



PCCP

Using ^1H and ^{13}C NMR Chemical Shifts to Determine Cyclic Peptide Conformations: A Combined Molecular Dynamics and Quantum Mechanics Approach

Journal:	<i>Physical Chemistry Chemical Physics</i>
Manuscript ID	CP-ART-03-2018-001616.R1
Article Type:	Paper
Date Submitted by the Author:	11-Apr-2018
Complete List of Authors:	Nguyen, Quynh Nhu; University of California, Davis, Chemistry Schwochert, Joshua; University of California Santa Cruz, Chemistry and Biochemistry Tantillo, Dean; University of California, Davis, Chemistry; Lokey, R.; University of California Santa Cruz, Chemistry and Biochemistry

SCHOLARONE™
Manuscripts



Journal Name

ARTICLE

Using ^1H and ^{13}C NMR Chemical Shifts to Determine Cyclic Peptide Conformations: A Combined Molecular Dynamics and Quantum Mechanics Approach

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

Q. Nhu N. Nguyen^{†a,b}, Joshua Schwochert^{†c}, Dean J. Tantillo^a, R. Scott Lokey^c

Solving conformations of cyclic peptides can provide insight into structure-activity and structure-property relationships, which can help in the design of compounds with improved bioactivity and/or ADME characteristics. The most common approaches for determining the structures of cyclic peptides are based on NMR-derived distance restraints obtained from NOESY or ROESY cross-peak intensities, and 3J -based dihedral restraints using the Karplus relationship. Unfortunately, these observables are often too weak, sparse, or degenerate to provide unequivocal, high-confidence solution structures, prompting us to investigate an alternative approach that relies only on ^1H and ^{13}C chemical shifts as experimental observables. This method, which we call Conformational Analysis from NMR and Density-functional prediction of Low-energy Ensembles (CANDLE), uses molecular dynamics (MD) simulations to generate conformer families and density functional theory (DFT) calculations to predict their ^1H and ^{13}C chemical shifts. Iterative conformer searches and DFT energy calculations on a cyclic peptide-peptoid hybrid yielded Boltzmann ensembles whose predicted chemical shifts matched the experimental values better than any single conformer. For these compounds, CANDLE outperformed the classic NOE- and 3J -coupling-based approach by disambiguating similar β -turn types and also enabled the structural elucidation of the minor conformer. Through the use of chemical shifts, in conjunction with DFT and MD calculations, CANDLE can help illuminate conformational ensembles of cyclic peptides in solution.

Introduction

Knowing the 3-dimensional conformation(s) of a cyclic peptide can provide insight into both its physicochemical properties¹⁻³ and its biological activity.⁴⁻⁶ There has been a great deal of interest in cyclic peptides as scaffolds in the development of drugs against difficult targets such as protein-protein interactions, based on the premise that large macrocycles are better suited to the inhibition of large binding surfaces.⁷ In these molecules, which often fail to meet common criteria for predicting “drug-likeness”, conformational states become increasingly important determinants of physicochemical properties.^{2, 8-12}

The most widely used techniques for small-molecule conformational analysis employ a combination of NMR spectroscopy and computational techniques.¹³⁻¹⁸ Most of these methods focus on inter-atomic distances and angles, using a

combination of experimentally derived constraints: 1) Nuclear Overhauser Effect (NOE) crosspeak volumes to provide interatomic distances; 2) residual dipolar couplings (RDCs) to provide relative bond vector orientations; 3) 3J correlations (coupling constants) to provide dihedral restraints *via* the Karplus relationship. When fed into a conformational search algorithm, these constraints can exclude many potential conformations. However, determining solution conformations of cyclic peptides and other macrocycles by NMR, especially in the ~800-1000 MW range, can present unique challenges. For example, in contrast to the situation with proteins, which have a high core-to-surface ratio, the number of informative, non-sequential (i.e., through-space) NOE cross-peaks in small macrocycles can be small, leading to under-determination of structures based on NOE data alone. In addition, the Karplus equation does not predict unique solutions for the same set of coupling constants.^{19, 20} Furthermore, for N-Me amides, key NH- α couplings are absent, preventing the use of these dihedral restraints for N-methylated residues.

Here we describe an NMR-based method for the analysis of solution conformations that relies solely on chemical shift information and functions independently of NOE- and 3J -based restraints. This method begins with conformational sampling using molecular dynamics (MD), followed by quantum mechanical (QM) geometry optimization and prediction of ^1H

^a Department of Chemistry ; UC Davis; 1 Shields Ave, Davis, CA 95616

^b Department of Chemistry, University of Oxford, Oxford, U.K.

^c Department of Chemistry, UC Santa Cruz, 1156 High St., Santa Cruz, CA 95064

† Authors contributed equally

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

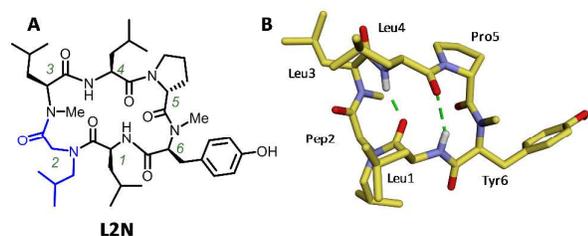


Fig. 1: Chemical structure and original NMR-based conformation of the major conformer of **L2N (L2N-major)**. (A) Structure of **L2N** showing residue numbering, with the peptoid residue (Pep2) in blue. (B) The NMR-derived solution structure of **L2N-major** determined previously using DISCON analysis based on distance restraints (from ROESY correlations) and dihedral restraints (from 3J coupling constants).

and ^{13}C NMR chemical shifts of individual conformations. QM-based chemical shift predictions have been applied extensively to structure elucidation in natural products,^{21–27} and have been used to analyze the conformations of cyclic peptides.²⁸ In the latter study, the relatively large deviations between experimental and predicted chemical shifts prompted us to develop a refined workflow in an effort to improve the accuracy of structure predictions including the ability to capture side chain rotamer populations. Here we describe a workflow that builds upon the study by Zaretsky, et al.,²⁸ but uses a different sampling approach designed to capture more subtle conformational motions, as well as a new correction procedure for addressing errors resulting from intramolecular CH—O interactions,^{29, 30} This variant of the previously reported method, which we call **Conformational Analysis from NMR and Density-functional prediction of Low-energy Ensembles (CANDLE)**, uses high-temperature molecular dynamics (MD) simulations to generate conformer families and density functional theory (DFT) calculations to predict their ^1H and ^{13}C chemical shifts. We applied this method to the solution structure of **L2N**, a cyclic peptide-peptoid hybrid that we had investigated previously in a study exploring the impact of peptoid substitutions on ADME properties in cyclic peptides. CANDLE yielded more conformational information than the classic NOE- and 3J -coupling-based approach, allowing us to distinguish between two very similar β -turn types where NOE crosspeaks were inconclusive. CANDLE also allowed us to elucidate the structure of a minor conformer for which classical distance and dihedral restraints failed to yield an unambiguous structure.

Methods

NMR. For peak assignments, we used a combination of HSQC, HMBC, TOCSY, and COSY spectra, which were acquired using the pulse sequences supplied with instrument (Varian CHEMPack). ROESY spectra used in conformational analysis were gathered at a mixing time of 200 ms, and sufficiently resolved crosspeaks were integrated with a focus on inter-residue correlations, as those are both the most structurally informative and are insensitive to TOCSY bleed-through. As a reference, the NOE between germinal proline delta protons

was set to 1.76 angstroms. Additional detail on the NMR methods used can be found in the Supporting Information.

DISCON. Previously, candidate solution conformations were identified using the Distribution of in Solution Conformations (DISCON) software.^{17, 31} DISCON first clusters conformations, based on inter-proton distances and dihedral angles, and then fits an ensemble of the cluster centers to the experimental NOE/ROE and 3J restraints. A pool of 250 conformations was generated and minimized via high temperature molecular dynamics using the CHARMM forcefield³² for input into DISCON. A set of 16, inter-residue ROEs and two 3J couplings were used as experimental constraints. For the major NMR conformation a single, dominant, conformation representing >94% of the ensemble was found.

Molecular Dynamics. Molecular dynamics (MD) simulations were used to generate input conformations for subsequent density functional theory (DFT) analysis. To prepare the starting structure for MD, a short dynamics cascade consisting of steepest decent followed by conjugate gradient minimization was carried out, culminating in a 14 ps 300 K dynamics simulation. High-temperature MD (HT-MD) simulations were run at 2000 K, using the CHARMM force field with a time step of 1 fs. For production runs, snapshots were taken every 20 ps over 5 ns, yielding pools of 250 conformations. For **L2N-minor**, snapshots were taken every 7.5 ps over 5 ns and only conformations containing a *cis* orientation about the Leu3-NMe amide bond were taken, generating a pool of 219 conformations. Conformations were minimized, clustered (*vide infra*), and taken on to QM single point energy calculations (*vide infra*). Room temperature MD (RT-MD) simulations were conducted at 300 K starting from the DFT optimized conformations over 25 ns with a timestep of 1 fs with snapshots take every 100 ps. The resulting pool of 250 conformations was then minimized and taken onto clustering analysis. Based on visual inspection, 15–20 high temperature clusters appeared to give degeneracy in the conformations sampled, but substantially increased the computational time for the DFT calculations. For larger peptides, increasing the sampling during the high temperature run and/or increasing the number of clusters may be necessary to adequately sample the backbone conformational space, especially in cases involving multiple peptoid, Pro, and *N*-Me residues for which all *cis-trans* isomers will need to be sampled. In this case, for the **L2N-minor** conformation, the use of NOESY-derived distance information can be useful to narrow the search space to only relevant *cis-trans* isomers.

Quantum Mechanics. Initial single point energy calculations were performed with *Gaussian09*³³ using either HF/3-21G^{34, 35} or B3LYP/6-31G(d) (*vide infra*).^{36–38} Full optimizations with frequency calculations (to confirm that structures were minima) were carried out using M06-2X/6-31G(d)^{38, 39} in the gas phase for all conformers within 5 kcal/mol of the lowest energy structure in each cluster (based on the B3LYP/6-31G(d) single point energies). NMR chemical shift calculations were performed using mPW1PW91/6-31+G(d,p)⁴⁰ in chloroform

(modeled using the SMD implicit continuum approach).⁴¹ Chemical shifts were determined by scaling absolute computed isotropic values using scaling factors available at the cheshiremr.info website (slope = -1.0803 for ¹H and -0.9726 for ¹³C; intercept = 31.7031 for ¹H and 194.9643 for ¹³C).⁴² See Supporting Information for additional details.

Clustering. Conformations resulting from the high temperature MD simulations were clustered by their backbone conformations. The ϕ and ψ angles of all amino acids for each snapshot were measured. Conformations were clustered based on backbone dihedral angles using the maximal dissimilarity partitioning method as implemented in *Discovery Studio 4.0* by Accelrys.⁴³ The number of clusters was fixed at of 10, 15, and 20, and visual inspection of the cluster centers indicated that when more than 10 clusters were used redundant clusters were formed, so 10 clusters were chosen for the single point energy analysis. Room temperature conformations were clustered by heavy-atom RMSD with a set number of 20 clusters. Representative members, as chosen by the initial maximal dissimilarity partitioning (cluster centers), were used for further analysis. All cluster centers, independent of the starting conformation from which they arose, were combined to create a pool of 60-80 conformations that was then aligned to the lowest energy conformation produced by a CHARMM energy minimization and again heavy-atom pairwise RMSD was calculated. Finally, this pool of conformations was clustered by RMSD to yield 20 representative conformers for further QM analysis.

Results and Discussion

Previously, we conducted a study examining the relationship between structure and passive membrane permeability in cyclic peptides and peptide-peptoid hybrids.⁴⁴ In the course of this work we found that the effect of peptide-to-peptoid substitutions on passive permeability was position-dependent, prompting us to investigate the solution conformation of **L2N** (Fig. 1A) in CDCl₃, a low-dielectric solvent that has been used to mimic the membrane environment and provide insight into the impact of conformation on permeability.^{8, 45} This compound, in which an N-Me residue at the *i* + 1 position of a β -turn was substituted for a peptoid residue with a leucine-like isobutyl side chain, was analyzed using NMR restraints in conjunction with the DISCON algorithm,^{14, 17, 46} which determines solution structures by fitting conformational ensembles to NMR-derived distance and dihedral restraints. While the major conformer of **L2N** (**L2N-major**) provided a convergent structure using DISCON (Fig. 1B), the NMR spectrum in CDCl₃ showed a minor conformer (33%) in slow exchange on the NMR time scale. Our attempts to solve the structure of the minor conformer (**L2N-minor**) using MD and DISCON through ³J-couplings and ROESY correlations were hampered by signal overlap, which limited the number of restraints that could be used to generate an unambiguous fit to the experimental data. We therefore selected **L2N** as a model system for this study because its structure had been determined using classical NOE- and ³J-based methods. In

addition, the minor conformer provided a good test of the feasibility of using chemical shifts alone to obtain conformational information in a case in which classical methods were unable to produce a high-confidence structure.

High-temperature MD (HT-MD) simulations were performed on **L2N** at 2000 K, a temperature which was high

A

Conf.	ΔG (kcal/mol)	MAD (ppm)	
		¹ H	¹³ C
HT-74	0.00	0.33	2.86
HT-115	1.29	0.33	2.63
HT-137	1.78	0.22	1.90
HT-185	1.97	0.19	2.41

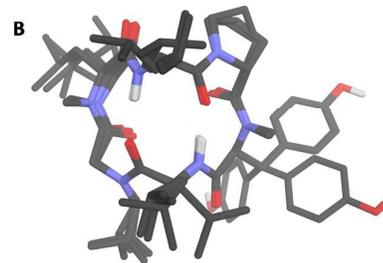


Fig. 2: Results of HT-MD simulations on **L2N** and comparison between predicted and experimental chemical shifts for the major conformer (**L2N-major**). (A) Table of lowest energy conformers optimized using M06-2X/6-31G(d). (B) Overlay of the four conformers.

enough to overcome amide *cis-trans* rotamer barriers for the N-Me and peptoid residues, but low enough to provide snapshots whose associated geometries could be minimized into reasonable structures. For each compound, the HT-MD simulation was performed for 5 ns using implicit solvent with a dielectric set to 4, and 250 conformers were generated by minimizing snapshots taken every 20 ps. Analysis and ranking of the conformations was initially performed by rank-ordering the entire set by their single-point energies (HF/3-21G), and conformers within 10 kcal/mol of the lowest energy conformer were optimized with M06-2X/6-31G(d). However, due to the relatively small size of the initial pool of HT-MD conformers and low accuracy of single-point energies, we were concerned that a simple ranking scheme might not capture the experimental backbone conformation, an error that would be propagated in the downstream analysis. Therefore, to ensure adequate representation of diverse backbones in the subsequent steps, in addition to ranking the entire conformer pool by their single-point energies, we clustered the pool into 10 distinct backbone families, and ranked the single-point energies within each family. The lowest-energy conformer from each backbone cluster was then geometry-optimized with M06-2X/6-31G(d). For each optimized conformer representing the 10 HT-MD clusters, backbone ¹³C and ¹H chemical shifts were calculated using mPW1PW91/6-31+G(d,p) (scaled). The deviations between predicted and experimental chemical shifts for each conformer (expressed simply as mean absolute deviations, MADs) were then compared to their QM energies.

HT-MD conformer generation for **L2N** and comparison with NMR of **L2N-major**

ARTICLE

Journal Name

Initially, single point energies were calculated for all 250 conformers generated from the HT-MD simulation. The seven conformers within 11 kcal/mol of the lowest energy conformer were optimized using M06-2X/6-31G. After optimization, the calculated free energies of the four lowest-energy conformers, *HT-74*, *HT-115*, *HT-137*, and *HT-185*, fell within a 2 kcal/mol free energy window (Fig. 2A). Their backbone conformations were virtually identical (Fig. 2B), and the mean absolute

deviations (MADs) between calculated and the experimental NMR chemical shifts for the backbone ^1H and ^{13}C atoms of **L2N-major** were between 0.19 and 0.33 for ^1H , and 1.90 and 2.86 for ^{13}C . When we used the clustering method to select a candidate from the 10 backbone clusters, HT-137 of cluster #1 gave the best match to the experimental backbone chemical shifts (see SI for

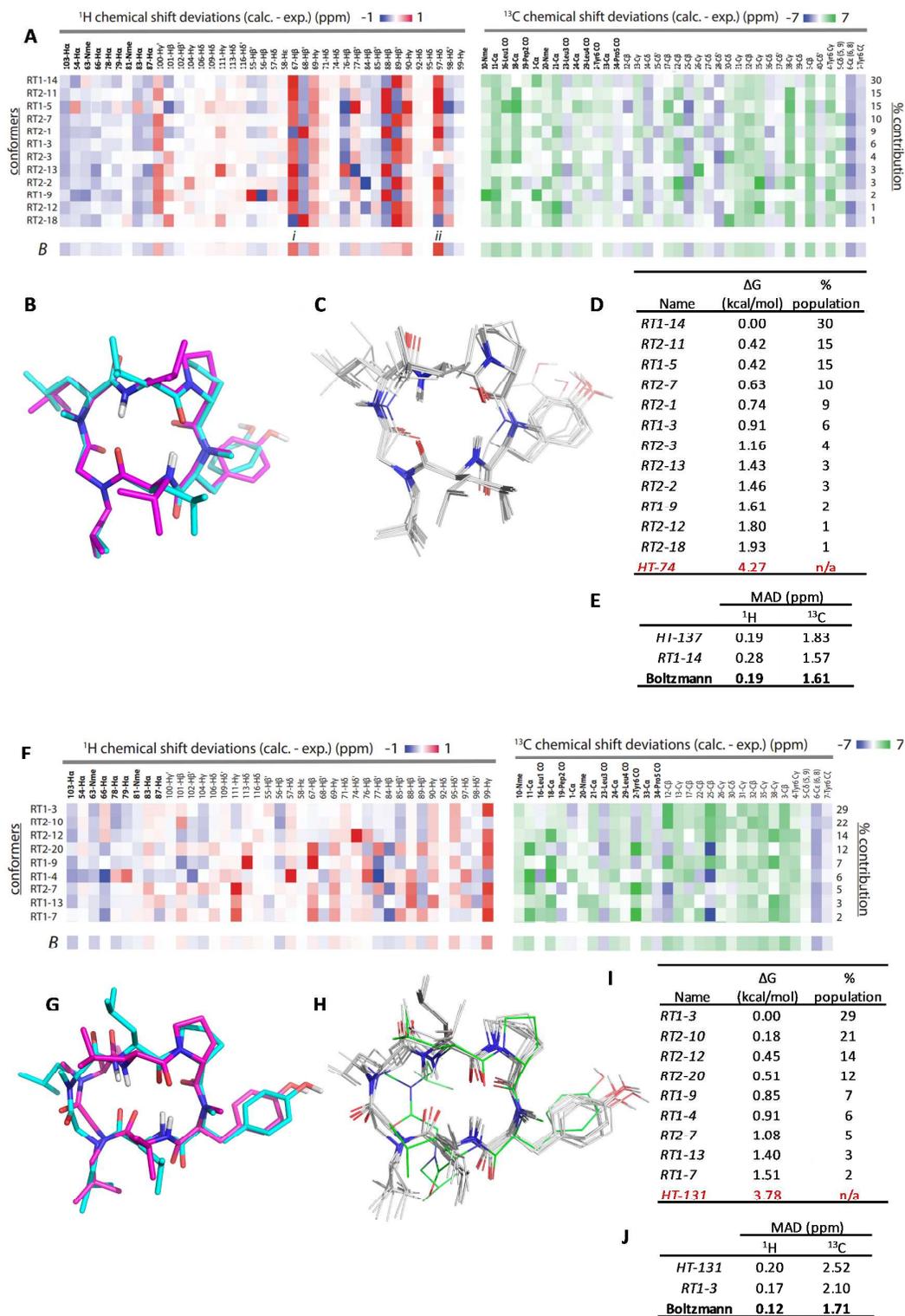


Fig. 4. Room-temperature MD (RT-MD) simulations used in structural refinements based on low-energy conformers from the HT-MD simulations of **L2N** (A-E) and **L2N-minor** (F-J). (A and F) Heat maps showing individual deviations between experimental and calculated ¹H and ¹³C chemical shifts for the RT-MD ensembles of **L2N** (A) and **L2N-minor** (F) with the Boltzmann averages denoted *B*. Outlier ¹H chemical shifts for Leu1-H β and Pro5-H δ of **L2N** denoted *i* and *ii*, respectively (A). (B and G) Overlay of the seed conformations from the HT-MD simulations (magenta) with the lowest energy structures from the RT-MD ensembles (cyan). (B) For **L2N**, HT-MD seed conformer *HT-137* (magenta) is compared with the lowest-energy RT-MD conformer, *RT1-14* (cyan). (G) For **L2N-minor**, HT-MD seed conformer *HT-131* (magenta) is compared with the lowest-energy RT-MD conformer, *RT1-3* (cyan). (C and H) Overlays of geometry-optimized, RT-MD-derived conformers from the Boltzmann distributions of **L2N** (C) and **L2N-minor** (H). For **L2N-minor** (H), the conformer in green (*RT1-4*) was a structural outlier with an extra *cis*-amide in the backbone. Given that *cis*-trans isomerism is unlikely on the NMR time scale, its relevance to the observed chemical shifts of **L2N-minor** is unclear. (D and I) Tables of relative energies of the lowest-energy RT-MD conformers and their relative contributions to the ensemble for **L2N** (D) and **L2N-minor** (I). (E and J) Tables comparing overall MAD scores for HT-MD seed conformers, lowest energy, RT-MD conformers, and Boltzmann weighted averages of RT-MD conformers for **L2N** (E) and **L2N-minor** (J).

more information on clusters' energy profiles). Thus, both the energy ranking and clustering methods arrived at essentially the same backbone for **L2N-major**.

The backbone conformation of **L2N-major** found previously using NOE-based methods and analysis with the DISCON algorithm was very similar to conformer *HT-89* from Cluster 6 in the HT-MD simulations (Fig. 3A-C and Tables S1 and S2). Conformer *HT-89* is both higher in calculated free energy (by 10 kcal/mol) and a worse fit to the NMR data than *HT-137* (Table S2), with the two conformers differing from each other in the β -turn type between Pep2 and Leu3. The DISCON structure has a Type II β -turn between Pep2 and Leu3, whereas CANDLE predicts a Type II' β -turn at this position. To determine why this turn inversion was missed in the original NOE-based analysis, we investigated the ability for interproton distances alone to discriminate between these conformations. The two conformations with the lowest error as determined by the DISCON algorithm, representing both β -turn types, were extracted and their interproton distances compared (Fig. 3D). There were only slight differences in the interproton distances of the atom pairs used in the NOE-based DISCON structure, highlighting the limitations in the use of NOE-derived distance restraints to distinguish between subtly different backbone conformations, such as between β -turn types. This result underscores the advantage of chemical shift analysis, which considers every carbon and proton in the structure, in place of (or in addition to) traditional NOE or *J*-coupling based methods.

RT-MD Refinement of **L2N-major**.

RT-MD refinement of **L2N-major**

Although the predicted backbone chemical shifts for conformer *HT-137* provided a good fit to the experimental values for **L2N-major**, there were significant deviations for some of the side chain chemical shifts. This was not unexpected, since the number of conformers selected from the HT-MD simulations was too small to sample all side chain rotamers for each backbone. Therefore, in order to exhaustively sample side chain dihedrals and further refine the solution structure of **L2N-major**, we used the three low-energy conformers from the HT-MD simulation that also had the lowest MAD scores, *HT-137*, *HT-185* and *HT-115*, as starting conformers to seed independent RT-MD simulations. The resulting RT-MD conformers were dereplicated by iterative clustering based on pairwise heavy atom RMSDs, and 20 cluster centers were optimized using M06-2X/6-31G(d). Chemical shifts were predicted for the lowest-energy conformers within a 2 kcal/mol window and compared with the experimental shifts for **L2N-major** (Fig. 4A). The geometry-optimized conformers resulting from the RT-MD simulation were identical at the backbone level to each other and to the seed conformers from which they were generated, differing only in the orientation of one or more side chains (Fig. 4B and C). The lowest energy conformers from the RT-MD simulation (*RT1-14*) was significantly lower in free energy than any of the HT-MD conformers used to seed the RT-MD simulation (Fig. 4D), indicating that the RT-MD refinement was able to identify

lower-energy conformers by more exhaustively sampling side chain rotamer space. However, NMR chemical shift predictions for all conformers within 2 kcal/mol (Gibbs free energy) of the lowest energy RT-MD conformer revealed several side chain protons with significant deviations (> 0.7 ppm) between predicted and experimental values.

To determine whether the large deviations for these shifts were due to incomplete sampling, two additional RT-MD simulations followed by QM optimizations were performed using *RT1-14* as the seed conformer. After the second simulation, several additional conformers within the same 2 kcal/mol window were identified and combined with the conformers from the first RT-MD simulation. After the 3rd RT-MD simulation, no additional low-energy conformers were found. Chemical shifts for the lowest energy conformers from the first two RT-MD runs were calculated, as well as the Boltzmann weighted average of each chemical shift for the ensemble (Fig. 4E). While the ^1H MAD score for the low-energy RT-MD conformer *RT1-14* (0.28 ppm) was higher than that of seed conformer *HT-137* (0.19 ppm), the ^{13}C MAD score for *RT1-14* (1.57 ppm) was lower than that of *HT-137* (1.83). Nonetheless, the Boltzmann-weighted MADs for the ensemble (0.19 ppm for ^1H and 1.61 ppm for ^{13}C) were better than either conformer alone. Therefore, while the ^1H and ^{13}C MAD scores of the RT-MD ensemble were only slightly better than those of *HT-137*, the 4.3-kcal/mol energy difference between the best HT-MD-derived conformer and the RT ensemble suggests that the ensemble provides a more accurate representation of the

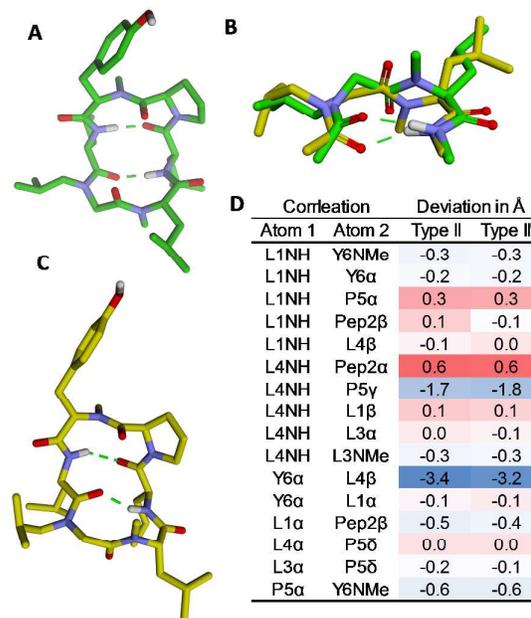


Fig. 3: Comparison between structure for **L2N-major** originally determined by classical NMR methods using DISCON and the best-fit CANDLE backbone obtained from fitting NMR chemical shifts to HT-MD-derived conformers in this study. (A) Conformer *HT-137* from CANDLE analysis of **L2N-major**. (B) Overlay of the two structures highlighting the β -turn about Pep2 and Leu3. (C) DISCON-derived conformation of **L2N-major**. (D) Table of ROE deviations for the two best-fit conformations by DISCON representing a Type II and Type II' β -turn about Pep2 and Leu3.

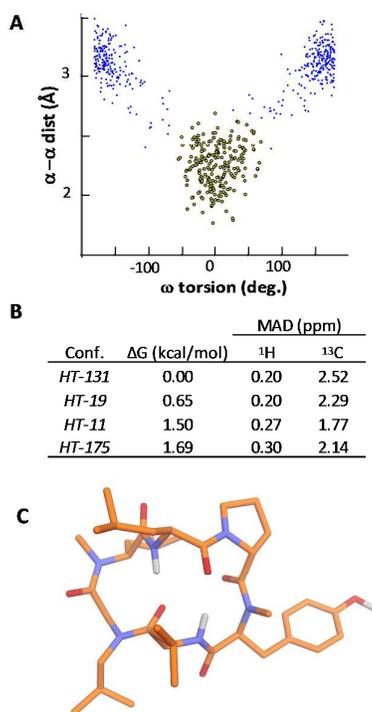


Fig. 5: Further HT-MD simulations on L2N to capture additional conformers with Pep2-Leu3 *cis*-amide dihedral. (A) Selection of 219 unminimized conformers containing Pep2-Leu3 *cis*-amide dihedral from 2nd HT-MD run based on Ha-Ha distance and ω dihedral angle. (B) Low-energy conformers with Pep2-Leu3 *cis*-amide dihedral, optimized using M06-2X/6-31G(d). (C) The conformer from the *cis*-amide series with the lowest free energy, HT-131.

solution structure.

Expansion of HT-MD simulations to sample likely *cis*-amide dihedral in L2N-minor

The ^1H and ^{13}C chemical shifts for the minor species in the NMR spectrum of L2N (L2N-minor) were fully assigned in chloroform, but there were only 13 ROE peaks that could be sufficiently resolved to be used as distance restraints for the DISCON analysis. DISCON did not give reproducible results when iterated, and the conformations that it generated for the NMR ensemble varied substantially between runs and cluster level. We believed that the low number of ROE correlations led to an underdetermined structure in which the DISCON algorithm failed to converge. Inspection of the ROESY spectrum did show a strong correlation between the α -protons of Pep3 and Leu2 (SI Fig. S4), which indicated a *cis*-dihedral about the amide bond between Pep2 and Leu3. In fact, nearly one third of the minimized snapshots from the HT-MD simulation of L2N had the *cis* configuration at Pep2-Leu3. To more completely sample the backbone conformational space within the population containing this *cis*-amide dihedral, we ran a new HT-MD simulation of L2N, sampling every 7.5 ps (rather than every 20 ps as in the first HT-MD run). The resulting set of 667 snapshots was filtered to select only those conformers with a *cis* Pep2-Leu3 amide bond, yielding 219 new conformers (Fig. 5A).

Initially, single point energies were computed using HF/3-21G, and full optimizations (M06-2X/6-31G in gas phase) were performed for conformers that were within 10 kcal/mol of the lowest energy structure. ^1H and ^{13}C shifts were computed, and the four conformers (HT-131, HT-19, HT-11, and HT-175; Fig. 5B) within 2 kcal/mol of the lowest were selected as backbone candidates for L2N-minor. Using the clustering approach, the 219 *cis* Pep2-Leu3 HT-MD-derived conformations were clustered by backbone similarity, single point calculations were recomputed with B3LYP/6-31G(d), and optimizations with M06-2X/6-31G(d) were performed for the lowest energy conformers representing each cluster. ^1H and ^{13}C shifts were calculated for the lowest energy structure of each cluster and compared to experimental values for the minor conformer. Unlike L2N-major, a single cluster was not identified as the likely backbone geometry for L2N-minor. Instead, the low-energy representatives of Clusters 1, 4, and 6 all had similar energies, backbone geometries, and similarly low MAD scores. The requirement for the Leu2-Pep3 amide bond to be *cis* likely limits the energetically-reasonable conformational space, leading to lower diversity in the HT-MD conformer pool and, in-turn, degenerate clusters at that cluster level. One of the four lowest energy conformer from the overall ranking, HT-131, was also the lowest energy representative of Cluster 4 (Fig. 5C). Encouragingly, all six conformations from the energy ranking and clustering methods had similar backbone conformations. Accordingly, we used the 4 conformations from the single-point energy ranking, HT-131, HT-19, HT-11, and HT-175 as the candidate backbone conformations with which to seed room temperature MD simulations for further refinement.

RT-MD refinement of L2N-minor

To further refine the structure of L2N-minor, HT-131, HT-11, HT-19, and HT-175 from the HT-MD simulations were used to seed the first round of RT-MD simulations. The lowest energy conformation from this round, RT1-3 was used to seed a second RT-MD simulation, and after clustering and geometry optimization, the RT-MD runs yielded 9 low-energy conformers. Chemical shifts were predicted for all assigned protons and carbons, including side chains (Fig. 4F). The lowest energy conformer from this set, RT1-3, had a backbone conformation that was similar, though not identical, to the backbone of the lowest-energy seed conformer (Fig. 4G). Surprisingly, one of the conformers, RT-1-4, had a second *cis*-amide bond between Leu1 and Pep2 (Fig. 4H, green lines). Further inspection of the seed conformations found that conformer HT-175 also contained this additional *cis*-amide, which led directly to conformer RT-1-4, the only conformation derived from HT-175 that was present in the room temperature ensemble. Despite its predicted contribution to the ensemble, RT-1-4 had the highest ^1H MAD and second highest ^{13}C MAD of the RT conformers; therefore, given that *cis-trans* isomers generally do not interconvert on the NMR time scale, the relevance of this conformer to the room temperature population is questionable. The other RT-MD

conformers had virtually identical backbones, differing only in terms of their side chain rotamers (Fig. 4H, grey lines).

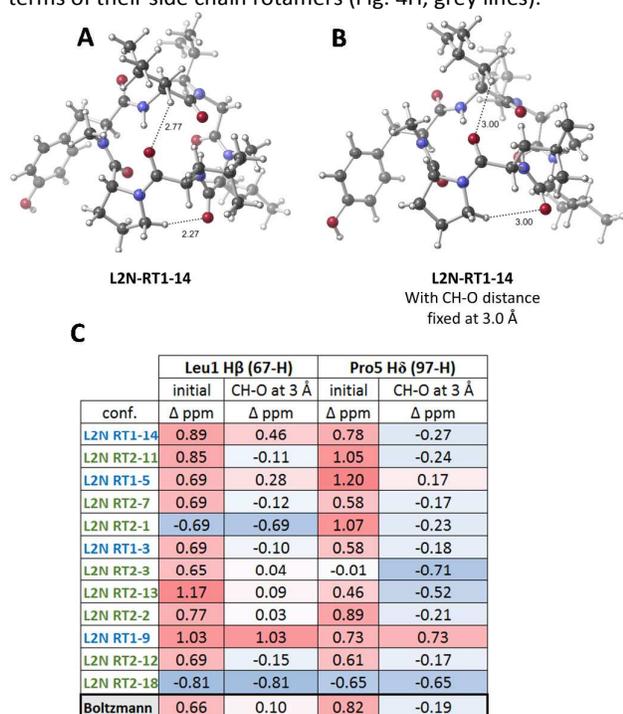


Fig. 6: CH-O corrections for L2N-major. (A) Geometry-optimized low energy conformer RT1-14. (B) RT1-14 with geometry optimization constraining the indicated distances at 3 Å. (C) Table showing ppm deviations before and after distance constraint and geometry optimization.

The lowest-energy RT-MD conformer was predicted to be 3.78 kcal/mol lower in free energy than the lowest-energy seed conformer (Fig. 4I), and its MAD scores were 0.17 (^1H) and 2.10 (^{13}C), compared to 0.20 and 2.52 for seed conformer HT-131 (Fig. 4J). The Boltzmann-weighted MAD scores for the ensemble were even lower: 0.12 (^1H) and 1.71 (^{13}C). As with L2N-major, the refined structures from the RT-MD simulations were lower in energy and had lower MAD scores than the seed conformer derived from the HT-MD simulations, and the Boltzmann distribution showed an even better fit to the NMR data than any single conformer.

Correcting for CH—O interactions

Some ^1H chemical shifts showed significant deviations between predicted and experimental values, even after multiple RT-MD simulations and Boltzmann averaging, suggesting that these deviations reflected inaccuracies in the chemical shift prediction rather than poor sampling. For L2N-major, there were significant deviations for Leu1-H β ($D = 0.66$ ppm) and Pro5-H δ ($D = 0.82$ ppm)(Fig. 4A, indicated in the Boltzmann average (B) as *i* and *ii*, respectively). In the solution ensemble predicted by CANDLE, both of these protons point directly toward neighboring carbonyl oxygens, leading us to hypothesize that our computations were overestimating the strength of intramolecular CH—O interactions and thus

underestimating the H \cdots O distance, perhaps due to our use of an implicit, rather than explicit, solvent model (Fig. 6).⁴⁷⁻⁵²

To evaluate this hypothesis, we fixed these two distances at 3.0 Å (an increase in 0.23 Å and 0.73 Å for Leu1-H β and Pro5-H δ , respectively), and then allowed the molecule to relax in a subsequent geometry optimization (using M06-2X/6-31G(d)) with these distance restraints turned on (Fig. 6 A and B). The chemical shifts for each conformer as well as the Boltzmann ensemble were then recalculated using the new geometries. Although the energy increased by ~ 2 kcal/mol when the H \cdots O distances were elongated, the chemical shift deviations of the outlier protons decreased significantly (from 0.66 to 0.10 ppm for Leu1-H β , and from 0.82 to -0.19 ppm for Pro5-H δ), resulting in a decrease in the overall MAD score for L2N from 0.19 to 0.16 ppm (Fig. 6C). This sensitivity to H \cdots O distances for NMR chemical shifts suggests that care should be taken in interpreting chemical shift predictions for relatively nonpolar protons when they are poised to engage in non-classical hydrogen bonding.

Conclusions

In summary, we have described a combined computational-experimental method, CANDLE, which can accurately predict energetically relevant conformational ensembles of cyclic peptides in solution, with mean absolute deviations between predicted and experimental NMR chemical shifts of less than 0.2 ppm for protons and 2 ppm for carbons. These deviations are significantly lower than those reported by Zaretsky, et al., who also used DFT-based chemical shift predictions applied to conformational ensembles of peptide macrocycles. A subtle but important distinction between this study and the one by Zaretsky, et al., is in our use of an iterative, two phase approach that combines HT-MD, which is used as an initial filter to exclude irrelevant backbone conformations, and RT-MD, which is used to refine the structures at both the backbone and side chain levels. Lower-level single-point energies can thus be used to rank initial conformers, saving higher-level geometry optimizations and free energy calculations for a more focused conformer pool.

For now, the application of CANDLE is limited to relatively small systems such as the hexapeptide presented here. The computational expense associated with QM-based geometry optimization and free energy calculations increases dramatically with increasing size, and for larger peptides, the number of candidate conformers derived from HT-MD (or any other MM-based approach) will grow exponentially. Nonetheless, the conformational space even for cyclic hexapeptides is vast, thus motivating the development of new methods such as CANDLE to enable data-driven ranking of computer-generated conformers.

Since the side chains of cyclic peptides are relatively flexible, NOESY or ROESY data rarely provide enough distance restraints to yield information on side chain rotamer populations. Additionally, in peptides containing multiple aliphatic sidechains, as it common for cell permeable cyclic-peptides, overlap in the 1.0 to 2.5 ppm range can lead to cross-

peak ambiguity where individual chemical shifts can still be assigned using heteronuclear experiments (HMBC, HMQC). Although homo- and heteronuclear 3J -coupling constants can be used to deduce side chain rotamer populations in peptides and proteins,^{53, 54} the degeneracy of the Karplus curve introduces some ambiguity even when multiple 3J -couplings are used to constrain the same dihedral angle. Since the chemical shifts of side chain protons are also sensitive to the c -dihedrals, in principle vicinal 3J -couplings could be used in conjunction with CANDLE to corroborate each method's predictions.

aqueous solution alone. Furthermore, the "closed" structure of CSA in organic solvents has been invoked to rationalize its membrane permeability,^{45, 57} although the mechanistic details of its membrane penetration could be further illuminated by more detailed knowledge of its conformational states in both low- and high-dielectric media.⁵⁸ In principle CANDLE could be useful in untangling the conformations of CSA in solution, provided that its conformational space can be adequately sampled. The sampling problem is particularly challenging for large macrocycles, especially ones such as CSA, which contain multiple N-Me groups. Enhanced MD sampling methods⁵⁹ and

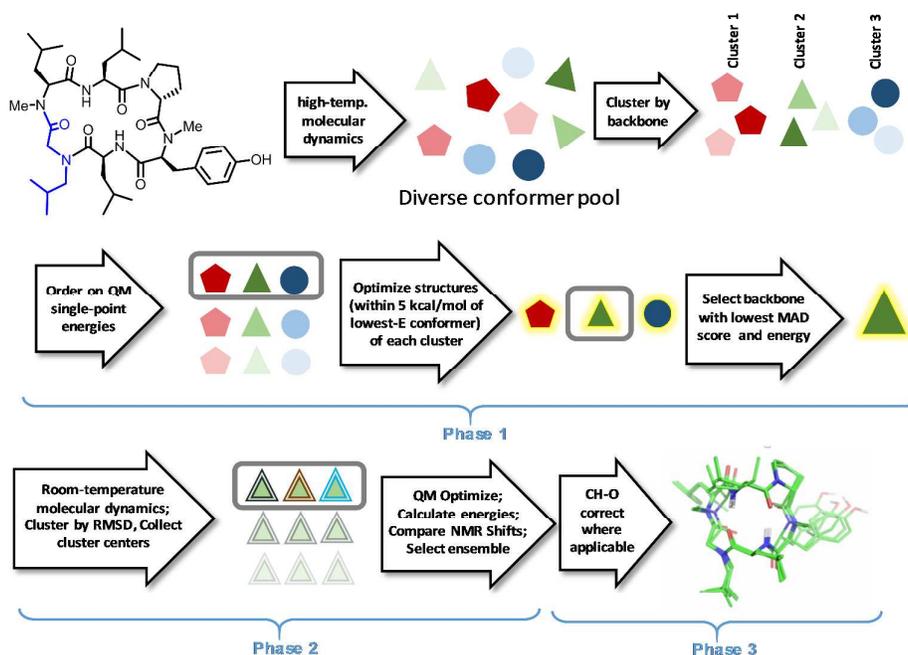


Fig 7: CANDLE workflow.

In **L2N-major**, whereas NOE-derived distances provided insufficient information to distinguish between similar β -turn types, CANDLE was able to disambiguate between the two conformer families. In principle, RDCs could complement NOE-based assignments to address this type of ambiguity, although the fact that the β -turn between Pep2 and Leu3 contains and NMe amide rather than an NH amide could limit the impact of RDCs in distinguishing between these similar backbone structures.

CANDLE requires the ability to assign all or most of the ^1H and ^{13}C chemical shifts in a spectrum. This is relatively straightforward even for complex molecules and/or very low-abundant conformers. Thus, in principle, CANDLE could be used to identify biologically relevant conformers even when present in low abundance. For example, the natural product cyclosporine A (CSA), exists as an ensemble of multiple conformations in water.⁵⁵ The "open" conformation of CSA bound to its target, cyclophilin, is presumably also present in the ensemble of conformers in aqueous solution,⁵⁶ although because of its complexity no NMR structures exist for CSA in

alternative conformational search algorithms, including low-mode MD,⁶⁰ Monte Carlo,⁶¹ kinematics-based approaches,^{62, 63} have been applied to increase the coverage of conformational space in macrocycles. Any of these approaches (or combinations thereof) could be used to generate the initial pool of backbone conformers in place of the MD-based approach described here, and such a modification may be required to achieve adequate coverage of conformational space for larger macrocycles. Alternatively, classical NMR-derived distance and dihedral information could be used to eliminate irrelevant backbone conformers from the initial pool. In some cases where large proton deviations can be attributed to CH-O interactions, additional calculations are proposed to show that the mismatch in chemical shifts are not due to the conformers being incorrectly identified, but rather are due to inaccurate CH-O distance predictions in the absence of explicitly modeled solvent molecules.

The CANDLE workflow includes these key features: assigning experimental NMR ^1H and ^{13}C chemical shifts for the compound of interest, running HT-MD simulations to generate

a diverse conformational library, clustering the conformations by backbone similarity, analyzing the energetic profile and NMR data with DFT calculations to find the best-match backbone, running RT-MD simulations to sample the side chain conformational landscape, and repeating the QM calculations and NMR analyses to obtain the final ensemble of relevant conformers (Fig. 7). CANDLE only requires the assignment of ^1H and ^{13}C resonances for a compound of interest, whereas techniques using NOE or RDCs begin with a full spectral assignment and then require additional multidimensional experiments and data processing. RDCs are technically challenging, requiring the use of a stretched gel.⁶⁴ While the NOESY or ROESY experiment can be easy to gather, the subsequent integration of through-space couplings can be tedious, compared to the relatively simple task of assigning chemical shifts. CANDLE thus represents a complimentary approach to traditional NMR-based methods, and may yield high-quality solution ensembles in cases where previous methods have been unsuccessful.

Conflicts of interest

R. Scott Lokey is co-founder of Circle Pharma, and Joshua Schwochert is co-founder of Unnatural Products.

Acknowledgements

Work conducted by Q.N. and D.T. was supported National Science Foundation (CHE-030089 [via XSEDE])

References

- 1 B. Over, P. McCarren, P. Artursson, M. Foley, F. Giordanetto, G. Gronberg, C. Hilgendorf, M. D. Lee, P. Matsson, G. Muncipinto, M. Pellisson, M. W. D. Perry, R. Svensson, J. R. Duvall and J. Kihlberg, *J. Med. Chem.*, 2014, **57**, 2746-2754.
- 2 T. Rezai, B. Yu, G. Millhauser, M. Jacobson and R. Lokey, *J. Am. Chem. Soc.*, 2006, **128**, 2510-2511.
- 3 J. G. Beck, J. Chatterjee, B. Laufer, M. U. Kiran, A. O. Frank, S. Neubauer, O. Ovadia, S. Greenberg, C. Gilon, A. Hoffman and H. Kessler, *J. Am. Chem. Soc.*, 2012, **134**, 12125-12133.
- 4 L. Doedens, F. Opperer, M. Cai, J. G. Beck, M. Dedek, E. Palmer, V. J. Hruby and H. Kessler, *J. Am. Chem. Soc.*, **132**, 8115-8128.
- 5 J. Boer, D. Gottschling, A. Schuster, M. Semmrich, B. Holzmann and H. Kessler, *J. Med. Chem.*, 2001, **44**, 2586-2592.
- 6 O. Demmer, A. O. Frank, F. Hagn, M. Schottelius, L. Marinelli, S. Cosconati, R. Brack-Werner, S. Kremb, H. J. Wester and H. Kessler, *Angew. Chem. Int. Ed. Engl.*, 2012, **51**, 8110-8113.
- 7 E. Marsault and M. L. Peterson, *J. Med. Chem.*, 2011, **54**, 1961-2004.
- 8 C. K. Wang, S. E. Northfield, B. Colless, S. Chaouis, I. Hamernig, R. J. Lohman, D. S. Nielsen, C. I. Schroeder, S. Liras, D. A. Price, D. P. Fairlie and D. J. Craik, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 17504-17509.
- 9 J. Schwochert, Y. Lao, C. R. Pye, M. R. Naylor, P. V. Desai, I. C. Gonzalez Valcarcel, J. A. Barrett, G. Sawada, M. J. Blanco and R. S. Lokey, *ACS Med. Chem. Lett.*, 2016, **7**, 757-761.
- 10 A. T. Bockus, K. W. Lexa, C. R. Pye, A. S. Kalgutkar, J. W. Gardner, K. C. Hund, W. M. Hewitt, J. A. Schwochert, E. Glassey, D. A. Price, A. M. Mathiowetz, S. Liras, M. P. Jacobson and R. S. Lokey, *J. Med. Chem.*, 2015, **58**, 4581-4589.
- 11 A. T. Bockus, J. A. Schwochert, C. R. Pye, C. E. Townsend, V. Sok, M. A. Bednarek and R. S. Lokey, *J. Med. Chem.*, 2015, **58**, 7409-7418.
- 12 A. Furukawa, C. E. Townsend, J. Schwochert, C. R. Pye, M. A. Bednarek and R. S. Lokey, *J. Med. Chem.*, 2016, **59**, 9503-9512.
- 13 I. Kolossvary and W. C. Guida, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 191-199.
- 14 X. N. Lai, C. P. Chen and N. H. Andersen, *Journal of Magnetic Resonance Series B*, 1993, **101**, 271-288.
- 15 J. G. Beck, A. O. Frank and H. Kessler, in *NMR of Biomolecules*, Wiley-VCH Verlag GmbH & Co. KGaA, 2012, DOI: 10.1002/9783527644506.ch19, pp. 328-344.
- 16 R. M. van Well, L. Marinelli, C. Altona, K. Erkelens, G. Segal, M. van Raaij, A. L. Llamas-Saiz, H. Kessler, E. Novellino, A. Lavecchia, J. H. van Boom and M. Overhand, *J. Am. Chem. Soc.*, 2003, **125**, 10822-10829.
- 17 O. Atasoylu, G. Furst, C. Risatti and A. B. Smith, 3rd, *Org. Lett.*, 2010, **12**, 1788-1791.
- 18 E. M. Larsen, M. R. Wilson, J. Zajicek and R. E. Taylor, *Org. Lett.*, 2013, **15**, 5246-5249.
- 19 M. T. Cung, M. Marraud and J. Neel, *Macromolecules*, 1974, **7**, 606-613.
- 20 M. Karplus, *J. Am. Chem. Soc.*, 1963, **85**, 2870-2871.
- 21 P. H. Willoughby, M. J. Jansma and T. R. Hoye, *Nat. Protoc.*, 2014, **9**, 643-660.
- 22 M. W. Lodewyk, C. Soldi, P. B. Jones, M. M. Olmstead, J. Rita, J. T. Shaw and D. J. Tantillo, *J. Am. Chem. Soc.*, 2012, **134**, 18550-18553.
- 23 D. J. Tantillo, *Nat. Prod. Rep.*, 2013, **30**, 1079-1086.
- 24 G. Saielli, K. C. Nicolaou, A. Ortiz, H. Zhang and A. Bagno, *J. Am. Chem. Soc.*, 2011, **133**, 6072-6077.
- 25 M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, *Chem. Rev.*, 2012, **112**, 1839-1862.
- 26 N. Grimblat and A. M. Sarotti, *Chemistry*, 2016, **22**, 12246-12261.
- 27 S. Di Micco, M. G. Chini, R. Riccio and G. Bifulco, *European J. Org. Chem.*, 2010, **2010**, 1411-1434.
- 28 S. Zaretsky, J. L. Hickey, M. A. St. Denis, C. C. G. Scully, A. L. Roughton, D. J. Tantillo, M. W. Lodewyk and A. K. Yudin, *Tetrahedron*, 2014, **70**, 7655-7663.
- 29 P. Maity, R. P. Pemberton, D. J. Tantillo and U. K. Tambar, *J. Am. Chem. Soc.*, 2013, **135**, 16380-16383.
- 30 J. R. Yates, T. N. Pham, C. J. Pickard, F. Mauri, A. M. Amado, A. M. Gil and S. P. Brown, *J. Am. Chem. Soc.*, 2005, **127**, 10216-10220.
- 31 D. A. Evans, M. J. Bodkin, S. R. Baker and G. J. Sharman, *Magn. Reson. Chem.*, 2007, **45**, 595-600.
- 32 F. A. Momany and R. Rone, *J. Comput. Chem.*, 1992, **13**, 888-900.
- 33 Gaussian 09, Revision D. 01 (Gaussian, Inc.: Wallingford, CT, USA, 2009). See supporting information for full citation.
- 34 C. C. J. Roothaan, *Reviews of Modern Physics*, 1951, **23**, 69-89.
- 35 M. S. Gordon, J. S. Binkley, J. A. Pople, W. J. Pietro and W. J. Hehre, *J. Am. Chem. Soc.*, 1982, **104**, 2797-2803.
- 36 A. D. Becke, *The Journal of Chemical Physics*, 1993, **98**, 5648-5652.
- 37 C. Lee, W. Yang and R. G. Parr, *Physical Review B*, 1988, **37**, 785-789.
- 38 R. Ditchfield, D. P. Miller and J. A. Pople, *The Journal of Chemical Physics*, 1971, **54**, 4186-4193.

- 39 Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215-241.
- 40 C. Adamo and V. Barone, *The Journal of Chemical Physics*, 1998, **108**, 664-675.
- 41 A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. B*, 2009, **113**, 6378-6396.
- 42 CHESHIRE. CHEMical SHift REpository with Coupling Constants Added Too. Available at <http://cheshirenmr.info/>.
- 43 M. Hassan, J. P. Bielawski, J. C. Hempel and M. Waldman, *Mol. Divers.*, 1996, **2**, 64-74.
- 44 J. Schwochert, R. Turner, M. Thang, R. F. Berkeley, A. R. Ponkey, K. M. Rodriguez, S. S. F. Leung, B. Khunte, G. Goetz, C. Limberakis, A. S. Kalgutkar, H. Eng, M. J. Shapiro, A. M. Mathiowetz, D. A. Price, S. Liras, M. P. Jacobson and R. S. Lokey, *Org. Lett.*, 2015, **17**, 2928-2931.
- 45 T. Rezai, B. Yu, G. L. Millhauser, M. P. Jacobson and R. S. Lokey, *J. Am. Chem. Soc.*, 2006, **128**, 2510-2511.
- 46 E. M. Larsen, M. R. Wilson, J. Zajicek and R. E. Taylor, *Org. Lett.*, 2013, **15**, 5246-5249.
- 47 F. A. Perras, D. Marion, J. Boisbouvier, D. L. Bryce and M. J. Plevin, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 7564-7567.
- 48 F. M. Raymo, M. D. Bartberger, K. N. Houk and J. F. Stoddart, *J. Am. Chem. Soc.*, 2001, **123**, 9264-9267.
- 49 P. Munshi and T. N. Guru Row, *J. Phys. Chem. A*, 2005, **109**, 659-672.
- 50 R. K. Castellano, *Curr. Org. Chem.*, 2004, **8**, 845-865.
- 51 G. R. Desiraju, *Acc. Chem. Res.*, 1996, **29**, 441-449.
- 52 R. C. Johnston and P. H.-Y. Cheong, *Org. Biomol. Chem.*, 2013, **11**, 5057-5064.
- 53 H. Kessler, C. Griesinger and K. Wagner, *J. Am. Chem. Soc.*, 1987, **109**, 6927-6933.
- 54 Y. Karimi-Nejad, J. M. Schmidt, H. Ruterjans, H. Schwalbe and C. Greisinger, *Biochemistry*, 1994, **33**, 5481-5492.
- 55 P. F. Augustijns, S. C. Brown, D. H. Willard, T. G. Consler, P. P. Annaert, R. W. Hendren and T. P. Bradshaw, *Biochemistry*, 2000, **39**, 7621-7630.
- 56 W. Lin, A. Quintero and Y. Zhang, *PLoS One*, 2016, **11**, e0153669.
- 57 C. K. Wang, J. E. Swedberg, P. J. Harvey, Q. Kaas and D. J. Craik, *J. Phys. Chem. B*, 2018, DOI: 10.1021/acs.jpcc.7b12419.
- 58 J. Witek, M. Muhlbauer, B. G. Keller, M. Blatter, A. Meissner, T. Wagner and S. Riniker, *Chemphyschem*, 2017, **18**, 3309-3314.
- 59 J. Witek, B. G. Keller, M. Blatter, A. Meissner, T. Wagner and S. Riniker, *J. Chem. Inf. Model.*, 2016, **56**, 1547-1562.
- 60 P. Labute, *J. Chem. Inf. Model.*, 2010, **50**, 792-800.
- 61 P. Hosseinzadeh, G. Bhardwaj, V. K. Mulligan, M. D. Shortridge, T. W. Craven, F. Pardo-Avila, S. A. Rettie, D. E. Kim, D. A. Silva, Y. M. Ibrahim, I. K. Webb, J. R. Cort, J. N. Adkins, G. Varani and D. Baker, *Science*, 2017, **358**, 1461-1466.
- 62 E. A. Coutsias, K. W. Lexa, M. J. Wester, S. N. Pollock and M. P. Jacobson, *J. Chem. Theory Comput.*, 2016, **12**, 4674-4687.
- 63 P. Hosseinzadeh, G. Bhardwaj, V. K. Mulligan, M. D. Shortridge, T. W. Craven, F. Pardo-Avila, S. A. Rettie, D. E. Kim, D.-A. Silva, Y. M. Ibrahim, I. K. Webb, J. R. Cort, J. N. Adkins, G. Varani and D. Baker, *Science*, 2017, **358**, 1461.
- 64 J.C. Freudenberger, P. Spitteller, R. Bauer, H. Kessler and B. Luy, *J Am Chem Soc*, 2004, **126**, 14690-14691.