Supporting Information

"The Rosetta all-atom energy function for macromolecular modeling and design"

Rebecca F. Alford, Andrew Leaver-Fay, Jeliazko R. Jeliazkov, Matthew J. O'Meara, Frank P. DiMaio, Hahnbeom Park, Maxim V. Shapovalov, P. Douglas Renfrew, Vikram K. Mulligan, Kalli Kappel, Jason W. Labonte, Michael S. Pacella, Richard Bonneau, Phillip Bradley, Roland L. Dunbrack Jr., Rhiju Das, David Baker, Brian Kuhlman, Tanja Kortemme, Jeffrey J. Gray

Major changes to the Rosetta energy function since 2000

The all-atom Rosetta energy function for proteins has undergone significant upgrades since the original implementation in 2000. These changes range from improved atomic parameters and models of hydrogen bonding to smoothing routines that eliminate errors during minimization. An overview of these advances is listed in **Table S1**.

Table S1: Major changes to the Rosetta Energy Function since 2000

Energy Term	Adjustment	Ref.
Lennard-Jones	Soften repulsive potential Atomic radii matched to crystal structures Shifted LJ Potential Extra soft repulsive potential Make derivatives continuous New well-depth parameters Incorporation of hydrogens in the `fa_atr` calculation	Kuhlman *et al.* 2000[1] Kuhlman *et al.* 2003[2] Tsai *et al.* 2003[3] Meiler & Baker, 2006[4] Scheffler 2006, Unpublished Park et al. 2016[5]
Solvation	Implementation of Lazaridis-Karplus Model Make derivatives continuous Anisotropic Solvation Model New atomic volume λ and ΔG^{free} parameters	Kuhlman *et al.* 2000[6] Sheffler 2006, Unpublished Yanover *et al.* 2011[7] Park *et al.* 2016[5]
Electrostatics	Knowledge-based Pair term Coulomb electrostatics for ligand interactions Coulomb electrostatics for nucleic acids Coulomb electrostatics for proteins Sigmoid dielectric model Avoidance of dipole splitting for local interactions New partial charges	Kuhlman *et al.* 2000[6] Meiler & Baker, 2006[4] Yanover *et al.* 2011[7] O'Meara *et al.* 2015[8] Park *et al.* 2016[5]
Hydrogen Bonding	Orientation-dependent hydrogen bond potential Favoring H-bonds in the sp^2 plane No H-bond environment dependence Weights on hydrogen bond donors and acceptors	Kortemme *et al.* 2003[9] O'Meara *et al.* 2015[8] Park *et al.* 2016[5]
Dunbrack Rotamers	Add 2002 backbone-dependent rotamer library Replace 2002 version with the 2010 smoothed rotamer library	Dunbrack *et al.* 2002[10] Shapovalov *et al.* 2011[11] Leaver-Fay *et al.* 2013[12]
Ramachandran & p_aa_pp	Interpolation with bicubic splines Correction for pre-proline backbone torsions	Leaver-Fay *et al.* 2013[12] Park *et al.* 2016[5]
Side-chain specific	Penalty for Tyr hydroxyl hydrogen leaving aromatic plane	O'Meara *et al.* 2015[8]
Design reference energy	Refit reference energies with OptE Refit reference energies with DualOptE	Leaver-Fay *et al.* 2011[12] O'Meara *et al.* 2015[8] Park *et al.* 2016[5]

Table S2: Rosetta revisions corresponding to major energy function updates

Version	Rosetta Revision[a]	Public Version[b]
Score12	pre-#55611	Pre-Rosetta 3.5
Talaris2013	#55611	Rosetta 3.5
Talaris2014	#58602	v2016.13-dev58602
REF2015	#59248	v2017.05-dev59248

[a] Internal code revision number available to member institutions of the Rosetta Commons.
[b] Download the public Rosetta release from http://www.rosettacommons.org.

Data for calibrating Rosetta energies to kcal/mol

The parameters and weights in *REF2015* were recently fit[5] such that Rosetta simulations reproduce high-resolution protein structural data and thermodynamic data for small molecules from Jorgensen *et al.*[13] Thus, the Rosetta energy is now expressed in kcal/mol. In support, **Figure S1** compares experimental data and Rosetta predictions of density, heat of vaporization (ΔH_{vap}) and heat capacity ($C_{p(l)}$) for seventeen molecules: ethane, propane, isobutene, cyclohexane, benzene, toluene, phenol, methanol, ethanol, 2-propanol, tert-Butyl alcohol, methane thiol, ethane thiol, dimethyl sulfide, acetamide, *N*-methylamide, *N*-methylformamide, dimethyl ether, ethanol and propanone.

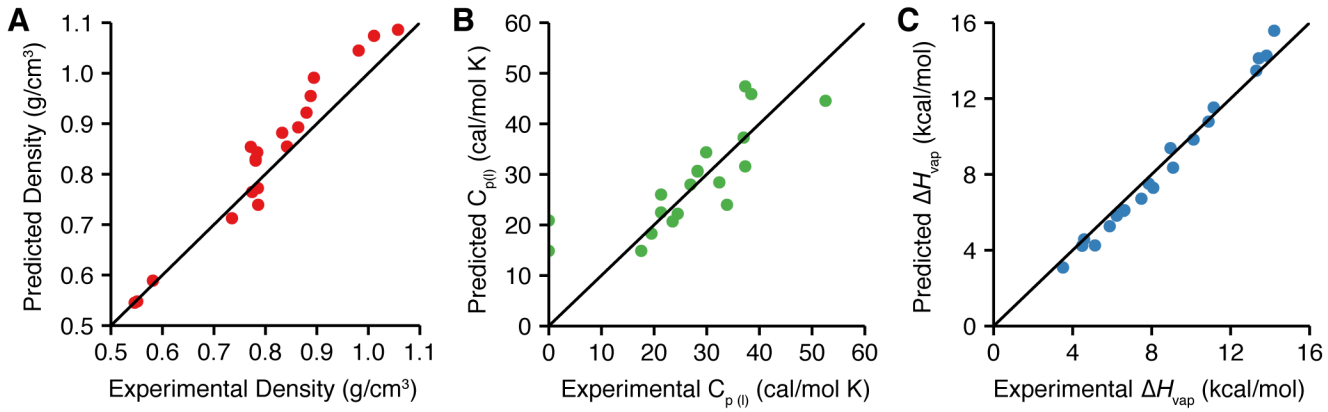

Figure S1: Comparison of Rosetta simulations with experimental thermodynamic data
Comparison between Rosetta predictions and experimental thermodynamic measurements for seventeen small molecules (A) Density (B) Heat Capacity and (C) Heat of vaporization.

Additional energy function details

Parameters for the Lennard-Jones and Lazaridis-Karplus energies

New experiments and numerical methods to optimize the energy function have led to updated atomic-parameters used by the Lennard-Jones[14,15] and Lazaridis-Karplus[16] potentials. The updated parameters are in the following Rosetta database files:

Table S3: Location of Rosetta atom type parameters

Parameter	Database File
Radii, ΔG^{free}, ϵ	`chemical/atom_type_sets/fa_standard/atom_properties.txt`
Partial charges	`chemical/residue_type_sets/residue_types/l-caa/*.params`
`lk_ball` weights	`chemical/atom_type_sets/fa_standard/extras/lk_ball_wtd.txt`
`hbond` parameters	`scoring/score_functions/hbonds/*`
`rama`	`scoring/score_functions/rama/*`
`p_aa_pp`	`scoring/score_functions/P_AA_pp/*`
`omega`	`scoring/score_functions/omega/*`
`fa_dun`	`rotamer/*`

Tables S4-S6 present a comparison between selected atomic parameters between the original source publication and the values in Rosetta energy functions, *Talaris2014* and *REF2015*.

Table S4: Atomic radii values from the *Neria* et al. force field, *Talaris2014*, and *REF2015*

Atom Type	Neria *et al.*[17] Radius (Å)	*Talaris2014* Radius (Å)	*REF2015*[5] Radius (Å)
CAbb	2.3650	2.0000	2.0112
CH1	2.3650	2.0000	2.0112
CH2	2.2350	2.0000	2.0112
CH3	2.1650	2.0000	2.0112
CNH2	--	2.0000	1.9922
COO	2.1000	2.0000	1.9649
CObb	--	2.0000	1.9649
aroC	2.1000	2.0000	1.9859
NH2O	--	1.7500	1.7632
Narg	1.6000	1.7500	1.7632
Nbb	1.6000	1.7500	1.7632
Nhis	1.6000	1.7500	1.7632
Nlys	1.6000	1.7500	1.7632
Npro	1.6000	1.7500	1.7632
Ntrp	1.6000	1.7500	1.7632
OCbb	1.6000	1.5500	1.5268
OH	1.6000	1.5500	1.5354
ONH2	--	1.5500	1.5760
OOC	--	1.5500	1.4492
S	0.0430	1.9000	2.0171
HNbb	--	1.0000	0.8773
Hapo	--	1.2000	1.4634
Haro	--	1.2000	1.3778
Hpol	0.8000	1.0000	0.8773

Table S5: Well-depth parameters from the *Neria* et al. force field, *Talaris2014*, and *REF2015*

Atom Type	Neria *et al.*[17] ϵ (kcal/mol)	*Talaris2014* ϵ (kcal/mol)	*REF2015*[5] ϵ (kcal/mol)
CAbb	0.0486	0.0486	0.0626
CH1	0.0486	0.0486	0.0626
CH2	0.1142	0.1142	0.0626
CH3	0.1811	0.1811	0.0626
CNH2	--	0.1200	0.0626
COO	0.1200	0.1200	0.0946
CObb	--	0.1400	0.1418
aroC	0.1200	0.1200	0.1418
NH2O	--	0.2834	0.0688
Narg	0.2384	0.2834	0.1617
Nbb	0.2384	0.2834	0.1617
Nhis	0.2384	0.2834	0.1617
Nlys	0.2384	0.2834	0.1617
Npro	0.2384	0.2834	0.1617
Ntrp	0.2384	0.2834	0.1617
OCbb		0.1591	0.1617
OH	0.1591	0.1591	0.1617
ONH2	--	0.1591	0.1424
OOC	--	0.2100	0.1619
S	0.0430	0.1600	0.1829
SH1	--	--	0.0999
HNbb	--	0.0500	0.4560
HS	--	--	0.4560
Hapo	--	0.0500	0.0050
Haro	--	0.0500	0.0508
Hpol	--	0.0500	0.0218

Table S6: ΔG^{free} parameters from Lazaridis & Karplus, *Talaris2014* and *REF2015*

Atom Type	Lazarids-Karplus[16] ΔG^{free} (kcal/mol)	*Talaris2014*, ΔG^{free} (kcal/mol)	*REF2015*[5] ΔG^{free} (kcal/mol)
CAbb	-0.2500	1.0000	2.5338
CH0	-0.2500	-0.2500	1.4093
CH1	-0.2500	-0.2500	-3.5384
CH2	0.5200	0.5200	-1.8547
CH3	1.5000	1.5000	7.2929
CNH2	--	0.0000	3.0770
COO	0.1200	-1.4000	-3.3326
CObb	--	1.0000	3.1042
aroC	0.8000	0.0800	1.7979
NH2O	--	-7.8000	-8.1016
Narg	-10.0000	-10.0000	-8.9684
Nbb	-7.8000	-5.0000	-9.9695
Nhis	-4.0000	-4.0000	-9.7396
Nlys	-20.000	-20.000	-20.865
Npro	-1.5500	-1.5500	-0.9846
Ntrp	-8.9000	-8.9000	-8.4131
OCbb	-10.0000	-5.0000	-8.0068
OH	-6.7000	-6.7000	-8.1335
ONH2	--	-5.8500	-6.5916
OOC	--	-10.0000	-9.2398
S	-4.1000	-4.1000	-1.7072
SH1	-2.7000	--	3.2916

Analytical form of the hydrogen bonding potential

To avoid expensive table lookups, the hydrogen bonding potential (**Eq. 21-22** in the main text) is given by component energies with simple analytical forms. For completeness, we detail these analytical forms below. The first two components, $E_{\text{hbond}}^{HA}(d_{HA})$ and $E_{\text{hbond}}^{HAD}(\theta_{HAD})$ are polynomial functions, $f_{\text{poly}}(P,x)$ where the polynomial P depends on the atom type of the acceptor and donor, and the order n varies from 6 to 10 (**Eq. S1**). The forms of $E_{\text{hbond}}^{HA}(d_{HA})$ and $E_{\text{hbond}}^{HAD}(\theta_{HAD})$ are given by **Eq. S2-3**.

$$f_{\text{poly}}(P,x) = C_0 + C_1 x + C_2 x^2 + \cdots + C_{n-1}x^{n-1} + C_n x^n \qquad \text{(S1)}$$

$$E_{\text{hbond}}^{HA}(d_{HA}) = F_{HA} \cdot f_{\text{poly}}(P, d_{HA}) \qquad \text{(S2)}$$

$$E_{\text{hbond}}^{HAD}(\theta_{HAD}) = F_{HAD} \cdot f_{\text{poly}}(P(\theta_{HAD})) \qquad \text{(S3)}$$

The third component, $E_{\text{hbond}}^{B_2BAH}(\rho, \phi_{B_2BAH}, \theta_{BAH})$, is dependent on the hybridization of the acceptor, ρ. For sp^2 hybridized acceptors, the potential is given as a combination of cosine and polynomial functions (**Eq. S4-5**) controlled by a cosine switch function (**Eq. S6**). The functional forms are also shown in **Fig. S2**.

$$F(\phi) = \begin{cases} \frac{d}{2}\cos(3(\pi-\phi)) + \frac{d-1}{2} & \frac{2\pi}{3} < \phi \\ \frac{m}{2}\cos\left(\frac{1}{l}\left(\pi - \frac{2\pi}{3}\phi\right)\right) + \frac{m-1}{2} & \frac{2\pi}{3} - l \le \phi \le \frac{2\pi}{3} \\ m - \frac{1}{2} & \phi < \frac{2\pi}{3} - l \end{cases} \qquad \text{(S4)}$$

$$G(\phi) = \begin{cases} d - \frac{1}{2} & \frac{2\pi}{3} < \phi \\ \frac{m-d}{2}\cos\left(\pi - \frac{1}{l}\left(\pi - \frac{2\pi}{3}\ \phi\right)\right) + \frac{m+d+1}{2} & \frac{2\pi}{3} - l \le \phi \le \frac{2\pi}{3} \\ m - \frac{1}{2} & \phi < \frac{2\pi}{3} - l \end{cases} \qquad \text{(S5)}$$

$$H(\phi) = \frac{\cos(2\phi)+1}{2} \qquad \text{(S6)}$$

For sp^3 hybridized acceptors, the potential is modeled as a composition of sine and cosine functions. If the acceptor is attached to a ring, the potential is modeled with a simple cosine function. The overall energy is given in **Eq. S7**.

$$E_{\text{hbond}}^{B_2BAH}(\rho, \phi_{B_2BAH}, \theta_{BAH}) = \begin{cases} H(\phi_{B_2BAH})F(\phi_{B_2BAH}) + \left(1 - H(\phi_{B_2BAH})G(\phi_{B_2BAH})\right) & \rho \sim \text{sp}^2 \\ f_{\text{poly}}(\cos(\theta_{BAH})) + \frac{1}{4}\left(1 + \cos(\phi_{B_2BAH})\right) & \rho \sim \text{sp}^3 \\ f_{\text{poly}}(\cos(\theta_{BAH})) & \rho \sim \text{ring} \end{cases} \qquad \text{(S7)}$$

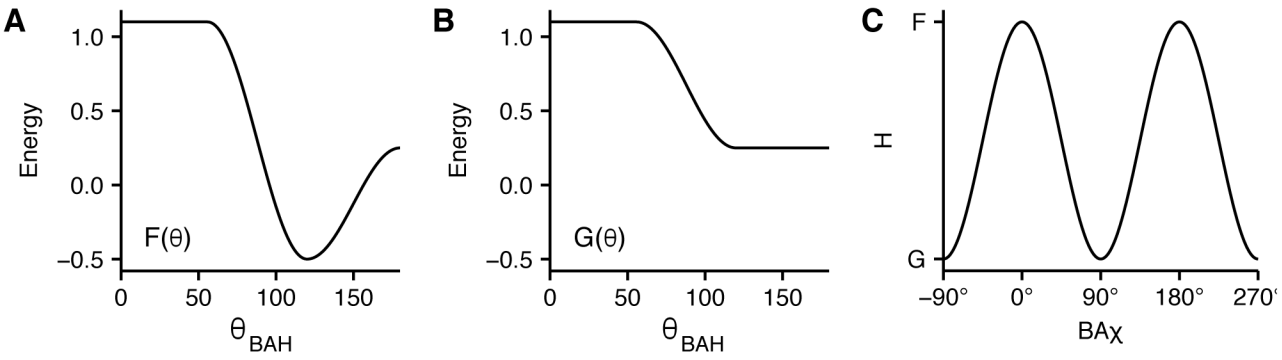

Figure S2: Analytic form of the hydrogen bonding BAχ potential for *sp2* hybridized acceptors
(A) Plot of the function $F(\theta)$ that models the energy of the BAH angle for an in-plane acceptor (B) Plot of the function $G(\theta)$ that models the energy of the BAH angle for an out-of-plane acceptor. (C) Switch function $H(\theta)$ that controls contributions from $F(\theta)$ and $G(\theta)$ at a specified value of the BAχ torsion, ϕ.

Analytical form of the disulfide bonding potential

Like the hydrogen bonding potential, the component disulfide bonding energies are defined by analytical forms. As defined by **Eq. 23** in the main text, the disulfide is computed given four component energies. First, the sulfur-sulfur distance energy $E_{\text{dslf}}^{SS}(d_{SS})$ is defined by **Eq. S8** given the sulfur-sulfur distance, d, mean distance $\overline{d_{ss}}$, standard deviation σ_{ss}, and fitting parameters α_d^{dslf}, ϵ_m and w_{SS}.

$$E_{\text{dslf}}^{SS}(d_{SS}) = w_{SS}\left(\left(\frac{d-\overline{d_{ss}}}{\sigma_{ss}}\right)^2 + \ln\left[\text{erf}\left(\alpha_d^{\text{dslf}}\left(\frac{d-\overline{d_{ss}}}{\sigma_{\text{dslf}}^{SS}}\right)\right) + \epsilon_m\right]\right) \qquad \text{(S8)}$$

Next, the energy of the angle formed by a C_β and two sulfur atoms $E_{\text{dslf}}^{CSS}(\theta_{CSS})$ is defined by **Eq. S9** given the angle θ and von Mises parameters A_{CSS}, w_{CSS}, κ_{CSS}, and μ_{CSS}.

$$E_{\text{dslf}}^{C_\beta SS}(\theta_{CSS}) = w_{CSS}\left(-\ln(A_{CSS}) - \kappa_{CSS}\cos(\theta - \mu_{CSS})\right) \qquad \text{(S9)}$$

The energy of the torsion formed by $C_{\beta 1}$, $C_{\beta 2}$ and the two sulfur atoms $E_{\text{dslf}}^{C_\beta SSC_\beta}(\phi_{CSSC})$ is defined by **Eq. S10** given the torsion angle ϕ and the von Mises parameters $A_{k,C_\beta SSC_\beta}$, $\kappa_{k,C_\beta SSC_\beta}$, $\mu_{k,C_\beta SSC_\beta}$ and ϵ_{m}.

$$E_{\text{dslf}}^{C_\beta SSC_\beta}(\phi_{CSSC}) = w_{C_\beta SSC_\beta}\ln\left(\sum_{k\leq 2}\exp\left(A_{k,C_\beta SSC_\beta} + \kappa_{k,C_\beta SSC_\beta}\cos\left(\phi - \mu_{k,C_\beta SSC_\beta}\right)\right) + \ \epsilon_{\text{m}}\right) \quad \text{(S10)}$$

Finally, the energy of the torsion formed by C_α, C_β and the two adjacent sulfur atoms $E_{\text{dslf}}^{C_\alpha C_\beta S,S}(\theta_{CCSS})$ is defined by **Eq. S11** given the torsion angle ϕ and the von Mises fitting parameters $A_{k,C_\alpha C_\beta SS}$, $\kappa_{k,C_\alpha C_\beta SS}$, $\mu_{k,C_\alpha C_\beta SS}$ and ϵ_{m}.

$$E_{\text{dslf}}^{C_\alpha C_\beta S,S}(\theta_{CCSS}) = -w_{C_\alpha C_\beta SS}\ln\left(\sum_{i\leq 3}\exp\left(A_{k,C_\alpha C_\beta SS} + \kappa_{k,C_\alpha C_\beta SS}\cos\left(\theta - \mu_{k,C_\alpha C_\beta SS}\right)\right) + \ \epsilon_{\text{m}}\right) \quad \text{(S11)}$$

Statistical potentials: Interpolation of energies rather than probabilities

The Rosetta energy function uses probabilities from the Dunbrack backbone-dependent rotamer library[18] to derive torsional energies E using the inverted Boltzmann relation the probability P (**Eq. S12**):

$$E = -kT \ln P \quad \text{(S12)}$$

Prior to 2012, the probabilities for the ϕ,ψ-dependent terms were stored on a 10° x 10° grid used for energy calculations. These probabilities were calculated using bilinear interpolation and then converted to energies using **Eq. S12** and the derivatives were calculated by linearly interpolating $1/P$ and dP/dx to compute $d(-\log P)/dx = -(1/P)\, dP/dx$ with $x = \phi$ or ψ. This method resulted in large inaccuracies because P can vary by orders of magnitude over very short ranges of ϕ and ψ. In addition, the linearly interpolated derivatives are constant between grid points, so that gradient-based minimization results in moving structures to the nearest grid point where the derivative changes sign. Therefore, it is more accurate to provide P and $E = -\ln P$ at each grid point and then interpolate the energies using bicubic interpolation.

Here we demonstrate why interpolating energies is better than interpolating the probabilities. **Figure S3** compares the different interpolation strategies for a toy problem: a one-dimensional probability distribution with a discrete rotamer modeled with the following von Mises function (**Eq. S13**). Here, the location constant $\mu = 180^{\circ}$, the concentration constant $\kappa = 20$, $x = \phi$ or ψ and $I_0(\kappa)$ is the modified Bessel function of order zero needed to normalize the distribution.

$$P(x) = \frac{\exp(\kappa - \cos(x-\mu))}{2\pi I_o(\kappa)} \quad \text{(S13)}$$

First, **Figure S3A** shows the probability distribution, P and its linear interpolation based on the 10° x 10° grid. Here, the difference between the curves demonstrate the effect of approximating P by linear interpolation. This effect would be more severe for steeper functions such as the Ramachandran probability density function. **Figure S3B** compares the $E = -\ln P$ calculation with two approaches to interpolating the function: interpolate P and then compute the energies versus compute the energies at the grid points and then interpolate. The second scenario clearly mitigates several errors which can be further improved using cubic rather than linear interpolation. Like the first panel, the benefits of cubic interpolation are clearer with steeper functions.

Figures S3C and **S3D** demonstrate that the effects of interpolating energies are more pronounced for the derivatives of P and E respectively. Previously, Rosetta computed the derivative of P as $dP/dx = [P(x+10) - P(x)]/10$ (**Fig. S3C**). The linear interpolation of this derivative includes noticeable artifacts. **Figure S3D** shows the four energy derivative curves: (1) the exact analytical expression $dE/dx = -(1/P)\, dP/dx$ where P is interpolated and dP/dx is the step function shown in **Fig. S3C** (2) $dE/dx = -(1/P)\, dP/dx$ where P and dP/dx are interpolated from the grid values, (3) $dE/dx = -(1/P)\, dP/dx$ where both P and dP/dx are interpolated from the grid values and (4) calculation of E and dE/dx on the grid followed by interpolation of dE/dx in between the grid points. The linear interpolation of dE/dx provides the closest match to the analytical expression. The current Rosetta energy function interpolates energies rather than probabilities: both P and E are stored in database files, dE/dx is calculated on the grid points, and then P, E and dE/dx are computed by bicubic spline interpolation.

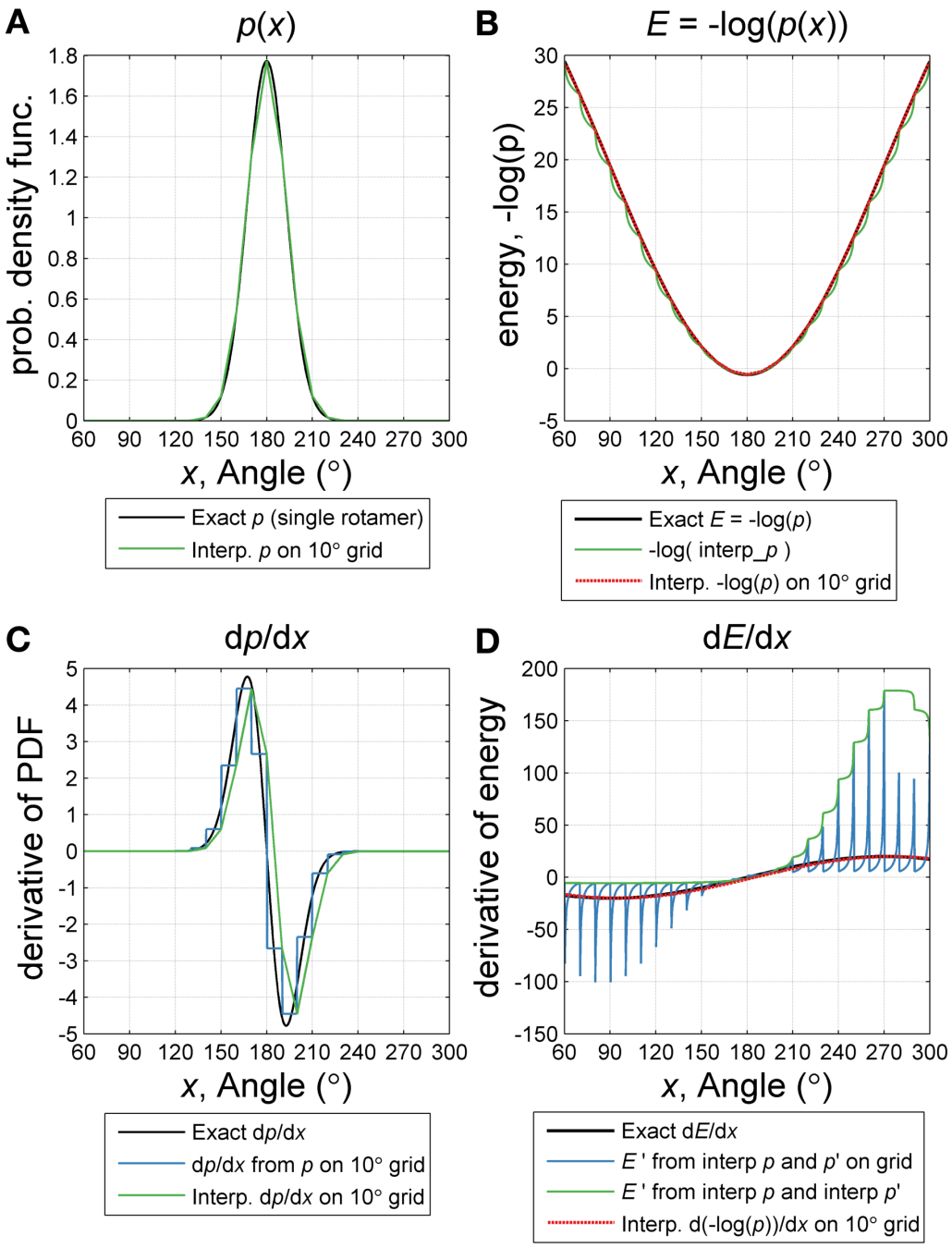

Figure S3. Approximating the energy and energy derivatives for torsional potentials
Comparison between the old and new approach of approximating the energy and energy derivatives using a toy example in one dimension. (A) Exact analytical expression of the probability distribution P(X) (black) compared to approximation of the grid (green). (B) Exact energy expression, -log p(x) (black) compared to interpolated probabilities (green) and interpolation on the grid (red). (C) Probability first-order derivatives: analytical expression (black), derivative approximation with no interpolation (blue), and derivative with linear interpolation (green). (D) Energy derivatives: exact (black), calculation as a step function (blue), calculation by linear interpolation (green), calculation from grid values (red).

Methods for energy-based analysis examples

ΔΔG of Mutation. The coordinate file for 1kgj was downloaded from the Protein Data Bank[19] and cleaned to remove any non-canonical amino acids. The PDB was refined with fast relax constrained to native coordinates using Cartesian-space refinement and the *REF2015* energy function using the following command line:

```
relax.linuxgccrelease –s 1kgj.pdb –use_input_sc \
–constrain_relax_to_start_coords –ignore_unrecognized_res –nstruct 1000 \
–relax:coord_constrain_to_sidechains –relax:ram_constraints false \
–relax:Cartesian –relax:min_type lbfgs_armijo_nonmonotone
```

After refinement, the lowest scoring model was used to generate five structures of the native conformation and five structures of the T193V mutated conformation using a Cartesian version of Rosetta's ddg protocol.[20]

```
cartesian_ddg.linuxgccrelease –s 1kgj_refined_lowest.pdb –ddg:mut_file \
$MUT_FILE –ddg:iterations 5 –optimization:default-max_cycles 200 –bbnbr 1 \
–relax:min_type lbfgs_armijo_nonmonotone –fa_max_dis 9.0
```

The energies were averaged for each ensemble of five structures. The $\Delta\Delta G$ was then calculated as the difference between the average energy of the mutated ensemble and the average energy of the native ensemble.

To determine which specific interactions underlie the observed differences in solvation, we first needed to identify which residue-pair interactions contribute most to the change in solvation energy. Because the mutation is taking place at residue 193, we can safely restrict our search to residue-pair interactions involving residue 193. Now, we use the PyRosetta[21] tool `print_residue_pair_energies()` to obtain a list of all residue pair interactions involving residue 193. Inspecting the output in `native_residue_pair_interactions.csv` and `mutant_residue_pair_interactions.csv` we can find a list of significant pair energy changes between residue 193 and other surrounding residues.

PyRosetta tools can also be used to analyze atom-pair interactions that contribute most strongly to the critical residue-pair interactions. The scoring machinery in Rosetta treats a residue (protein amino acid, sugar monosaccharide, nucleic acid base) as the simplest unit for calculating pairwise energies. All two body energy terms must define `residue_pair_energy()` to calculate the pairwise energy between two residues. For energies such as hydrogen bonding this is necessary because scoring an individual hydrogen bond using the distance and orientation dependent potential described in the main text requires knowledge of not only the donor hydrogen and the acceptor atoms but also the acceptor and donor base atoms to calculate an energy. However, for other terms in the Rosetta score function (such as Lennard Jones attraction/repulsion, implicit solvation, and electrostatics) the `residue_pair_energy()` method simply sums up all of the pairwise interactions between all atoms in each of the residues. These atom pair energies are not normally reported by the scoring function, however in some situations they can assist in pinpointing which specific atom pair interactions are influencing the residue pair energy most strongly.

The PyRosetta toolkit provides two tools for analyzing specific atom pair energies. First, the `etable_atom_pair_energies()` method takes two residues (`res1`, `res2`) and atom indices specifying one atom on each residue (`atom_index_1`, `atom_index_2`) and calculates atom pair

energies for Lennard Jones attractive/repulsive, implicit solvation, and electrostatics using a specified score function (`sfxn`).

The second tool, `print_atom_pair_energy_table()`, is designed to output energies for all pairwise atom pair interactions between two specified residues. For ease of viewing this tool outputs the pairwise energies as a table formatted in a .csv file. The tool takes a `score_type` and score function (`sfxn`) as inputs in addition to two specified residues (`res1`, `res2`) and a specified `output_filename`.

Docking. The coordinate file for 1ztx was downloaded from the Protein Data Bank and cleaned to remove any non-canonical amino acids. The structure was first refined to remove significant clashes in the structure using the following command line:

```
relax.linuxgccrelease -s 1ztx_unbound.pdb -relax:ramp_constriants false \
-relax:constrain_relax_to_start_coords -ex1 -ex2 -use_input_sc -flip_HNQ \
-no_optH false
```

Next, the structure was prepacked and then docked using the procedure described in Chaudhury *et al.*[22] using the *REF2015* energy function.

```
docking_prepack_protocollinux -s 1ztx_relaxed.pdb -partners LH_G \
-dock_rtmin -docking:sc_min

docking_protocol.linuxgccrelease —s 1ztx_unbound_prepacked.pdb —native \
1ztx_native.pdb -ignore_unrecognized_res -ex1 -ex2aro -dock_pert 3 8 \
-partners LH_G -nstruct 1000
```

Finally, the interface scores were extracted from the output score file for analysis.

Energy terms for biomolecules other than proteins

An active research area is the development of energy functions compatible for biomolecules other than proteins containing the 20 canonical amino acids. So far, this has involved two approaches: (1) generalizing terms to score non-l amino acids and (2) developing new terms to accommodate other biomolecules. Below, we provide details of the main non-protein energy functions currently being developed in Rosetta.

Generalizing the Existing Energy Terms

The physically-derived terms in the Rosetta energy function capture forces that are general to all biomolecules. Therefore, these terms were generalized to be compatible will D-amino acids, nucleic acids, carbohydrates, and other biomolecules.

Table S7: Summary of energy term compatibility with other biomolecules

Term	Can score
`fa_atr`	All molecules
`fa_rep`	All molecules
`fa_intra_rep`	All molecules
`fa_sol`	All molecules
`lk_ball`	All molecules
`fa_inra_sol`	All molecules
`fa_elec`	All molecules
`hbond_sr_bb`	All molecules
`hbond_lr_bb`	All molecules
`hbond_bb_sc`	All molecules
`hbond_sc`	All molecules
`dslf_fa13`	L-, D-, and mixed D/L disulfide bonds between cysteine or cysteine-like residues (e.g., homocysteine, penacillamine)
`rama_prepro`	Glycine, canonical L-amino acids, their D-counterparts, and similar alpha-amino acids that can use canonical rama tables.
`p_aa_pp`	Glycine, canonical L-amino acids, their D-counterparts, and similar alpha-amino acids that can use canonical rama tables.
`omega`	All α-amino acids, or β-amino acids.
`fa_dun`	All polymer building blocks.
`pro_close`	L- and D-proline.
`yhh_planarity`	L- and D-tyrosine.
`ref`	Glycine, canonical L-amino acids, and their D-counterparts.

Compatibility with D-amino acids

To make the energy terms compatible with D-amino acids, several modifications were made to the torsional terms.[23] First, the ϕ, ψ values were inverted in the `rama_prepro`, `omega`, and `p_aa_p` terms to accommodate the chirality of the backbone. Accordingly, the derivatives were inverted to ensure that mirror-image structures energy-minimize identically. Second, the `fa_dun` score term was modified to

invert main chain and side-chain torsional values. Special amino acid-specific score terms, such as `pro_close` and `yhh_planarity`, were updated to recognize D-proline and D-tyrosine, respectively. The `dslf_fa13` term was symmetrized to ensure that mirror-image conformations of mixed D/L disulfides score identically. Finally, the `ref` term was altered to ensure that D-amino acids have a reference energy penalty or bonus identical to that of their L-counterparts. All other score terms were compatible with arbitrary molecules without modification.

Energy terms for non-canonical amino acids

Toward the goal of designing proteins with non-canonical amino acids, Renfrew *et al.* implemented an energy function with terms derived from molecular mechanics. This energy function, called `mm_std`, removes the terms that depend on residue identity (i.e. `rama_prepro`, `p_aa_pp`, `omega`, and `fa_dun`) and replaces them with terms that capture the internal and torsional energy preferences: `mm_lj_intra_rep`, `mm_lj_intra_atr`, and `mm_twist`. The `ref` term is replaced by either a term that explicitly models the unfolded state, (`unfolded`), or a pair of terms that capture the change in energy experienced by an atom of a specific type going from an unfolded to folded environment (`split_unfolded_1b` and `split_unfolded_2b`). These terms were developed toward the goal of designing proteins containing non-canonical alpha-amino acid residues. It has also been used to model oligo-oxypiperizines (OOPs),[24] hydrogen bond surrogates (HBS), oligo-peptoids,[25] and hybrid molecules.

Intra-residue van der Waals interactions are calculated between atom pairs from the same residue using a Lennard-Jones 6-12 potential. Like `fa_rep` and `fa_atr`, the potential is divided between attractive (`mm_lj_intra_atr`) and repulsive (`mm_lj_intra_rep`) components that can be weighted separately. The two terms have the same functional form as the inter-reside terms (Eq. 3 and 4 in the main text) but with the following differences. The summed atomic radii, σ_{ij}, and the geometric mean of atomic well-depths, ϵ_{ij}, are based on the CHARMM 24[26] parameters. The terms are applied to all atom pairs in a residue with a bond separation of 3 or more. Some atom-type pairs have different parameters when separated by 3 bonds (and involved in a proper torsion) and when separated by 4 of more bonds, but no connectivity weight is applied. Both attractive and repulsive energies are calculated for hydrogens. The attractive potential is not smoothed and consequently is evaluated to 8 Å.

The **torsional term**, called `mm_twist` (**Eq. S14**), is a molecular mechanics torsion term. It is evaluated for all atom quads involved in proper torsions. To match the intra-residue van der Waals term the parameters for K_θ and n come from CHARMM 24. A given set of 4 atoms types may have multiple K_θ and n parameters that are summed in a Fourier series to more accurately describe the rotation about the central bond of the torsion.

$$E_{\text{twist}} = \sum_i K_\theta(1 - \cos(n\theta)) \quad \text{(S14)}$$

Explicit Unfolded State Energy (EUSE) represents the unfolded energy of the protein and compensates for the difficultly in packing large side chains (**Eq. S15**). The `ref` term is fit during the weight optimization protocol which is only trained on protein data and therefore incompatible with non-protein residues. The EUSE is the sum over each residue and each term in the energy function where $U(\mathrm{AA}_r, t)$ is the unfolded reference value of residue type, AA_r, of residue, r, and energy term t. The unfolded reference values are the Boltzmann weighted average energies of the central residue of 5-mer fragments of high quality protein structures. The central residue of each fragment was mutated to the residue of choice, repacked and scored and the Boltzmann weighted average for each energy term, t, for each residue type is stored. For peptoids, only XXGPX fragments were used to mimic an oligo-peptoid environment.[27]

$$E_{\text{unfolded}} = \sum_r \sum_t W_t U(AA_r, t) \quad \text{(S15)}$$

Two-Component Reference Energy (TCRE) is a reference energy that compensates for some of the shortcomings of the EUSE; primarily the dependence of the EUSE on short peptide fragments which limits the types of oligomer chemistry to those that contain an α-amino acid backbone (e.g. OOPs, HBS, peptoids; **Eq. S16**). The one-body component is the sum over each residue and each one-body energy term in the energy function where $R_{1\text{B}}(\text{AA}_r, t1\text{b})$ is the one-body reference value of the residue type, AA_r, of residue, r, and one-body scoring term $t1b$. The one-body reference values are the unweighted $t1b$ energies for each energy term, taken from lowest energy conformation of that residue type in the context of a didpeptide model system. The two-body component is the sum over each atom and each two-body energy term in the energy function where $R_{2\text{B}}(T_i, t2\text{b})$ is the two-body reference value for atom type, T_i, of atom, i, and two-body energy term $t2b$. The two-body reference values are the median $t2b$ energy of an atom of type T_i in the context of a folded protein.

$$E_{\text{TCRE}} = \overbrace{\left(\sum_r \sum_{t1b} W_{t1\text{b}} R_{1\text{B}}(\text{AA}_r, t1\text{b})\right)}^{\text{one-body}} + \overbrace{\left(\sum_i \sum_{t2b} W_{t2\text{b}} R_{2\text{B}}(T_i, t2\text{b})\right)}^{\text{two-body}} \quad \text{(S16)}$$

Reference values were determined using structures from the Top8000 database.[28] The effect is to produce a single reference value for a residue type just like the `ref` and `unfolded` terms. The term is a measure of the difference between the base energy of inherent to a peptide sequence and the average interaction that sequence would make with itself when folded. Currently W_{t1b} and W_{t1b} are set to the weight of that term in the energy function but could be modified.

Energy terms for carbohydrates

To model realistic carbohydrate geometries, Rosetta implements the `sugar_bb` term which rewards preferred glycosidic torsion angles.[29] The `sugar_bb` term is a mixture of functions specific to glycosidic torsions and linkage types. For most torsion/linkage types, Rosetta uses the CHarbohydrate-Intrinsic (CHI) energy functions developed from quantum mechanical calculations with isomers of *O*-linked tetrahydropyran oligomers.[30,31] The data were fit to Gaussian functions and matched with statistical data. Together, they are used to compute the energy, given as a function of some torsion angle *x* in degrees, magnitude of the Gaussian distribution *a*, midpoint of the distribution *b*, the intercept of the distribution *d*, and a constant *c* which is twice the square width of the distribution (**Eq. S17**).

$$E_{\text{sugar\_bb}} = d + \sum_i a_i e^{-(x-b_i)^2/c_i} \quad \text{(S17)}$$

For ω torsions, the energy is instead modeled using a series of parabolic functions with coefficients fit to statistical data and centered around the ideal staggered and Gauche conformations. This energy is defined as a function of the torsion angle *x* (in degrees), a constant to define the parabola width, *k*, the vertex of the parabola θ, and the energy difference relative to the minimum *b* (**Eq. S18**). This function approximates the so-called Gauche effect.

$$E_{\text{sugar\_bb}} = k(x-\theta)^2 + b \quad \text{(S18)}$$

The `sugar_bb` score per residue is the sum of each function for each glycosidic torsion in the residue. **Table S8** lists the functional form for each torsion and linkage type. (The functions assume that D-sugars are in the $^4C_1$ chair conformation and that L-sugars are in the $^1C^4$ chair conformation.)

Table S8: Functional form of the sugar backbone energy for each torsion and linkage type

Angle	Ax./eq. designation	Stereoisomer	Exocyclic	Range	Functional form
ϕ	axial (α)	D	—	−180°–180°	Gaussian
	eqiuatorial (β)	D	—	−180°–180°	Gaussian
	axial (α)	L	—	−180°–180°	Gaussian, x=−ϕ
	equatorial (β)	L	—	−180°–180°	Gaussian, x=−ϕ
ψ	ax. (parent at odd O)	D (parent)	no	0–360°	Gaussian
	eq. (parent at odd O)	D (parent)	no	0–360°	Gaussian
	ax. (parent at even O)	D (parent)	no	0–360°	Gaussian
	eq. (parent at even O)	D (parent)	no	0–360°	Gaussian
	ax. (parent at odd O)	L (parent)	no	0–360°	Gaussian, 360°−ψ
	eq. (parent at odd O)	L (parent)	no	0–360°	Gaussian, 360°−ψ
	ax. (parent at even O)	L (parent)	no	0–360°	Gaussian, 360°−ψ
	eq. (parent at even O)	L (parent)	no	0–360°	Gaussian, 360°−ψ
	axial (α)	D	yes	0–360°	Gaussian
	equatorial (β)	D	yes	0–360°	Gaussian
	axial (α)	L	yes	0–360°	Gaussian, 360°−ψ
	equatorial (β)	L	yes	0–360°	Gaussian, 360°−ψ
ω	axial (parent O4)	D (parent)	yes	0–120°	parabolic
	axial (parent O4)	D (parent)	yes	120°–240°	parabolic
	axial (parent O4)	D (parent)	yes	240°–360°	parabolic
	eq. (parent O4)	D (parent)	yes	0–120°	parabolic
	eq. (parent O4)	D (parent)	yes	120°–240°	parabolic
	eq. (parent O4)	D (parent)	yes	240°–360°	parabolic
	axial (parent O4)	L (parent)	yes	0–360°	parabolic, 360°−ω
	axial (parent O4)	L (parent)	yes	120°–240°	parabolic, 360°−ω
	axial (parent O4)	L (parent)	yes	240°–360°	parabolic, 360°−ω
	eq. (parent O4)	L (parent)	yes	0–120°	parabolic, 360°−ω
	eq. (parent O4)	L (parent)	yes	120°–240°	parabolic, 360°−ω
	eq. (parent O4)	L (parent)	yes	240°–360°	parabolic, 360°−ω

Energy terms for nucleic acids

The Rosetta energy function captures van der Waals and electrostatic forces general to all biomolecules. However, these terms do not capture rules specific to the geometry and pairing of nucleic acid bases. Therefore, Das and coworkers have implemented terms to explicitly capture these rules.

Electrostatics. The standard Rosetta electrostatic potential (`fa_elec`) disfavors Watson-Crick base pairs due to repulsion between the fixed positive charges on the hydrogen atoms in close proximity in G-C and A-U pairs. To alleviate this problem, Rosetta uses two modified terms to evaluate electrostatics involving RNA bases. First, electrostatic interactions between phosphate atoms are evaluated using the standard `fa_elec` potential (**Eq. 10** in the main text), via a term called `fa_elec_rna_phos_phos`. Second, electrostatic interactions between RNA bases are captured using the `stack_elec` term.[32] This term scales the `fa_elec` potential as a function of the angle (κ_i) between the normal to the plane of the base (z_i) and the vector $d_{i,j}$ between base heavy atoms i and j in residues r_1 and r_2, respectively (Figure S4). The equation for `stack_elec` is given by **Eq. S19**.

$$E_{\text{stack\_elec}} = \sum_{r_1<r_2} \sum_{i,j} f(\kappa_i, \kappa_j) E_{\text{fa\_elec}} \qquad \text{(S19)}$$

The scaling function $f(\kappa_i, \kappa_j)$ suppresses the electrostatic energy to zero when the bases are coplanar and maintains the full value of the energy when the bases are stacked (**Eq. S20; Fig. S4B**).

$$f(\kappa_i, \kappa_j) = cos^2(\kappa_i) + cos^2(\kappa_j) \qquad \text{(S20)}$$

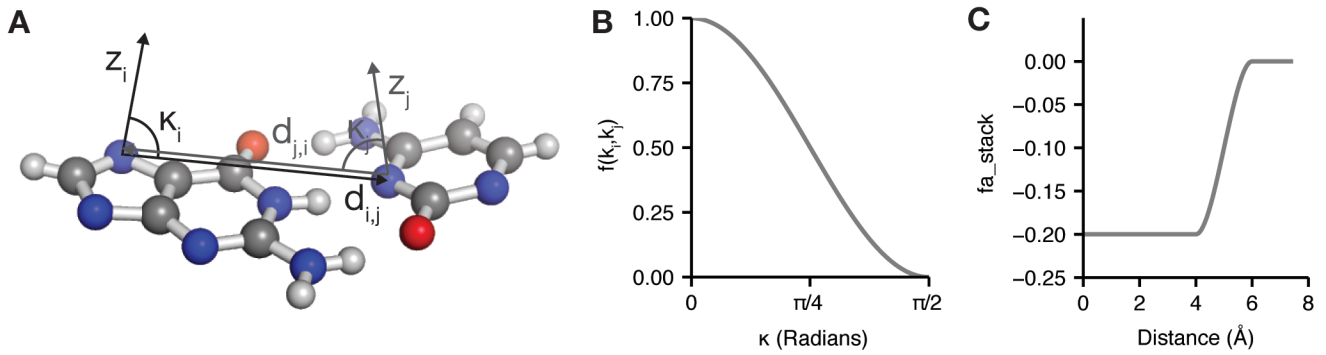

Figure S4. Electrostatic and stacking energies for RNA.
(A) `fa_stack` and `stack_elec` are scaled as a function of the angle, κ_i, between the normal to the base, z_i, and the distance vector between atoms i and j. (B) The scaling function takes the form $f(\kappa_i) = \cos^2(\kappa_i)$, such that the weight is equal to 1.0 when the bases are stacked and 0 when they are coplanar. (C) The `fa_stack` energy for stacked bases (when $f(\kappa_i) = 1.0$).

Base stacking. $\pi - \pi$ stacking interactions are not explicitly captured by `fa_atr`; thus, Rosetta includes an additional stacking bonus term, called `fa_stack`.[33] The `fa_stack` term applies a constant bonus for base atoms less than 4 Å from each other to reward neighboring stacked bases. Like the `stack_elec` term, `fa_stack` also depends on the angle (κ_i) between the normal to the plane of the base (z_i) and the distance vector from atoms i to j ($d_{i,j}$), such that stacked, but not coplanar bases receive this bonus (**Eq. S19: Fig. 4C**). The potential is smoothed to zero between 4 Å and 6 Å using a smoothing function given in **Eq. S21-S23**.

$$E_{\text{fa\_stack}} = \sum_{r_1<r_2} \sum_{i,j} f(\kappa_i, \kappa_j) g(|d_{i,j}|) \qquad \text{(S21)}$$

$$g(|d_{i,j}|) = \begin{cases} -0.2, & |d_{i,j}| \leq 4.0 \\ -0.2h(|d_{i,j}|) & 4.0 < |d_{i,j}| < 6.0 \\ 0.0, & |d_{i,j}| \geq 6.0 \end{cases} \qquad \text{(S22)}$$

$$h(|d_{i,j}|) = -0.2\left[2\left(\frac{|d_{i,j}|-4}{2}\right)^3 - 3\left(\frac{|d_{i,j}|-4}{2}\right)^2 + 1\right] \qquad \text{(S23)}$$

RNA torsions. Like carbohydrates and non-canonical amino acids, nucleic acids require a separate term to evaluate specific torsional energies. For RNA, the `rna_torsion` term evaluates the energies for the nucleic acid backbone and side chain torsions: α, β, γ, δ, ε, ζ, ν_1, ν_2, χ, O2'. The torsional energies are computed as a function of the frequency of some general torsion A found in RNA structures in the PDB (**Eq. S24, Fig. S5**).

$$E_{\text{rna\_torsion}} = \sum_k -\ln(P(A_k)) \qquad \text{(S24)}$$

To accommodate special cases, separate potentials were derived for each of the δ, ε, ν_1, ν_2, χ, O2' torsions depending on whether the sugar pucker is 2'-endo or 3'-endo. Additionally, a separate χ potential was derived for purines and pyrimidines. For ζ, there are three separate potentials depending on whether the α torsion of the following residue is gauche$^-$, trans, or gauche$^+$. Additionally, a set of four harmonic restraints, together comprising `rna_sugar_close`, are applied to ensure that the RNA sugar ring remains closed: a bond distance restraint between atoms O4' and C1', and three angle restraints for the O4'-C1'-C2', C4'-O4'-C1', and O4'-C1'-first base atom angles.

Solvation. The full atom RNA potential contains an orientation-dependent desolvation penalty for polar atoms (`geom_sol`). The penalty is equal to the sum of the values of the orientation-dependent Rosetta hydrogen bonding energies for virtual water molecules placed at the positions of each occluding atom. The form of this term is given by **Eq. S25**.

$$E_{\text{geom\_sol}} = \sum_{r_1<r_2} \sum_{i,j} E_{\text{hbond}}(r_i - v_j) \qquad \text{(S25)}$$

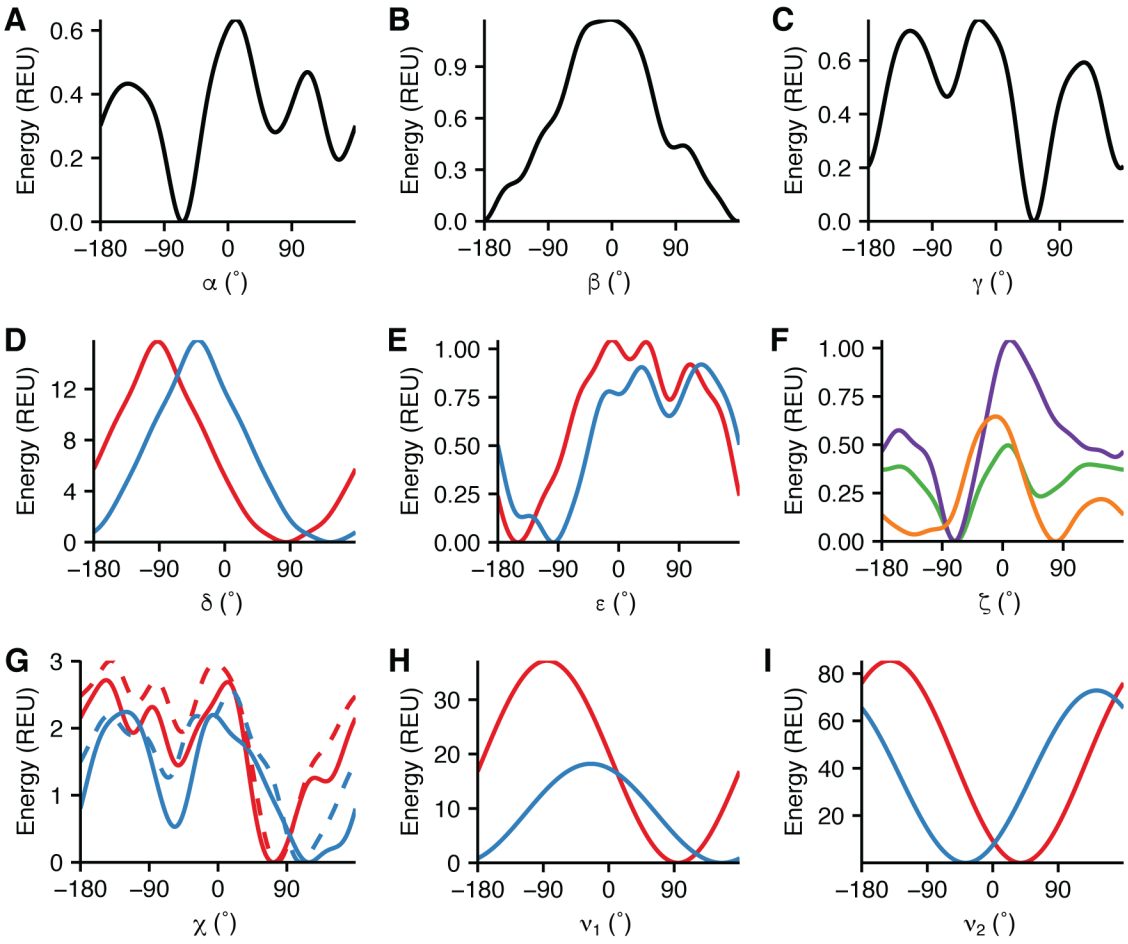

Figure S5. Torsion potentials for RNA

RNA torsional potential for (A) α, (B) β, (C) γ, (D) δ, (E) ε, (F) ζ when the α torsion of the following residue is gauche$^-$ (orange), trans (cyan), or gauche$^+$ (purple) (G) χ for purines (lighter red and blue) and pyrimidines (darker red and blue), (H) ν_1, (I) ν_2. Potentials when the sugar pucker is C2'-endo are shown in red and C3'-endo shown in blue.

References

(1) Kuhlman, B.; Baker, D. Native Protein Sequences Are close to Optimal for Their Structures. *Proc Natl Acad Sci U S A* **2000**, *97* (19), 10383–10388.

(2) Kuhlman, B.; Dantas, G.; Ireton, G. C.; Varani, G.; Stoddard, B. L.; Baker, D. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science (80-.).* **2003**, *302* (5649), 1364–1368.

(3) Tsai, J.; Bonneau, R.; Morozov, A. V.; Kuhlman, B.; Rohl, C. A.; Baker, D. An Improved Protein Decoy Set for Testing Energy Functions for Protein Structure Prediction. *Proteins Struct. Funct. Bioinforma.* **2003**, *53* (1), 76–87.

(4) Meiler, J.; Baker, D. ROSETTALIGAND: Protein-Small Molecule Docking with Full Side-Chain Flexibility. *Proteins Struct. Funct. Bioinforma.* **2006**, *65* (3), 538–548.

(5) Park, H.; Bradley, P.; Greisen, P.; Liu, Y.; Kim, D. E.; Baker, D.; DiMaio, F. Simultaneous Optimization of Biomolecular Energy Function on Features from Small Molecules and Macromolecules. *J. Chem. Theory Comput.* **2016**, *12* (12), 6201–6212.

(6) Kuhlman, B.; Baker, D. Native Protein Sequences Are close to Optimal for Their Structures. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (19), 10383–10388.

(7) Yanover, C.; Bradley, P. Extensive Protein and DNA Backbone Sampling Improves Structure-Based Specificity Prediction for C2H2 Zinc Fingers. *Nucleic Acids Res.* **2011**, *39* (11), 4564–4576.

(8) O'Meara, M. J.; Leaver-Fay, A.; Tyka, M. D.; Stein, A.; Houlihan, K.; DiMaio, F.; Bradley, P.; Kortemme, T.; Baker, D.; Snoeyink, J.; Kuhlman, B. Combined Covalent-Electrostatic Model of Hydrogen Bonding Improves Structure Prediction with Rosetta. *J. Chem. Theory Comput.* **2015**, *11* (2), 609–622.

(9) Kortemme, T.; Morozov, A. V; Baker, D. An Orientation-Dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein-Protein Complexes. *J. Mol. Biol.* **2003**, *326* (4), 1239–1259.

(10) Dunbrack, R. L.; Cohen, F. E.; Cohen, F. E. Bayesian Statistical Analysis of Protein Side-Chain Rotamer Preferences. *Protein Sci.* **1997**, *6* (8), 1661–1681.

(11) Shapovalov, M. V; Dunbrack, R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19* (6), 844–858.

(12) Leaver-Fay, A.; O'Meara, M. J.; Tyka, M.; Jacak, R.; Song, Y.; Kellogg, E. H.; Thompson, J.; Davis, I. W.; Pache, R. A.; Lyskov, S.; Gray, J. J.; Kortemme, T.; Richardson, J. S.; Havranek, J. J.; Snoeyink, J.; Baker, D.; Kuhlman, B. Scientific Benchmarks for Guiding Macromolecular Energy Function Improvement. *Methods Enzymol.* **2013**, *523*, 109–143.

(13) William L. Jorgensen, *; David S. Maxwell, and; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. **1996**.

(14) Lennard-Jones, J. On the Determination of Molecular Fields II: From the Variation of Viscosity of a Gas with Temperature. *R. Soc. London, Ser. A, Contain. Pap. a Math. Phys. Character* **1924**, *106*, 464–477.

(15) Lennard-Jones, J. On the Determination of Molecular Fields I: From the Variation of Viscosity of a Gas with Temperature. *R. Soc. London, Ser. A, Contain. Pap. a Math. Phys. Character* **1924**, *106*, 441–462.

(16) Lazaridis, T.; Karplus, M. Effective Energy Function for Proteins in Solution. *Proteins* **1999**, *35* (2), 133–152.

(17) Neria, E.; Fischer, S.; Karplus, M. Simulation of Activation Free Energies in Molecular Systems. *J. Chem. Phys.* **1996**, *105* (5), 1902–1921.

(18) Shapovalov, M. V; Dunbrack Jr., R. L. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive Kernel Density Estimates and Regressions. *Structure* **2011**, *19*

(6), 844–858.
(19) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
(20) Kellogg, E. H.; Leaver-Fay, A.; Baker, D. Role of Conformational Sampling in Computing Mutation-Induced Changes in Protein Structure and Stability. *Proteins* **2011**, *79* (3), 830–838.
(21) Chaudhury, S.; Lyskov, S.; Gray, J. J. PyRosetta: A Script-Based Interface for Implementing Molecular Modeling Algorithms Using Rosetta. *Bioinformatics* **2010**, *26* (5), 689–691.
(22) Chaudhury, S.; Berrondo, M.; Weitzner, B. D.; Muthu, P.; Bergman, H.; Gray, J. J. Benchmarking and Analysis of Protein Docking Performance in Rosetta v3.2. *PLoS One* **2011**, *6* (8), e22477.
(23) Bhardwaj, G.; Mulligan, V. K.; Bahl, C. D.; Gilmore, J. M.; Harvey, P. J.; Cheneval, O.; Buchko, G. W.; Pulavarti, S. V. S. R. K.; Kaas, Q.; Eletsky, A.; Huang, P.-S.; Johnsen, W. A.; Greisen, P. J.; Rocklin, G. J.; Song, Y.; Linsky, T. W.; Watkins, A.; Rettie, S. A.; Xu, X.; Carter, L. P.; Bonneau, R.; Olson, J. M.; Coutsias, E.; Correnti, C. E.; Szyperski, T.; Craik, D. J.; Baker, D. Accurate de Novo Design of Hyperstable Constrained Peptides. *Nature* **2016**, *538* (7625), 329–335.
(24) Lao, B. B.; Drew, K.; Guarracino, D. A.; Brewer, T. F.; Heindel, D. W.; Bonneau, R.; Arora, P. S. Rational Design of Topographical Helix Mimics as Potent Inhibitors of Protein–Protein Interactions. *J. Am. Chem. Soc.* **2014**, *136* (22), 7877–7888.
(25) Drew, K.; Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Chou, F.-C.; Lyskov, S.; Bullock, B. N.; Watkins, A.; Labonte, J. W.; Pacella, M.; Kilambi, K. P.; Leaver-Fay, A.; Kuhlman, B.; Gray, J. J.; Bradley, P.; Kirshenbaum, K.; Arora, P. S.; Das, R.; Bonneau, R. Adding Diverse Noncanonical Backbones to Rosetta: Enabling Peptidomimetic Design. *PLoS One* **2013**, *8* (7), e67051.
(26) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.
(27) Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Kirshenbaum, K.; Bonneau, R. A Rotamer Library to Enable Modeling and Design of Peptoid Foldamers. *J. Am. Chem. Soc.* **2014**, *136* (24), 8772–8782.
(28) Richardson, J. S.; Keedy, D. A.; Richardson, D. C. In Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map. *World Sci. Publ. Co. Pte. Ltd Singapore* **2013**, 46–61.
(29) Labonte, J. W.; Aldof-Bryfogle, J.; Schief, W. R.; Gray, J. J. Residue-Centric Modeling and Design of Saccharide and Glycoconjugate Structures. *J. Comput. Chem.* **2017**, *38* (5), 276–287.
(30) Nivedha, A. K.; Thieker, D. F.; Makeneni, S.; Hu, H.; Woods, R. J. Vina-Carb: Improving Glycosidic Angles during Carbohydrate Docking. *J. Chem. Theory Comput.* **2016**, *12* (2), 892–901.
(31) Nivedha, A. K.; Makeneni, S.; Foley, B. L.; Tessier, M. B.; Woods, R. J. Importance of Ligand Conformational Energies in Carbohydrate Docking: Sorting the Wheat from the Chaff. *J. Comput. Chem.* **2014**, *35* (7), 526–539.
(32) Chou, F.-C.; Kladwang, W.; Kappel, K.; Das, R. Blind Tests of RNA Nearest-Neighbor Energy Prediction. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (30), 8430–8435.
(33) Sripakdeevong, P.; Kladwang, W.; Das, R. An Enumerative Stepwise Ansatz Enables Atomic-Accuracy RNA Loop Modeling. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (51), 20573–20578.