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.

Table 2
Distributions, in decreasing order, of seven most populated conformers of **1** and **2** in fitting runs for LES-MD/AMBER and EDMC/ECEPP ensembles, depending of various α. The sets for $\alpha = 0.01$ chosen as representative for discussion and the figures are in are in italic.

	No.[a]	*X*	No.[a]	*X*	No.[a]	*X*	No.[a]	*X*	No.[a]	*X*	No.[a]	*X*
α	0.0		0.0001		0.001		*0.01*		0.1		1.0	
Analog 1												
AMBER												
1	18	0.167	26	0.236	23	0.015	*333*	*0.004*	333	0.002	1039	0.001
2	26	0.154	23	0.134	801	0.012	*56*	*0.003*	262	0.002	688	0.001
3	23	0.151	18	0.101	931	0.010	*1067*	*0.003*	1067	0.002	333	0.001
4	245	0.109	2	0.023	333	0.010	*787*	*0.002*	688	0.001	262	0.001
5	145	0.033	37	0.012	245	0.009	*498*	*0.002*	755	0.001	1110	0.001
6	2	0.010	53	0.007	787	0.009	*1174*	*0.002*	1029	0.001	755	0.001
7	37	0.009	11	0.007	355	0.008	*664*	*0.002*	787	0.001	9	0.001
ECEPP												
1	664	0.281	650	0.248	650	0.243	*650*	*0.243*	664	0.223	1	0.331
2	682	0.206	664	0.226	664	0.232	*664*	*0.230*	650	0.223	664	0.035
3	650	0.167	724	0.110	724	0.107	*724*	*0.110*	724	0.108	650	0.035
4	252	0.109	682	0.082	682	0.088	*682*	*0.087*	682	0.092	682	0.029
5	309	0.103	252	0.069	252	0.068	*252*	*0.069*	252	0.066	434	0.023
6	255	0.046	434	0.051	434	0.048	*615*	*0.048*	639	0.043	496	0.019
7	251	0.032	693	0.045	639	0.047	*639*	*0.047*	434	0.035	724	0.016
Analog 2												
AMBER												
1	8	0.310	935	0.292	243	0.026	*470*	*0.004*	28	0.002	28	0.001
2	35	0.289	691	0.225	935	0.022	*402*	*0.004*	1106	0.002	323	0.001
3	361	0.160	8	0.104	609	0.019	*1222*	*0.003*	323	0.002	912	0.001
4	13	0.076	276	0.048	402	0.017	*401*	*0.003*	912	0.002	337	0.001
5	700	0.060	700	0.043	8	0.015	*822*	*0.003*	422	0.002	421	0.001
6	87	0.022	1122	0.039	691	0.014	*748*	*0.003*	421	0.002	709	0.001
7	327	0.015	361	0.034	1222	0.014	*276*	*0.002*	337	0.001	1106	0.001
ECEPP												
1	361	0.134	361	0.171	361	0.168	*361*	*0.152*	361	0.123	752	0.017
2	575	0.129	575	0.117	752	0.125	*752*	*0.114*	752	0.109	361	0.015
3	752	0.119	752	0.114	768	0.114	*575*	*0.102*	575	0.097	575	0.011
4	768	0.102	768	0.110	575	0.113	*768*	*0.091*	768	0.073	518	0.009
5	579	0.086	352	0.079	126	0.077	*126*	*0.076*	126	0.060	459	0.008
6	126	0.068	126	0.077	352	0.077	*579*	*0.065*	518	0.044	388	0.006
7	354	0.068	388	0.075	388	0.068	*518*	*0.055*	579	0.031	137	0.006

[a] Running numbers of the conformers in the generated ensembles.

$f = 50$ kcal mol^{-1} rad^{-2} during the entire simulation), ca 10% of them converted into *cis*, presumably, due to the decrease of the "intensity" of the energy hypersurface by the factor of 5, typical of the 5-copy-LES MD, see above. Thus the energy-filtering resulted in 938 and 1058 conformers and subsequent elimination of the *cis* peptides, respectively, in 822 and 900 conformers in the ensembles of **1** and **2**.

2.3. Ensemble fit [7]

The conformational equilibria of **1** and **2** were found by fitting the computed NOE- and *J*-values to the corresponding experimental values, using the maximum entropy principle as implemented by Groth et al. [7]. Briefly, the algorithm, being a part of the freely distributed software termed ANALYZE [14], consists of the following three steps: (i) search of the conformational space in order to find conformations with reasonably low energy; (ii) calculation of the NOE and vicinal coupling constants for each of the low-energy conformations; and (iii) determination of the statistical weights of the conformers using the maximum-entropy method, in order to obtain the best fit of the averaged NOE intensities and $^3J_{NH\alpha H}$ coupling constants to the experimental quantities. ANALYZE implements the MORASS program [15] and the Karplus equation parameterized by Bystrov [16] for the theoretical calculations of the NOE intensities and $^3J_{NH\alpha H}$ values, respectively, and averaged NOE intensities and $^3J_{NH\alpha H}$ values are functions of the weights $x_1, x_2, \ldots, x_n$ which could thus be found by least-square fitting of the former to their respective experimental NOE and $^3J_{NH\alpha H}$ quantities, as given by the objective function $\Phi_{\min}(V_0, x_1, x_2, \ldots, x_n)$, in Eq. (1) [7]:

$$\Phi_{\min}(V_0, x_1, x_2, \ldots, x_n) = \sum_{k=1}^{\#NOEs} w_k \times [NOE_k^{exp} - \overline{NOE_k}(V_0, x_1, x_2, \ldots, x_n)]^2 + \sum_{l=1}^{\#Js} w_l [J_l^{exp} - \bar{J}_l(x_1, x_2, \ldots, x_n)]^2 + \sum_{i=1}^{2} \frac{1}{\sigma_{A_i}^2}(A_i - A_{i0})^2 + \sum_{l=1}^{2} \frac{1}{\sigma_{B_i}^2}(B_i - B_{i0})^2 + \sum_{i=1}^{2} \frac{1}{\sigma_{C_i}^2}(C_i - C_{i0})^2 \quad (1)$$

where $\overline{NOE_k} = V_0 \sum_i^n x_i NOE_{ik}$, $\bar{J}_l = \sum_i^n x_i J_{il}$, $x_1, x_2, \ldots, x_n$ are the sought weights, k and l are the indices summing up the differences between the measured and computed NOE- and *J*-values, respectively, and V_0, w_k and w_l are adjustable parameters enabling a proper balance between the NOE- and *J*-contributions to the objective function. One has to add that the default scaling factors w_k conform to the NOE weights pertaining to meticulously estimated NOE standard deviations at ~0.02 from thorough model measurements of NOEs [7]. It was shown that distributions of the resolved conformers are

Table 3
Summary of steps and numbers of the conformers after each step.

	AMBER				ECEPP/3		
	Used after MD	>20 kcal/mol	No *cis* peptide	No C7ax-like	Total conf.	Accepted conf.	>10 kcal/mol
Analog 1	1250	938	822	686	13,150	3000	822
Analog 2	1250	1058	900	761	12,734	3000	778

not sensitive to these scaling factors; e.g. compare Fig. 8 in [7], showing effects of NOE weights varied over 3 orders. The last three terms account for the fact that the computed *J*-values carry errors arising from the intrinsic uncertainties of the Karplus–Bystrov equation, $^3J_{NH\alpha H} = A\cos^2\Theta + B\cos\Theta + C$; ($A_0$, B_0, C_0) = (9.4, −1.1, 0.4) [16]. Thus, in fitting, these terms allow some adjustments of the respective coefficients, relative to their standard values A_0, B_0 and C_0, within the limits set by the values of $\sigma_A = \sigma_B = \sigma_C = 2$ Hz. There are two sets of the *A*, *B* and *C* coefficients, specific for non-Gly-$J_{NH\alpha H}$ given above, and for Gly/NH–CH_2 *J*. In the latter case the Karplus equation applies to the measured $\Sigma^3J_{NH\alpha H}$ and (A_0, B_0, C_0) = (−9.4, −1.1, 14.9) [16]. Accordingly, the running index *i* covers two possibilities and the last three terms thus contribute in an increases of a number of the sought parameters by 3 (non-glycine-type *J*s) or 6 if complete sets of measured $^3J_{NH\alpha H}$ types is used in fitting.

A typical minimization of Φ solves to a few contributing conformations, while the weights of the remaining ones get close to zero. The authors of the method maintain that this is accompanied with apparent lack of transposal of these errors into Φ, for more see [7]. One should arrive at a typical overfit, effectively fitting a good share of noise.

The maximum entropy approach [7] introduces an entropy term: $-\alpha\sum_i^n x_i \log x_i$, which is subtracted from the minimized sum of squares:

$$\Psi_{\min}(V_0, x_1, x_2, \ldots, x_n) = \Phi_{\min}(V_0, x_1, x_2, \ldots, x_n) + \alpha\sum_i^n x_i \log x_i \quad (2)$$

The entropy term is the smallest when the weights x_1, x_2, …, x_n are equal. This would take place if conformation-sensitive experimental data would not exist, a state approached to a loose extent by the starting LES-MD/AMBER ensembles, consisting of hundreds of conformers with very flatly distributed populations, see Section 2.2. Weight differentiation comes only from the Φ term that includes experimental information. Hence, a typical procedure is to choose the entropy coefficient α so that the mean errors in the fitted quantities are comparable with the error estimates, or equivalently, χ^2 would equal to the number of measured data. It is a natural approach, because the expected measures of misfit

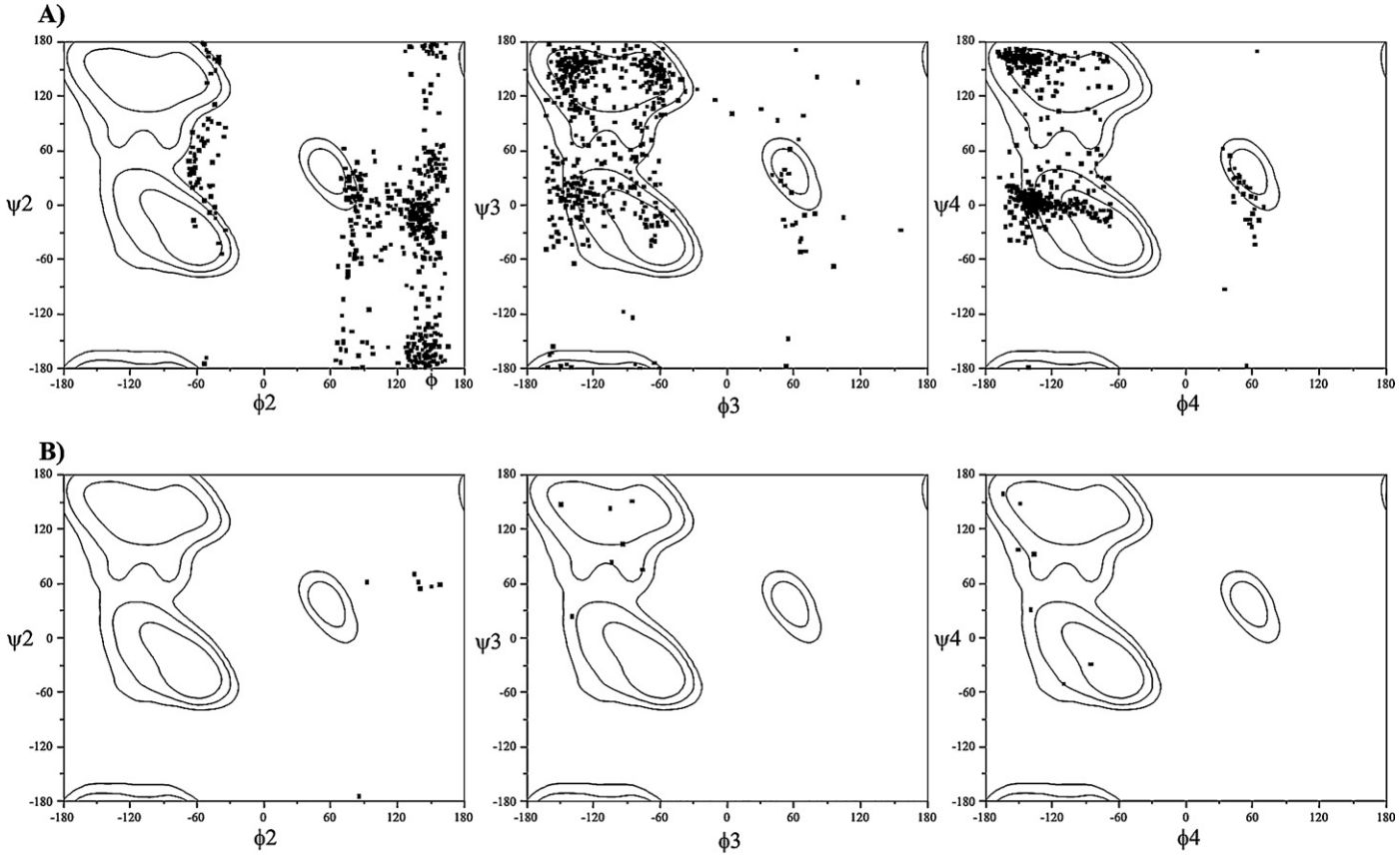

Fig. 1. Ramachandran plots for **1**. (A) For the 822 starting conformers to the LES-MD/AMBER fit. (B) For the resultant 7 conformers from the EDMC/ECEPP fit. Panel (A) clearly indicates numerous conformers in the "forbidden" C7ax-type regions, corresponding to $(\varphi,\psi) \approx (-80 \pm 30, 70 \pm 30)$ for L-Phe[3], Lys[4], and the opposite – for D-Dab[2]. 136 conformers were found having at least one residue in the bad C7ax turn, see text.

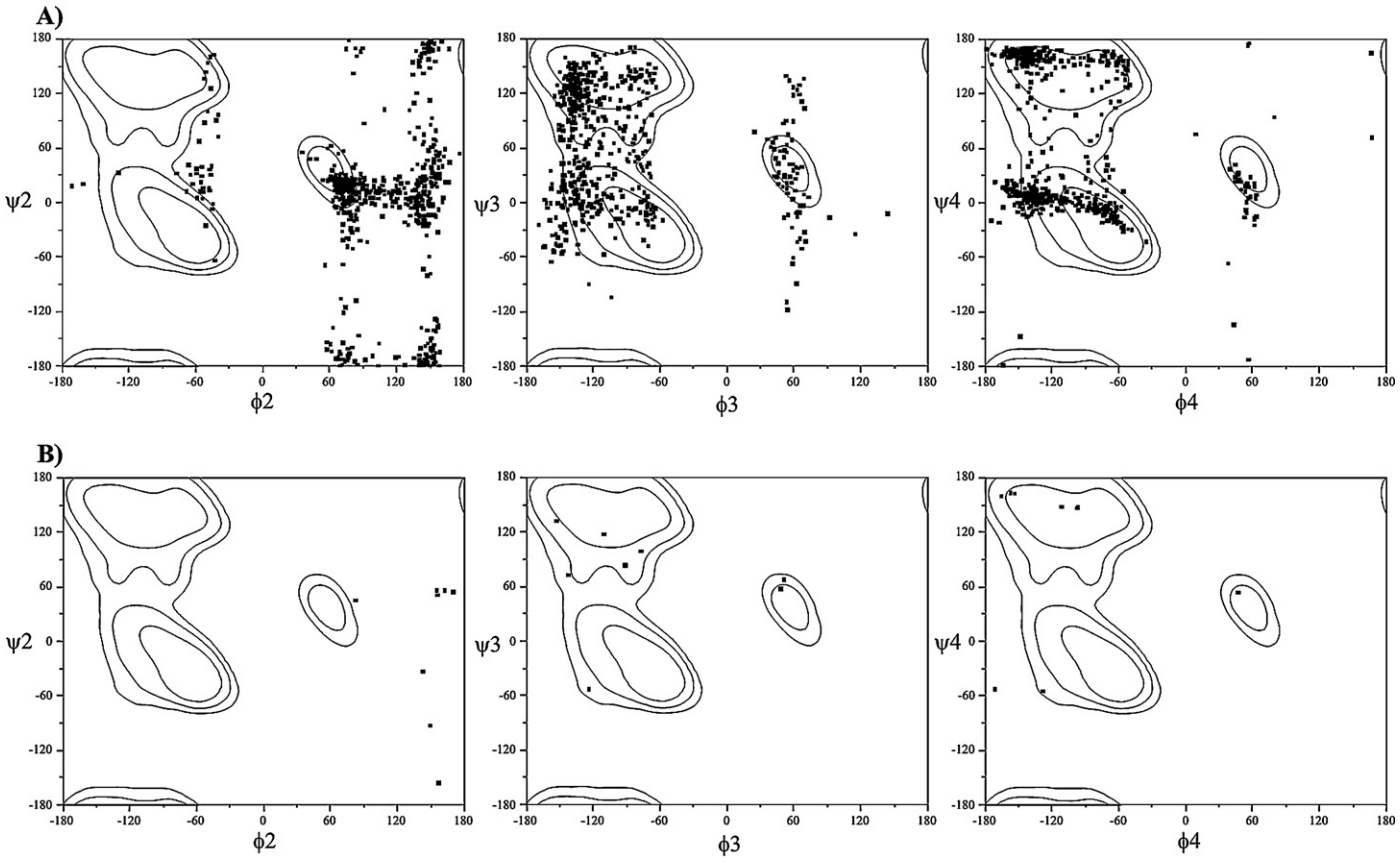

Fig. 2. Ramachandran plots for **2**. (A) For the 900 starting conformers to the LES-MD/AMBER fit. (B) For the resultant 10 conformers from the EDMC/ECEPP fit. Panel (A) clearly indicates numerous conformers in the "forbidden" C7ax-type regions, corresponding to $(\varphi,\psi)\approx(-80\pm30,70\pm30)$ for L-Phe[3], Orn[4], and the opposite – for D-Dap[2]. 139 conformers were found having at least one residue in the bad C7ax turn, see text.

should be equal to the estimated experimental inaccuracy, because if the agreement between theory and experiment is better one starts fitting the noise. Thus, the reasonable fit is approached by trial-and-error search of a value of α which leads to the expected parameter errors in minimization of Ψ. Typically, elimination of the overfit is accompanied by a considerable increase of the number of contributing conformers with simultaneous dispersion of the conformational equilibrium into flatly populated conformers. Once the resulting conformer populations x_i plus Karplus coefficients exceed the number of the NMR observables (restraints), the system gets undetermined (having more unknowns than restraining parameters/equations) and any answer out of infinity number of solutions becomes algebraically proper. Another, one could say – operational, cost of introducing the entropy relies on exchange of linear regression into a non-linear one, for more details, see [7,14].

2.4. Evaluation of the equilibria

The molecular structures were analyzed using PTRAJ module of AMBER 8.0 [13] and drawn using the graphic MOLMOL [17] and PYMOL [18] programs. Selected representatives of conformational families were overlaid using all heavy atoms making the 16- and 14-membered restraining rings.

3. Results

Earlier use of EDMC/ECEPP ensemble by some of us [8] resulted with 7 structures for **1** and 11 structures for **2** at $\alpha = 0$ (no entropy term), at the condition that the contributions smaller than 3% of the total population were ignored, which appeared equivalent to the choice that ~94% of the total population were accepted in the fit. Clearly, none of the restraining experimental *J*s and NOEs were not provided in that paper [8], therefore we now make up for that omission providing the experimental *J*s in Table 1 and the values of the NOEs in Supplementary Tables S1 and S2. In this work, we have done four more fits to this ensemble, for $\alpha = (0, 0.0001, 0.001, 0.1$ and $1.0)$ and similar fits for the LES-MD/AMBER ensembles, holding in all these fits the confidence level to 80%, i.e. to this percentage accepting conformational families, contributing in decreasing weights to the fit. The results of all series of the fits are shown in Table 1.

The table clearly demonstrates that both ensemble types do fit in profoundly different ways, as expected, compare the results in [7]. In the LES-MD/AMBER ensemble the objective function Ψ takes up the values not higher than 2 and the resolved σ_J values, while increasing with "addition" of entropy, barely approach 0.2 Hz at $\alpha = 1$, despite putting $\sigma_A = \sigma_B = \sigma_C = 2$ Hz, see above. This implies a far-reaching overfit throughout all the values α tested.

On the contrary, in the case of the EDMC/ECEPP ensemble, the objective function Ψ assumes the values as high as ~40 and ~10 for **1** and **2**, respectively, at the default scaling of the NOE terms in Φ [7], compare legends to Eq. (1). The quite high values of Ψ at default NOE weights may indicate that the ensemble is poor. Indeed, the apparent resolved σ_{NOE} values in Table 1 exceed by an order its estimate (~0.02, see Section 2), especially for **1**. As this is also the case to a lesser extent for the fit of **1** to the LES-MD/AMBER ensemble, it may pertain also to the quality of the experimental NOEs used in this fit. Judging upon the standard deviations of the coupling constants σ_J, there is not much overfit throughout the

Table 4
Summary of the cluster analyses of the results of fitting the LES-MD/AMBER ensembles at $\alpha = 1.0$, with no *cis*-peptides and C7ax-type turns.

Cluster/family	Analog 1, 686 conformers		Analog 2, 761 conformers	
	No. of conformers	Mol fraction	No. of conformers	Mol fraction
1	62	0.097	117	0.161
2	65	0.096	107	0.145
3	50	0.074	85	0.116
4	44	0.066	69	0.093
5	43	0.065	67	0.090
6	42	0.062	49	0.067
7	36	0.053	30	0.039
8	33	0.050	27	0.035
9	19	0.029	24	0.033
10	19	0.028	10	0.014
11	18	0.027	6	0.009
12	14	0.022		
13	14	0.021		
14	13	0.020		
15	13	0.019		
16	12	0.018		
17	10	0.015		
18	9	0.013		
19	9	0.013		
20	4	0.006		
21	4	0.006		
Total	533 (77.7%)	0.801	591 (77.5%)	0.801

whole series of tested coefficients $\alpha = (0, 0.0001, 0.001, 0.01, 0.1$ and $1.0)$, see Table 1. Notably our current results at $\alpha = 0.1$/confidence level 80% (Table 1) appear equivalent to those in the former work at $\alpha = 0$/confidence level ~94% [8].

Another clear difference, perhaps related to those described above, in the behavior of both fitted ensemble-types is relative insensitivity of the EDMC/ECEPP ensembles to the varied α, contrary to LES-MD/AMBER ensembles showing just the opposite. This manifests in two aspects, viz. the results of fitting the EDMC/ECEPP ensembles return comparatively small numbers (7–10) of conformers in equilibrium, in a relatively wide range of α ($\alpha = 0$–0.1), see Table 1, and the identities and populations of these conformers (e.g. in decreasing order) are stable in this whole range of α, see Table 2.

Both features do not hold in the LES-MD/AMBER ensembles: with increasing α in the comparable range, the numbers of the returned conformers in the equilibrium grow rapidly, very soon exceeding the numbers of the restraining experimental parameters (the sum of *J*s + NOEs + Karplus–Bystrov coefficients), i.e. the solution becomes algebraically undetermined systems. For instance, one can see that at $\alpha = 0.01$ the fitting returns already as many as 372 and 299 conformers for **1** and **2**, respectively, see Table 1. Unsurprisingly, their identities and populations are not retained and vary in the comparable range of α, see Table 2.

Such behavior of the LES-MD/AMBER ensemble poses a serious question, because the resolved equilibria, while still deeply in the overfit (see unrealistic $\sigma_J = 0.001$ at $\alpha = 0.01$ in Table 1), simultaneously provide equilibria contributed with hundreds of conformers, whose population orders are not maintained with varied α. Indeed, at $\alpha = 2$, although the overfit is almost reduced, the number of conformers in equilibrium reaches almost their input number reduced by the confidence level/cut off at 80% (not shown). How one can handle such results and can one attribute to them any physical significance?

Inspired by the works of Nikiforovich et al. [19,20], we decided to consider, instead of single values of the statistical weights of conformers, their distributions within groups of similar structures, attributing cumulative weights to the groups. Before doing so, also inspired by these works, we inspected the positions of the residues 2–4 of **1** and **2** on the Ramachandran maps, see Figs. 1 and 2. As expected, the resultant EDMC/ECEPP conformers did not exhibit conformations with residues contributing in the "forbidden" C7ax-like areas [21]. On the contrary, the LES-MD/AMBER ones revealed 136 (16.5%) and 139 (15.4%) conformers for **1** and **2**, respectively, having at least one residue in the forbidden area, corresponding to the conformers defined as the C7ax-type [22], see details in Table 3. One may notice that in AMBER calculations ΔE(C7ax versus C7eq) for an isolated alanyl dipeptide is quite small, equal to ~1.2 kcal/mol [22], contrary to the *ab initio* estimates of this difference to 2.1–2.9 kcal/mol [23], so it is clear that AMBER underestimates this difference, clearly, even more so in the LES regime.

Accordingly, having eliminated all the "forbidden" conformers, we eventually resulted with 686 and 761 conformers for **1** and **2**, respectively, for which we repeated the fitting, see Table S3. As seen, the elimination of the "bad" conformers affected neither the general results nor statistics. Subsequently, to select the groups, we did a few cluster analyses (using Molmol [17]) for the results at $\alpha = 0.01$ and $\alpha = 1.0$ (438 and 533 conformers, respectively, for **1**, see Table S3), using varying upper cut-off limits (RMSD), to eventually choose 1.6 Å as the criterion of distinction between the families. In spite of this quite tolerant RMSD value, we obtained 21 and 11 families for **1** and **2**, respectively, at $\alpha = 1.0$, see Table 4, and a similar result for $\alpha = 0.01$ (not shown) indicating that the conformations are relatively uniformly distributed within the whole ensemble, and poorly discernible into distinct families. At the same time, the conformer contents of the clusters, the cluster orders, the cumulative populations in the clusters (families) and their population ranking remained relatively stable and did not depend much on the value of α.

To get an idea about these clusters, we calculated the "mean coordinates" for any family and picked up a *true* conformation in each cluster having the lowest RMSD relative to this mean, to consider it as a representative of the given family. The backbone angles of these conformers typical of 21 and 11 clusters are given in Table 5 while their shapes are given in Supplementary Figs. S1 and S2, for **1** and **2**, respectively. For comparison, in Table 5 are also shown the backbone angles of the conformers resulting from fitting the EDMC/ECEPP ensembles at $\alpha = 0.01$, see Table 1.

4. Discussion

Our analysis of the data in Table 1 indicates the following.

Table 5
The peptide backbone angles of the representative conformers for the clusters resulting from the LES-MD/AMBER ensembles fits at $\alpha = 1.0$. The conformers contributing to the equilibria resulting from the EDMC/ECEPP/3 fits are also given. Matching conformers (see text) for **1** and **2** are shown in bold.

	LES-MD/AMBER							EDMC/ECEPP/3							
	Conformer running no. in the ensemble	ψ^1	φ^2	ψ^2	φ^3	ψ^3	φ^4	ψ^4	Conformer running no. in the ensemble	ψ^1	φ^2	ψ^2	φ^3	ψ^3	φ^4	ψ^4
Analog 1																
1	692	145	73	17	−135	161	−124	170	**650**	**162**	**148**	**51**	**−141**	**140**	**−133**	**139**
2	**873**	**170**	**156**	**29**	**−131**	**169**	**−136**	**159**	664	136	141	−23	−79	76	−86	−34
3	977	−77	63	−138	−145	33	−142	166	724	37	85	−77	45	53	−69	−33
4	874	−52	151	−25	−146	125	−146	8	682	31	92	52	−89	77	−89	−31
5	889	55	−62	23	−132	156	−114	104	252	152	138	62	−95	104	−151	98
6	228	119	144	−155	−138	24	−139	1	615	−44	92	62	−77	76	−137	92
7	379	−85	143	1	−135	164	−157	9	639	65	148	60	−146	139	−135	146
8	668	−89	141	−97	−67	107	−132	−3								
9	362	149	−46	23	−66	−173	51	13								
10	1154	−117	142	−173	−136	11	−93	150								
11	204	161	73	−12	−146	156	−144	163								
12	195	155	84	−166	−12	116	−139	161								
13	483	171	151	−134	66	−37	−112	150								
14	558	138	147	−111	66	−50	40	55								
15	1109	138	150	172	−162	117	−151	9								
16	260	152	−62	52	−134	52	62	−35								
17	449	−101	136	−14	−118	50	−148	166								
18	375	127	145	−29	−75	131	−163	171								
19	773	178	158	−167	−123	19	−96	2								
20	534	130	68	−172	−160	−16	−131	9								
21	591	123	157	−52	40	34	−150	153								
Analog 2																
1	688	177	153	71	−129	153	−123	8	**361**	**149**	**157**	**−156**	**51**	**68**	**−154**	**163**
2	11	−80	89	11	−133	18	−54	−29	752	158	155	56	−125	−52	47	54
3	501	128	114	10	−144	127	−138	7	575	161	170	55	−111	118	−166	160
4	69	−66	138	8	−114	75	−129	0	768	155	82	46	−78	99	−172	−53
5	729	−76	76	0	−116	2	−59	157	126	132	156	51	−143	73	−158	164
6	745	−103	67	14	−133	105	−141	−6	579	109	142	−33	−92	84	−112	149
7	508	−97	150	113	−73	−14	−146	163	518	158	162	56	−153	132	−129	−55
8	507	−65	144	143	−122	−5	−77	−9	352	−48	149	−93	48	58	−98	148
9	642	129	−50	155	−136	14	−140	−3	354	−48	149	−92	47	58	−98	148
10	**1218**	**−106**	**152**	**168**	**51**	**50**	**−133**	**157**	577	108	143	−33	−91	84	−112	149
11	982	143	147	−172	53	55	−144	158								

Firstly, neither representative conformer (both from the LES-MD/AMBER and EDMC/ECEPP fits) exhibits any regular β-turn-type fold [24]. This is a surprise in view of the observation by Groth et al., who in similar analyses observed predominance of the βII'-turns in restrained cyclic tetrapeptide analogs of enkephalins [7]. Secondly, the specific conformers predominant in equilibrium, returned from the EDMC/ECEPP fits in this work, compare favorably with those found in the earlier work (compare Table 5 with Table 5 in [8]), despite using $\alpha = 0$ in the former and $\alpha = 0.01$ in this work. Specifically, three and six most populated conformers do repeat for **1** and **2**, respectively, in both works. Thirdly, similar repetitiveness does hold between the EDMC/ECEPP and LES-MD/AMBER results. As seen from Table 5, one can match only one pair per each compound, viz. for **1**, Conformer_650(EDMC/ECEPP) matches Conformer_873(LES-MD/AMBER) and for **2**, Conformer_361(EDMC/ECEPP) in some way matches Conformer_1218(LES-MD/AMBER), see Graphical Abstract. These results argue in favor of a conclusion that both ensemble types lead to mutually incompatible results. Fourthly, on the whole, there seems an agreement in the all but one (ψ_2) "internal" backbone angles: φ_2, ψ_2, φ_3, ψ_3, φ_4 as to the sign among compounds **1** and **2** and the resultant equilibrium sets, independently of the starting ensemble. Thus, φ_2 and ψ_3 tend to be prevailingly positive while φ_3 and φ_4 – negative. Although doubtless correlated with other angles, ψ_2 differs in this cumulative appearance from the others in that in the LES-MD/AMBER results, it tends to by negative for **1** and positive for **2** while in the EDMC/ECEPP results it does not manifest a tendency.

In view of the above results, picking up most populated conformers arisen from LES-MD/AMBER ensemble appears quite arbitrary. Both types of the ensembles, although each conformationally well diversified, fit to quite different equilibria in the regression using maximum entropy principle and the same experimental NOE and *J*-values as constraints. This is a disappointing message. Intriguing is startlingly small standard deviation in *J*-couplings: $\sigma_J \approx 0.01$ Hz up to $\alpha = 0.01$ for LES-MD/AMBER ensemble, contrary to $\sigma_J = 0.8$–1.1 Hz at $\alpha = (0$–$1.0)$ for the EDMC/ECEPP ensemble, i.e. relatively insensitive to the entropy coefficient. 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 hardly know an answer.

All in all, there is more good-quality NMR data on restrained peptides needed to add to the above results. Currently, we do similar comparative studies on performance of the maximum entropy approach, depending on starting ensembles, using NMR data in water for 8 enkephalin analogs.

Acknowledgements

Support from the Polish Ministry of Science and Higher Education, grant DS/8372-4-0138-11 (for JC and SŁ) is acknowledged. The calculations were carried out in the Academic Computer Center in Gdansk TASK, Poland. We thank for discussions and advice to Adam

Liwo and Cezary Czaplewski. Referees' remarks and suggestions are greatly appreciated.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.10.004.

References

[1] X. Daura, I. Antes, W.F. van Gunsteren, W. Thiel, A.E. Mark, The effect of motional averaging on the calculation of NMR-derived structural properties, Proteins 36 (1999) 542–555.
[2] R. Bürgi, J. Pitera, W.F. van Gunsteren, Assessing the effect of conformational averaging on the measured values of observables, J. Biomol. NMR 19 (2001) 305–320.
[3] M.P. Williamson, Peptide structure determination by NMR, in: C. Jones, B. Mulloy, A.H. Thomas (Eds.), Spectroscopic Methods and Analyses: NMR, Mass Spectrometry, and Metalloprotein Techniques, Methods in Molecular Biology, vol. 17, Springer, 1993, pp. 69–85.
[4] K.A. Feenstraa, C. Peter, R.M. Scheek, W.F. van Gunsteren, A.E. Mark, A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides, J. Biomol. NMR 23 (2002) 181–194.
[5] A. Zieleniak, S. Rodziewicz-Motowidło, Ł. Rusak, N.N. Chung, C. Czaplewski, E. Witkowska, P.W. Schiller, J. Ciarkowski, J. Izdebski, Deltorphin analogs restricted via a urea bridge: structure and opioid activity, J. Pept. Sci. 14 (2008) 830–837.
[6] D.A. Pearlman, How is an NMR structure best defined? An analysis of molecular dynamics distance-based approaches, J. Biomol. NMR 4 (1994) 1–16.
[7] M. Groth, J. Malicka, C. Czaplewski, S. Ołdziej, L. Łankiewicz, W. Wiczk, A. Liwo, Maximum entropy approach to the determination of solution conformation of flexible polypeptides by global conformational analysis and NMR spectroscopy – application to DNS^1-c-[D-A_2bu^2,Trp^4,Leu^5]-enkephalin and DNS^1-c-[D-A_2bu^2,Trp^4, D-Leu^5]enkephalin, J. Biomol. NMR 15 (1999) 315–330.
[8] K. Filip, M. Oleszczuk, J. Wójcik, N.N. Chung, P.W. Schiller, D. Pawlak, A. Zieleniak, A. Parcińska, E. Witkowska, J. Izdebski, Cyclic enkephalin and dermorphin analogues containing a carbonyl bridge, J. Pept. Sci. 11 (2005) 347–352.
[9] D.R. Ripoll, H.A. Scheraga, On the multiple-minima problem in the conformational analysis of polypeptides. II. An electrostatically driven Monte Carlo method-tests on poly(L-alanine), Biopolymers 27 (1988) 1283–1303.
[10] D.R. Ripoll, H.A. Scheraga, The multiple-minima problem in the conformational analysis of polypeptides. III. An electrostatically driven Monte Carlo method: tests on enkephalin, J. Protein Chem. 8 (1989) 263–287.
[11] A. Roitberg, R. Elber, Modeling side chains in peptides and proteins: application of the locally enhanced sampling technique and the simulated annealing methods to find minimum energy conformations, J. Chem. Phys. (1991) 9277–9287.
[12] C. Simmerling, M.R. Lee, A.R. Ortiz, A. Kolinski, J. Skolnick, P.A. Kollman, Combining MONSSTER and LES/PME to predict protein structure from amino acid sequence: application to the small protein CMTI-1, J. Am. Chem. Soc. 122 (2000) 8392–8402.
[13] D.A. Case, T.A. Darden, T.E. Cheatham III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, K.M. Merz, B. Wang, D.A. Pearlman, M. Crowley, S. Brozell, V. Tsui, H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J.W. Caldwell, W.S. Ross, P.A. Kollman, AMBER 8, University of California, San Francisco, 2004, and references therein.
[14] A. Liwo, C. Czaplewski, ANALYSE (1999): Free Access ICM, Interdisciplinary Centre for Mathematical and Computational Modeling, University of Warsaw: http://www.icm.edu.pl/kdm/ANALYZE or CBSU@Cornell University: http://cbsu.tc.cornell.edu/software/analyze/index.htm.
[15] C.B. Post, R.P. Meadows, D.G. Gorenstein, On the evaluation of interproton distances for three-dimensional structure determination by NMR using a relaxation rate matrix analysis, J. Am. Chem. Soc. 112 (1990) 6796–6803.
[16] V.F. Bystrov, Spin–spin coupling and the conformational states of peptide systems, Progr. NMR Spectrosc. 10 (1976) 41–81.
[17] R. Koradi, M. Billeter, K. Wüthrich, MOLMOL: a program for display and analysis of macromolecular structures, J. Mol. Graph. 14 (1996) 52–55.
[18] The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC. (2002).
[19] G.V. Nikiforovich, O. Prakash, C.A. Gehrig, V.J. Hruby, Conformations of the dermenkephalin backbone in DMSO solution by a new approach to the solution conformations of flexible peptides, J. Am. Chem. Soc. 115 (1993) 3399–3406.
[20] G.V. Nikiforovich, K.E. Kövér, W.-J. Zhang, G.R. Marshall, Cyclopentapeptides as flexible conformational templates, J. Am. Chem. Soc. 122 (2000) 3262–3273.
[21] G.J. Kleywegt, T.A. Jones, Phi/Psi-chology: Ramachandran revisited, Structure 15 (1996) 1395–1400.
[22] C.A. Schiffer, J.W. Caldwell, R.M. Stroud, P.A. Kollman, Inclusion of solvation free energy with molecular mechanics energy: alanyl dipeptide as a test case, Protein Sci. 1 (1992) 396–400.
[23] W.D. Cornell, I.R. Gould, P.A. Kollman, The effects of basis set and blocking groups on the conformational energies of glycyl and alanyl dipeptides: a Hartree–Fock and MP2 study, J. Mol. Struct.: Theochem. 392 (1997) 101–109.
[24] E.G. Hutchinson, J.M. Thornton, A revised set of potentials for beta-turn formation in proteins, Protein Sci. 3 (1994) 2207–2216.