

Expanding Molecular Modeling and Design Tools to Non-Natural Sidechains

David Gfeller,^[a] Olivier Michielin,^{*[a,b,c]} and Vincent Zoete^{*[a]}

Protein–protein interactions encode the wiring diagram of cellular signaling pathways and their deregulations underlie a variety of diseases, such as cancer. Inhibiting protein–protein interactions with peptide derivatives is a promising way to develop new biological and therapeutic tools. Here, we develop a general framework to computationally handle hundreds of non-natural amino acid sidechains and predict the effect of inserting them into peptides or proteins. We first generate all structural files (pdb and mol2), as well as parameters and topologies for standard molecular mechanics software (CHARMM and Gromacs). Accurate predictions of rotamer probabilities are provided using a novel combined knowledge and physics based strategy. Non-natural sidechains are useful to

increase peptide ligand binding affinity. Our results obtained on non-natural mutants of a BCL9 peptide targeting beta-catenin show very good correlation between predicted and experimental binding free-energies, indicating that such predictions can be used to design new inhibitors. Data generated in this work, as well as PyMOL and UCSF Chimera plug-ins for user-friendly visualization of non-natural sidechains, are all available at <http://www.swisssidechain.ch>. Our results enable researchers to rapidly and efficiently work with hundreds of non-natural sidechains. © 2012 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.22982

Introduction

Protein–protein interactions are fundamental to most biological and biochemical processes. However, targeting protein–protein interactions to develop therapeutics has remained a significant challenge in recent years.^[1] This can be partly understood by the relatively flat and surface exposed binding interfaces with distant hotspots that are not very well suited for standard small molecule high-throughput screening experiments. Recent experimental advances have provided an ever-increasing amount of information about protein–protein interactions, in particular interactions mediated by short peptides binding to a protein, as often found in signaling pathways.^[2] Structural studies have enabled characterizing a very large number of these complexes.^[3] Moreover, high-throughput techniques, such as peptide arrays,^[4] phage-display,^[5] or ribosome-display,^[6] reveal unprecedented insights into binding specificity.^[7] However, the use of natural peptides as inhibitors is often restricted by their sensitivity to protease-mediated degradation. In addition, several naturally occurring interactions (either between proteins of the same organism, or between host and pathogens) have already been optimized along evolution, suggesting that peptides restricted to the set of 20 natural amino acids will hardly be appropriate for inhibiting or competing with these interactions.

Peptido-mimetics, and especially incorporation of non-natural sidechains, is a promising strategy to harness the current wealth of data about natural peptides towards the development of more potent inhibitors.^[8] Some remarkable experimental advances to genetically encode non-natural sidechains,^[9] such as the use of amber codon,^[10] tRNA acylation,^[11] or engineered quadruplet-decoding ribosomes^[12]

are currently expanding peptide screening technologies to include non-natural sidechains. Correct interpretation of these results at the atomic level requires structural and biochemical information about these sidechains, such as rotamers. Moreover, as the number of possible non-natural sidechains is huge, rational structure-based design of non-natural peptides provides a powerful alternative to high-throughput screening.^[8] As such, computer-aided and modeling strategies are promising tools to help narrowing-down the list of ligands to be experimentally tested. *In silico* approaches are also particularly appropriate for non-natural sidechains insertion, since this corresponds to a relatively small extrapolation from experimental structural data. In particular, mutating one natural amino acid to a non-natural one is likely to leave unchanged the binding mode of the rest of the peptide.

Most existing molecular modeling software, such as FoldX^[13] or TINKER,^[14] as well as molecular dynamics (MD) software, such as Gromacs^[15] or CHARMM,^[16] only contain a limited set of amino acid sidechains in addition to the 20 natural ones

This article was published online on 14 April 2012. An error was subsequently identified. This notice is included in the online and print versions to indicate that both have been corrected 24 April 2012.

[a] D. Gfeller, O. Michielin, V. Zoete
Swiss Institute of Bioinformatics (SIB), Quartier Sorge, Bâtiment Génopode,
CH-1015 Lausanne, Switzerland
E-mail: olivier.michielin@unil.ch, vincent.zoete@unil.ch

[b] O. Michielin
Ludwig Institute for Cancer Research, Centre Hospitalier Universitaire
Vaudois, Lausanne, Switzerland

[c] O. Michielin
Pluridisciplinary Center for Clinical Oncology (CePO), Centre Hospitalier
Universitaire Vaudois, Lausanne, Switzerland

© 2012 Wiley Periodicals, Inc.

(typically consisting of phosphorylated or methylated residues, as well as a few others). To bridge this gap and expand the sidechain chemical alphabet that can be used in molecular modeling and *in silico* drug design studies, we built structural files (mol2, pdb) for 209 non-natural sidechains. Parameter and topology files to run standard MD simulations with these non-natural sidechains in CHARMM and Gromacs have been generated using direct mapping from existing parameters supplemented by data retrieved from the SwissParam^[18] web service when necessary. Rotamer probabilities are predicted using a novel algorithm combining energy calculations with statistical analysis of experimental data. We show that accurate binding free-energy predictions can be obtained with our data for non-natural sidechains. These results are made available at <http://www.swisssidechain.ch> and the non-natural sidechains can be inserted into peptide or protein structures and visualized using our PyMOL and UCSF Chimera^[19] plug-ins.

Results and Discussion

Building non-natural sidechains

The chemical alphabet for non-natural amino acid sidechains is nearly infinite. Here, to maximize the applicability of the results, we focused on amino acid sidechains with structural information in the Protein Data Bank (PDB), as well as commercially available ones. Non-natural amino acids that modify the backbone, such as β -homo, cyclic or aromatic backbones, or proline derivatives, were not included, since they are more likely to perturb the overall conformation of peptides or proteins and are therefore less amenable to molecular modeling studies. This resulted in a total of 209 non-natural sidechains, with 141 being present in the PDB. For the latter, pdb files were downloaded directly from this database. For the rest we used the tools of ChemAxon (MarvinSketch) and the UCSF Chimera modeling software.^[19] Mol2 and smiles files were generated with OpenBabel. We next generated parameter and topology files for CHARMM and Gromacs. For 33% of the sidechains, topology files, especially partial charges, could be readily generated by analogy with natural sidechains. Examples of such cases include for instance allo-threonine, or ornithine (ORN), a shorter lysine, where partial charges of THR and LYS, respectively, could be directly mapped. For other sidechains, we used the SwissParam web service^[18] to retrieve both partial charges and missing force-field parameters (see Material and Methods).

Rotamer predictions

Amino acid sidechains are highly flexible and can adopt several different conformations. In practice, it is known that some conformations are more likely than others. For instance, dihedral angles along carbon chains are known to adopt preferentially values peaked around -60 , 60 , and 180° , referred to as rotamers. VAL, for example, has three different rotamers, while LEU has 3×3 since it contains two freely rotating dihedral angles with three preferred conformations each. The probability of each rotamer depends on the nature of the chemical bonds and interactions with other atoms of the sidechain,

interactions with the backbone of the protein, and interactions with other atoms found in the vicinity of the sidechain. Backbone dependent rotamer libraries capture the former two aspects by providing rotamer probability distributions for each value of backbone dihedral angles ϕ and ψ (typically using a $10^\circ \times 10^\circ$ grid on these angles).^[20] Backbone independent rotamer libraries instead provide rotamer probabilities without incorporating dependencies on backbone dihedral angles. Such data are crucial to select conformations that are physically reasonable when refining experimental structures or mutating *in silico* a sidechain.^[21,22] For natural sidechains, rotamers are computed by running statistics on existing structures in the PDB.^[20,23–25] In the absence of sufficient experimental data for non-natural sidechains, we designed a novel strategy to predict rotamer probabilities, combining physical and statistical approaches. We first validate this strategy on natural sidechains, and then use it to predict rotamer probabilities of our non-natural sidechains.

Physics-based approach

Rotamer probabilities have often been predicted using molecular mechanics or quantum mechanics calculations.^[26–28] Here, we rely on MD simulations. A tri-alanine peptide was used, where the second residue is mutated to all other sidechains (see Material and Methods). Simulations were carried out for a total of 200ns per sidechain with CHARMM^[16] using the FACTS implicit solvent model.^[29] Rotamer probabilities are estimated as frequencies of conformations corresponding to each rotamer along the trajectories (see Material and Methods). This way of computing rotamer probabilities is similar to the way experimental rotamer probabilities are computed (the only difference being that the average is done over all existing structures, rather than over a trajectory). Similar statistics as in Ref. [23] have been applied to smoothen and extrapolate probabilities to low-sampling regions (see Material and Methods).

To assess the capability of computed rotamer probabilities to reproduce experimental ones, we first tested our approach on all natural sidechains (expect ALA, GLY, and PRO). For residues such as PHE, ILE, LEU, or VAL, good correlations are obtained between predicted and experimental rotamer probabilities (see Fig. 1, gray bars). However, for many other sidechains, such as GLU, poor correlations are observed. We also attempted to predict rotamer probabilities by scanning all possible values for sidechain and backbone dihedral angles (using grids of 10° on each dihedral angle) and computing the energy (see Material and Methods). Using the FACTS solvent model to estimate energies of these conformations, similar results are obtained as in the case of MD-based probabilities, while *in vacuo* energy calculations give much worse results for all sidechains (see Supporting Information Fig. S1). It should also be noted that for long sidechains, energy calculations tend to become quite expensive. Even for sidechains with only two dihedral angles, more than a million (36^4) minimizations and energy calculations need to be carried out, while for sidechains such as ARG, this number exceeds 2×10^9 if a 10° degree grid is defined on each dihedral angle. To investigate

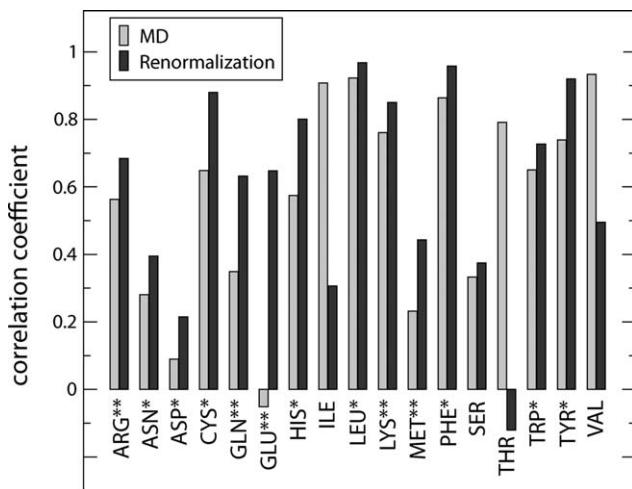


Figure 1. Cross-validation analysis of rotamer predictions. The gray bars show the correlation between experimental and predicted rotamer probabilities using MD simulations. Dark bars show the correlation between experimental and predicted rotamers after renormalization of the first dihedral angles. * indicates sidechains satisfying the first dihedral angle renormalization criterion. ** indicates sidechains satisfying the first two dihedral angles renormalization criterion.

the effect of the choice of the solvent model we performed MD simulations with the GBMV2 model^[30] for natural sidechains. Although the latter is 10 times more computationally demanding, overall, similar results are obtained. The only exception is for GLU which is better predicted with GBMV2 before renormalization (see below) but similarly after renormalization (see Supporting Information Fig. S2). Because of the large number of non-natural sidechains to handle, we used FACTS for the rest of this study.

In general better agreement with experimental data is observed for hydrophobic sidechains compared to polar sidechains. This is likely so because polar sidechains are more sensitive to long-range interactions, which are more difficult to accurately sample in simulations or energy calculations. Moreover, the effect of the solvent is more difficult to reliably estimate for polar sidechains compared to hydrophobic ones. Overall, it appears that purely physics-based approaches using either MD simulations or regular sampling of dihedral angles perform well on some sidechains, such as VAL or ILE, but fail to accurately predict other experimentally observed rotamer probabilities.

Knowledge-based approach

To improve on these results, we observed that several properties of rotamer libraries derived from experimental data are conserved among different natural amino acid sidechains. First, similar sidechains have very similar rotamers. This is most striking when comparing PHE and TYR, where a correlation coefficient of 0.98 is found between their experimental rotamer probabilities. Such a high correlation was never obtained between our calculations and the experimental rotamers, neither for PHE nor for TYR (see Fig. 1, gray bars). This observation suggests that simple derivatives of PHE are better predicted by mapping the rotamers of PHE, rather than trying to

compute them from MD simulations or energy calculations. Less expectedly, we also found a remarkable conservation among the experimental probabilities $P_{\text{exp}}(r_1)$ of the rotameric state of the first dihedral angle χ_1 for all sidechains starting with a linear three-carbon chain $C\alpha-C\beta-C\gamma$ aside from ASP and ASN (i.e., ARG, GLN, GLU, HIS, LEU, LYS, MET, PHE, TYR, TRP), and for CYS (see Fig. 2A). For instance, the average correlation coefficient of $P_{\text{exp}}(r_1)$ between these 11 sidechains is 0.85 in backbone dependent rotamer libraries. Similar results are obtained for backbone independent rotamer libraries (Fig. 2B). Moreover, the rotamer probabilities $P_{\text{exp}}(r_1, r_2)$ for the rotameric state of χ_1 and χ_2 are very well conserved for longer sidechains (ARG, GLN, GLU, LYS and MET), as shown in Figures 2C and 2D. These observations indicate that $P_{\text{exp}}(r_1)$ or $P_{\text{exp}}(r_1, r_2)$ are conserved between many different sidechains and could be used to help computing rotamer probabilities of new sidechains. In this work, we use the average of $P_{\text{exp}}(r_1)$ over ARG, CYS, GLU, GLN, HIS, LEU, LYS, MET, PHE, TYR, and TRP (referred to as $P_{\text{av}}(r_1)$) and we renormalize with this quantity the rotamer probabilities computed along MD trajectories for sidechains starting with a linear chain of three carbon or sulfur atoms (see Materials and Methods). For instance, for a sidechain with two flexible dihedral angles, the final probability is computed as $P_{\text{final}}(r_1, r_2) = P_{\text{av}}(r_1) P(r_1, r_2)/P(r_1)$, where $P(r_1, r_2)$ is the probability of the rotameric state of the two dihedral angles based on MD simulations and $P(r_1)$ is the probability of the rotameric state of first dihedral angle in the MD simulation. Similarly, we used the average of $P_{\text{exp}}(r_1, r_2)$ over ARG, GLU, GLN, LYS and MET (referred to as $P_{\text{av}}(r_1, r_2)$) to renormalize the computed rotamer probabilities of sidechains starting with a linear chain of four carbon or sulfur atoms. The cross-validation analysis (see Material and Methods) of Figure 1 (dark bars) shows a very clear improvement after renormalization. For all natural sidechains without branching at $C\beta$ (and excluding SER as well) the renormalized rotamer probabilities correlate better with experimental ones (average correlation coefficient increasing from 0.49 to 0.72). Reversely, for most sidechains that do not start with a linear chain of three carbon or sulfur atoms, and especially for the ones with a branching at $C\beta$ (ILE, THR, VAL), renormalizing rotamer probabilities leads to very poor correlations with experimental data (Fig. 1). Similar results are obtained for backbone independent rotamer predictions (Supporting Information Fig. S3) or when focusing only on (ϕ, ψ) dihedral angles corresponding to alpha helices or beta strands (Supporting Information Fig. S4).

This enables us to define renormalization criteria that are summarized in Figure 3. First, for simple derivatives of natural sidechains without additional dihedral angles, $P_{\text{exp}}(r_1, \dots, r_n)$ of the corresponding natural sidechain can be simply mapped. Second, for derivatives of natural sidechains with additional dihedral angles (e.g., 2-amino-4-ethyl sulfanyl butyric acid) the dihedral angles shared with the natural sidechain are renormalized by rotamer probabilities of the corresponding natural sidechain. Third, rotamer probabilities of sidechains starting with a linear chain $C\alpha-C\beta-C\gamma/S\gamma$ are renormalized by $P_{\text{av}}(r_1)$. Fourth, rotamer probabilities of sidechains starting with a linear chain $C\alpha-C\beta-C\gamma-C\delta/S\delta$ are renormalized by $P_{\text{av}}(r_1, r_2)$.

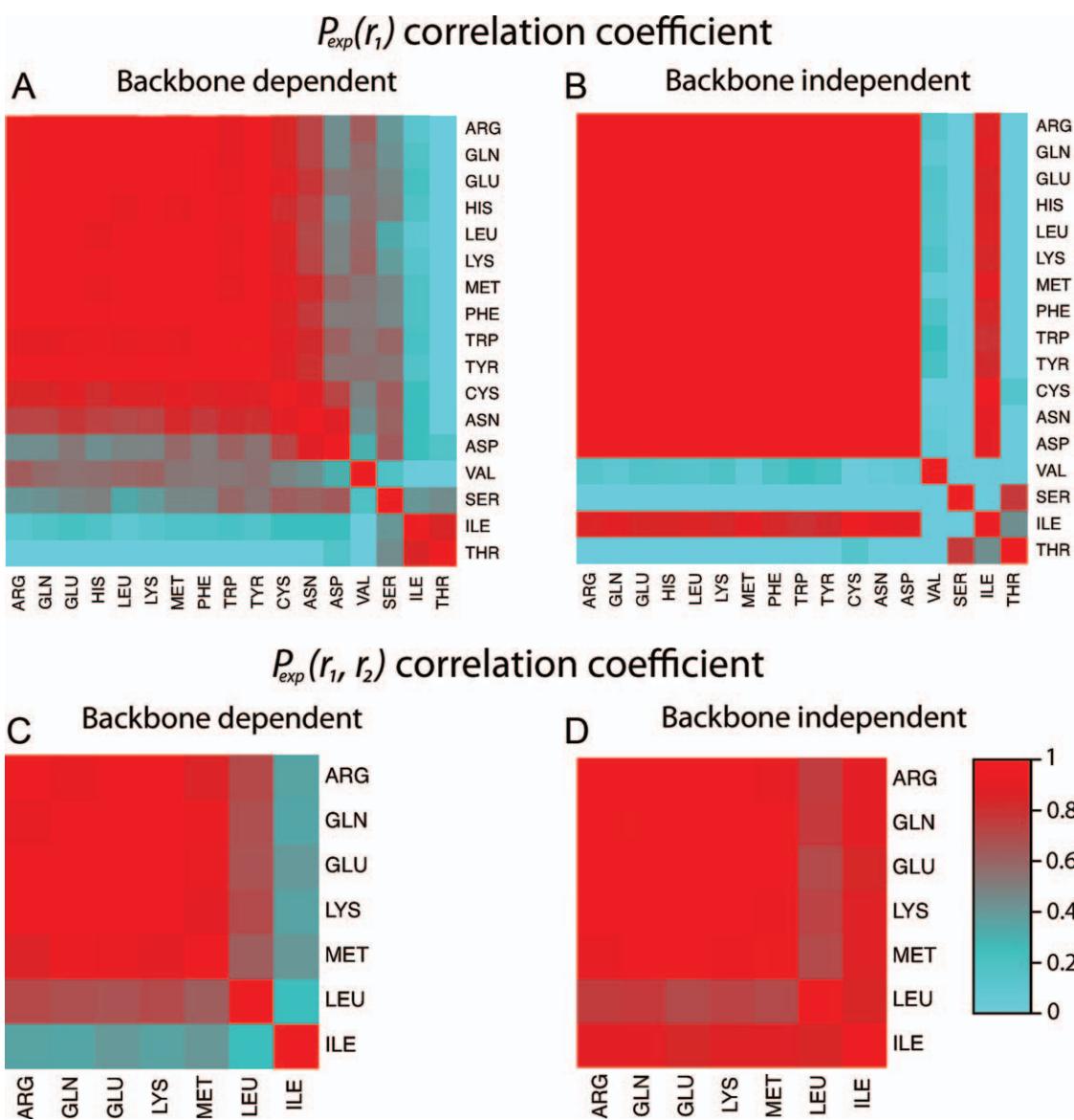


Figure 2. Conservation of the first dihedral angle distribution in natural sidechains. a) and b): heatmap of correlation coefficients between experimental rotamer probabilities $P_{\text{exp}}(r_1)$ of the first dihedral angle for natural sidechains, both for backbone dependent a) and backbone independent b) rotamer libraries. c) and d): heatmap of correlation coefficients between rotamer probabilities $P_{\text{exp}}(r_1, r_2)$ of the first two dihedral angles for the seven sidechains with dihedral angles χ_1 and χ_2 of the same type (i.e., with three main energy minima at $-60, 60$, and 180).

We note that some of the discrepancies between rotamer probabilities of natural sidechains computed from MD simulations or energy calculations and experimental ones might be explained by biases in X-ray structures due to secondary structure elements. For instance, alpha-helices are known to create restraints on some sidechain conformations compared to small unstructured peptides.^[27] We do not exclude that one of the effects of the renormalization scheme is to introduce some of these biases into our computed rotamer probabilities. As our goal is not to extensively analyze the conformational space of free sidechains on model systems such di- or tri-alanine, but rather to provide accurate estimates of the most likely conformations in the context of proteins, incorporating these biases proves to be a useful feature.

A more fine-grained version of natural sidechain rotamer libraries was recently released.^[25] We observe similar cross-validation performance of our predictions for natural sidechains with this library (see Supporting Information Fig. S5). However the additional details of this new library give rise to a much larger number of parameters that need to be inferred, resulting in higher risks of inaccuracies when predicting rotamer probabilities. In addition, this generates several Gigabytes of data that become hardly tractable for standard visualization tools such as PyMOL or UCSF Chimera^[19] when dealing with hundreds of sidechains. For this reason, we focused on the 2002 version of the Dunbrack rotamer library, which is also the most widely used in visualization and modeling software.

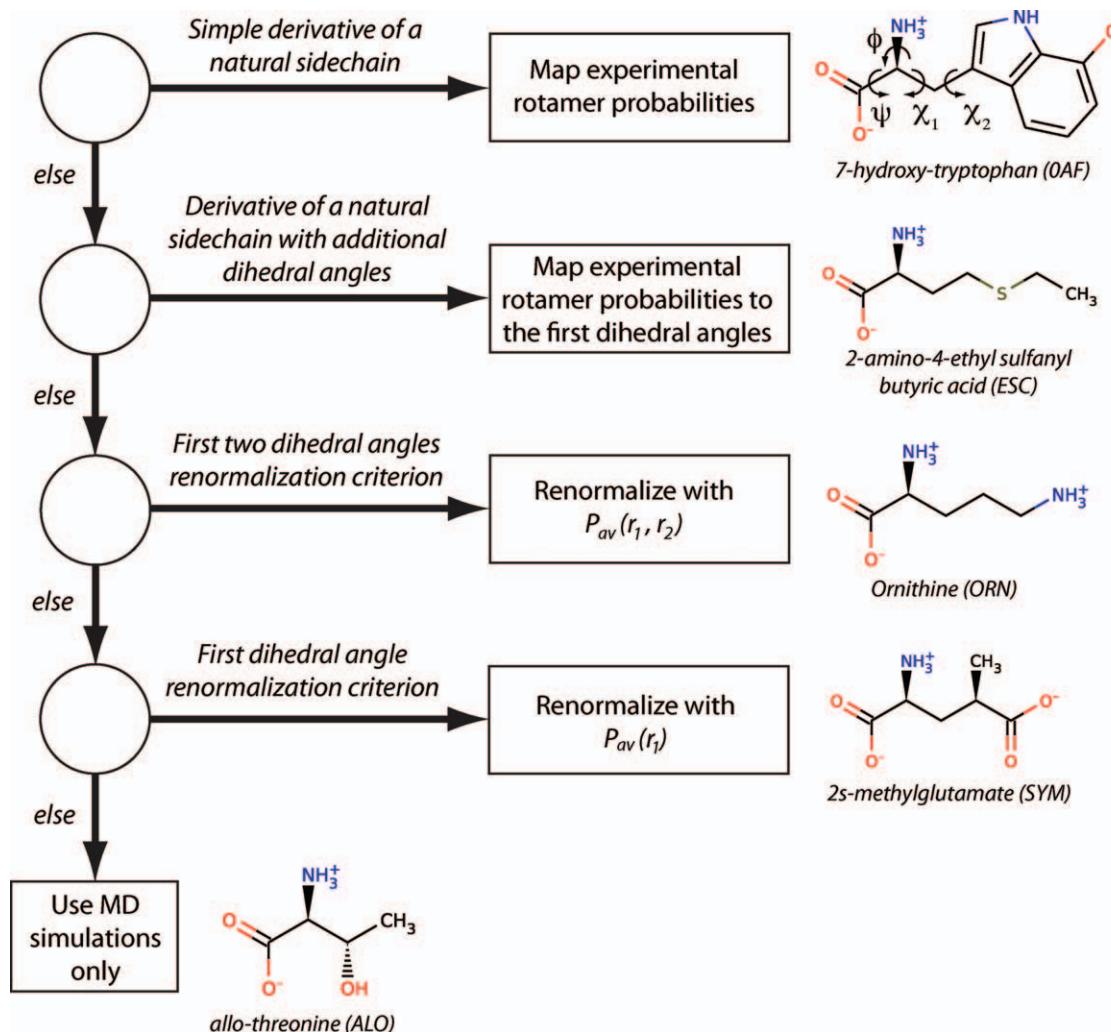


Figure 3. Decision diagram for renormalizing rotamer probabilities. For each possible output, an example of non-natural sidechain is displayed for illustration with its full name and the PDB code in parentheses. The first 2D structure recalls how dihedral angles are defined. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Rotamer predictions for non-natural sidechains

We used the combined physics and knowledge-based strategy to predict the rotamer probabilities of all 209 non-natural sidechains studied in this work. For simple derivatives of natural sidechains, such as hydroxy-tryptophan or chloro-phenylalanine, without any additional free dihedral angle, we simply mapped the rotamer probabilities of the corresponding natural sidechains (see Material and Methods). In total, rotamer probabilities could be mapped in this way for more than one third of the non-natural sidechains studied in this work (see Supporting Information Table S1). All other non-natural sidechains were inserted into a tri-alanine peptide with neutral termini and subjected to a total of 200ns MD simulations each. For sidechains satisfying our renormalization criteria (see Supporting Information Fig. 3 and Table S1), the computed rotamer probabilities were then renormalized as described above.

To validate our predictions for non-natural sidechains, we compared the predicted rotamers to structural data available

for non-natural sidechains in the PDB. First, for each occurrence of non-natural sidechains in experimental X-ray structures, we retrieved its rotamer and computed the rank of this rotamer based on our backbone dependent predicted probabilities (backbone independent rotamer libraries where used for N- or C-terminal amino acids). Ranks were normalized between 0 and 1, and the distribution of ranks was computed over the whole PDB. Figure 4 shows that the vast majority of experimental conformations correspond to rotamers with high ranks ($p < 10^{-15}$, one-sided t -test). For instance, 27% of the non-natural sidechains in these structural data adopt a conformation corresponding to the best predicted rotamer and 85% of them adopt a conformation ranking among the top 50% of the predicted rotamers.

Second, we predicted *in silico* the experimental conformation of these sidechains in the context of the protein crystal structure. Toward this goal we used our UCSF Chimera plug-in. Each occurrence of a non-natural sidechain in a crystal structure of the PDB (see Supporting Information Table S4 for the

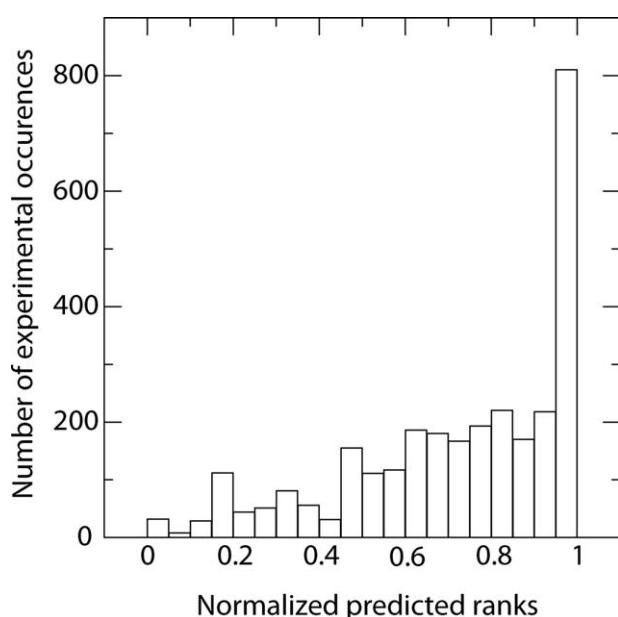


Figure 4. Rotamer distribution in existing structural data of non-natural sidechains. The probability and the corresponding rank of each experimental structural conformation of all non-natural sidechains was computed based on the predicted backbone dependent rotamer probabilities. The histogram shows the distribution of normalized ranks for these experimental structures. Clearly, experimental structures are given a high rank, corresponding to a good probability, in our predicted rotamer libraries.

complete list of structures) was mutated to itself and the predicted conformation was automatically generated with the UCSF Chimera^[19] software in the same way as for natural sidechains. The test was conducted for 3381 occurrences of 111 non-natural sidechains in 1440 different protein structures. There, conformations were selected according to the following default criteria of UCSF chimera (in this order): lowest number of clashes, highest number of H-bonds and highest rotamer probability. Longer sidechains are notoriously more difficult to predict than shorter ones, due to the large number of degrees of freedom.^[22] To quantify the accuracy of the predictions, we computed for each non-natural sidechain with experimental structural data the fraction of correctly predicted rotameric state up the *i*th dihedral angle (N_i). For instance, N_2 for ORN stands for the fraction of ORN residues for which the rotameric state of the first two dihedral angles was correctly predicted. The final score is defined as $S = \frac{1}{C} \sum_{i=1}^C N_i$ (see Ref. ^[22] for a related approach), where C stands for the total number of free dihedral angles considered to define rotamers. Table 1 shows the average of the S score and N_i values over all non-natural sidechains with the same number of dihedral angles (the raw data for each non-natural sidechain can be found in Supporting Information Table S4). Despite the relatively simple approach used here to place the sidechains, compared to more advanced algorithms such as SCWRL4,^[22] we obtain a reasonable degree of accuracy. For instance, more than half of the sidechains with two dihedral angles could be accurately placed (see Table 1), although most of them have between six and nine different possible rotameric states.

We also attempted to compare our results with the very few publicly available data for rotamer libraries of non-natural sidechains. The latest release of Rosetta^[17] (release 3.3) currently includes rotamers for Norvaline (NVA) and Homoleucine. Sufficient structural data are only available for the first one, so we used the rotamers provided in Rosetta with the selection protocol of UCSF Chimera to predict the conformation of NVA. Slightly better results were obtained with our rotamer library ($S = 0.85$ with our data and $S = 0.77$ with the rotamer library provided in Rosetta), although the limited testing set (13 occurrences in nine structures) does not allow us to draw general conclusions. Moreover, when directly comparing the two rotamer libraries, a correlation coefficients of 0.65 is obtained for the rotamer probabilities over all ϕ and ψ angles. This correlation rises to 0.8 when restricting the comparison to alpha-helix and beta-sheet regions of the Ramachandran (see Materials and Methods).

These results confirm that our predicted rotamers correspond to reasonable values and can be used to accurately explore *in silico* the different conformations of non-natural sidechains.

Binding free-energy estimation for non-natural sidechains

We next tested the ability of our non-natural sidechains parameterization to correctly reproduce experimentally measured binding free energy differences. As a benchmark, we used a recent study by Kawamoto et al.^[31] of the interaction between BCL9 and beta-catenin, where 12 mutants with non-natural sidechains have been tested (Table 2). BCL9 is known to bind beta-catenin on a different binding site than the one targeted by E-cadherin or TCF4, and this interaction is critical for beta-catenin transcriptional activity. Using the crystal structure of the complex between beta-catenin and a peptide derived from BLC9^[32] (PDB code: 2GL7), we computationally inserted the mutations tested in Ref. ^[31] with the help of our PyMOL plug-in. We then estimated binding free energy differences between wild-type and mutants with the MM-GBSA algorithm^[33] starting from different rotamers (see Material and Methods). Our results display a good correlation between predicted and experimental binding free energies ($r = 0.75$ when considering only the rotamer giving lowest predicted binding free-energy and $r = 0.78$ when averaging over all rotamers, see Fig. 5A and Table 2). This correlation is similar to the typical values obtained when estimating binding free-energies of natural amino acid mutants.^[33] In particular, the best mutation, corresponding to PHE374 mutated to 2-Naphthalene (NAL), is correctly predicted (residue numbering follows the one in the PDB structure 2GL7). Simple mutation of this residue to NAL without allowing structural rearrangements generated some minor clashes with the backbone carbonyl group of GLN177. However, along the simulation, these residues rapidly find new conformations that resolve all clashes, without affecting any of the other important interactions. The corresponding predicted binding mode is displayed in Figure 5B. The increase in affinity can be understood by the additional hydrophobic contacts of the sidechain of NAL compared to the wild-type

Table 1. Evaluation of correctly placed non-natural sidechains in X-ray data.

Number of free dihedral angles	Number of non-natural sidechains	Total number of occurrences in X-ray data	$\langle S \rangle$	$\langle N_1 \rangle$	$\langle N_2 \rangle$	$\langle N_3 \rangle$	$\langle N_4 \rangle$
1	9	295	0.417	0.417			
2	60	593	0.717	0.865	0.568		
3	25	1825	0.548	0.82	0.482	0.341	
4	17	668	0.464	0.775	0.611	0.298	0.172

The averages are done over all sidechains with a given number of dihedral angles (the total number of such sidechains is indicated in column 2). Column 3 shows the total number of mutations tested in this work. Raw data for each sidechain can be found in Supporting Information Table S4.

Table 2. List of all BCL9 non-natural mutants plotted in Figure 5.

Mutants	Name	Ki[μM]	RT*ln(Ki/Ki _{wt}) [kcal/mol]	Predicted ΔΔG [kcal/mol] (rotamer with lowest energy)	Predicted ΔΔG [kcal/mol] (Boltzmann average over all rotamers)
Wild-type		28.8	0	0	0.00
L366HLEU	Homoleucine	259	1.31	2.64	4.43
L366NLE	Norleucine	180	1.09	2.65	4.57
L366BUG	Tertleucine	400	1.57	15.64	15.64
I369HLEU	Homoleucine	20	-0.22	1.84	3.61
I369NLE	Norleucine	44	0.25	1.14	3.08
I369BUG	Tertleucine	122	0.86	7.88	7.88
L373HLEU	Homoleucine	66	0.49	0.11	2.05
L373NLE	Norleucine	88	0.67	0.03	1.91
L373BUG	Tertleucine	133	0.91	7	7.00
F374HPE	Homophenyl-alanine	54	0.37	0.88	2.27
F374ALN	1-naphthyl-alanine	16.8	-0.32	-0.79	0.67
F374NAL	2-naphthyl-alanine	8.41	-0.73	-4.21	-2.74

Ki values correspond to the affinities measured in Ref. [31]. Positions are numbered according to the PDB structure 2GL7.
[Corrections made here after initial online publication.]

phenylalanine. Beta-catenin – BCL9 interaction further provides an interesting example of the use of non-natural sidechains, since none of the 50 experimentally tested mutations to natural sidechains significantly improved binding affinity.^[31] Moreover, the sequence of BCL9 peptide, and especially residues on the binding interface, is well conserved in distant organisms.^[32] This suggests that the wild-type is already quite optimized and only incorporation of novel chemical building blocks, here non-natural sidechains, could offer room for affinity improvement.

Conclusion

Expanding structural and molecular modeling tools to non-natural sidechains is key both to interpret experimental data with these sidechains, as well as to make de novo structure-based predictions. Our work aims to fill the existing lack of computational tools that enable representing and analyzing non-natural sidechains incorporated into peptides or proteins.

To explore the dynamical aspects of non-natural sidechains, we have generated parameters for CHARMM and Gromacs modeling software. Since non-natural sidechains are mostly used to insert them into existing polypeptide chains, we tried as much as possible to directly infer topologies and parameters from the CHARMM protein force field to maximize compatibility with the rest of the protein. For the remaining parameters, we used the SwissParam web service^[18] that is based on MMFF force field.^[34] As such, we stress that our data

are mostly designed first for visual inspection of non-natural sidechain mutants, including most favorable rotamers, as well as short simulations used for instance to better sample the structural environment when estimating binding free-energy differences. This kind of parameterization was previously shown to provide accurate results with small-molecules in complex with proteins.^[18] More extensive simulations, such as analyzing the folding of polypeptides, might be more sensitive to the choices of parameters. Thus, we cannot exclude that in some of these cases combining parameters from the CHARMM force field together with MMFF might prove less accurate. Nevertheless, our results on beta-catenin–BCL9 interaction show that accurate binding free-energy predictions can be achieved and that the results of such calculations can be useful to structurally interpret experimental data.

Our novel strategy to predict rotamer probabilities combining physics-based and knowledge-based analysis enabled us to generate rotamer libraries for non-natural sidechains that can be used by standard visualization tools. Our method is computationally very efficient compared with more detailed energy calculations,^[27,28] and therefore allowed us to predict rotamer probabilities for hundreds of non-natural sidechains. Moreover, it is particularly appealing for long sidechains, which mostly benefit from the renormalization scheme and are more difficult to handle with purely physics-based approaches due to the large number of degrees of freedom. To enable user-friendly visualization of these new sidechains, we have developed plug-ins for PyMOL and UCSF Chimera that can be

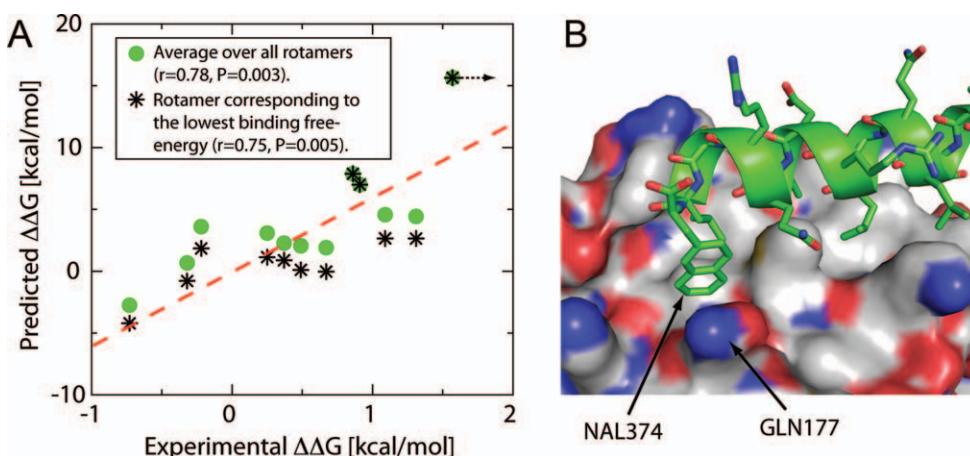


Figure 5. a) Binding free energy predictions for BCL9 non-natural mutants interacting with beta-catenin. $\Delta\Delta G$ predicted with the MM-GBSA algorithm^[33] for the 12 non-natural mutants studied in Ref. ^[31] are compared with their experimentally measured $\Delta\Delta G$. Black stars indicate the results for the rotamer with the lowest predicted binding free energy. Green circles show the Boltzmann average over all rotamers. The arrow indicates that only the lower bound for experimental binding affinity was measured. b) Snapshot of the stable conformation reached in the simulation for PHE374NAL mutant.

downloaded at <http://www.swisssidechain.ch>. In particular, users can toggle between different rotamers to evaluate the most favorable ones and minimize steric clashes.

In addition to increasing binding affinity, non-natural sidechains may also help preventing degradation by allowing peptides to escape residue-specific proteases. Other powerful ways of making peptides more resistant to protease degradation consist for instance in incorporating D-amino acids. Apart from rotamer probability predictions, topologies and parameters generated in this work can be readily transferred to D-enantiomers. In particular, binding free-energy can be computed in the same way as we did for BCL9 mutants. Furthermore, current technologies enable screening large random libraries of cyclic^[35] or bi-cyclic peptides^[36] with natural sidechains that are highly resistant to proteolytic degradation. Insertion of non-natural sidechains into these ligands is a promising way to affinity mature them, for which structure-based methods relying on the data generated in this work would be particularly well suited.

Material and Methods

Building structural files

Considering all non-natural sidechains present in the PDB database, as well as a several others that are commercially available, pdb and mol2 files were generated for a total of 209 sidechains. Three-letter code of the PDB was used when available. Otherwise a four-letter code was chosen such as to provide an intuitive abbreviation of the sidechain full name. Atom names were chosen according to the Greek alphabet order C α -C β -C γ ... Here again we tried to follow as much as possible the atom names used in the PDB. However, some of them had to be renamed for consistency, especially the ones that did not use standard names for backbone atoms. Unless specified,

the protonated form corresponds to pH = 7. All amino acids are provided in their L-form.

Generating topologies and parameters

Parameters and topologies for CHARMM and Gromacs (using the CHARMM22 protein force-field^[37]) are provided for all non-natural sidechains. Since our main goal when modeling non-natural sidechains is to probe *in silico* their insertion into natural peptides or proteins, we tried to be as close as possible to the existing CHARMM force field for natural amino acids. For sidechains similar to natural ones (e.g.,

ORN) topology files were manually built by similarity with existing ones. For other sidechains we used the SwissParam web service^[18] that provides topologies for any organic molecule based on MMFF.^[34] In all cases, only chemical groups without analogs among existing sidechains, as well as immediately neighboring atoms, were given partial charges derived from SwissParam. For instance, backbone atoms were always parameterized in the same way as natural amino acids in the CHARMM force field. Moreover, hydrogen atoms within —CH, —CH₂, or —CH₃ groups were always given a charge of 0.09 and the carbon charges were updated accordingly to maximize compatibility with the original CHARMM force field. For 34% of the sidechains, new atom types had to be defined, either when the atom itself did not exist in CHARMM22 (e.g., iodine or bromine), or when the atom was part of an uncharacterized chemical group (e.g., sulfur atoms within aromatic rings). Force-field parameters required to describe some non-natural sidechains (i.e., force constants and equilibrium values for bonds, angles, etc...) and not present in CHARMM were also retrieved from the SwissParam web service, as well as the list of internal coordinates for CHARMM topology files. For simulations in Gromacs, hydrogen database files have also been generated.

Rotamer definitions

Rotamers were defined considering all flexible dihedral angles along the sidechains (see Supporting Information Table S1). Both backbone dependent and backbone independent rotamers are considered. For backbone dependent rotamers, a 10° × 10° grid was defined on the (ϕ , ψ) plane, resulting in a total of 36 × 36 = 1296 bins. Similar bins on dihedral angles were used as in the Dunbrack 2002 rotamer library^[20] to define the rotamers on natural sidechains (see Supporting Information Table S2). For instance all sidechains starting with a linear chain C α -C β -C γ have three bins on the first dihedral

angle χ_1 ($r_1 = 1$ if $0 \leq \chi_1 < 120$, 2 if $120 \leq \chi_1 < 240$, and 3 if $-120 \leq \chi_1 < 0$). For some non-natural sidechains (e.g., asymmetric phenylalanine derivatives such as 3-methyl-phenylalanine, PDB code: APD), additional or different bins had to be included (see Supporting Information Tables S1 and S2 for the full list of bins of each kind of dihedral angles). Rotamer probabilities are denoted by $P(r_1, \dots, r_n)$ where r_i stands for the different bins on the i th dihedral angle and n for the total number of freely rotating dihedral angles used in the rotamer definition. For clarity, we do not explicitly write the backbone (ϕ, ψ) dependences. When considering rotamer probability for the first dihedral angle χ_1 , $P(r_1)$ stands for $\sum_{\{r_2\}} \dots \sum_{\{r_n\}} P(r_1, r_2, \dots, r_n)$. Similarly $P(r_1, r_2)$ stands for $\sum_{\{r_3\}} \dots \sum_{\{r_n\}} P(r_1, r_2, \dots, r_n)$. $P_{\text{exp}}(r_1, \dots, r_n)$ stands for experimental probabilities retrieved from the Dunbrack 2002 rotamer library.^[20] Recently, a new version of the Dunbrack rotamer library was published^[25] that includes several additional bins on some non-rotameric dihedral angles (e.g., c2 of ASN and ASP). When comparing our predictions for natural sidechains to this new library (Supporting Information Fig. S5) the same new bin definitions, including χ_1 dependencies on some dihedral angle bins, were used as in Ref. ^[25].

Rotamer predictions

Our rotamer predictions rely on a combined physics-based (using MD simulation to infer rotamer probabilities) and knowledge-based (exploiting conserved features of experimentally determined rotamer libraries) strategy.

Physics-based approach

To infer rotamer probabilities from molecular calculations, we used MD trajectories and evaluated the probability of each rotamer along the trajectories. Simulations were carried out on a tri-alanine peptide where the residue in the middle was mutated to each non-natural sidechain analyzed in this work. The N-terminal was acetylated and the C-terminal was amidated to prevent spurious interactions with or between charged termini along the simulation. Fifty-nanosecond second simulations were run at 300 K with CHARMM using the FACTS solvent model^[29] and starting from four different (ϕ, ψ) conformations corresponding to the main energy basins of the rama-chandran plot ((−139, 135), (−70, −27), (60, 40), and (60, −150)). Snapshots were recorded every picosecond and rotamer frequencies were computed as the frequency of observation along the trajectory in each bin of each dihedral angle defining the different rotamers. A similar smoothing based on Bayesian statistics with Dirichlet priors as in Ref. ^[23] was applied to these frequencies to allow for a fair comparison with rotamer probabilities derived from experimental structures. The average dihedral angles for each rotamer were also estimated from the dihedral angles along MD trajectories, using Bayesian statistics with Gaussian priors as in Ref. ^[23].

Alternate strategies have been used to predict rotamers, such as computationally scanning conformations of free dihedral angles and evaluating their energies.^[27] To compare with

these approaches we used the CHARMM software to evaluate energies of sidechain conformations for all natural sidechains (excluding ALA, GLY, and PRO). Starting from a minimized structure (5000 steps of steepest descent), backbone (ϕ, ψ) as well as sidechain dihedral angles were then restrained to given values by adding constraints of 5000 kcal/mol² on dihedral angles. Structures were subsequently minimized with 100 steps of steepest descent and the total energy was evaluated, not including the contribution due to the artificial constraints. Bins of 10 degrees were used on ϕ and ψ as well as on side-chain dihedral angles. To prevent excessive computations, the χ_1 dihedral angle was only sampled at (−60, 60, 180) for GLN, GLU, MET, and the χ_1, χ_2 angles were sampled at (−60, 60, 180) for ARG and LYS. Energy calculations were done both in *vacuo* ($\varepsilon = 1$) and with the FACTS solvent model ($\varepsilon_{\text{int}} = 1$, $\varepsilon_{\text{solv}} = 80$). Probabilities were derived from energies using Boltzmann statistics $P(r_1, \dots, r_n) = \frac{1}{Z} \sum_{c \in S(r_1, \dots, r_n)} \exp(-E(c)/kT)$, where $S(r_1, \dots, r_n)$ stands for the ensemble of conformations with rotamer (r_1, \dots, r_n) , $E(c)$ for the energy of a particular conformation, and Z for the appropriate normalizing factors. For backbone dependent rotamers, Z corresponds to a sum over all conformations with given (ϕ, ψ), while for backbone independent it includes all possible conformations. Comparison between MD-based and energy-based rotamer predictions can be found in Supporting Information Figure S1.

Knowledge-based approach

The main idea of this approach is to make use of the remarkable degree of conservation of the first dihedral angle rotamer distributions among sidechains starting with a linear chain of carbon atoms (see Fig. 2). For sidechains satisfying the χ_1 renormalization criterion (see Supporting Information Table S1 and Result section), the probabilities $P(r_1, \dots, r_n)$ computed with CHARMM are renormalized by the experimental averaged probability distribution $P_{\text{av}}(r_1)$, as defined in the main text. For instance, if $P(r_1, r_2)$ describes the computed rotamer probability for a sidechain with two degrees of freedom that satisfy the χ_1 renormalization criterion, the final rotamer probability is given by $P_{\text{final}}(r_1, r_2) = P_{\text{av}}(r_1)P(r_1, r_2)/P(r_1)$. For simple derivatives of natural sidechains that contain additional dihedral angles (e.g., methionine derivative 2-amino-4-ethyl sulfanyl butyric acid in Fig. 3), the first dihedral angles (χ_1, χ_2 , and χ_3 in this example) were renormalized with the rotamer probabilities of the corresponding natural sidechain. We note that for para-substituents of PHE, the bins on χ_2 were chosen such as to consider the introduced asymmetry (see Supporting Information Table S2) and the experimental rotamer probabilities of PHE were divided by 2 before distributing them between the symmetric bins. Moreover, for ortho- substituents on PHE or TRP, the χ_1 renormalization scheme was used, rather than a direct mapping, since the substituents are likely to have a stronger effect on χ_2 .

Comparing rotamer probabilities

To compare rotamer probabilities (either experimental probabilities among different sidechains, or experimental versus

predicted probabilities), we used the correlation coefficient between all rotamer probabilities. This measure is especially appropriate because it is not too sensitive to small variations among very low probabilities, which are often less meaningful. When comparing with experimental rotamer probabilities, we restrict our analysis to (ϕ, ψ) bins with at least 10 experimental data points in the rotamer library of Ref. [20], to avoid biasing the results because of the low coverage and smoothing applied to the data. Supporting Information Figure S4 shows the results of the cross-validation study for (ϕ, ψ) bins corresponding to alpha helices or beta sheets (i.e., $-70 \leq \phi \leq -40$ and $-50 \leq \psi \leq -20$, or $-160 \leq \phi \leq -110$ and $110 \leq \psi \leq 160$).

Cross-validation

For cross-validation of the knowledge-based approach, we did not include $P_{\text{exp}}(r_1)$ of the corresponding sidechain in the average experimental distribution $P_{\text{av}}(r_1)$ used to renormalize the computed probabilities. For instance, when renormalizing the computed $P(r_1, r_2)$ for LEU, we used the average probability distribution $P_{\text{av}}(r_1)$ over ARG, CYS, GLN, GLU, HIS, LYS, MET, PHE, TYR and TRP, excluding experimental data for LEU. Similarly, we excluded $P_{\text{exp}}(r_1, r_2)$ of the sidechain under consideration from $P_{\text{av}}(r_1, r_2)$ when renormalizing $P(r_1, r_2, r_3)$ for MET, GLU and GLN, as well as $P(r_1, r_2, r_3, r_4)$ for ARG and LYS.

Binding free-energy predictions

Twelve mutants of a BCL9 peptide involving non-natural sidechains were studied in Ref. [31] and binding affinity for beta-catenin was measured for all of them. The mutants were structurally modeled with our PyMOL plug-in, starting from the X-ray structure of BCL9 bound to beta-catenin^[32] (PDB: 2GL7). For each mutant, we used all rotamers except the ones corresponding to very big clashes. To estimate binding free-energy, we used the method of Ref. [33]. Structures were first subjected to 500 steps of steepest descent minimization in Gromacs. MD simulations were then run for 1ns in Gromacs with the CHARMM22 force-field^[38] supplemented by our newly derived parameters, and frames were extract every 10ps. Binding free-energy (ΔG) of the complex for each frame was estimated using the MM-GBSA algorithm^[33] and results were averaged over all 100 frames for each mutant and for the wild-type. Only contributions from residues directly on the binding interface were considered. The final $\Delta\Delta G$ predictions correspond to the binding free-energy differences between the wild-type and each mutant. For the average over all rotamers, we used Boltzmann weights according to the following formula:

$$\Delta\Delta G_{\text{final}} = -RT \ln \left(\frac{1}{C} \sum_{i=1}^C \exp \left(\frac{-\Delta\Delta G_i}{RT} \right) \right),$$

where C stands for the total number of rotamers (rotamers excluded because of large clashes are assumed to have a very large $\Delta\Delta G$ and not to contribute to the sum of exponentials).

Acknowledgments

We thank Aurélien Grosdidier and Michel Cuendet for insightful discussions about this work. We are thankful to Michel Aguet and Patrick Rodriguez for pointing out the BCL9-beta catenin interaction study. DG is financially supported by an EMBO long-term fellowship.

Keywords: non-natural sidechain • rotamer • peptide • molecular modeling • drug design

How to cite this article: D. Gfeller, O. Michielin, V. Zoete, J. Comput. Chem. **2012**, *33*, 1525–1535. DOI: 10.1002/jcc.22982

Additional Supporting Information may be found in the online version of this article.

- [1] J. A. Wells, C. L. McClendon, *Nature* **2007**, *450*, 1001.
- [2] T. Pawson, P. Nash, *Science* **2003**, *300*, 445.
- [3] A. J. Barr, E. Ugochukwu, W. H. Lee, O. N. King, P. Filippakopoulos, I. Alfano, P. Savitsky, N. A. Burgess-Brown, S. Muller, S. Knapp, *Cell* **2009**, *136*, 352.
- [4] U. Wiedemann, P. Boisguerin, R. Leben, D. Leitner, G. Krause, K. Moellering, R. Volkmer-Engert, H. Oschkinat, *J. Mol. Biol.* **2004**, *343*, 703.
- [5] R. Tonikian, X. Xin, C. P. Toret, D. Gfeller, C. Landgraf, S. Panni, S. Paoletti, L. Castagnoli, B. Currell, S. Seshagiri, H. Yu, B. Winsor, M. Vidal, A. R. Davidson, M. B. Gerstein, G. D. Bader, R. Volkmer, G. Cesareni, D. G. Drubin, P. M. Kim, S. S. Sidhu, C. Boone, *PLoS Biol.* **2009**, *7*, e1000218.
- [6] C. Zahnd, P. Amstutz, A. Pluckthun, *Nat. Methods* **2007**, *4*, 269.
- [7] D. Gfeller, F. Butty, M. Wierzbicka, E. Verschueren, P. Vanhee, H. Huang, A. Ernst, N. Dar, I. Stagliar, L. Serrano, S. S. Sidhu, G. D. Bader, P. M. Kim, *Mol. Syst. Biol.* **2011**, *7*, 484.
- [8] S. A. Sievers, J. Karanicolas, H. W. Chang, A. Zhao, L. Jiang, O. Zirafi, J. T. Stevens, J. Munch, D. Baker, D. Eisenberg, *Nature* **2011**, *475*, 96.
- [9] J. Xie, P. G. Schultz, *Nat. Rev. Mol. Cell. Biol.* **2006**, *7*, 775.
- [10] F. Tian, M. L. Tsao, P. G. Schultz, *J. Am. Chem. Soc.* **2004**, *126*, 15962.
- [11] H. Murakami, A. Ohta, H. Ashigai, H. Suga, *Nat. Methods* **2006**, *3*, 357.
- [12] H. Neumann, K. Wang, L. Davis, M. Garcia-Alai, J. W. Chin, *Nature* **2010**, *464*, 441.
- [13] J. Schymkowitz, J. Borg, F. Stricher, R. Nys, F. Rousseau, L. Serrano, *Nucleic Acids Res.* **2005**, *33*, W382.
- [14] C. Wang, P. Bradley, D. Baker, *J. Mol. Biol.* **2007**, *373*, 503.
- [15] Van Der D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.
- [16] B. R. Brooks, C. L. Brooks, III, A. D. Mackerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.
- [17] A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker, P. Bradley, *Methods Enzymol.* **2011**, *487*, 545.
- [18] V. Zoete, M. A. Cuendet, A. Grosdidier, O. Michielin, *J. Comput. Chem.* **2011**, *32*, 2359.
- [19] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, *J. Comput. Chem.* **2004**, *25*, 1605.
- [20] R. L. Dunbrack, Jr. *Curr. Opin. Struct. Biol.* **2002**, *12*, 431.
- [21] J. W. Ponder, F. M. Richards, *J. Mol. Biol.* **1987**, *193*, 775.
- [22] G. G. Krivov, M. V. Shapovalov, R. L. Dunbrack, Jr., *Proteins* **2009**, *77*, 778.
- [23] R. L. Dunbrack, Jr., F. E. Cohen, *Prot. Sci.* **1997**, *6*, 1661.

- [24] S. C. Lovell, J. M. Word, J. S. Richardson, D. C. Richardson, *Proteins* **2000**, *40*, 389.
- [25] M. V. Shapovalov, R. L. Dunbrack, Jr., *Structure* **2011**, *19*, 844.
- [26] J. Mendes, H. A. Nagarajaram, C. M. Soares, T. L. Blundell, M. A. Carrodon, *Biopolymers* **2001**, *59*, 72.
- [27] P. D. Renfrew, G. L. Butterfoss, B. Kuhlman, *Proteins* **2008**, *71*, 1637.
- [28] G. Revilla-Lopez, J. Torras, D. Curco, J. Casanovas, M. I. Calaza, D. Zanuy, A. I. Jimenez, C. Cativiela, R. Nussinov, P. Grodzinski, C. Aleman, *J. Phys. Chem. B* **2010**, *114*, 7413.
- [29] U. Haberthur, A. Caflisch, *J. Comput. Chem.* **2008**, *29*, 701.
- [30] M. S. Lee, M. Feig, F. R. Salsbury, Jr., C. L. Brooks, III, *J. Comput. Chem.* **2003**, *24*, 1348.
- [31] S. A. Kawamoto, A. D. Thompson, A. Coleska, Z. Nikolovska-Coleska, H. Yi, S. Wang, *Biochemistry* **2009**, *48*, 9534.
- [32] J. Sampietro, C. L. Dahlberg, U. S. Cho, T. R. Hinds, D. Kimelman, W. Xu, *Mol. Cell.* **2006**, *24*, 293.
- [33] V. Zoete, O. Michielin, *Proteins* **2007**, *67*, 1026.
- [34] T. A. Halgren, *J. Comput. Chem.* **1996**, *17*, 490.
- [35] S. W. Millward, S. Fiacco, R. J. Austin, R. W. Roberts, *ACS Chem. Biol.* **2007**, *2*, 625.
- [36] C. Heinis, T. Rutherford, S. Freund, G. Winter, *Nat. Chem. Biol.* **2009**, *5*, 502.
- [37] J. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evans, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B* **1998**, *102*, 3586.
- [38] P. Bjelkmar, P. Larsson, M. A. Cuendet, B. Hess, E. Lindahl, *J. Chem. Theory Comput.* **2010**, *6*, 459.

Received: 17 November 2011

Revised: 7 February 2012

Accepted: 10 March 2012

Published online on 14 April 2012