiles as a function of separation distance between the center of mass of hydrophobic amino acid side chain analogs.

Figure 7. Averaged R_g as a function of charge patterning parameter κ for 30 variants of $(EK)_{25}$ peptide. Values from HyRes simulations are uniformly shifted up by 4 Å to roughly account for larger side chain beads. ABSINTH results are taken from Ref. ⁷⁰.

Figure 8. Residual helicity profiles of (A) (AAQAA)₃ and (B) KID at 300 K calculated using the HyRes model.

Figure 9. Averaged residue helicity profiles and distributions of R_g of ACTR and NCBD calculated using HyRes.

Figure 10.

Reversible helix-coil transition rates for each residue of (AAQAA)₃ at atomistic level. For each replica, we first computed helix/coil transition rate as number of reversible helix/coil transitions divided by simulation time. Results for each simulation shown here are averaged values over all replicas. Note that only the second half of each trajectory was used in these calculations.

Figure 11.

Convergence rate of (AAQAA)₃ residual helicity profiles obtained from MSES/HyRes, MSES/G , and T-REX simulations. At each time point t , we first computed residual helicity profile using trajectory from the beginning to time t . RMSD(t) is calculated between the profile at time t and the averaged one calculated from all six simulations shown in Figure S12.

Table 1T-REX and MSES Simulations of (AAQAA)₃.

Protocol	Run	Simulation time (ns/replica)	Exchange acceptance ratio
T-REX	Control	100	0.61
	Folding	100	0.61
MSES/G	Control	50	0.55
	Folding	50	0.55
MSES/HyRes	Control	50	0.51
	Folding	50	0.51

Table 2

Ensemble averaged end-to-end distance (d_{e2e}), radius of gyration (R_g), and solvent accessible surface area (SASA) of Gly₁₀ simulated using HyRes, GBSW, CHARMM27, CHARMM36, Amber ff12SB and ABSINTH.

Model	$\langle d_{e2e} \rangle$ (Å)	$\langle R_g \rangle$ (Å)	$\langle \text{SASA} \rangle$ (Å ²)
HyRes	13.65	6.15	1000
GBSW	10.65	5.23	899
CHARMM27 [§]	13.30	5.65	880
CHARMM36 [§]	18.14	7.22	1010
Amber ff12SB [§]	14.78	6.31	934
ABSINTH	16.54	6.94	959

[§]Results are from Drake and Pettitt's work⁸².

Table 3

Lennard-Jones parameters for all side chain beads.

Residue	CG bead	ϵ_i (kcal/mol)	$r_i^{\min}/2$ (Å)	$r_i^{\min}/2$ (1-4) (Å)
Ala	CB	-0.308	2.12	2.12
Val	CB	-0.62	2.75	2.75
Leu	CB	-0.9	2.96	2.96
Ile	CB	-0.772	2.97	2.97
Met	CB	-0.636	2.98	3.68
Asn	CB	-0.18	2.65	3.25
Asp	CB	-0.148	2.61	3.11
Gln	CB	-0.204	2.89	3.89
Glu	CB	-0.14	2.85	3.95
Cys	CB	-0.532	2.47	2.77
Ser	CB	-0.188	2.32	2.72
Thr	CB	-0.228	2.62	2.62
Pro	CB	-0.212	2.77	2.77
Lys	CB	-0.05	2.78	3.48
Lys	CC	-0.05	2.36	3.06
Arg	CB	-0.135	2.78	3.18
Arg	CC	-0.135	2.54	2.94
His	CB	-0.108	2.34	2.64
His	CC	-0.081	2.18	2.48
His	CD	-0.081	2.11	2.41
Phe	CB	-0.22	2.64	2.94
Phe	CC	-0.22	2.33	2.63
Phe	CD	-0.22	2.33	2.63
Tyr	CB	-0.197	2.64	2.94
Tyr	CC	-0.197	2.33	2.63
Tyr	CD	-0.0984	2.45	2.75
Trp	CB	-0.168	2.42	2.72
Trp	CC	-0.084	2.24	2.54
Trp	CD	-0.168	2.09	2.39
Trp	CE	-0.168	2.33	2.63
Trp	CF	-0.168	2.33	2.63

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

ε_i (in RT) before rescaling obtained from deconvoluting the MJ matrix.

ε_i	AA(i)	C	M	F	I	L	V	W
		4.63	5.51	7.53	6.65	7.73	5.37	5.31
ε_i								
ε_i	AA(i)	Y	A	G	T	S	N	Q
		4.3	2.75	2.14	2.08	1.74	1.67	1.87
ε_i								
ε_i	AA(i)	D	E	H	R	K	P	
		1.4	1.34	2.68	1.87	1	1.94	
ε_i								

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 5

Residue pairs with the deviation $\Delta \varepsilon'_{i,j} = \varepsilon'_{i,j}(\text{MJ}) - \varepsilon'_{i,j}(\text{LB})$ greater than 0.5 RT.

AA _i	C	F	L	W	Y	D
AA _j	C	K	K	P	K	R
$\Delta \varepsilon'_{i,j}$	0.81	0.62	0.59	0.52	0.53	0.67
AA _i	D	E	E	R	K	
AA _j	K	R	K	K	K	
$\Delta \varepsilon'_{i,j}$	0.5	0.69	0.64	-0.78	-0.88	