Journal of Molecular Graphics and Modelling 32 (2012) 67–74

Contents lists available at SciVerse ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Ensemble fits of restrained peptides' conformational equilibria to NMR data. Dependence on force fields: AMBER/8 ff03 versus ECEPP/3

Jerzy Ciarkowski[a,*], Sylwia Łuczak[a], Dawid Jagieła[a], Emilia Sikorska[a], Jacek Wójcik[b], Marta Oleszczuk[c], Jan Izdebski[d]

[a] Faculty of Chemistry, University of Gdańsk, Sobieskiego 18, 80-952 Gdańsk, Poland
[b] Institute of Biochemistry and Biophysics, Polish Academy of Sciences, 02-106 Warsaw, Poland
[c] Department of Biochemistry, University of Alberta, Canada T6G 2H7, Canada
[d] Peptide Laboratory, Faculty of Chemistry, Warsaw University, 02-093 Warsaw, Poland

ARTICLE INFO

Article history:
Received 6 June 2011
Received in revised form 4 October 2011
Accepted 14 October 2011
Available online 20 October 2011

Keywords:
Conformational equilibria
Peptides
Molecular dynamics
Monte Carlo
Statistical ensemble fit
Maximum entropy method

ABSTRACT

Two variants of NMR-based conformational analyses of flexible peptides are compared using two examples meeting the formula Tyr-D-Daa-Phe-Daa-NH_2 (Daa = diamino acid): **1** combining D-Dab^2 (α,γ-diaminobutyryl) with Lys^4, and **2** – D-Dap^2 (α,β-diaminopropionyl) with Orn^4. The ω-amino groups of D-Daa^2 and Daa^4 are coupled with)C=O into the urea, restraining **1** and **2** with 16- and 14-membered rings and leading to potent and impotent μ/δ opioid peptides, respectively. To the current task, we took from an earlier work (*Filip et al, J. Pept. Sci. 11 (2005) 347–352*) the NMR NOE- and *J*-data in H_2O/D_2O; and the selection of the ensembles of **1** and **2**, 822 and 788 conformational families, respectively, obtained by using the EDMC/ECEPP3 method. Here, we generated ensembles of **1** and **2** using AMBER molecular dynamics in explicit water to eventually selected 686 and 761 conformers for **1** and **2**, respectively. We did numbers of fits for both types of the conformational ensembles of **1** and **2** to their NOE- and *J*-data using a common method i.e. maximum entropy approach (*Groth et al, J. Biomol. NMR 15 (1999) 315–330*). Both types of the well structurally diversified ensembles fit to quite different equilibria in regressions to common experimental NOE- and *J*-restraints using maximum entropy principle, which is a disappointing message. Intriguing is startlingly small standard deviation in *J*-couplings: $\sigma_{J_{NH\alpha H}} \approx 0.01$ Hz for LES-MD/AMBER ensemble, contrary to $\sigma_{J_{NH\alpha H}} = 0.8 - 1.1$ Hz for the EDMC/ECEPP ensemble, over the wide range of entropy, i.e. relatively insensitive to it. A similar feature is not the case when comparing σ_{NOE} in both methods. Hence, at minute entropy contributions, it follows that *J* does or does not transpose "overfitted" into the final σ_J in the AMBER or ECEPP ensemble, respectively. Could this be an effect of softness of the AMBER flexible-valence force field compared to ECEPP rigid-geometry, and its effect on ensemble sampling? We do not know an answer.

1. Introduction

Contrary to proteins which tend to make structures, peptides make equilibria in solutions. The resolved conformational equilibria of small peptides depend on fitting procedures/algorithms that in diverse ways handle motional or ensemble averaging and/or use various force fields – both features being critical in fitting structures of flexible solutes to their NMR spectra [1–4]. Our observations (unpublished and Ref. [5]) hint that diverse equilibria are obtained using time- [6] or ensemble-averaged [7] fits of peptides to common NMR spectra. Hence, to consider the issue of divergent conformational results for common experimental data in a more disciplined way, we compare in this work two variants of fitting conformational equilibria of peptides to their NMR data [8] using the same ensemble-averaging procedures [7] but different force fields and ways of generating statistical ensembles. To this aim, we chose two cyclic deltorphin analogs Tyr-D-Daa-Phe-Daa-NH_2 as models, having the combinations of D-Dab^2 (α,γ-diaminobutyryl) with Lys^4, **1**, and of D-Dap^2 (α,β-diaminopropionyl) with Orn^4, **2**, with their side chain ω-NH_2 groups locked into the urea using carbonyl, restraining them to 16- and 14-membered rings and leading to potent and impotent μ/δ opioid peptides, respectively [8]. The conformational equilibria of **1** and **2** are fitted to their NMR NOE- and *J*-data in H_2O/D_2O: (i) in ensembles generated with EDMC-ECEPP/3 methodology implementing rigid-valence geometry [7] and; (ii) in ensembles generated by extensive AMBER molecular dynamics (MD) using flexible-valence geometry. In both cases method-specific means

* Corresponding author. Tel.: +48 58 523 5330.
E-mail address: jurek@chem.univ.gda.pl (J. Ciarkowski).

1093-3263/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2011.10.004

to enhance sampling were employed: while in the former a rigid-valence ECEPP force field does not support MD, its standard Monte Carlo (MC) method with minimization was enhanced with the "electrostatically driven" option (EDMC) intended to overcome higher barriers on the energy hypersurface in sampling [9,10]. In the latter, the AMBER flexible-valence MD was done in the locally enhanced-sampling mode(s) (LES-MD/AMBER) aimed to artificially reduce energy barriers [11,12]. The common fitting procedures implement the maximum entropy principle to circumvent overfit.

One has to add that the only known earlier similar comparison, using two enkephalin analogs as examples at the introduction and description of this method [7], was biased. Notably, the AMBER-derived ensembles were dependent on the conformational analyses of the EDMC/ECEPP/3 ensembles, because the most populated fitted conformers from the latter were used as the starting structures in the generation of the former [7]. Current comparison is free of such cause–result relationship.

The overfit is a false positive, important when a physical quantity or property with inherent statistical uncertainty (e.g. equilibrium) is fitted to parameters (here NOE and J) charged with empirical errors, typically ignored in the fitting. In the procedure implemented by Groth and coworkers [7] the overfit is dispersed by introducing into the objective function the scaled entropy term favoring uniform distribution of conformers, with the scaling factor α; see Section 2 and Eq. (11) in Ref. [7]. Typically, for $\alpha = 0$ (no entropy term), a fit results with a very few conformations having major weights while the standard deviations (e.g. σ_{NOE}, σ_J) of the fitting parameters get unduly small. This is a consequence of the fact that the measured σ_{NOE} and σ_J do not transpose into the objective function and we arrive at a typical overfit, effectively fitting some noise. On the opposite side, e.g. $\alpha \geq 1$, an overabundance of conformers is found, objective function grows larger and more realistic, χ^2 test gets excellent. For flexible peptides, a task is to get by trial-and-error a handy conformational distribution meeting χ^2 criterion. The results of fitting conformational equilibria in both ensembles with variable entropy contributions are compared and discussed.

2. Materials and methods

2.1. Synthesis, biology, NMR and the ensemble from electrostatically driven Monte Carlo (EDMC) in ECEPP/3 force field [8]

Synthesis, bioactivity as well as the recording and interpretation of the NMR spectra of **1** and **2** have been described and the authors [8] provided us with their NMR NOE- (Supplementary Tables S1 and S2) and J-data in H_2O/D_2O (Table 1), and with the selection of the ensembles of **1** and **2**, obtained using the EDMC/ECEPP/3 methodology [7], followed by filtering off the conformers whose energies exceeded the relative energy 10 kcal/mol. This resulted in 822 and 788 conformations stored for conformational fitting, from the initially sampled 13,500 and 12,734 conformers of **1** and **2**, respectively.

2.2. Conformational ensemble from molecular dynamics with locally enhanced sampling in AMBER ver. 8 (LES-MD/AMBER) [13]

The entities with the urea bridge have been already parameterized [5]. Starting structures were built and their energies minimized first *in vacuo* and then in the cube (1000.0 Å^3) of TIP3P water as described [5]. Subsequently **1** and **2** were submitted to the 4 ns molecular dynamics (MD) in the periodic boundary conditions (1000.0 Å^3) water cube, at constant pressure using the implemented in AMBER8 [13] locally enhanced sampling (LES) procedure for a more efficient sampling of conformational space. LES was set up for 5 copies, decreasing energy barriers ~5-foldly [11,12]. The *ff03* (default) force-field was used. The time step was 2 fs and the coordinates were saved every 2000 steps (every 4 ps). The saved sets were put together and every 4th set of the coordinates was collected resulting in 1250 conformers per each **1** and **2**. Both complete sets of 1250 conformers were energy-minimized and afterwards those having energies above 20 kcal/mol relative to the lowest-energy conformer inside each set, were rejected. While doing so, we noticed that despite putting standard restraints [13] on peptide bonds to maintain them *trans* (force constant

Table 1
Measured and computed values of $J_{NH\alpha H}$ and other measures of the performance of the maximum-entropy fitting in **1** and **2**. The sets for $\alpha = 0.01$ chosen as representative for discussion and the figures are in italic.

Fit to 80% population		AMBER						ECEPP					
α		0.0	0.0001	0.001	*0.01*	0.1	1.0	0.0	0.0001	0.001	*0.01*	0.1	1.0
$J_{NH\alpha H}$ [Hz]	J_{exp}	J_{calc}						J_{calc}					
Analog 1													
Dab[2]	8.50	8.500	8.499	8.499	*8.486*	8.422	8.350	9.400	9.977	9.974	*9.979*	9.954	9.902
Phe[3]	7.90	7.900	7.901	7.899	*7.899*	7.907	7.985	8.779	8.727	8.724	*8.739*	8.690	8.305
Lys[4]	8.70	8.700	8.700	8.698	*8.686*	8.606	8.427	7.851	8.011	8.028	*8.038*	8.114	9.064
σ_J [Hz]		0.000	0.000	0.001	*0.002*	0.011	0.186	0.876	1.055	1.049	*1.054*	1.014	0.868
σ_{NOE} [a]		0.132	0.132	0.132	*0.132*	0.132	0.132	0.322	0.318	0.318	*0.318*	0.319	0.826
Φ		1.755	1.755	1.755	*1.756*	1.757	1.781	38.328	37.575	37.574	*37.579*	37.611	42.684
Entropy			−4.356	−6.256	*−6.617*	−6.681	−6.709		−2.260	−2.264	*−2.257*	−2.463	−4.116
RMSD [Å][b]		0.990	0.976	0.990	*0.985*	0.981	0.979	0.631	0.672	0.672	*0.697*	0.664	0.779
No. of conformers		116	173	372	*533*	588	639	4	7	7	*7*	8	87
Analog 2													
Dap[2]	7.80	7.802	7.800	7.801	*7.808*	7.863	8.089	9.108	8.995	8.957	*9.024*	9.094	9.767
Phe[3]	8.50	8.503	8.501	8.499	*8.489*	8.430	8.232	8.839	8.819	8.858	*8.806*	8.843	8.599
Orn[4]	8.00	7.995	8.001	8.000	*7.996*	7.968	7.876	7.962	7.750	7.751	*7.990*	8.059	8.478
σ_J [Hz]		0.003	0.001	0.001	*0.008*	0.058	0.239	0.780	0.729	0.714	*0.728*	0.774	1.170
σ_{NOE} [a]		0.050	0.050	0.050	*0.051*	0.051	0.053	0.177	0.178	0.178	*0.178*	0.178	0.194
Φ		0.174	0.174	0.176	*0.180*	0.182	0.211	9.794	9.869	9.881	*9.835*	9.890	11.974
Entropy			−2.523	−5.981	*−6.689*	−6.773	−6.798		−2.580	−2.629	*−2.823*	−3.460	−6.286
RMSD [Å][b]		0.863	0.834	0.912	*0.872*	0.865	0.868	0.735	0.721	0.721	*0.762*	0.723	0.722
No. of conformers		4	8	299	*580*	650	693	8	8	8	*10*	18	401

[a] 31 and 27 NOEs were fitted in **1** and **2**, respectively.
[b] Averaged root-mean-square deviation over the atoms of the 16- and 14-membered rings in **1** and **2**, respectively.