Supplementary information

Optimization of rotamers prior to template minimization improves stability predictions made by computational protein design

James A. Davey and Roberto A. Chica* Department of Chemistry, University of Ottawa, Ottawa, ON, K1N 6N5

* To whom correspondence should be addressed: rchica@uottawa.ca

Supplementary material included:

Figure S1. Gβ1 sequences

Figure S2. Enrichment profiles for SSD calculations excluding ROM-minimized rotamers

Figure S3. Template backbone comparison

Figure S4. Contact map for ROM_{NMR} templates

Figure S5. Enrichment profiles for SSD calculations using ROM templates prepared by energy minimization of backbone (bbROM) or side-chain (scROM) atoms only

Figure S6. Enrichment profiles for amino acid biased single-state design with the minimized crystal structure template

Figure S7. Enrichment profiles for amino acid biased single-state design with the minimized NMR structure template

Figure S8. Enrichment profiles for configuration biased single-state design with the minimized crystal structure template

Figure S9. Enrichment profiles for configuration biased single-state design with the minimized NMR structure template

Figure S10. Sequence motifs constructed from the top 100 ranked sequences predicted by SSD using a variety of MIN and ROM templates

Table S1. Sequence enrichment results for SSD calculations including or excluding ROMminimized rotamers

Table S2. Sequence enrichment results for SSD calculations using ROM templates prepared by energy minimization of backbone (bbROM) or side-chain (scROM) atoms only

Table S3. Sequence enrichment results for biased calculations

Table S4. Sequence binning results for biased calculations



Figure S1. G\beta1 sequences. Amino acid residues found at each design position in wild-type (WT) and mutant G β 1 sequences (1-84) are listed. The 84 mutant sequences are classified into

one of four stability groups: 24 sequences displaying stability greater than or approximately equal to the WT (stabilizing, green), 12 sequences of lower stability than the WT (destabilizing, yellow), 24 sequences that do not fold (unfolded, red), and 24 sequences postulated to adopt an alternate non-native fold (non-native, blue).



optimized and energy minimized crystal (ROM_{XTAL}) and NMR (ROM_{NMR}) templates are shown as bars. Each bar is colored according to the proportion of sequences from each stability group found in the top 24, with stabilizing, destabilizing, unfolded, and non-native sequences colored green, yellow, red, and blue, respectively. ROM templates were prepared from the wild-type (WT, black) and 84 mutant G β 1 sequences that are numbered and colored according to their stability group. Enrichment profiles of several ROM_{NMR} templates prepared from non-native sequences do not contain 24 sequences, and are shown as bars of reduced length.

A ROM _{XTAL}	XTAL	NMR	Stabilizing	Destabilizing	Unfolded	Non-native
Stabilizing	0.47 ± 0.02	1.16 ± 0.01	0.14 ± 0.04	0.14 ± 0.04	0.16 ± 0.04	0.20 ± 0.03
Destabilizing	0.47 ± 0.01	1.15 ± 0.01	0.14 ± 0.04	0.10 ± 0.03	0.13 ± 0.04	0.20 ± 0.04
Unfolded	0.48 ± 0.02	1.16 ± 0.01	0.16 ± 0.04	0.13 ± 0.04	0.12 ± 0.05	0.23 ± 0.03
Non-native	0.47 ± 0.02	1.16 ± 0.02	0.20 ± 0.03	0.20 ± 0.04	0.23 ± 0.03	0.17 ± 0.05
В			abilizing	estabilizing	nfolded	on-native
B ROM _{NMR}	XTAL	NMR	Stabilizing	Destabilizing	Unfolded	Non-native
B ROM _{NMR} Stabilizing	XTAL 1.01 ± 0.02	NMR 0.74 ± 0.03	Stabilizing	90.0 + 95.0 + 95.0	0.29 ± 0.08	Non-native
B ROM _{NMR} Stabilizing Destabilizing	XTAL 1.01 ± 0.02 1.00 ± 0.02	NMR 0.74 ± 0.03 0.73 ± 0.02	0.24 ± 0.08 0.26 ± 0.06	Destabilizing = 02.0 = 00.0 = 00.0	pepoju 0.29 ± 0.08 0.30 ±	Non-mative 0.01 ± 0.06
B ROM _{NMR} Stabilizing Destabilizing Unfolded	XTAL 1.01 ± 0.02 1.00 ± 0.02 0.96 ± 0.02	NMR 0.74 ± 0.03 0.73 ± 0.02 0.72 ± 0.03	0.24 ± 0.08 0.06 0.00 0.00 ± 0.00	Destabilizing 0.26 ± 0.00 0.20 ± 0.00	Paplogun 0.29 ± 0.08 0.30 ± 0.09 0.22 ± 0.10	0.01 ± 0.06 0.31 ± 0.06 0.34 ± 0.06

Figure S3. Template backbone comparison. Backbone RMSD (N-C_{α}-C=O) between pairs of ROM_{XTAL} (A) and ROM_{NMR} (B) templates are reported as the average and standard deviation for ROM templates grouped according to the stability of their seed sequence (stabilizing, destabilizing, unfolded, and non-native). Average backbone RMSD values are colored blue, green, yellow, orange, or red if they fall within the 0.10-0.19 Å, 0.20-0.29 Å, 0.30-0.39 Å, 0.40-0.49 Å, or > 0.5 Å ranges, respectively. The backbones of ROM templates are also compared to that of the crystal (XTAL) and NMR seed structures.



Figure S4. Contact map for ROM_{NMR} templates. van der Waals interaction energies between residues at designed positions and all other residues in ROM_{NMR} templates are averaged by stability group and by residue identity within each stability group. Residues found at each designed position in ROM_{NMR} templates prepared from stabilizing, destabilizing, unfolded, and non-native sequences are colored green, yellow, red, and blue, respectively, with the wild-type residue indicated in bold. Designed position residues are boxed separately from residues whose identity does not vary between templates. Interaction energies are colored according to their strength ranging from 0 kcal/mol (white, no interaction) to -3.5 kcal/mol (dark purple, strong favorable interaction).



Figure S5. Enrichment profiles for SSD calculations using ROM templates prepared by energy minimization of backbone (bbROM) or side-chain (scROM) atoms only. The top 24

sequences (excluding wild type) predicted by single-state design using rotamer optimized and energy minimized crystal (ROM_{XTAL}) and NMR (ROM_{NMR}) templates are shown as bars. Only the backbone or side-chain atoms were minimized in the case of bbROM or scROM templates, respectively, with all other atoms fixed during minimization. Each bar is colored according to the proportion of sequences from each stability group found in the top 24, with stabilizing, destabilizing, unfolded, and non-native sequences colored green, yellow, red, and blue, respectively. ROM templates were prepared from the wild-type (WT, black) and 84 mutant G β 1 sequences that are numbered and colored according to their stability group. Enrichment profiles of several ROM_{NMR} templates prepared from non-native sequences do not contain 24 sequences, and are shown as bars of reduced length.



Figure S6. Enrichment profiles for amino acid biased single-state design with the minimized crystal structure template. The top 24 predicted sequences (excluding wild type) are shown as bars. Each bar is colored according to the proportion of sequences from each

stability group found in the top 24, with stabilizing (1-24), destabilizing (25-36), unfolded (37-60), and non-native (61-84) sequences colored green, yellow, red, and blue, respectively. Various bias weights in kcal/mol were applied to the scoring function to favor (negative bias weights) or disfavor (positive bias weights) the amino acid sequence of the wild-type (WT, black) or of one of the 84 G β 1 mutants. Enrichment profiles that do not contain 24 sequences are shown as bars of reduced length.

	-1000	-100	-10	-4.2	-2.8	-1.4	No Bias	+1.4	+2.8	+4.2	+10	+100	+1000
WT	<u>.</u>												
2													
3	-												
4													
6													
7													
8	•												
9													
10				_							_		
12													
13	-												
14	1												
15				_									
17													
18	-												
19													
20													
22													
23													
24	<u> </u>												
25	-			_							_		
27													
28	-												
30	-												
31	<u>.</u>												
33													
34													
36	1 -												
37	-												
38	1												
39	12												
41													
42	•												
43													
44				_									
46	-												
47													
48													
49													
51	-												
52	1 C												
53	1												
55													
56	-												
57	-												
58 59													
60													
61													
62													
63 64													
65													
66	•												
67													
69				_							_		
70													
71	-												
72	L												
74													
75	•												
76	1 C												
77													
78 79													
80													
81													
82													
83 84													

Figure S7. Enrichment profiles for amino acid biased single-state design with the minimized NMR structure template. The top 24 predicted sequences (excluding wild type) are shown as bars. Each bar is colored according to the proportion of sequences from each stability group found in the top 24, with stabilizing (1-24), destabilizing (25-36), unfolded (37-60), and

non-native (61-84) sequences colored green, yellow, red, and blue, respectively. Various bias weights in kcal/mol were applied to the scoring function to favor (negative bias weights) or disfavor (positive bias weights) the amino acid sequence of the wild-type (WT, black) or of one of the 84 G β 1 mutants. Enrichment profiles that do not contain 24 sequences are shown as bars of reduced length.



Figure S8. Enrichment profiles for configuration biased single-state design with the minimized crystal structure template. The top 24 predicted sequences (excluding wild type) are shown as bars. Each bar is colored according to the proportion of sequences from each

stability group found in the top 24, with stabilizing (1-24), destabilizing (25-36), unfolded (37-60), and non-native (61-84) sequences colored green, yellow, red, and blue, respectively. Various bias weights in kcal/mol were applied to the scoring function to favor (negative bias weights) or disfavor (positive bias weights) the rotamer configuration for the wild-type (WT, black) or one of the 84 G β 1 mutants following rotamer optimization on the minimized crystal structure. Enrichment profiles that do not contain 24 sequences are shown as bars of reduced length.

	-1000	-100	-10	-4.2	-2.8	-1.4	No Bias	+1.4	+2.8	+4.2	+10	+100	+1000
WT													
2													
3	-		_										
4													
5 6		_											
7													
8	-												
9													
11													
12	-												
13													
14		_											
16													
17	-												
18													
20													
21	-												
22	-	-	_										
23		-											
25													
26											_		
27													
29	-												
30													
32													
33													
35	-												
36	-				_	_				_	_	_	
37													
39													
40	-												
41													
42													
44													
45													
46													
48													
49	-												
50													
52													
53	-												
54													
55													
57	-												
58								_		_			
59 60	-												
61													
62													
63													
64		-											
66													
67													
68													
70													
71					-	_	-	_		_	_		
72													
73 74													
75													
76													
77		-											
79													
80													
81													
82													
84													

Figure S9. Enrichment profiles for configuration biased single-state design with the minimized NMR structure template. The top 24 predicted sequences (excluding wild type) are shown as bars. Each bar is colored according to the proportion of sequences from each stability group found in the top 24, with stabilizing (1-24), destabilizing (25-36), unfolded (37-60), and

non-native (61-84) sequences colored green, yellow, red, and blue, respectively. Various bias weights in kcal/mol were applied to the scoring function to favor (negative bias weights) or disfavor (positive bias weights) the rotamer configuration for the wild-type (WT, black) or one of the 84 G β 1 mutants following rotamer optimization on the minimized NMR structure. Enrichment profiles that do not contain 24 sequences are shown as bars of reduced length.



Figure S10. Sequence motifs constructed from the top 100 ranked sequences predicted by SSD using a variety of MIN and ROM templates. Side-chain rotamers of core residues (positions 3, 5, 7, 20, 26, 30, 34, 39, 52, and 54) were optimized on each fixed backbone template using hydrophobic amino acids (A, V, L, I, and F). The wild-type Y amino acid was also allowed at position 3. For ROM templates prepared from stabilizing, destabilizing, unfolded, and non-native sequences, sequences were optimized on each template and ranked based on their backbone drift modelling score (see text for details). The size of each letter is proportional to the frequency of occurrence of each amino-acid type at each position in the top 100 ranked sequences. Sequences motifs comprise substitutions that are included (grey) or not (black) in the 85 G β 1 seed sequences listed in Figure S1.

	Sequences in Top 24										
Template	Stabilizing	Destabilizing	Unfolded	Non-native							
ROM _{XTAL} excluding ROM-minimized rotamers											
WT	16	1	5	2							
Stabilizing	17 ± 2	2 ± 2	4 ± 2	2 ± 2							
Destabilizing	11 ± 1	8 ± 3	4 ± 3	0							
Unfolded	13 ± 4	1 ± 3	9 ± 4	1 ± 1							
Non-native	11 ± 3	1 ± 1	0	12 ± 3							
ROM _{XTAL} including ROM-m	inimized rot	amers									
WT	14	4	4	2							
Stabilizing	17 ± 2	2 ± 2	4 ± 2	2 ± 2							
Destabilizing	11 ± 1	9 ± 3	4 ± 4	0							
Unfolded	10 ± 3	2 ± 3	12 ± 4	0							
Non-native	10 ± 4	0	0	13 ± 4							
ROM _{NMR} excluding ROM-m	inimized rot	amers									
WT	7	2	15	0							
Stabilizing	13 ± 5	1 ± 1	9±5	1 ± 2							
Destabilizing	7 ± 3	1 ± 1	15 ± 3	0							
Unfolded	2 ± 3	2 ± 0	20 ± 3	0							
Non-native	3 ± 6	0	3 ± 5	7 ± 5							
ROM _{NMR} including ROM-m	inimized rota	amers									
WT	7	2	15	0							
Stabilizing	13 ± 5	1 ± 1	10 ± 5	1 ± 1							
Destabilizing	10 ± 4	8 ± 3	6 ± 7	0							
Unfolded	3 ± 3	3 ± 2	19 ± 5	0							
Non-native	3 ± 5	0	2 ± 4	10 ± 3							

Table S1. Sequence enrichment results for SSD calculations including or excluding ROM-minimized rotamers

	Sequences in Top 24									
Template	Stabilizing	Destabilizing	Unfolded	Non-native						
scROM _{XTAL}										
WT	13	5	5	1						
Stabilizing	13 ± 2	5 ± 1	5 ± 3	0						
Destabilizing	10 ± 1	9 ± 2	5 ± 3	0						
Unfolded	10 ± 1	4 ± 2	10 ± 3	0						
Non-native	11 ± 3	4 ± 3	2 ± 2	7 ± 2						
bbROM _{XTAL}										
WT	12	7	3	2						
Stabilizing	11 ± 1	7 ± 1	6 ± 1	1 ± 0						
Destabilizing	10 ± 1	7 ± 1	6 ± 1	1 ± 0						
Unfolded	10 ± 1	6 ± 1	7 ± 1	1 ± 0						
Non-native	11 ± 1	6 ± 1	5 ± 2	2 ± 1						
scROM _{NMR}										
WT	14	2	8	0						
Stabilizing	16 ± 3	1 ± 0	7 ± 3	0						
Destabilizing	11 ± 5	8 ± 3	5 ± 6	0						
Unfolded	3 ± 3	3 ± 2	17 ± 5	0						
Non-native	1 ± 2	1 ± 1	7 ± 4	9 ± 1						
bbROM _{NMR}										
WT	5	2	17	0						
Stabilizing	5 ± 2	2 ± 1	17 ± 2	0						
Destabilizing	11 ± 2	1 ± 1	12 ± 2	0						
Unfolded	5 ± 3	2 ± 1	17 ± 2	0						
Non-native	7 ± 4	0	10 ± 3	3 ± 3						

Table S2. Sequence enrichment results for SSD calculations using ROM templates prepared by energy minimization of backbone (bbROM) or side-chain (scROM) atoms only

Table S3. Sequence enrichment results for biased calculations

	Sequences in Top 24								
Template	Stabilizing	Destabilizing	Unfolded	Non-native					
No Bias									
MIN _{XTAL}	14	4	4	2					
MIN _{NMR}	11	1	8	4					
MIN _{XTAL} Amir	no Acid Bias	a							
Stabilizing	16 ± 2	4 ± 1	4 ± 2	0					
Destabilizing	12 ± 1	8 ± 2	4 ± 3	1 ± 1					
Unfolded	11 ± 1	3 ± 2	10 ± 2	0					
Non-native	10 ± 1	3 ± 2	1 ± 1	11 ± 2					
MIN _{NMR} Amin	o Acid Bias	a							
Stabilizing	13 ± 2	1 ± 1	8 ± 1	1 ± 1					
Destabilizing	11 ± 2	1 ± 1	8 ± 2	2 ± 1					
Unfolded	10 ± 2	2 ± 1	10 ± 2	1 ± 1					
Non-native	8 ± 2	1 ± 1	4 ± 2	11 ± 1					
MIN _{XTAL} Conf	figuration Bi	as ^b							
Stabilizing	15 ± 2	5 ± 1	4 ± 1	0					
Destabilizing	11 ± 1	8 ± 2	4 ± 3	0					
Unfolded	10 ± 3	4 ± 2	9 ± 3	0					
Non-native	11 ± 1	3 ± 1	1 ± 1	10 ± 2					
MIN_{NMR} Conf	iguration Bia	as ^b							
Stabilizing	13 ± 1	1 ± 1	8 ± 1	2 ± 1					
Destabilizing	11 ± 2	1 ± 1	8 ± 2	3 ± 2					
Unfolded	11 ± 1	2 ± 1	10 ± 1	2 ± 1					
Non-native	7 ± 1	1 ± 1	6 ± 1	10 ± 2					

^aA 100 kcal/mol bias was applied in favor of amino acid identity at each design position ^bA 10 kcal/mol bias was applied in favor of rotamer configuration at each design position

	MIN _{XTAL}							MIN _{NMR}					
	Success	True	False	False	True	Cut-off	Success	True	False	False	True	Cut-off	
	Rate	Positive	Negative	Positive	Negative	(kcal/mol)	Rate	Positive	Negative	Positive	Negative	(kcal/mol)	
No Bias			0		0	, , ,							
WT	79%	8	16	2	58	-71.5	71%	2	22	2	58	-63.6	
Amino A	cid Bias ^a												
WT	73%	1	23	0	60	-171.5	73%	1	23	0	60	-162.7	
1	76%	6	18	2	58	-142.1	75%	3	21	0	60	-133.7	
2	86%	16	8	4	56	-117.4	81%	8	16	0	60	-119.3	
3	92%	17	7	0	60	-125.6	85%	11	13	0	60	-116.2	
4	79%	9	15	3	57	-140.5	75%	3	21	0	60	-134.0	
5	76%	5	19	1	59	-156.1	73%	1	23	0	60	-148.5	
6	79%	6	18	0	60	-142.0	76%	4	20	0	60	-133.9	
7	87%	17	7	4	56	-115.1	82%	9	15	0	60	-115.3	
8	75%	17	7	14	46	-90.4	80%	8	16	1	59	-111.6	
9	83%	11	13	1	59	-140.7	75%	3	21	0	60	-134.2	
10	71%	16	8	16	44	0°	80%	9	15	2	58	-105.5	
11	89%	18	6	3	57	-115.6	81%	8	16	0	60	-118.6	
12	77%	5	19	0	60	-155.2	74%	2	22	0	60	-147.9	
13	82%	9	15	0	60	-141.9	79%	6	18	0	60	-134.0	
14	82%	14	10	5	55	-126.4	75%	3	21	0	60	-132.3	
15	73%	1	23	0	60	-171.3	73%	1	23	0	60	-162.8	
16	76%	4	20	0	60	-155.5	73%	1	23	0	60	-148.5	
17	82%	9	15	0	60	-142.0	76%	4	20	0	60	-134.2	
18	81%	8	16	0	60	-139.2	76%	4	20	0	60	-132.7	
19	81%	9	15	1	59	-141.5	76%	4	20	0	60	-131.4	
20	81%	9	15	1	59	-140.2	75%	5	19	2	58	-122.2	
21	75%	3	21	0	60	-157.2	75%	3	21	0	60	-148.5	
22	75%	3	21	0	60	-156.3	74%	2	22	0	60	-148.4	
23	79%	6	18	0	60	-141.3	76%	4	20	0	60	-131.7	
24	75%	3	21	0	60	-155.4	74%	2	22	0	60	-146.7	
Configura	ation Bias ^₅												
WT	73%	1	23	0	60	-141.5	73%	1	23	0	60	-133.7	
1	77%	9	15	4	56	-117.8	74%	2	22	0	60	-113.7	
2	88%	14	10	0	60	-111.0	76%	4	20	0	60	-97.8	
3	86%	16	8	4	56	-98.2	85%	11	13	0	60	-97.1	
4	79%	10	14	4	56	-117.9	74%	2	22	0	60	-113.7	
5	76%	5	19	1	59	-128.7	75%	3	21	0	60	-116.8	
6	81%	13	11	5	55	-98.4	76%	4	20	0	60	-107.1	
(74%	16	8	14	46	-90.4	81%	8	16	0	60	-96.8	
8	80%	12	12	5	55	-109.3	76%	4	20	0	60	-93.0	
9	80%	1	17	0	60	-121.3	74%	2	22	0	60	-113.7	
10	75%	14	10	11	49	-97.8	76%	4	20	0	60	-93.8	
11	75%	16	8	13	47	-90.5	77%	5	19	0	60	-96.5	
12	76%	4	20	0	60	-131.3	73%	2	22	1	59	-123.7	
13	86%	12	12	0	60	-100.6	79%	6	18	0	60	-106.6	
14	81%	11	13	3	57	-113.9	79%	6	18	0	60	-97.9	
15	73%	1	23	0	60	-141.5	73%	1	23	0	60	-133.7	
10	//%	0	18	1	59	-128.7	7.5%	3	21	0	60	-110./	
1/	ŏ1%	ð 1	16	0	60	-119.0	74%	2	22	0	00	-113.7	
10	89%	15	9	0	6U	-110.1	7.1%	5	19	0	60	-102.7	
19	74%	0	13	9	51	-100.5	79%	0	10	0	60	-90.9	
∠0 21	01% 770/	9	10	0	59	-121.0	7 / %	2	19	0	60	-101.0	
∠ I 22	76%	3	19	0	60	-110.0	7/0/	3	21	0	60	-121.0	
22	70%	4	20 19	0	60	-110.0	74%	2	22	0	60	-110.0	
20	75%	3	21	0	60	-110.0	76%	4	20	0	60	-112.4	
24	1570	3	21	0	00	-131.3	1570	3	21	U	00	-112.4	

Table S4. Sequence binning results for biased calculations

 ^aA 100 kcal/mol bias was applied in favor of amino acid identity at each design position
 ^bA 10 kcal/mol bias was applied in favor of rotamer configuration at each design position
 ^cA cut-off value of 0 kcal/mol was assigned to sequence 10 because single-state design with an amino acid bias could not favorably score the wild-type sequence