Side-Chain Conformational Preferences Govern Protein−Protein Interactions

Andrew M. Watkins,[‡] Richard Bonneau,[*,∥] and Paramjit S. Arora[*,‡]

Departments of [†]Biology, [‡]Chemistry, and [§]Computer Science, New York University, 100 Washington Square East, New York, New York 10003, United States
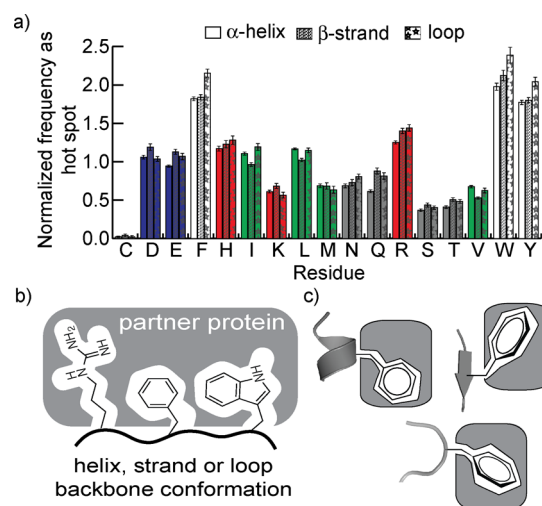
[∥]Center for Computational Biology, Simons Foundation, New York, New York 10010, United States

S *Supporting Information*

ABSTRACT: Protein secondary structures serve as geometrically constrained scaffolds for the display of key interacting residues at protein interfaces. Given the critical role of secondary structures in protein folding and the dependence of folding propensities on backbone dihedrals, secondary structure is expected to influence the identity of residues that are important for complex formation. Counter to this expectation, we find that a narrow set of residues dominates the binding energy in protein−protein complexes independent of backbone conformation. This finding suggests that the binding epitope may instead be substantially influenced by the side-chain conformations adopted. We analyzed side-chain conformational preferences in residues that contribute significantly to binding. This analysis suggests that preferred rotamers contribute directly to specificity in protein complex formation and provides guidelines for peptidomimetic inhibitor design.

Figure 1. Surprisingly, the frequency at which a given amino acid is a hot spot residue divided by its prevalence overall is approximately constant across helices, strands, and loops (a). Cartoon representation of the important residues at protein−protein interfaces (b). In this work, we find that different rotamers are favored by hot spot residues in a secondary structure specific manner (c).

Protein secondary structures scaffold side-chain functionality to mediate complex formation.[1] Mimicry of protein secondary structures has led to the development of successful inhibitors of protein−protein interactions (PPIs).[2−7] To categorize complexes mediated by secondary structures and motivate subsequent design of interfacial peptidomimetics, we and others have analyzed entries in the Protein Data Bank (PDB) and cataloged high affinity secondary structure elements at interfaces.[8−11] These efforts used computational alanine scanning analysis to identify hot spot residues, residues whose mutation to alanine results in an estimated $\Delta\Delta G$ of binding >1 kcal/mol.[12−15] The data set of secondary structures mediating PPIs offers a curated starting point for the development of peptidomimetic inhibitors. This data set provides insights into fundamental weak forces that mediate protein complex formation. In our earlier work, we found that alanine mutagenesis scanning estimates agreed with observations that the aromatic residues and arginine are most likely to be hot spot residues.[14,16] Surprisingly, these secondary structure-specific data sets reveal that hot spot residue frequencies are independent of the backbone environment (Figure 1). Thus, β-branched residues are no more enriched as hot spots on β-strands than they are on α-helices, and residues with high helical propensity are not overrepresented as hot spots on interfacial helices.

The lack of correlation between hotspot residue propensity and backbone conformation prompted a question: How do

secondary structures featuring largely similar hot spot residues dictate specific interactions in proteins? The three-dimensional epitope that interacts with the binding partner is likely influenced by various factors, including the identity and the positioning of key residues. Side chain dihedral angles govern atomic positions and thus contribute directly to the binding epitope.[17] We hypothesized that the backbone-dependent side-chain conformational preferences of hot spot residues in particular may provide an additional factor that governs the binding energetics of protein interfaces.[18−21] We sought to examine if a particular secondary structure displaying similar side-chain functionality might be able to specifically target different receptors by adopting distinct side-chain conformations (Figure 1c).

We performed a residue level analysis of protein−protein interfaces with the goal of defining the contribution of side-chain conformation to molecular recognition. Specifically, we asked if hot spot residues favor particular side-chain conformations in a secondary structure specific manner. We began by assessing the

frequency of side-chain rotamers on hot spot residues and the differences between these frequencies and non-hot spot residues, with the hypothesis that rotamers that are enriched in hot spot residues may be critical for binding affinity or specificity.[20] The distribution of rotamer states for different residue types is well established and known to be backbone dependent, providing an important foundation for our analysis.[18,22−27] Since we are examining native structures, we used the Dunbrack backbone-dependent rotamer libraries as our standard for comparison.[25] Without subgrouping by secondary structure, we typically saw moderate enrichments to common rotameric states (Figure S1). When we subdivided the results based on secondary structure, however, we found stronger enrichments that included uncommon rotamers. For example, the *anti* (*a*) χ_1 bin is enriched for Phe hot spot residues on loops, the *gauche−* (*g−*) χ_1 bin is enriched only for strand Phe, and the *gauche+* (*g+*) χ_1 bin dominates for helical Phe (Figure S2). The results suggest that the preferred side-chain conformation for hot spot residues is backbone dependent. A summary of these data for χ_1 is provided in Table S1. A detailed description of data analysis for all figures is available in the supplement.

We examined the average $\Delta\Delta G$ for residues with each of these secondary structure and χ_1 combinations. We adjusted the $\Delta\Delta G$ to account for scoring terms that would penalize rotamers with poor intraresidue interactions, since we are interested in the strength of interchain interactions given that a particular rotamer has been adopted. Helical Phe residues had an average adjusted $\Delta\Delta G$ of 3.7 ± 0.06 for *g+*, 1.9 ± 0.03 for *g−*, and 1.3 ± 0.01 for *a* (Figure 2). Strand and loop Phe possessed entirely different signatures from helical Phe, favoring *g−* and equally *g+* and *a* conformations, respectively. Isoleucine's loop preferences resemble phenylalanine's helix preferences, but its helix and strand profiles are unique.
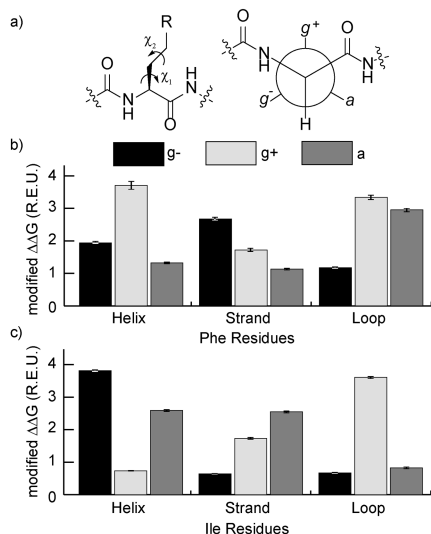
Figure 2. (a) The χ_1 and χ_2 angles of an amino acid side chain describe the relative orientation of N−Cα−Cβ−Xγ and Cα−Cβ−Xγ−Yδ dihedrals. The *g−*, *g+*, and *a* rotamers for χ_1 (i.e., −60°, 60°, and 180°) correspond to the three staggered conformations on a Newman projection of the Cα−Cβ bond. (b,c) The average modified $\Delta\Delta G$ for Phe (b) and Ile (c) residues of a given secondary structure type with a given χ_1 rotamer (in Rosetta energy units (R.E.U.), ~1 kcal/mol). $\Delta\Delta G$ is modified to omit the Dunbrack scoring term that favors certain rotamers.

Further rotamers of particular interest include helical *g+* leucine and *g+* conformations of residues with two or more χ angles on helices in general (Figure S3). These are very difficult to obtain on a helical backbone due to clashes with adjacent helical turns but form remarkably strong interactions when they do appear. In contrast, the single-χ residues have diverse conformations of interest, such as helical valine in any conformation but *a* and *g−* strand serine. Any of these rotamers may be attractive targets for inhibition by "topographical" mimetics, which possess compatible residue positioning but a distinct backbone that may stabilize such rotamers.

Surprisingly many hot spot residues adopt off-rotamer side-chain conformations, which are quite uncommon otherwise.[28] The absolute value of the angular deviation in off-rotamer conformations is often moderate (10−15°), but the energetic difference can be substantial. Rosetta models rotamer energy distributions surrounding ideal conformations as Gaussians with mean and standard deviation fit to PDB statistics; for example, the energy wells for common helical rotamers have standard deviations of 8°. The Dunbrack rotamer energy term (*fa_dun*) therefore penalizes each χ deviation of 10−15° by 1−2 R.E.U.[29] To reach the 1.0 R.E.U. hot spot threshold, an off-rotamer residue must exhibit significant favorable interchain interactions in order to account for that penalty. Nonetheless, we find that leucine and isoleucine appear as hot spot residues at comparable rates whether on or off rotamer (Figure 3; full data in Figure S4).
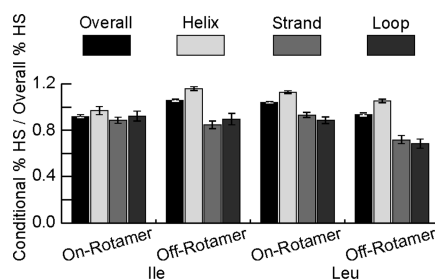
Figure 3. Off-rotamer states (those more than 10° from a *g+*, *g−*, or *a* rotamer well minimum) appear as hot spot (HS) residues at a comparable rate to on-rotamer states, those within 10° of a rotamer well minimum. Measured here is the percentage of residues that are HS under a particular condition (e.g., on-rotamer and helical) over the percentage overall for that amino acid.

The percentage of off-rotamer states that are hot spot residues remains relatively constant and roughly independent of backbone conformation. Such a result is contrary to expectations, as a penalty of even 0.5 R.E.U. would reduce a given leucine's overall likelihood of being a hot spot residue from 28.8% to 15.6%.

This similarity between hot spot residue occurrences for on- and off-rotamer states holds for rotameric χ angles, which govern the rotation about an sp^3−sp^3 σ bond. In contrast, the conformational distribution for an sp^2−sp^3 σ bond, e.g., χ_2 of Phe, is better modeled by a continuous distribution than by a set of rotamer wells (Figure S5). For such "non-rotameric" χ angles, hot spot amino acids tend to enrich already well populated states.

These particular side-chain conformational preferences may make significant contributions to a peptide binding epitope. To analyze this possibility, we examined the LxxLL binding motif, which is commonly observed in interactions between nuclear receptors and coactivators. LxxLL helices contain three leucines at positions *i*, *i* + 3, and *i* + 4, with varying residues at the x positions. While the N- and C-terminal flanking residues and the

residues at the x position can provide selectivity in the interactions of these motifs,[30] the three leucines typically make the highest affinity interactions with the partner protein. We were drawn to LxxLL motifs because the relative leucine sequence and the helical backbone conformation are conserved over a range of protein complexes, allowing an opportunity to isolate the role of side-chain conformations on recognition.

We analyzed and categorized the leucine rotamers in the high-resolution structures of protein interfaces containing the LxxLL helix (Table 1) and found that the side-chain dihedral patterns

Table 1. Conformations of LxxLL Helical Binding Motifs, Represented as $(\chi_1, \chi_2)^a$

conformations of the i, i + 3, and i + 4 leucine residues	corresponding PDB entries
$(g-, a)$; $(g-, a)$; $(g-, a)$	*1bsx*, 1ixm, 2bnx, 2qm4, 3q9d, 3ter, *2prg*, *3h0a*, *4j24*
$(g-, a)$; $(a, g+)$; $(g-, a)$	2qm4, 4mcw
$(a, g+)$; $(g-, a)$; $(g-, a)$	1b9m, *1k4w*, *1n4h*, *1rdt*, *1rjk*, *1ymt*, 2bjn, *2p1t*, 2pv7, 2zla, 3a6m, 3ech, *3gyt*, *3kwy*, *3l3x*, 3nqo, 3nrv, *4dk7*, *4e2j*, *4giz*, *4rwv*
$(a, g+)$; $(g-, a)$; $(a, g+)$	*1ot7*, 3bro
$(a, g+)$; $(a, g+)$; $(g-, a)$	2ip2, 2izx, 2vzg, *2xfx*, 3bdd, 3c7j
$(-150, a)$; $(g-, a)$; $(g-, a)$	*1m2z*, 3gn8
$(-150, -150)$; $(g-, a)$; $(g-, a)$	*1zdt*, 3oll
$(-105, a)$; $(g-, a)$; $(g-, a)$	3g8i, *3l0j*, *4jyg*
off rotamer states[c]	3tos, *2izy*, 1kbh, 1t63, 1t7f, *1zky*, 2a3i, 2q7j, 2xhs, 3a2h, 3bqd, 3fxv, 3hlv, 3k22, 3okh, 3vt3, 4lsj, 4q13, 4qjr

[a]Only one representative example is listed for PDB structures with multiple copies of one LxxLL-containing peptide or for one LxxLL-peptide/receptor complex represented by multiple structures. [b]PDB codes in italics are nuclear receptor/peptide complexes; others are diverse structures that incidentally possess an LxxLL motif. [c]We denote off rotamer states as those in which at least one residue, but often two or three, are 20° or further from a g−, g+, or a rotamer well.
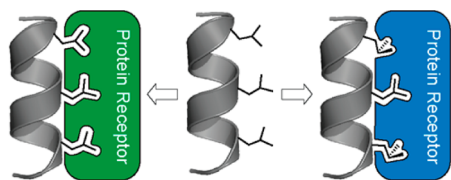

Figure 4. One mechanism for binding promiscuity in LxxLL helices. Helices (gray) with three leucine residues interact with proteins (green, blue) whose interfaces demand different patterns of rotamers.

for the three leucines vary for different complexes (Figure 4). Conformational plasticity is known to be essential for promiscuity in protein and ligand recognition.[19,31] We reasoned that the LxxLL motif demonstrates conformational plasticity at the side-chain rotamer level that enables the motif to recognize this wide array of target proteins: the LxxLL motif is able to bind to different receptors largely because the leucine residues can access different rotamer geometry.[21,32]

Of the above complexes, the majority are formed by nuclear hormone receptors and coactivator peptides. The crystal structures demonstrate the side-chain conformational diversity inherent in the complexes: the off-rotamer conformations are

exclusively the purview of these complexes, and of the 27 observed conformational states, 21 are only observed in receptor/coactivator complexes. These observations support the hypothesis that nuclear receptor specificity is mediated largely by the recognition of specific rotameric states (Figure S6). Structures of the same receptor/peptide complex consistently exhibited the same rotamers, even if they are uncommon conformations (e.g., the 10 complexes of DRIP/vitamin D3 receptor, and the 10 of NRC2/glucocorticoid receptor, feature consistent leucine rotamers). This consistency also suggests that these subtle geometrical distinctions are not due to crystal structure artifacts; the observed side-chain conformations are indeed preferred. Furthermore, the same peptide employs different rotamers when forming complexes with distinct receptors. Figure 5 illustrates the distinct rotameric states that
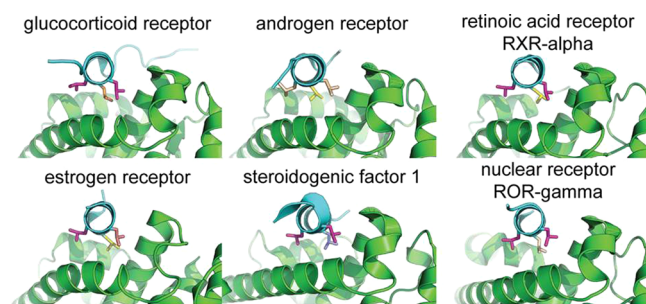
Figure 5. Three leucines in LxxLL helices prefer distinct side-chain geometries in complex with nuclear receptors. PDB codes: 1m2z (GR), 1t63 (AR), 1zky (ER), 1zdt (SF1), 2p1t (RXRα), 3l0j (RORγ).

nuclear receptor coactivator 2 (NRC2) assumes in its complexes with six protein targets. At least one of the three leucine residues differs in rotamer geometry in each complex.

To obtain a quantitative description of the strength of these rotameric preferences, we calculated energetics of three LxxLL motifs in their complexes with three different proteins (2prg, 4giz, 4j24). We applied flexible backbone docking and refinement algorithms using RosettaScripts[33] to optimize the bound conformations (details in SI). For each sequence, we conducted simulations in which we constrained just the leucine side chains of these peptide complexes to their own native values or to the native values of the other LxxLL motifs in question (Figure 6). Each sequence demonstrated a strong preference in binding energy (2−4 R.E.U.) for its own native rotamers, despite conformational similarities: 2prg and 4j24 both have three $(g-, a)$ leucines, while the N-terminal leucine of 4giz is $(a, g+)$. These protein interfaces therefore recognize preferred rotamers with extremely fine detail. In no case were the energies the result of
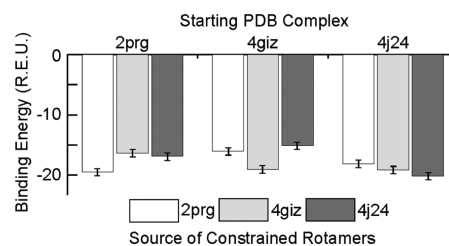
Figure 6. Binding energy scores of peptide−protein complexes simulated with constraints enforcing either the LxxLL motif's native rotamers or those of the peptide from a different complex, measured in R.E.U.

explicit clashes; residues around the constrained side chains were able to repack to accommodate the constraints, but the resulting complexes were energetically inferior.

Beyond biophysical insight, these results suggest concrete technological applications. The fact that LxxLL motifs populate distinct rotamer states to bind to different receptors suggests that amino acids with constrained side chains may deliver highly specific binders. Despite the key biological role of LxxLL motifs in transcription, it has been difficult to produce inhibitors that specifically target the chosen hormone receptor.[34,35] A preorganized side chain might have a lower entropic penalty and therefore better binding affinity. Off-rotamer states, for example, evoke the internal dihedral angles preferred by small alkane rings; residues such as Ile for which off-rotamer states are particularly common may benefit from cyclobutyl or cyclopentyl mimicry. Uncommon rotamers may be obtained by other means. Cyclopropane amino acids developed by Martin,[36] previously applied to Phe, can offer either $g-$ or $g+$ constraints, as can the β-substituted Phe, Trp, and Tyr derivatives developed by Hruby.[37] Amino acid variants with unsaturated side chains have also been described.[38] Further development of such noncanonical amino acids offers a potentially fruitful route for the design of specific PPI inhibitors.

Our investigation into the role of side-chain rotamers on protein−protein complex formation was originally motivated by the finding that the identity of hot spot residues does not directly correlate with backbone ϕ and ψ dihedrals: interfacial α-helices, β-strands and loops all similarly favor aromatic residues as hot spots. We find persuasive support for the hypothesis that side-chain rotamers contribute significantly to the desired level of specificity in protein complex formation. We find that some rotamers contribute appreciably more to binding than others in a residue and secondary structure specific manner. This result suggests that noncanonical amino acid residues with constrained side chains may offer an exciting avenue for new classes of secondary structure mimics as PPI inhibitors. We obtained further support for this hypothesis by analyzing the well-studied LxxLL motif in coactivators of nuclear receptors, demonstrating that the receptors dictate particular sets of rotamers in the peptides that bind them.

■ ASSOCIATED CONTENT

ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.6b04892.

Methods and supplementary figures, including a detailed account of all data analysis (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*arora@nyu.edu
*bonneau@nyu.edu

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Jones, S.; Thornton, J. M. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13.
(2) Sheng, C.; Dong, G.; Miao, Z.; Zhang, W.; Wang, W. *Chem. Soc. Rev.* **2015**, *44*, 8238.
(3) Pelay-Gimeno, M.; Glas, A.; Koch, O.; Grossmann, T. N. *Angew. Chem., Int. Ed.* **2015**, *54*, 8896.
(4) Jayatunga, M. K. P.; Thompson, S.; Hamilton, A. D. *Bioorg. Med. Chem. Lett.* **2014**, *24*, 717.
(5) Arkin, M. R.; Tang, Y.; Wells, J. A. *Chem. Biol.* **2014**, *21*, 1102.
(6) Azzarito, V.; Long, K.; Murphy, N. S.; Wilson, A. J. *Nat. Chem.* **2013**, *5*, 161.
(7) Modell, A. E.; Blosser, S. L.; Arora, P. S. *Trends Pharmacol. Sci.* **2016**, *37*, 702.
(8) Watkins, A. M.; Arora, P. S. *ACS Chem. Biol.* **2014**, *9*, 1747.
(9) Bullock, B. N.; Jochim, A. L.; Arora, P. S. *J. Am. Chem. Soc.* **2011**, *133*, 14220.
(10) Jochim, A. L.; Arora, P. S. *Mol. BioSyst.* **2009**, *5*, 924.
(11) Gavenonis, J.; Sheneman, B. A.; Siegert, T. R.; Eshelman, M. R.; Kritzer, J. A. *Nat. Chem. Biol.* **2014**, *10*, 716.
(12) Cunningham, B. C.; Wells, J. A. *Science* **1989**, *244*, 1081.
(13) Kortemme, T.; Kim, D. E.; Baker, D. *Sci. Signaling* **2004**, *2004*, pl2.
(14) Bogan, A. A.; Thorn, K. S. *J. Mol. Biol.* **1998**, *280*, 1.
(15) Lo Conte, L.; Chothia, C.; Janin, J. *J. Mol. Biol.* **1999**, *285*, 2177.
(16) Moreira, I. S.; Fernandes, P. A.; Ramos, M. J. *Proteins: Struct., Funct., Genet.* **2007**, *68*, 803.
(17) Ferrante, A. *Immunol. Res.* **2013**, *56*, 85.
(18) Dunbrack, R. L., Jr.; Karplus, M. *Nat. Struct. Biol.* **1994**, *1*, 334.
(19) Gaudreault, F.; Chartier, M.; Najmanovich, R. *Bioinformatics* **2012**, *28*, i423.
(20) Guharoy, M.; Janin, J.; Robert, C. H. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 3219.
(21) Kirys, T.; Ruvinsky, A. M.; Tuzikov, A. V.; Vakser, I. A. *Proteins: Struct., Funct., Genet.* **2012**, *80*, 2089.
(22) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2000**, *40*, 389.
(23) Scouras, A. D.; Daggett, V. *Protein Sci.* **2011**, *20*, 341.
(24) Hintze, B. J.; Lewis, S. M.; Richardson, J. S.; Richardson, D. C. *Proteins: Struct., Funct., Genet.* **2016**, DOI: 10.1002/prot.25039.
(25) Shapovalov, M. V.; Dunbrack, R. L., Jr. *Structure* **2011**, *19*, 844.
(26) Xiang, Z.; Honig, B. *J. Mol. Biol.* **2001**, *311*, 421.
(27) Taghizadeh, M.; Goliaei, B.; Madadkar-Sobhani, A. *Mol. BioSyst.* **2015**, *11*, 2000.
(28) Petrella, R. J.; Karplus, M. *J. Mol. Biol.* **2001**, *312*, 1161.
(29) 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. *Methods Enzymol.* **2013**, *523*, 109.
(30) Plevin, M. J.; Mills, M. M.; Ikura, M. *Trends Biochem. Sci.* **2005**, *30*, 66.
(31) James, L. C.; Roversi, P.; Tawfik, D. S. *Science* **2003**, *299*, 1362.
(32) Ruvinsky, A. M.; Kirys, T.; Tuzikov, A. V.; Vakser, I. A. *J. Mol. Biol.* **2011**, *408*, 356.
(33) Raveh, B.; London, N.; Zimmerman, L.; Schueler-Furman, O. *PLoS One* **2011**, *6*, e18934.
(34) Ravindranathan, P.; Lee, T. K.; Yang, L.; Centenera, M. M.; Butler, L.; Tilley, W. D.; Hsieh, J. T.; Ahn, J. M.; Raj, G. V. *Nat. Commun.* **2013**, *4*, 1923.
(35) Phillips, C.; Roberts, L. R.; Schade, M.; Bazin, R.; Bent, A.; Davies, N. L.; Moore, R.; Pannifer, A. D.; Pickford, A. R.; Prior, S. H.; Read, C. M.; Scott, A.; Brown, D. G.; Xu, B.; Irving, S. L. *J. Am. Chem. Soc.* **2011**, *133*, 9696.
(36) Martin, S. F.; Dorsey, G. O.; Gane, T.; Hillier, M. C.; Kessler, H.; Baur, M.; Matha, B.; Erickson, J. W.; Bhat, T. N.; Munshi, S.; Gulnik, S. V.; Topol, I. A. *J. Med. Chem.* **1998**, *41*, 1581.
(37) Hruby, V. J.; Li, G.; Haskell-Luevano, C.; Shenderovich, M. *Biopolymers* **1997**, *43*, 219.
(38) Jiang, J.; Ma, Z.; Castle, S. L. *Tetrahedron* **2015**, *71*, 5431.