Can Simulations and Modeling Decipher NMR Data for Conformational Equilibria? Arginine-Vasopressin

*Elke Haensele,[‡,†] Noureldin Saleh,[‡] Christopher M. Read,[§] Lee Banting,[†] David C. Whitley,[†] and Timothy Clark.[‡,]**

[‡] Computer-Chemie-Centrum der Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany.

[†] School of Pharmacy and Biomolecular Sciences, University of Portsmouth, Portsmouth PO1 2DT, United Kingdom.

[§] School of Biological Sciences, University of Portsmouth, Portsmouth PO1 2DY, United Kingdom.

e-mail: Tim.Clark@fau.de

Abstract: Arginine vasopressin (AVP) has been suggested by molecular-dynamics (MD) simulations to exist as a mixture of conformations in solution. The ^{1}H and ^{13}C NMR chemical shifts of AVP in solution have been calculated for this conformational ensemble of the ring conformations (identified from a 23 μs molecular-dynamics simulation). The relative free energies of these conformations were calculated using classical metadynamics simulations in explicit water. Chemical shifts for representative conformations were calculated using density-functional theory. Comparison with experiment and analysis of the results suggests that the ^{1}H

chemical shifts are most useful for assigning equilibrium concentrations of the conformations in this case. ^{13}C chemical shifts distinguish less clearly between conformations and the distances calculated from the nuclear Overhauser effect do not allow the conformations to be assigned clearly. The ^{1}H chemical shifts can be reproduced with a standard error of less than 0.24 ppm (< 2.2 ppm for ^{13}C). The combined experimental and theoretical results suggest that AVP exists in an equilibrium of approximately 70% *saddle*-like and 30% *clinched open* conformations. Both newly introduced statistical metrics designed to judge the significance of the results and Smith and Goodman's DP4 probabilities are presented.

Introduction

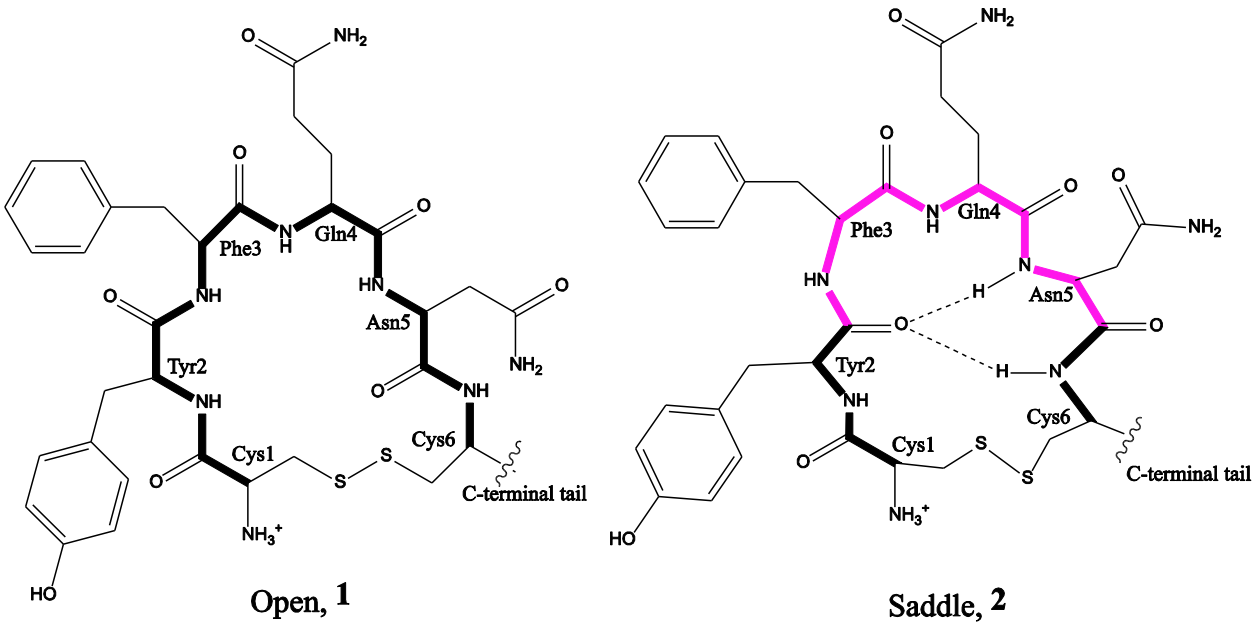
Many biologically important molecules, especially peptide hormones, exist as an equilibrium mixture of two or more conformations in solution.[1,2] Identifying these conformations and their relative free energies is important because, as long as the conformations in solution are competitive in energy then each is a candidate as the biologically active conformation, which need not be the same in all receptors.

X-ray crystallography usually only provides single snapshots that give little insight into dynamic equilibria, so that nuclear magnetic resonance (NMR) spectroscopy becomes the experimental method of choice. Unfortunately, the most common technique used to determine structures in solution, using the nuclear Overhauser effect (NOE)[3] is often not sufficient to determine even a single structure uniquely, and even less so for conformational equilibria. In this respect, the r^{-6} distance dependence of the NOE (r is the internuclear distance) prevents simple averaging of the structures and renders interpretation more difficult, even when molecular-dynamics (MD) simulations are used as the basis for ensemble calculations.[4]

Chemical shifts are not often used to determine conformations in solution because they are not directly related to interatomic distances. A reliable technique for calculating chemical shifts for a given geometry is thus needed and density-functional theory (DFT) calculations now provide such a technique at a reasonable computational cost.[5,6] When a regression equation was used to convert atomic screening to chemical shifts, accuracies of ± 0.15 ppm and ± 2.2 were obtained for ^{1}H and ^{13}C chemical shifts, respectively.[7] Unfortunately, what might naively be considered the most informative chemical shifts in peptides and proteins, those of the acidic ($pK_a \approx 15$) amide NH-protons often involved in hydrogen bonds, are also strongly affected by exchange

phenomena in aqueous solution.[8,9] These effects increase their chemical shifts compared to those calculated for the pure NH protonation state in continuum water. The inclusion of explicit water molecules in the DFT calculations can improve the results,[8,9] but in the case of vasopressin, a nonapeptide, this would lead to extensive sampling problems and make the technique computationally intractable. A further difficulty is that the superficially attractive technique of calculating the chemical shifts of the possible conformations in the equilibrium and fitting a linear combination to the experimental chemical shifts by regression lacks predictive power because the calculated chemical shifts of the conformations are strongly correlated, so that least-squares fits are seldom unique. This means that, although the fitted results are good, the coefficients of the individual conformations may not necessarily be meaningful because of their strongly correlated chemical shifts. This problem is most visible in bagging regression models, where the coefficients obtained in the different component models vary widely, but is also inherent in partial least squares models, where it is less obvious. These problems have been addressed by Smith and Goodman,[10,11] who used chemical shifts exclusively to distinguish between pairs of diastereomers and proposed improved metrics to overcome the fitting problem. Unfortunately, most of their metrics were designed to assist assignment of spectra to pairs of chiral molecules for which both experimental spectra are available. However, their DP4 probabilities[11] are applicable in the present case, as demonstrated by Nazarski et al.,[12] but even using these probabilities as a metric does not solve the problem of linearly dependent descriptors. We have therefore resorted to MD simulations to avoid the fitting problem. We have investigated the use of MD simulations and DFT chemical-shift calculations combined with NMR experiments to assign the conformational equilibrium in solution for 8-arginine-vasopressin (AVP), a flexible peptide hormone.

AVP is the human form of vasopressin, a peptide hormone of the vasopressin family. Vasopressin-related peptides, which include vasopressin, oxytocin, urotensin-II and a variety of other non-human tocins, are G-protein coupled receptor ligands that share the common feature of a six-residue ring closed by a disulfide bridge. Although the peptides are very closely related, the conformation of the six-residue ring differs in X-ray crystal structures of AVP (1YF4),[13] 8-lysine-vasopressin (1JK4)[14] and oxytocin (1NPO),[15] suggesting that multiple bioactive conformations may be operative, depending on the binding site.



Scheme 1: The *open* and *saddle* conformation-types for AVP. The ring backbone bonds are shown as broad lines and the β-turns in magenta.

The ring conformations for these peptide hormones can be classified broadly into *open* and *saddle* types, shown in scheme 1. The *open* ring conformations, **1**, such as that found in PDB-entry 1YF4, do not feature transannular hydrogen bonds and exhibit a flat, open ring structure. In contrast, the *saddle* conformations, **2**, (PDB entries 1JK4 and 1NPO) feature a ring that is folded

with possible transannular hydrogen bonds, resulting in a saddle-like shape that features well-defined β-turns at residues 3 and 4 and/or 4 and 5.

NMR studies of AVP[16,17] have concentrated on the *cis*/*trans*-isomerization across the Cys6-Pro7 peptide bond and have assumed only folded (*saddle*) ring conformations. The *trans*-isomer predominates in solution, although the *cis*-isomer can be identified in the NMR spectrum. It will not be discussed here because the *cis*/*trans*-interconversion is slow on the NMR time scale. Recent extensive molecular-dynamics (MD) simulations[18] suggest that AVP exists in an equilibrium between several interconverting ring-conformations in aqueous solution. The NMR studies summarized in Table 1 of reference 17 indicate that the ring can adopt diverse structures, all of which, however, have been interpreted as containing well-defined turns, as found in the saddle conformation. Exact knowledge of the ring conformational equilibrium is, however, important, as the biologically relevant conformations of AVP have not been identified. We therefore now report a combined theoretical (MD simulations, density-functional theory (DFT) modeling) and NMR study of the conformational equilibrium of AVP in aqueous solution that compares chemical shifts and interatomic distances calculated without experimental input with data derived from experiments.

Methods

Complete computational and experimental details are given in the Supporting Information; the procedure will only be described briefly here. Measured chemical-shift and NOE data are compared directly with those predicted essentially without experimental input. These predictions are based on:

1) Identifying the relevant conformations of AVP in solution from extended time-scale, unconstrained MD simulations:

Our previous[18] 11 μs MD simulation of AVP in solution was extended to 23 μs to improve sampling. Even this simulation, however, proved insufficient to deduce equilibrium concentrations of individual conformations, as identified using DASH.[19] We therefore, used the conformations identified in the 23 μs simulation to define the path variable for subsequent metadynamics simulations.[20]

2) Calculating the relative free energies of these conformations in solution using metadynamics:

The single path variable used for the metadynamics is simply a numerical assignment to one of the five most prevalent conformations found in the long MD simulation. These conformational assignments are made using the root-mean-square deviation (RMSD) from the individual conformations. This criterion allowed over 90% of the frames from the 23 μs simulation to be assigned. In order to make the collective variable as "physical" as possible, the numbering of the

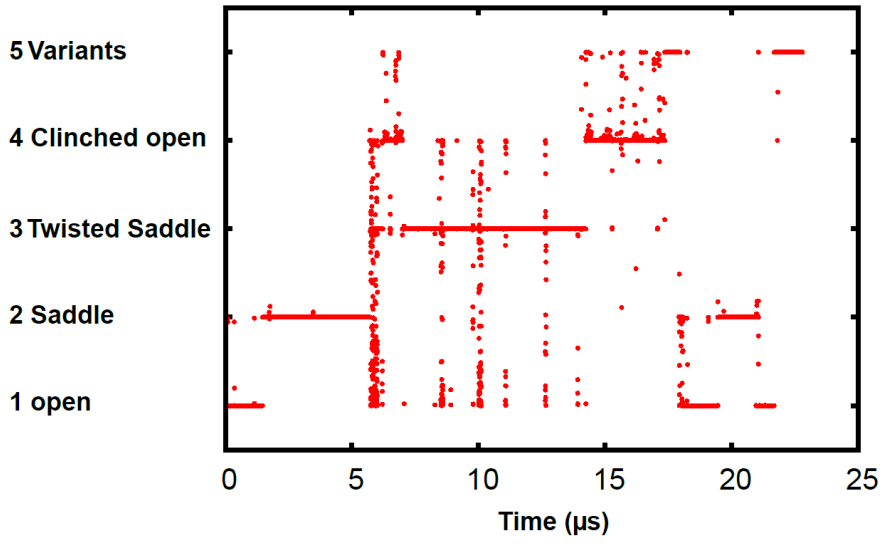


Figure 1: The numerical order of the conformations used in the metadynamics collective variable: the conformational assignments are plotted against simulation time for the five most populated DASH states observed in the 23 μs unconstrained simulation. The conformations 1-4 can interconvert as follows: 1-2, 2-3, 3-4. The direct 4-5 interconversion is also seen but conformations 5 were not included (see text).

conformations was chosen so that the 23 μs simulation exhibited transitions between adjacently numbered conformations, thus ensuring that paths between neighboring conformations exist.

Figure 1 shows the numerical assignment of the conformations. The "variants" cluster of conformations, which proved to be least stable and only occurred after the original 11 μs simulation, was not included in the further analysis (for details, see the Supporting Information).

3) Calculating geometries for cluster centers and NMR chemical shifts with DFT:

Cluster centers for the four most populated ring-conformations (including two different tail conformations for saddle and clinched open to give a total of six representative structures) were taken from the 23 μs simulation and optimized with Gaussian09[21] at the B3LYP[22,23]/6-31G(d)[24] level using the standard polarizable continuum model (PCM) for water.[25] The optimized geometries are given in the Supporting Information, Dispersion corrections were not used as we do not expect them to be appropriate for PCM calculations in a polar solvent. Note that this neglect of dispersion corrections can only affect the DFT-optimized geometries because the relative DFT energies are not used in the analysis. Relative free energies include dispersion because they were obtained exclusively from force-field based simulations with explicit solvent. Magnetic shieldings were calculated on the optimized structures using the gauge-independent atomic orbital (GIAO) technique[26] at the B3LYP/6-31G(d) level with PCM water. The regression technique for converting calculated isotropic magnetic shielding to chemical shifts in solution[7] was extended to enable B3LYP/6-31G(d) calculations with PCM-water to reproduce ^{1}H and ^{13}C chemical shifts in D_2O relative to (3-trimethylsilyl)propane sulfonic acid (DSS). Details of the training set and the results are given in the Supporting Information. The regression equations are:

$$\delta\left({}^{1}\mathrm{H}\right) = -0.9912\sigma_{H} + 32.05$$
$$\delta\left({}^{13}\mathrm{C}\right) = -1.0833\sigma_{C} + 203.97 \quad (1)$$

where δ is the chemical shift and σ the calculated isotropic atomic magnetic shielding, both in ppm. The root-mean-square deviations from experiment for the training set are 0.18 ppm for ^{1}H and 1.96 ppm for ^{13}C.

Chemical shifts for each optimized cluster-center conformation were calculated using Eq. (1) and ensemble chemical shifts (denoted as "equilibrium" in the following) obtained by linear combination of the individual shifts according to the calculated equilibrium concentrations. $^{1}H^{N}$ chemical shifts were not included, as in practice, these are subject to wide variation by hydrogen bonding, pH and solvent-based environmental changes and are generally not reproduced well by calculations on a single protonation state.

4) Direct comparison of experimental and calculated spectra or measurements:

The ensemble NMR spectra calculated in step (3) can be compared with experimental data. We have assigned the ^{1}H, and ^{13}C chemical-shifts almost fully, in two different aqueous solution conditions at pH 4.7 and pH 6.0. The former pH is that given on dissolving the peptide in H_2O and the latter was chosen to be compatible with the MD simulations. To complete the set of known experimental NMR data we report for the first time ^{15}N shifts at natural abundance. NOESY and TOCSY NMR spectra gave NOEs and facilitated assignment (see the Supporting Information for details).

Both the quality of fit between the calculated and experimental parameters and whether the fit for the calculated equilibrium mixture of conformations is better than that for any of the individual contributing conformations serve to validate the approach. This is often not a

straightforward analysis problem,[10,11,12] so that we have defined two statistical metrics below that are designed to test the significance of the differences in correlations of the chemical shifts calculated for individual conformations with the experimental data.

Results and Discussion

Unconstrained Molecular Dynamics

A 23 μs unrestrained MD simulation of Arg^8-vasopression was performed with explicit water-solvation at 300 K using the Amber ff99SB force field[27] (details are given in the Supporting Information). The conformational space was clustered using DASH[19] and compared with the conformations (clusters) found in the first 11 μs of the simulation.[18] These main clusters, *open, saddle, clinched open,* and *twisted saddle*, dominated the simulation (Figure 2). They have been described in detail.[18] Even after 23 μs, the simulation exhibited too few interconversions between the main clusters to estimate reliable equilibrium populations directly. Thus, we chose the representatives (cluster centers) of the four main clusters to calculate their free energies and relative populations with metadynamics. A fifth cluster, *variants*, which occurred for the first and only time at the end of the 23 μs simulation, was also added to the selection. A description of this cluster of conformations is given in the Supporting Information. The conformational clusters *open, saddle, clinched open,* and *twisted saddle* represent 86.4% of all conformations found for AVP in the simulation, and *variants* 7.4% to give a total of 93.8% that can be assigned to the five clusters. The rest are transient states not discussed here further. We showed[18] previously that the tail moves independently of the ring conformation of AVP, adopting either *folded* or *extended* conformations, which interconvert frequently and rapidly. Thus, it is possible to take the individual populations of these tail conformations directly from the 23 μs MD simulation.

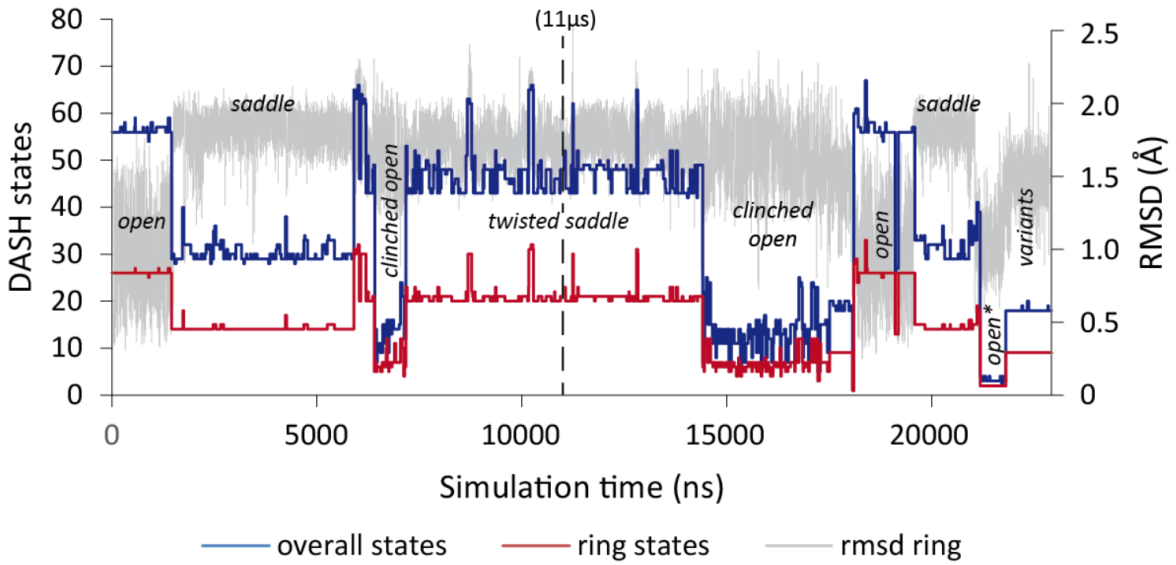


Figure 2: RMSD of $C^{\alpha}_{1\text{-}6}$ (gray), and the corresponding clusters of ring and overall conformations of 23 μs unrestrained MD simulation of Arg^{8}-vasopressin. The main clusters (ring conformations) are labeled.

Metadynamics

A well-tempered metadynamics simulation[28] using four walkers converged within 200 ns to give the relative free energies of the five conformations shown in Table 1.

These results can be compared with those obtained by least squares fitting the calculated NMR chemical shifts to observations, although the latter, as outlined above, may not be significant. The comparison therefore serves at best as a rough test as to whether the equilibrium concentrations obtained from the simulations are similar to those that would give the best fit. Figure 3 shows the equilibrium concentrations calculated from free-energy differences obtained in the metadynamics simulations and those obtained by fitting two different regression models to the experimental chemical shifts. As the calculated chemical shifts for the individual

conformations correlate strongly, fitting does not yield a robust statistical model, as demonstrated by the scatter in the fitted results. Whereas the regression models differ as to whether the *saddle* or *clinched open* conformation is the most prevalent in the solution equilibrium, the metadynamics results indicate that the population of the saddle conformation is highest.

Table 1: Equilibrium populations and relative free energies (ΔΔG) from the metadynamics simulation. The ΔΔG values are converged to approximately ± 0.2 kcal mol^{-1}. The equilibrium concentrations are given at 298K. Errors are based on ± 0.2 kcal mol^{-1} energetic uncertainty and are given as ± one standard deviation.

	Saddle	**Clinched open**	**Twisted saddle**	**Open**	**Variants**
ΔΔG (kcal mol^{-1})	0.0	0.5	3.0	2.0	3.5
% at equilibrium (5 conformations)	68.5	29.5	0.4	1.4	0.2
% at equilibrium (4 conformations)	68.7±3.9	29.5±4.0	0.4±0.1	1.4±0.5	

The fitted equilibrium concentration can serve, however, as a reality check for the metadynamics results. The metadynamics equilibrium is quite compatible with the optimum PLS-fits for this dataset (Figure 3), which is encouraging, and we emphasize once more that, in contrast to the regression data, those calculated for the metadynamics equilibrium use essentially no experimental data. The exception is the standard set of chemical shifts used to obtain Eq. (1) to convert shielding to ppm. However, the training dataset (given in the Supporting Information) only contains small organic molecules, which can be considered independent of AVP. The conformations were identified from the 23 μs MD-simulation, the chemical shifts calculated for

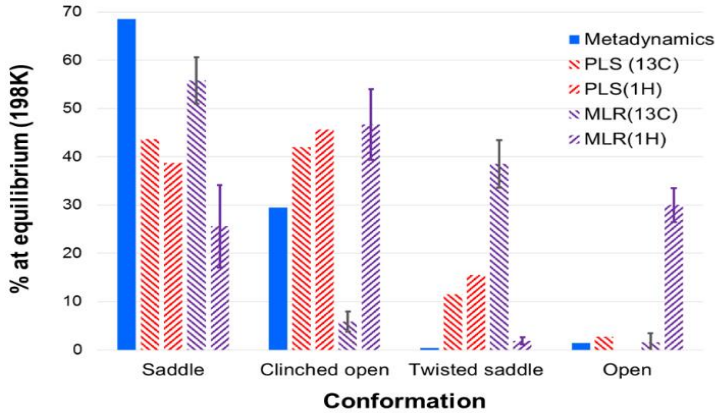


Figure 3: Calculated equilibrium concentrations (%, 298K) for the saddle, clinched open, open and twisted saddle conformations. The fitted values are taken from partial least squares (PLS) and bagged multiple linear regression (MLR) fits. The variants conformations are not found to be significant. Bagged MLR and PLS calculations were performed with SAR-caddle.[29] The error bars given for the bagged MLR results are the standard deviations of five fitting runs.

B3LYP/6–31G(d)-optimized geometries and the equilibrium calculated from the free energies obtained from the metadynamics simulations.

Figure 4 shows the B3LYP/6-31G(d) (in PCM water) optimized structure of the major saddle conformation. The C-terminal tail adopts two conformations.[16] The extended conformation, which positions the guanidinium moiety of Arg^{8} close to the ring was present in the 23 μs MD simulation for approximately 73% of the occurrence time for the *saddle* conformation (Fig. 4a). The folded tail conformation (Fig. 4b) makes up the remaining 27%. In this case, error estimates are difficult because probable errors depend on how well the simulation has converged, which is unknown. We estimate from the length and convergence of the simulation that the above concentrations have uncertainties of at most ±5%. The equilibrium between these two tail

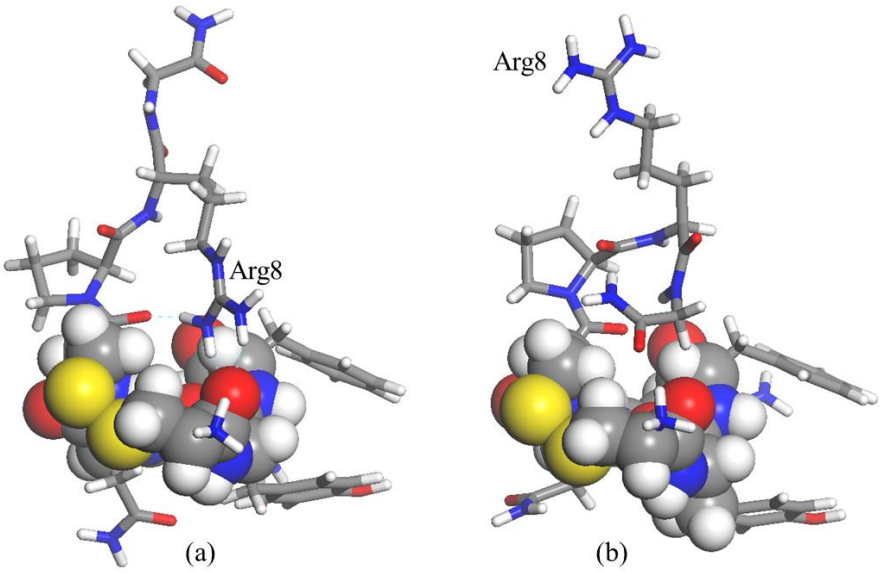


Figure 4: Optimized structures of the saddle conformation obtained at the B3LYP/6-31G(d) level in PCM water solvent. The ring atoms as spheres: (a) the extended tail conformation, (b) the folded equivalent.

conformations is fast on the simulation time scale, so that we can refine the calculation of the NMR chemical shifts by treating the saddle conformation as a 73:27 mixture of the two conformations shown in Figure 4. The *clinched open* conformation is treated similarly (63% extended: 37% folded, see the Supporting Information). This results in some improvements in the agreement between calculations and experiment, as shown in Table 2 below.

Figure 5 shows plots of the results of the final computational model compared with experiment

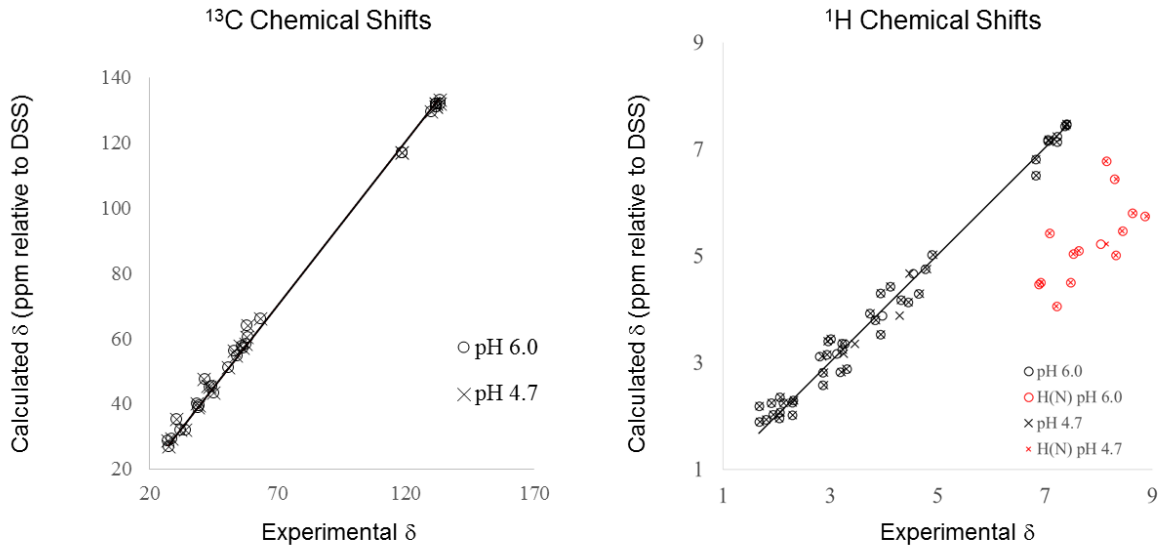


Figure 5: Plots of the calculated vs. experimental ^{13}C and ^{1}H chemical shifts using the equilibrium model for both ring and tail conformations. δ (ppm) are relative to DSS (3-(trimethylsilyl)propane sulfonic acid. The NH-protons are outliers because of hydrogen bonding and exchange effects.[8,9]

for both ^{13}C and ^{1}H chemical shifts.

Is the statistical analysis for the calculated equilibrium significant? This question has been addressed several times in the literature.[12,30,31] These studies have been summarized by Smith and Goodman,[10,11] who also proposed improved metrics for judging the goodness of fit between calculated and experimental chemical shifts. As outlined above, many of their metrics were designed to assist assignment of spectra to pairs of diastereomers for which both experimental spectra are available and are inapplicable in this case. We have resorted to conventional metrics such as mean signed (MSE) and unsigned error (MUE), coefficient of determination (R^2) and root-mean-square error (RMSE) as a specific test of the significance of the conclusions, but have

also defined a weighted RMSE (WRMSE) in the spirit of Smith and Goodman and have used their DP4 probability[11] as an additional check.

The WRMSE is defined as;

$$WRMSE = \frac{\sqrt{\sum_{i=1}^{N} \left(\hat{y}_i - y_i\right)^2 \sigma_i}}{\sqrt{\sum_{j=1}^{N} \sigma_j}} \tag{2}$$

where $\hat{y}_i$ and y_i are the predicted and observed chemical shifts for atom *i*, respectively, and σ_i is the standard deviation of the calculated chemical shifts for atom *i* over all conformations. WRMSE is equivalent to RMSE if all σ_i are equal and otherwise weights the contributions of the atoms that display a wide range of chemical shifts between the conformations more heavily than those with little variation.

A second specific test of the significance of the conclusions is the mean absolute error expressed in units of the standard deviation over all conformations, Δ_σ :

$$\Delta_\sigma = \frac{\sum_{i=1}^{N} \frac{\left|\hat{y}_i - y_i\right|}{\sigma_i}}{N} \tag{3}$$

Δ_σ expresses the significance of the MUE in terms of the total spread of calculated chemical shifts for the individual conformations. Ideally $\Delta_\sigma \leq 1$ indicates that on average the deviation between experimental and calculated results is below the standard deviation between the different conformations; the model can discriminate between conformations. We arbitrarily assign a limit of $\Delta_\sigma \leq 1.5$ to indicate reliable discrimination between conformations. The results are shown in Table 2.

Table 2: Statistics of the comparison of ^{13}C and ^{1}H chemical shifts for AVP at pH 6.0 and 4.7 in aqueous solution. The best performing model is indicated in bold for each parameter. The amide protons are omitted, as outlined in the text.

Conformation		MSE	MUE	RMSE	WRMSE	Δ_σ	R^2
Ring	**Tail**						
				^{13}C, pH 6.0			
Saddle	extended	0.87	1.69	2.33	2.74	1.40	0.9965
	folded	**0.52**	1.75	2.52	3.18	1.26	0.9958
	equilibrium	0.78	1.62	2.23	2.68	1.32	0.9968
Clinched open	extended	0.74	2.27	3.15	3.75	1.71	0.9936
	folded	0.78	2.18	2.94	3.48	1.71	0.9943
	equilibrium	0.76	2.16	2.98	3.56	1.65	0.9942
Twisted saddle	extended	0.73	1.70	2.23	2.66	1.42	0.9969
Open	extended	1.18	2.49	3.72	5.24	1.93	0.9807
Equilibrium	extended	0.84	1.55	2.19	2.50	1.34	0.9969
Equilibrium	equilibrium	0.78	**1.46**	**2.12**	**2.45**	**1.26**	**0.9971**
				^{13}C, pH 4.7			
Saddle	extended	0.95	1.73	2.37	2.74	1.47	0.9964
	folded	**0.60**	1.76	2.55	3.19	**1.31**	0.9957
	equilibrium	0.85	1.66	2.26	2.68	1.39	0.9967
Clinched open	extended	0.82	2.36	3.28	3.93	1.79	0.993
	folded	0.85	2.28	3.09	3.67	1.80	0.9938
	equilibrium	0.83	2.26	3.13	3.74	1.74	0.9937
Twisted saddle	extended	0.80	1.78	2.32	2.75	1.50	0.9966
Open	extended	1.25	2.56	3.77	5.29	2.00	0.9904
Equilibrium	extended	0.91	1.63	2.28	2.58	1.43	0.9967

Equilibrium	equilibrium	0.85	**1.54**	**2.20**	**2.54**	1.35	**0.9969**
			^{1}H, pH 6.0				
Saddle	extended	0.06	0.22	0.31	0.37	1.02	0.9706
	folded	0.02	0.31	0.38	0.42	1.43	0.9571
	equilibrium	0.05	0.23	0.29	0.33	1.03	0.9748
Clinched open	extended	0.05	0.22	0.28	0.30	1.13	0.9773
	folded	-0.02	0.33	0.43	0.51	1.57	0.9441
	equilibrium	0.02	0.20	0.25	0.26	1.11	0.9800
Twisted saddle	extended	-0.03	0.42	0.58	0.79	1.62	0.9486
Open	extended	**0.01**	0.30	0.48	0.68	1.09	0.9674
Equilibrium	extended	0.05	0.20	0.26	0.30	0.93	0.9793
Equilibrium	equilibrium	0.04	**0.18**	**0.23**	**0.25**	**0.93**	**0.9832**
			^{1}H, pH 4.7				
Saddle	extended	0.04	0.23	0.32	0.37	1.03	0.9692
	folded	**0.00**	0.31	0.38	0.43	1.43	0.9562
	equilibrium	0.02	0.23	0.29	0.34	1.04	0.9735
Clinched open	extended	0.02	0.22	0.29	0.31	1.12	0.9753
	folded	-0.04	0.33	0.43	0.51	1.54	0.9446
	equilibrium	**0.00**	0.21	0.26	0.27	1.12	0.9789
Twisted saddle	extended	-0.03	0.42	0.58	0.79	1.62	0.9496
Open	extended	0.01	0.30	0.48	0.68	1.09	0.9672
Equilibrium	extended	0.03	0.21	0.27	0.31	0.95	0.9777
Equilibrium	equilibrium	0.02	**0.19**	**0.24**	**0.26**	**0.94**	**0.9820**

Surprisingly, the ^{1}H chemical shifts give the clearest and most consistent picture; they indicate that the experimental ^{1}H shifts are best in agreement with the equilibrium model that uses metadynamics free-energy differences for the ring conformations and equilibrium concentrations

from the unconstrained simulation for the faster tail equilibrium. This model is quite consistently the best for ^{13}C; only Δ_σ at pH 4.7 indicates the saddle conformation with a folded tail to fit better than the calculated equilibrium. However, WRMSE is always larger than RMSE and $\Delta\sigma$ approximately 1.3, so that we must conclude that the ^{13}C chemical shifts are not sensitive enough to conformation to allow us to assign values to the conformational equilibrium unequivocally.

The situation for the ^{1}H chemical shifts is clearer; with the exception of the MSE, all metrics indicate that the model that uses the metadynamics free energies for the ring conformations and the distributions of the tail conformations from the 23 μs unconstrained simulation matches the experimental data best. Most importantly, in contrast to the ^{13}C results, WRMSE is close to RMSE for the equilibrium models and $\Delta\sigma$ is less than one.

Table 3 shows that Smith and Goodman's DP4 probabilities[11] provide very strong support for these conclusions.

Table 3: DP4 probabilities for the AVP conformations at pH 6.0 and 4.7 in aqueous solution. The best performing model is indicated in bold. The probabilities were calculated using the data from the Supporting Information with the DP4 app.[11] The amide protons are omitted, as outlined in the text.

Conformation		**pH 6**			**pH 4.7**		
Ring	**Tail**	**^{13}C**	**^{1}H**	**^{13}C and ^{1}H**	**^{13}C**	**^{1}H**	**^{13}C and ^{1}H**
Saddle	Extended	1.7	0.0	0.0	4.7	0.0	0.0
	Folded	0.0	0.0	0.0	0.1	0.0	0.0
Clinched open	Extended	0.0	0.0	0.0	0	0.0	0.0
	Folded	0.0	0.0	0.0	0	0.0	0.0
Twisted saddle	Extended	0.3	0.0	0.0	0.2	0.0	0.0
Open	Extended	0.0	0.0	0.0	0.0	0.0	0.0
Saddle	Equilibrium	4.2	0.0	0.0	12.7	0.0	0.0
Clinched open	Equilibrium	0.0	0.2	0.0	0.0	0.5	0.0
Equilibrium	Extended	19.8	0.2	0.0	19.1	0.2	0.1
	Equilibrium	**74.0**	**99.6**	**100.0**	**63.2**	**99.3**	**99.9**

The DP4 probabilities lead to exactly the same conclusions as the metrics reported in Table 2. The conformational model that considers the equilibrium distributions of both the ring and the tail fits the experimental data best and ^{1}H chemical shifts allow firmer conclusions than ^{13}C. However, the DP4 probabilities also allow tentative conclusions to be reached from the ^{13}C chemical shifts; the equilibrium conformational model gives a 60-75% probability of being correct, although this probability is close to 100% for ^{1}H.

As outlined above, the differences in the statistical metrics would not be as convincing if they were based on a fitting procedure. However, as the identification of possible conformations, the calculation of equilibrium concentrations and the chemical-shift calculations are all *ab initio*, in the sense that they are completely independent of experimental data (with the exception of the regression equations (1)), we consider the quality of the agreement between experimental and calculated chemical shifts to be significant. RMSEs lower than 0.24 ppm for ^{1}H (without ^{1}NH) and 2.2 for ^{13}C are as good as, or better than, those reported previously using a variety of techniques,[3,4,6,7] and these values are only slightly larger than the standard errors obtained for the training set of small molecules (0.18 and 1.96 ppm for ^{1}H and ^{13}C, respectively). In order to strengthen these conclusions, we have carried out a sensitivity analysis to see how sensitive WRMSE and $\Delta\sigma$ are to the equilibrium concentrations. For this analysis, we used both a binary mixture of the majority saddle and clinched open conformations (WRMSE and $\Delta\sigma$) and the full equilibrium with four components (WRMSE′ and $\Delta\sigma'$).

Sensitivity Analysis

Figure 6 shows the dependence of WRMSE and $\Delta\sigma$ on the percentage of the saddle conformation in the binary mixture. Both react quite sensitively to the concentrations at

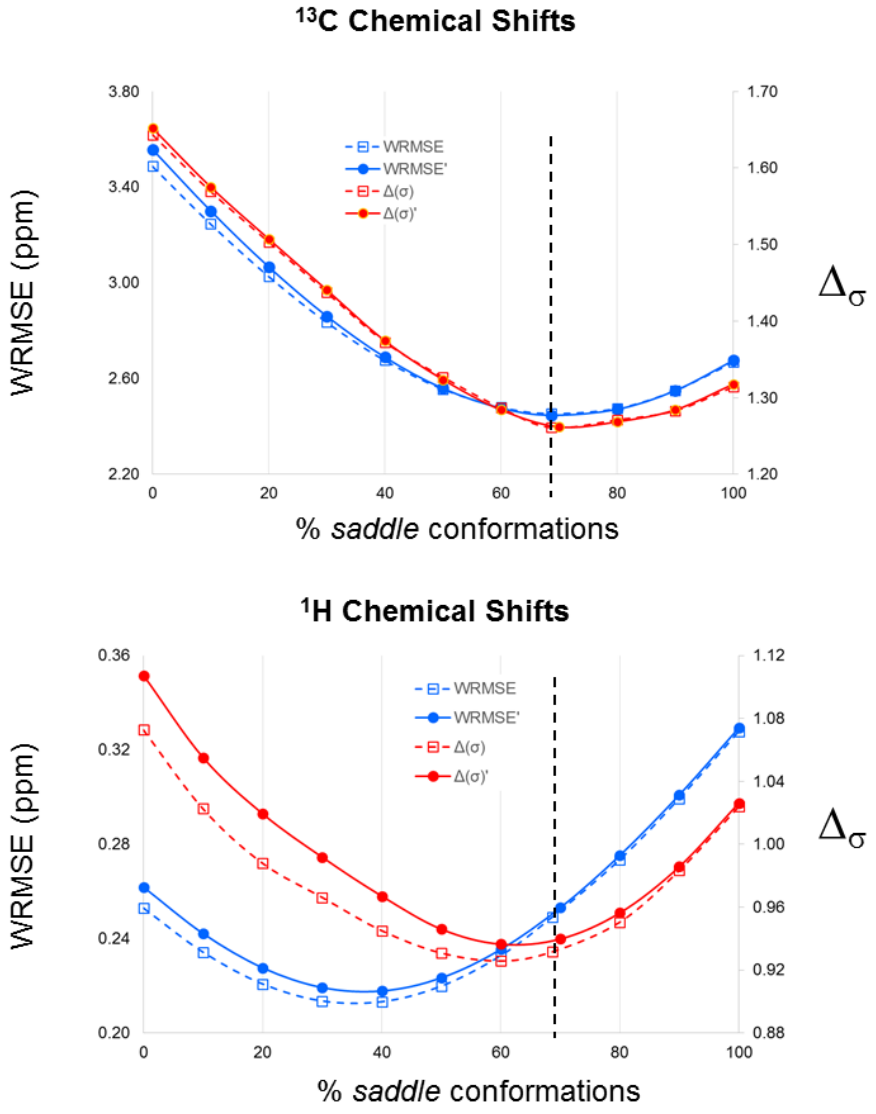


Figure 6: The dependence of WRMSE and $\Delta\sigma$ on the concentrations in mixtures of *saddle* and *clinched open* conformations at pH 6.0. The vertical dashed lines indicate the metadynamics equilibrium. WRMSE and $\Delta\sigma$ refer to the binary mixture and WRMSE′ and $\Delta\sigma'$ to the four-component equilibrium. The corresponding plots for pH 4.7 are very similar.

equilibrium and exhibit clear minima. For ^{13}C, the two curves correspond closely with a common minimum at the metadynamics values of 70% *saddle* and 30% *clinched open*. The two metrics agree less well for the ^{1}H data; WRMSE gives a minimum at approximately 35% *saddle* and $\Delta\sigma$

at approximately 60%. As three of four metrics give minima close to the metadynamics prediction, we feel that Figure 6 supports our conclusions.

Nuclear Overhauser Effect

An independent check of the conformational assignment compares the interatomic distances provided by nuclear Overhauser effect (NOE) data with those obtained from the simulations (details of the calculations are given in the Supporting Information). The correlation obtained for the observed NOEs and the statistics of the agreement between experiment and simulation are shown in Table 4. At pH 4.7, highest R^2 (0.622) is found for the saddle conformation with folded tail but this model is not favored clearly by any other metric. The metadynamics equilibrium considering the tail conformation is always close to the best values found but the differences are not significant. All conformations perform similarly (there are, for instance, five conformations with an RMSE of 0.39 Å). The saddle and clinched open conformations with the extended tail conformation give the best coefficients of determination (0.553) but the data are in general inconclusive. The small number of NOE distances available at pH 6.0 also does not allow a definitive conformational determination but tend towards *clinched open* with the extended tail conformation. Thus, the NOE simulations are compatible with the chemical-shift results but not definitive. These results illustrate the difficulties pointed out by Zagrovic and van Gunsteren[2] that NOE studies can, in fact, often be ambiguous; especially for highly flexible structures where intramolecular hydrogen bond distances may "average" by fast equilibria of different conformations.

Table 4: Statistics of the comparison of calculated and observed NOE distances for AVP at pH 6.0 and 4.7 in aqueous solution. The best performing model is indicated in bold for each

parameter. Details of the derivation of both experimental and simulated distances are given in the Supporting Information.

Conformation		pH 4.7				pH 6.0			
Ring	**Tail**	**MSE**	**MUE**	**RMSE**	**R^2**	**MSE**	**MUE**	**RMSE**	**R^2**
Saddle	extended	0.33	0.56	0.74	0.549	-0.12	0.36	0.45	**0.553**
	folded	**-0.33**	**0.52**	**0.68**	**0.622**	-0.16	0.37	0.48	0.084
	equilibrium	**0.33**	0.55	0.71	0.575	-0.11	0.32	**0.39**	0.176
Clinched open	extended	0.41	0.56	0.81	0.370	**-0.06**	**0.31**	**0.39**	**0.553**
	folded	-0.38	0.56	0.80	0.417	-0.07	0.32	0.41	0.514
	equilibrium	0.40	0.56	0.80	0.395	**-0.06**	**0.31**	**0.39**	0.543
Twisted saddle	extended	0.37	0.55	0.75	0.527	-0.13	0.35	0.45	0.340
Open	extended	0.32	0.56	0.72	0.551	-0.11	0.35	0.44	0.337
Equilibrium	extended	0.36	0.54	0.72	0.533	-0.11	0.32	**0.39**	0.366
Equilibrium	equilibrium	0.36	0.53	0.71	0.557	-0.12	0.33	0.40	0.344

Figure 7 shows the correlation between experimental at pH 4.7 and r^{-6} time-averaged interatomic distances for the metadynamics equilibrium.

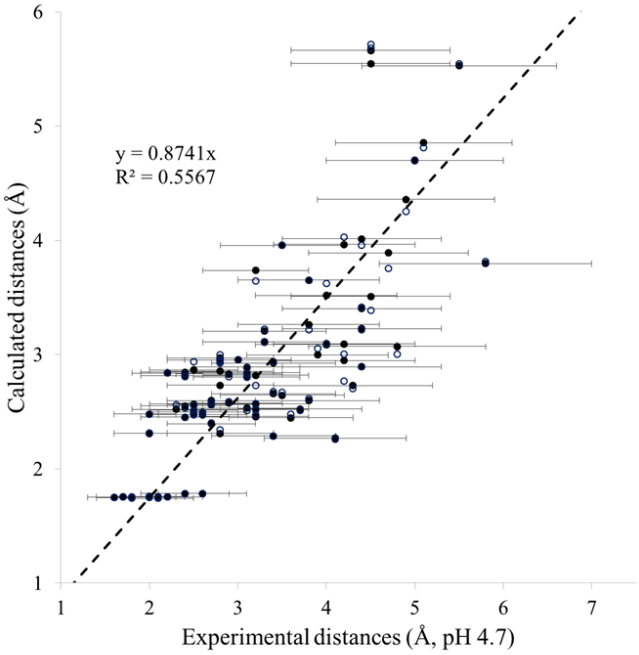


Figure 7: Plots of the calculated vs. experimental interatomic distances at pH 4.7 (cf. Figure S3).

Conclusions

We have reported an attempt to assign conformations for the equilibrium structures of AVP in aqueous solution by simulating the equilibrium and comparing calculated chemical shifts directly with experiment. This procedure avoids fitting and uses only minimal unconnected experimental data to parameterize the regression equation for the calculated chemical shifts. Our models reproduce the experimental data very well (RMSE < 0.24 ppm for ^{1}H and < 2.2 ppm for ^{13}C) but the question remains as to whether the agreement is significant enough to allow conclusions about the equilibrium mixture of conformations.

The proton NMR results present the strongest argument, even though amide protons cannot be included because they are shifted from the calculations for the pure NH-protonation state by exchange. The ^{13}C data are reproduced well, but the diagnostic metrics are not as clear,

indicating that the ^{13}C chemical shifts are less sensitive to conformation than ^{1}H and therefore less suitable for our purpose.

The calculated equilibrium concentrations are, however, comparable to those found for an optimal fit, so that we can be confident that they are close to reality, although the regression models suffer from strongly correlated descriptors.

We conclude that the conformational equilibrium for AVP in aqueous solution consists of approximately 70% *saddle*, 30% *clinched open* conformations and that the free-energy penalty for *clinched open* as a biologically active conformation is approximately 0.5 kcal mol^{-1}. It is conceivable that the folded, saddle-like type of conformations comprises a higher amount of *twisted saddle* than predicted by metadynamics. In fact, the representative conformations of *saddle* and *twisted saddle* are closely related; they only differ in the turn type of the β-turn at residues 3 and 4. This is also reflected in a very high correlation of the ^{13}C chemical shifts for saddle and twisted saddle (R^2=0.997) in contrast to ^{1}H (R^2=0.944). A similar sensitivity analysis to that shown in Figure 6 indicates that the ^{13}C data are compatible with mixtures from 10% to 50% *twisted saddle* in the saddle-like component of the equilibrium and the ^{1}H data with approximately 70:30 *saddle*:*twisted saddle*. We are currently unable to resolve this discrepancy between long unbiased simulation and metadynamics. In any case, all data are consistent with the conservative conclusion that the equilibrium consists of 70% *saddle*-type and 30% *open*-type conformations (Scheme 1).

One important result of this work is to show that modern MD-simulations and DFT calculations provide data that can be compared directly with experiment without fitting. In this respect, as suggested by Smith and Goodman,[10,11] chemical shifts prove to be more useful than NOEs and, surprisingly, in this example ^{1}H chemical shifts present a clearer picture than ^{13}C, as

also found by Nazarski et al.[12] Smith and Goodman's DP4 probabilities[11] suggest very clearly that, of those considered, our equilibrium model agrees best with experiment.

The methodology used does not require the unconstrained MD simulation to be long enough to be able to determine equilibrium concentrations. Its function is to identify the conformations (and the transitions between them) for subsequent determination of the free energy differences, here metadynamics simulations. For this reason, and for economy of computer time, we have used the cluster centers for each conformation, rather than calculating shifts for a large number of snapshots in an ensemble model.

Acknowledgments

This work was supported by the *Deutsche Forschungsgemeinschaft* within GRK 1910 *Medicinal Chemistry of Selective GPCR ligands* and by a large grant of computer time on SuperMUC at the *Leibniz Rechenzentrum München* (project pr94to).

Associated Content

Supporting Information Available: Computational and experimental (NMR) details and Gaussian Archive Entries for the B3LYP/6-31G(d)-optimized geometries. This material is available free of charge via the Internet at http://pubs.acs.org.

Notes and references

1. Hagler, A. T.; Osguthorpe, D. J.; Dauber-Osguthorpe, P.; Hempel, J. C., Dynamics and Conformational Energetics of a Peptide Hormone: Vasopressin. *Science* **1985**, *227*, 1309-1315.

2. Hruby, V. J.; Al-Obeidi, F.; Kazmierski, W., Emerging Approaches in the Molecular Design of Receptor-Selective Peptide Ligands: Conformational, Topographical and Dynamic Considerations. *Biochem. J.* **1990**, *268*, 249-262.

3. Campbell, A. P.; Sykes, B. D., The 2-Dimensional Transferred Nuclear Overhauser Effect - Theory and Practice. *Ann. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 99-122.

4. Zagrovic, B.; van Gunsteren, W. F., Comparing Atomistic Simulation Data with the NMR Experiment: How Much Can NOEs Actually Tell Us? *Proteins: Struct., Funct., Bioinf.* **2006**, *63*, 210-218.

5. Flaig, D.; Maurer, M.; Hanni, M.; Braunger, K.; Kick, L.; Thubauville, M.; Ochsenfeld, Benchmarking Hydrogen and Carbon NMR Chemical Shifts at HF, DFT, and MP2 Levels. *J. Chem. Theory Comput.* **2014**, 10, 572-578.

6. Zhu, T.; He, X.; Zhang, J. Z. H., Fragment Density Functional Theory Calculation of NMR Chemical Shifts for Proteins with Implicit Solvation. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7837-7845.

7. van Eikema Hommes, N. J. R.; Clark, T., Regression Formulae for ab initio and Density Functional Calculated Chemical Shifts. *J. Mol. Model.* **2005**, *11*, 175-185.

8. Zhu, T.; Zhang, J. Z. H.; He, X., Automated Fragmentation QM/MM Calculation of Amide Proton Chemical Shifts in Proteins with Explicit Solvent Model. *J. Chem. Theory Comput*. **2013**, *9*, 2104-2114.

9. Exner, T. E.; Frank, A.; Onila, I.; Möller, H. M., Toward the Quantum Chemical Calculation of NMR Chemical Shifts of Proteins. 3. Conformational Sampling and Explicit Solvents Model. *J. Chem. Theory Comput.* **2012**, *8*, 4818-4827.

10. Smith S. G.; Goodman J. M., Assigning the Stereochemistry of Pairs of Diastereoisomers Using GIAO NMR Shift Calculation. *J. Org. Chem.* **2009**, *74*, 4597-4607.

11. Smith S. G.; Goodman J. M., Assigning Stereochemistry to Single Diastereoisomers by GIAO NMR Calculation: The DP4 Probability. *J. Am. Chem. Soc.* **2010**, *132*, 12946-12959.

12. Nazarski, R. B.; Pasternak, B.; Leśniak, S., Three-Component Conformational Equilibria of some Flexible Pyrollidin-2-(thi)ones in Solution by NMR Data (δ_C, δ_H, and $^nJ_{HH}$) and their DFT Predictions: A Confrontation of Different Approaches. *Tetrahedron* **2011**, *67*, 6901-6916.

13. Ibrahim B. S.; Pattabhi, V., Trypsin Inhibition by a Peptide Hormone: Crystal Structure of Trypsin-Vasopressin Complex. *J. Mol. Biol.* **2005**, *348*, 1191-1198.

14. Wu, C. K.: Hu, B.; Rose, J. P.; Liu, Z. J.; Nguyen, T. L.; Zheng, C.; Breslow E.; Wang, B. C., Structures of an Unliganded Neurophysin and its Vasopressin Complex: Implications for Binding and Allosteric Mechanisms. *Prot. Sci.* **2001**, *10*, 1869-1880.

15. Rose, J. P.; Wu, C. K.; Hsiao, C. D.; Breslow E.; Wang, B. C., Crystal Structure of the Neurophysin-Oxytocin Complex. *Nature Struct. Biol*. **1996**, *3*, 163-169.

16. Schmidt, J. M.: Ohlenschlager, O.; Ruterjans, H.; Grzonka, Z.; Kojro, E.; Pavo I.; Fahrenholz, F., Conformation of [8-Arginine]Vasopressin and V1 Antagonists in Dimethyl Sulfoxide Solution Derived from Two-Dimensional NMR Spectroscopy and Molecular Dynamics Simulation. *Eur. J. Biochem.* **1991**, *201*, 355-371.

17. Sikorska E.; Rodziewicz-Motowidło S., Conformational Studies of Vasopressin and Mesotocin using NMR Spectroscopy and Molecular Modelling Methods. Part I: Studies in Water. *J. Pept. Sci.* **2008**, *14*, 76-84.

18. Haensele, E.; Banting, L.; Whitley D. C.; Clark, T., Conformation and Dynamics of 8-Arg-Vasopressin in Solution. *J. Mol. Model.* **2014**, *20*, 2485.

19. Salt, D. W.; Hudson, B. D.; Banting, L.; Ellis M. J.; Ford, M. G., DASH: a Novel Analysis Method for Molecular Dynamics Simulation Data. Analysis of Ligands of PPAR-gamma. *J. Med. Chem*. **2005**, *48*, 3214-3220.

20. Barducci, A.; Bonomi, M.; Parrinello, M., Metadynamics, *WIRES Comp. Mol. Sci.* **2011**, *1*, 826-843.

21. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, M.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A. J.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09, Revision C.01*; Gaussian, Inc.: Wallingford CT, **2010**.

22. Becke, A. D., Density-Functional Thermochemistry. III. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

23. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B*. **1988**, *37*, 785-789.

24. Ditchfield, R.; Hehre, W. J.; Pople, J. A., Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724-728.

25. Tomasi, J.; Mennucci, B.; Cammi, R., Quantum Mechanical Continuum Solvation Models. *Chem. Rev.* **2005**, *105*, 2999-3094.

26. Wolinski, K.; Hinton, J. F.; Pulay, P. J., Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations, *J. Am. Chem. Soc.* **1990**, *112*, 8251-8260.

27. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C., Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006,** *65*, 712-725.

28. Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti C.; Parrinello, M., Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110*, 3533-3539.

29. *SAR-caddle*, Cepos Insilico Ltd., Kempston, UK, 2014; http://www.ceposinsilico.de/products/sar-caddle.htm (accessed May 3rd 2016)

30. Barone, G.; Duca, D.; Silvestri, A.; Riccio, R.; Bifulco, G., Structure Validation of Natural Products by Quantum-Mechanical GIAO Calculations of ^{13}C NMR Chemical Shifts. *Chem. Eur. J.* **2002**, *8*, 3233-3239.

31. Barone, G.; Duca, D.; Silvestri, A.; Gomez-Paloma, L.; Riccio, R.; Bifulco, G., Determination of the Relative Stereochemistry of Flexible Organic Compounds by ab initio Methods: Conformational Analysis and Boltzmann-Averaged GIAO ^{13}C NMR Chemical Shifts. *Chem. Eur. J.* **2002**, *8*, 3240-3245.

ToC Graphic