Molecular Diversity, **4:** 199–207, 1998.
KLUWER/ESCOM

Comparative conformational analysis of peptide libraries

Saul G. Jacchieri
Fundação Antônio Prudente, Rua Prof. Antônio Prudente 211, São Paulo, SP 01509-090, Brazil
(E-mail: jacchier@node1.com.br)

Received 14 June 1999; Accepted 18 October 1999

Key words: comparative conformational analysis, peptide libraries, probabilities of occurrence of structural motifs, sequence and conformational spaces

Summary

Six computer-based combinatorial libraries, including tetrapeptide sequences (generated with five amino acids) and conformations (generated with five main chain and three side chain rotamers), were obtained and sequence-conformation probabilities were calculated with a molecular and statistical mechanics procedure. The structural motifs α-helix, β-sheet, 3_{10}-helix, reverse turn I and γ-turn were focused in these calculations. It is shown that sequence-conformation-probability surfaces provide a broad view of structural changes accompanying changes in sequence. Numerical indices are defined to enable comparisons between frequencies of occurrence of these structural motifs in peptide libraries and in a database of low sequence identity protein structures. Fine details of sequence-conformation-probability surfaces show the effect of point mutations. Broad comparisons between different regions of these surfaces indicate how to select the occurrence of structural motifs in the combinatorial synthesis of peptide chains.

Introduction

A common objective of the de novo design of polypeptide chains is to find amino acid sequences that adopt beta hairpins in solution [1]. Even more complex features may be sought, such as the buried and exposed surfaces, the end-to-end distance and an increase in the probability of occurrence of a certain conformation.

Although the chemical synthesis of peptide libraries [2] may create polypeptide chains with the desired properties [3], it would still be necessary to select these chains with a high throughput screening of structural features [4–6].

A related possibility is to make use of computer-based peptide libraries. Combinatorial approaches have been recently applied to the de novo design of a zinc finger [7] and protein ligands [8], as well as to the search [9] of lead compounds in the pharmaceutical industry. In the present work, combinatorial libraries are essentially tables of amino acid sequences, oligopeptide structure coordinates and the corresponding internal energies. As will be shown, screening of structural properties in these libraries is a straightforward procedure.

In this work we generate tetrapeptide libraries obtained by combining amino acid residues belonging to sets of five amino acids resulting in large data sets, typically containing 5^8 points in the sequence and conformational spaces.

Although the generation (as it is presently done) of all rotamer combinations (from a previously defined set) corresponding to all tetrapeptide amino acid residue combinations is a shot-gun strategy, the results may be rationalized by classifying in conformational families the chain conformations obtained in this systematic search. We may then determine which conformational families are adopted by which tetrapeptide sequences.

Theory

Indexing sequences and main chain conformations

Let us denote by $\{a_1, a_2, a_3, a_4, a_5\}$ the combinatorial library of tetrapeptide chains obtained with the

Table 1. Tetrapeptide sequences and main chain conformations corresponding, respectively, to n_s and n_c values

n_s, n_c	Sequence[1]	Conformation[2]
1	LLLL	$\alpha\alpha\alpha\alpha$
2	LLLK	$\alpha\alpha\alpha\beta$
3	LLLE	$\alpha\alpha\alpha\gamma$
4	LLLI	$\alpha\alpha\alpha\delta$
5	LLLV	$\alpha\alpha\alpha\epsilon$
⋮		
25	LLVV	$\alpha\alpha\epsilon\epsilon$
26	LKLL	$\alpha\beta\alpha\alpha$
27	LKLK	$\alpha\beta\alpha\beta$
28	LKLE	$\alpha\beta\alpha\gamma$
29	LKLI	$\alpha\beta\alpha\delta$
30	LKLV	$\alpha\beta\alpha\epsilon$
⋮		
125	LVVV	$\alpha\epsilon\epsilon\epsilon$
126	KLLL	$\beta\alpha\alpha\alpha$
127	KLLK	$\beta\alpha\alpha\beta$
128	KLLE	$\beta\alpha\alpha\gamma$
129	KLLI	$\beta\alpha\alpha\delta$
130	KLLV	$\beta\alpha\alpha\epsilon$
⋮		
621	VVVL	$\epsilon\epsilon\epsilon\alpha$
622	VVVK	$\epsilon\epsilon\epsilon\beta$
623	VVVE	$\epsilon\epsilon\epsilon\gamma$
624	VVVI	$\epsilon\epsilon\epsilon\delta$
625	VVVV	$\epsilon\epsilon\epsilon\epsilon$

[1] {L,K,E,I,V} peptide library. [2] Set of α, β, γ, δ and ϵ main chain rotamer conformations defined in the Methods section.

amino acids a_1, a_2, a_3, a_4 and a_5. There are $5^4 = 625$ tetrapeptides in this library. Each tetrapeptide chain $a_i a_j a_k a_l$ is identified by the numeral n_s ($1 \leq n_s \leq 625$) according to the equations

$$i = int\left(\frac{n_s - 1}{5^3}\right) + 1$$

$$j = int\left(\frac{n_s - 1}{5^2}\right) - int\left(\frac{n_s - 1}{5^3}\right) * 5 + 1$$

$$k = int\left(\frac{n_s - 1}{5}\right) - int\left(\frac{n_s - 1}{5^2}\right) * 5 + 1$$

$$l = n_s - int\left(\frac{n_s - 1}{5}\right) * 5 \qquad (1)$$

Let us also attribute $\rho = 5$ main chain rotamer conformations to each monomer residue in the library. Each tetrapeptide chain may then adopt 625 conformations $c_i c_j c_k c_l$ identified by the numeral n_c. Replacing n_S by n_C in Equation 1, the monomer residue rotamer combinations corresponding to these 625 conformations are generated. The {A,M,F,G,P} library is an exception because, as shown in the Methods section, two additional rotamers are adopted for proline monomer residues. In that case, 5^n must be replaced by 7^n in Equation 1, (n = 1, 2 or 3) in order to generate 7^4 conformations.

The indexing of tetrapeptide sequences and main chain conformations by n_S and n_C is illustrated in Table 1 for the peptide library {L,K,E,I,V} and five monomer residue rotamer conformations, defined in the Methods section, denoted by the greek letters α, β, γ, δ and ϵ.

Calculation of sequence and main chain conformation probabilities

P_{SC} is the probability of adoption of sequence n_s by conformation n_c. The variables n_s, n_c and P_{sc} may be plotted forming patterns of structural preferences that arise as the sequence and conformational spaces are swept.

To calculate P_{SC} it is necessary to consider that by adopting ρ monomer residue main chain and ρ_s side chain rotamers, $(\rho.\rho_s)^4$ different chain conformations are obtained for each tetrapeptide. Considering, as stated before, $n_a = 5$ amino acid replacements per amino acid residue we reach a total of $(n_a.\rho.\rho_s)^4$ different sequences and conformations. In the present work $\rho = 5$ and $\rho_s = 3$, so that over thirty million tetrapeptide sequences and conformations are obtained per tetrapeptide library.

To each sequence and conformation is attributed the statistical weight $\exp(-E_{n_c}^{\sigma_1,\sigma_2,\sigma_3,\sigma_4}/RT)$, where σ_i ($1 \leq \sigma_i \leq 3$) is the side chain rotamer conformation of residue i and E is the internal energy calculated with a matrix algorithm [10] and the ECEPP/2 force field [11].

Z_s is the partition function of tetrapeptide chain n_s obtained with the equation

$$Z_s = \sum_{n_c=1}^{625} \sum_{\sigma_1=1}^{3} \sum_{\sigma_2=1}^{3} \sum_{\sigma_3=1}^{3} \sum_{\sigma_4=1}^{3} \exp(-E_{n_c}^{\sigma_1,\sigma_2,\sigma_3,\sigma_4}/RT) \qquad (2)$$

On the other hand, z_{SC}, given by

Table 2. Comparison between theoretical results and results derived from experimental data for the peptide libraries and structural motifs in the study

	α-Helix		β-Sheet		3_{10}-Helix		Reverse turn I		γ-Turn	
	$I_{e/t}$	$I_{t/e}$	$I_{e/t}$	$I_{t/e}$	$I_{e/t}$	$I_{t/e}$	$I_{e/t}$	$I_{t/e}$	$I_{e/t}$	$I_{t/e}$
{A,L,V,M,I}	0.99	0.91	0.57	0.76	1.00	0.05	1.00	0.09	0	0
{A,M,F,G,P}	0.75	0.72	0.67	0.46	0.67	0.02	0.56	0.12	1.00	0.01
{D,R,C,S,T}	0.94	0.45	0.89	0.33	0.67	0.01	0.94	0.35		
{E,D,K,R,H}	0.96	0.76	1.00	0.20	1.00	0.01	0.94	0.25		
{L,K,E,I,V}	1.00	0.98	0.66	0.77	1.00	0.04	1.00	0.23	0.50	0.01
{N,Q,H,Y,W}	0.98	0.32	0.91	0.20	0.00	0.00	0.80	0.03		
Average	0.94	0.69	0.78	0.45	0.72	0.02	0.87	0.18	0.50	0.003

The indices $I_{e/t}$ and $I_{t/e}$ are defined in the Theory section. γ-Turns belonging to the libraries {D,R,C,S,T}, {E,D,K,R,H} and {N,Q,H,Y,W} were not found to occur in a database [14] of protein structures.

Table 3. Comparison between the extents of formation of the structural motifs in the study in six peptide libraries and in a database [14] of low sequence identity proteins and sequences. γ-Turns belonging to the libraries {D,R,C,S,T}, {E,D,K,R,H} and {N,Q,H,Y,W} were not found to occur in a database [14] of protein structures

	α-Helix		β-Sheet		3_{10}-Helix		Reverse turn I		γ-Turn	
{A,L,V,M,I}	0.47[1]	0.34[2]	0.31	0.15	0.01	0.24	0.02	0.25	0.002	0.16
{A,M,F,G,P}	0.17	0.25	0.06	0.13	0.005	0.20	0.03	0.15	0.002	0.15
{D,R,C,S,T}	0.13	0.60	0.07	0.38	0.005	0.34	0.08	0.41		
{E,D,K,R,H}	0.25	0.51	0.03	0.26	0.002	0.28	0.05	0.19		
{L,K,E,I,V}	0.61	0.54	0.34	0.24	0.01	0.39	0.05	0.38	0.003	0.25
{N,Q,H,Y,W}	0.08	0.37	0.04	0.25	0.002	0.31	0.01	0.33		
Average ratio	0.69[3]		0.73		0.02		0.15		0.01	

[1] Results derived from experimental data.
[2] Present calculations.
[3] Average ratio of columns 1 and 2.

$$z_{sc} = \sum_{\sigma_1=1}^{3}\sum_{\sigma_2=1}^{3}\sum_{\sigma_3=1}^{3}\sum_{\sigma_4=1}^{3} \exp(-E_{n_c}^{\sigma_1,\sigma_2,\sigma_3,\sigma_4}/RT) \quad (3)$$

is the statistical weight of main chain conformation n_c for all considered side chain conformations. Finally,

$$P_{sc} = \frac{z_{sc}}{Z_s} \quad (4)$$

is the calculated probability that tetrapeptide n_s assumes main chain conformation n_c. There are 625 z_{SC}'s and p_{SC}'s for each tetrapeptide chain. Thus, by averaging side chain conformations in Equation 3, the total number of sequences and conformations has decreased to 5^8.

Comparison with experimental data

A rigorous validation of the present calculations requires a comparison with large numbers of crystallographic or NMR structures of tetrapeptide chains or of tetrapeptide fragments belonging to longer chains.

The main source of polypeptide structures is the Protein Data Bank [12]. We have developed a theoretical protocol, described in detail elsewhere [13], that enables the evaluation of P_{sc} values from a database [14] of low sequence identity protein structures. For each tetrapeptide conformation a list of 625 P_{sc} values is obtained. A P_{sc} belonging to this list is considered a local maximum if it is equal or greater than one fifth of the absolute maximum. By local maximum we mean a cluster of relatively large P_{sc} values. This definition is not necessarily equivalent to the mathematical definition of an extreme of a continuous function.

Similarly, a list of 625 calculated P_{sc} values is obtained for each tetrapeptide conformation. In this case the definition of a local P_{sc} maximum is more involved. If the actual P_{sc} value is less than one fifth of the absolute maximum it is considered a local maximum if it is greater than $P_{s-1,c}$, $P_{s-2,c}$, $P_{s+1,c}$ and

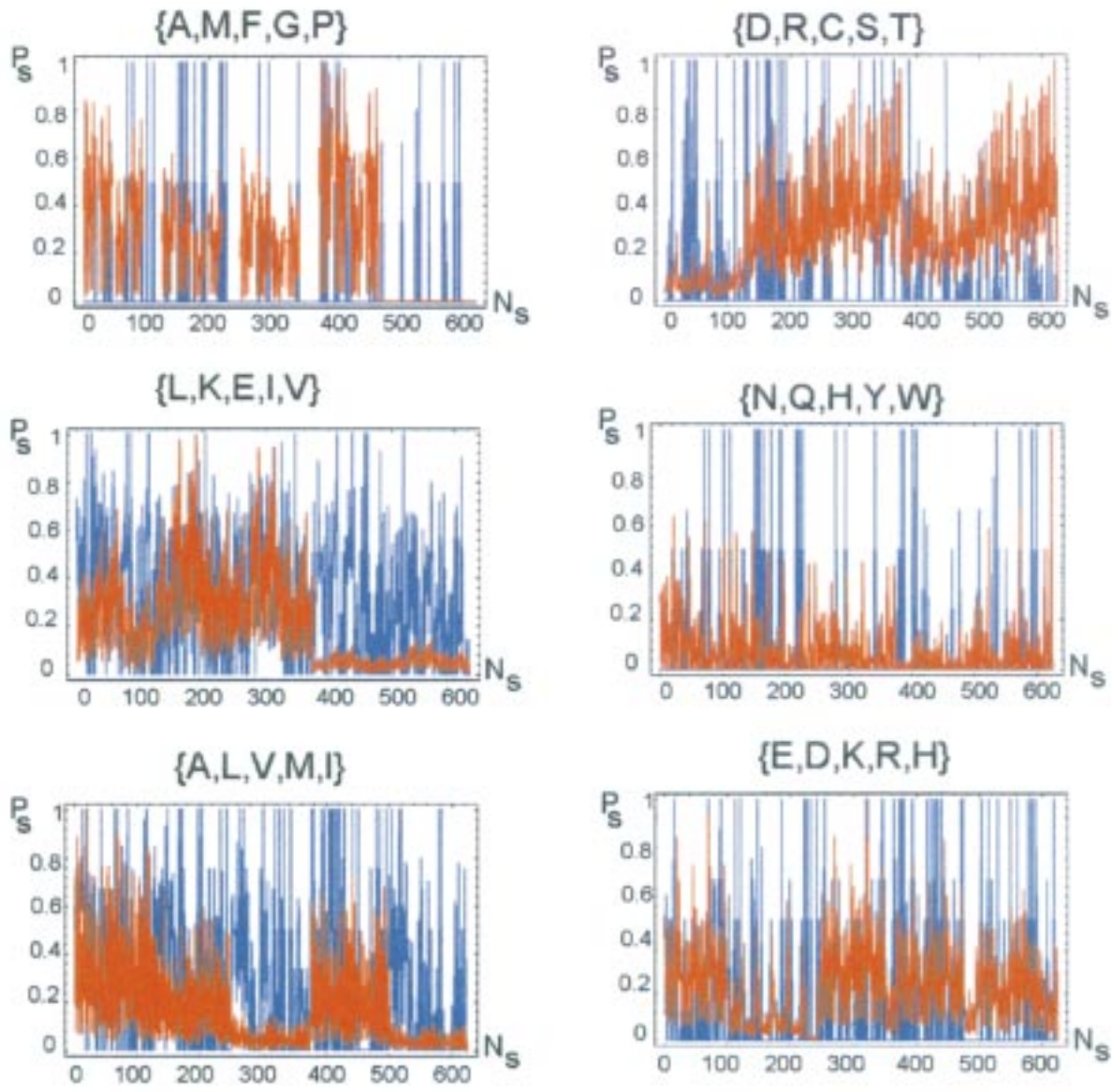

Figure 1. Depicted in red, cross sections of $P_{SC} \times n_s \times n_c$ surfaces, corresponding to the α-helical main chain conformation, belonging to the peptide libraries {A,M,F,G,P}, {D,R,C,S,T}, {L,K,E,I,V}, {N,Q,H,Y,W}, {A,L,V,M,I} and {E,D,K,R,H}. Also shown, depicted in blue, the same cross sections obtained from an analysis [13] of a database [14] of protein sequences and structures with less than 25% sequence identity. P_s axis in arbitrary units.

$P_{s+2,c}$. If this condition is fulfilled, then $P_{s-1,c}$ is also considered a local maximum if it is greater than $P_{s-2,c}$ and P_{s-3c}, and $P_{s+1,c}$ is also considered a local maximum if it is greater than $P_{s+2,c}$ and $P_{s+3,c}$. If P_{sc} is equal to or greater than one fifth of the absolute maximum it is considered a local maximum if it is greater than $P_{s-1,c}$ and $P_{s+1,c}$. If this condition is fulfilled then $P_{s+1,c}$ and $P_{s-1,c}$ are also considered local maxima if they are equal to or greater than one fifth of the absolute P_{sc} maximum.

The adoption of these criteria is based on the principle that proximity in the n_s axis is equivalent to a similarity in amino acid composition, as an examination of Table 1 will show. Thus, clusters of local P_{sc} maxima corresponding to tetrapeptide chains having similar amino acid compositions are identified.

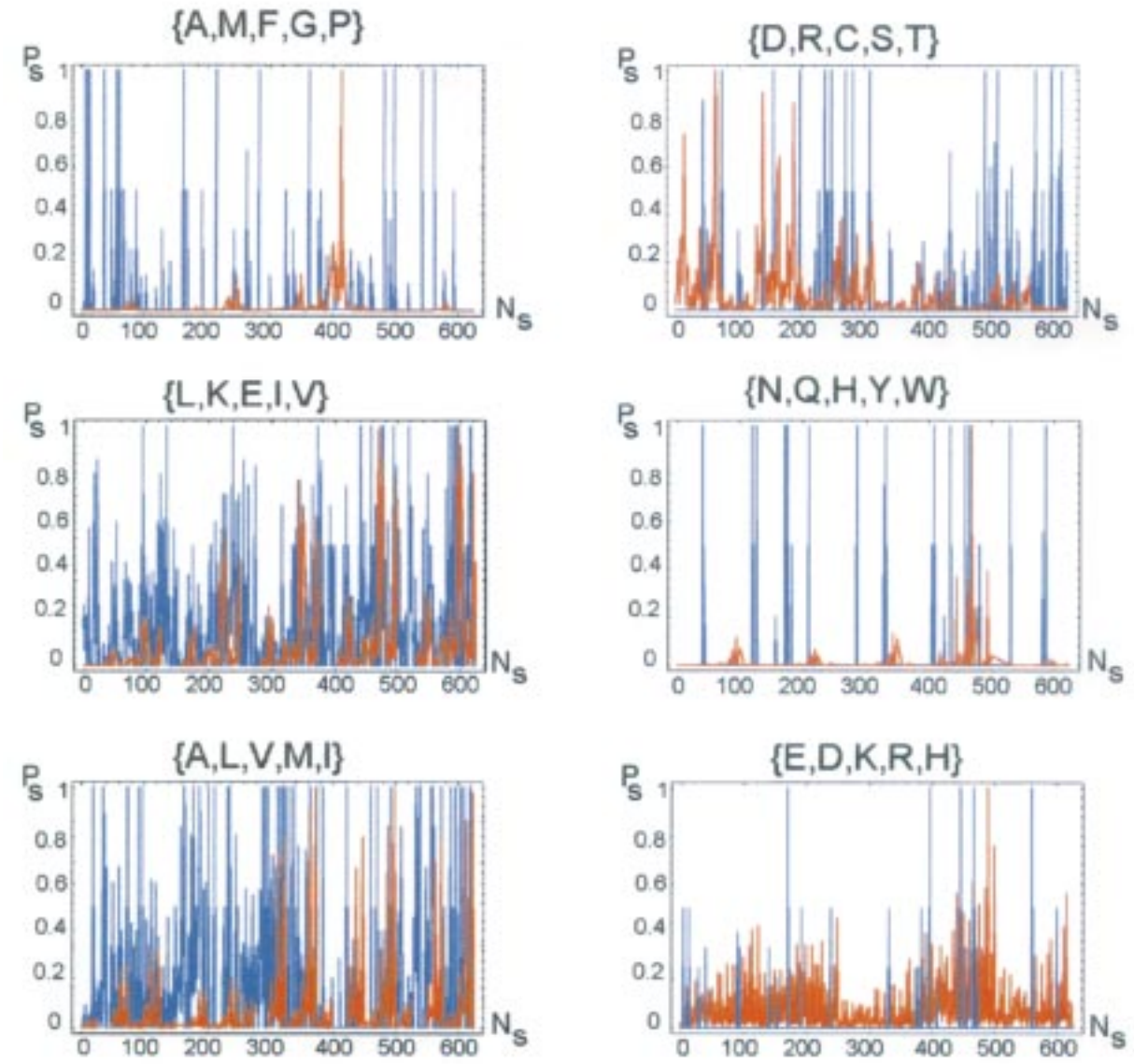

Figure 2. Depicted in red, cross sections of $P_{SC} \times n_s \times n_c$ surfaces, corresponding to the β-sheet main chain conformation, belonging to the peptide libraries {A,M,F,G,P}, {D,R,C,S,T}, {L,K,E,I,V}, {N,Q,H,Y,W}, {A,L,V,M,I} and {E,D,K,R,H}. Also shown, depicted in blue, the same cross sections obtained from an analysis [13] of a database [14] of protein sequences and structures with less than 25% sequence identity. P_s axis in arbitrary units.

Having tabulated P_{sc} theoretical and experimental local maxima we want to establish a comparison between the two sets. If an experimental P_{sc} value is a local maximum and any of the calculated $P_{s-2,c}$, $P_{s-1,c}$, P_{sc}, $P_{s+1,c}$ and $P_{s+2,c}$ is also a local maximum then this proximity is considered a coincidence between theory and experiment. The fulfillment of this criterion ensures that the combinatorial synthesis of an n_s interval comprehending five tetrapeptide sequences will create at least one tetrapeptide with the desired structural propensity.

Finally, the following definitions are also necessary. The number of experimental P_{sc} local maxima coincident with theoretical P_{sc} local maxima divided by the total number of P_{sc} experimental local maxima is defined as the $I_{e/t}$ index. The number of theoretical P_{sc} local maxima coincident with experimental P_{sc} local maxima divided by the total number of theor-

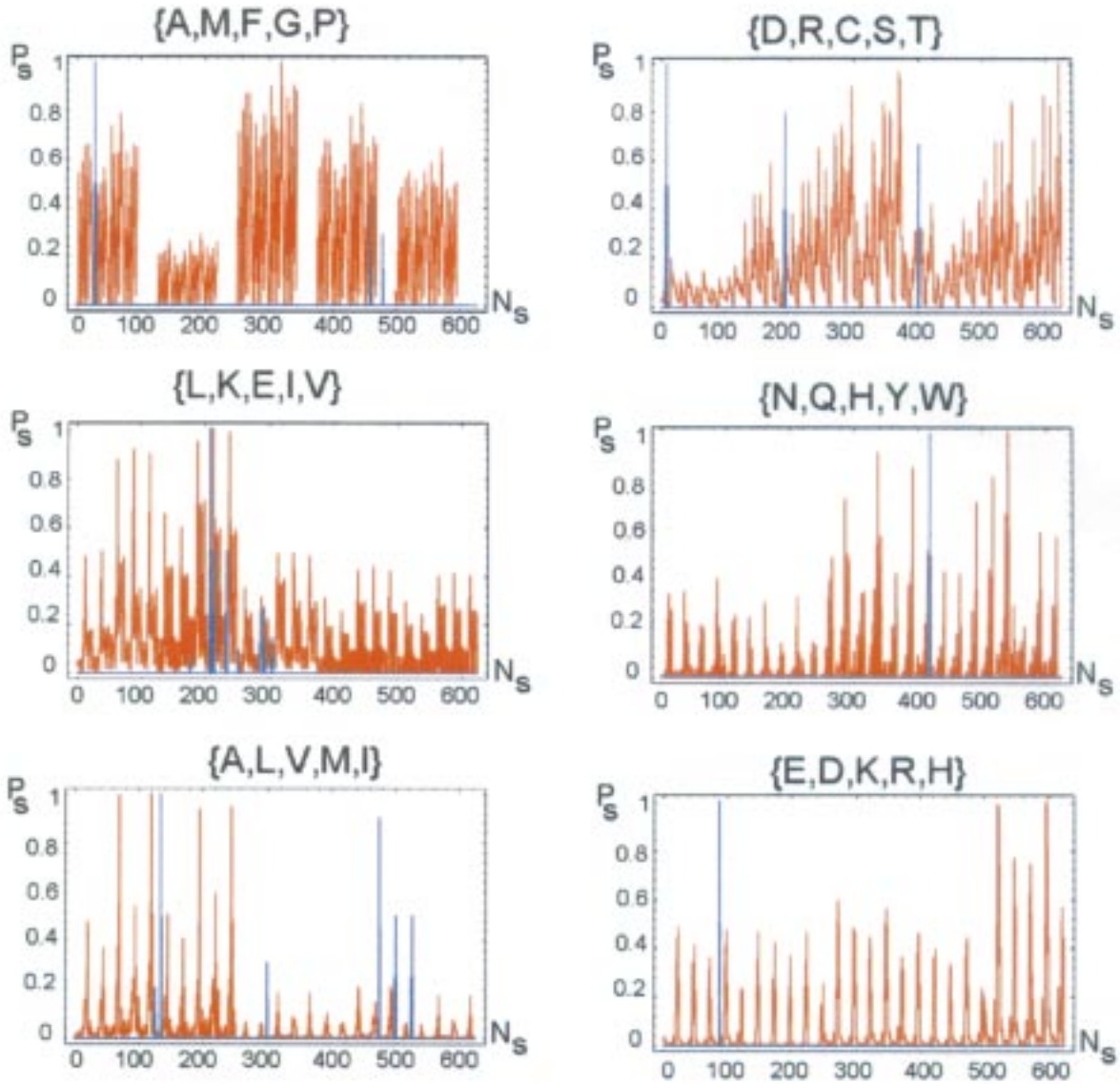

Figure 3. Depicted in red, cross sections of $P_{SC} \times n_s \times n_c$ surfaces, corresponding to the 3_{10}-helix main chain conformation, belonging to the peptide libraries {A,M,F,G,P}, {D,R,C,S,T}, {L,K,E,I,V}, {N,Q,H,Y,W}, {A,L,V,M,I} and {E,D,K,R,H}. Also shown, depicted in blue, the same cross sections obtained from an analysis [13] of a database [14] of protein sequences and structures with less than 25% sequence identity. P_s axis in arbitrary units.

etical P_{sc} local maxima is defined as the $I_{t/e}$ index. The total number of P_{sc} experimental local maxima divided by 625 is the experimental evaluation of the extent of formation of conformation n_c by the peptide library. The total number of P_{sc} theoretical local maxima divided by 625 is the theoretical evaluation of the extent of formation of conformation n_c by the peptide library.

Methods

Five main chain rotamers (α: ϕ = –57°, ψ = –47°; β: ϕ = –139°, ψ = 135°; γ: ϕ = –60°, ψ = –30°, δ: ϕ = –90°, ψ = 0° and ϵ: ϕ = 70°, ψ = –60°; for proline residues ζ: ϕ = –75°, ψ = 158°; ω: ϕ = –75°, ψ = 149° replaced the α and β main chain rotamers) were employed. For the side chain conformations we have made use of the gauche minus, trans and gauche plus

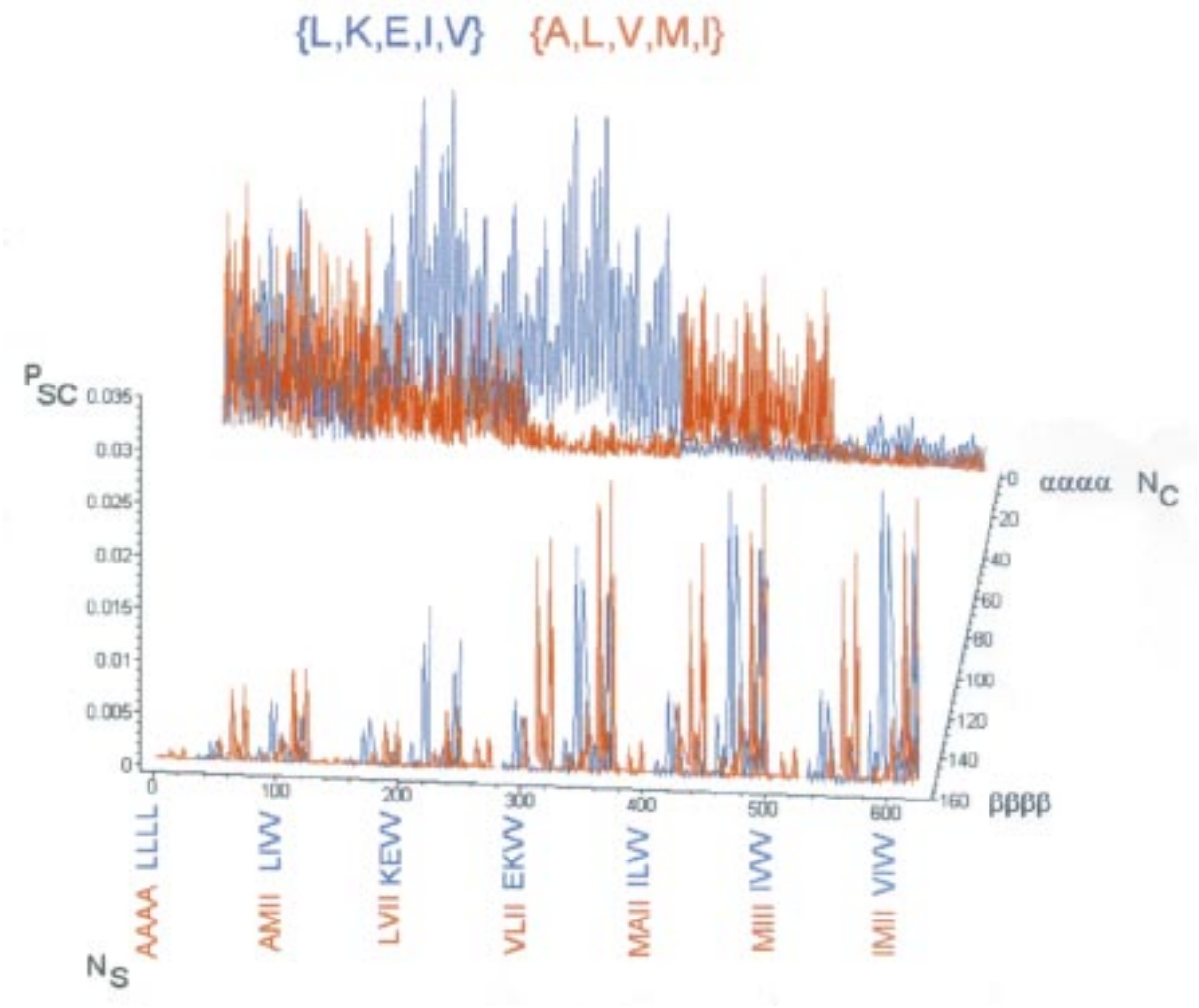

Figure 4. Comparison of α-helical and β-sheet regions of $n_c \times n_s \times P_{sc}$ surfaces corresponding to the {L,K,E,I,V}, depicted in blue, and {A,L,V,M,I}, depicted in red, peptide libraries.

rotamers of the Ponder and Richards [15] classification (the numerals 1, 2 and 3 indicate, respectively, the gauche minus, trans and gauche plus rotamers of χ_1). Gas phase energies were calculated with the ECEPP/2 force field [11]. All calculations were carried out on an Origin 200 Silicon Graphics workstation.

Results and discussion

As discussed below, the simultaneous exploration of the sequence and conformational spaces enables a broad view of structural changes accompanying changes in sequence. With such an analysis it should be possible to predict, based on amino acid composition, the structural diversities of different peptide libraries.

To compare the present calculations and experimental data we have established a comparison with calculations carried out with a methodology described elsewhere [13] in which frequencies of occurrence of structural motifs in the Protein Data Bank are evaluated. Probabilities equivalent to P_{SC} obtained from these calculations will be referred to as data derived from experimental data or simply as experimental data.

The coincidence between P_{SC} maxima in the theoretical and experimental data sets, rather than their relative magnitudes is considered relevant. This choice is justified by reasoning that, although both the occur-

rence and relative magnitudes of P_{SC} maxima indicate structural preferences, the latter may be overwhelmed by interactions with the external environment, which in the present study is everything beyond the tetrapeptide domain. The high $I_{e/t}$ values shown in Table 2 also add to the validity of comparing only the locations of P_{SC} maxima.

The ability of the present calculations to predict whether a peptide library may form a particular structural motif is measured by the index $I_{e/t}$ defined in the Theory section. An $I_{e/t}$ close to 1 indicates a coincidence between n_s regions where that structural motif is found and predicted to occur. The ability to precisely predict the n_s regions that form a structural motif is measured by the $I_{t/e}$ index, as $I_{t/e}$ near unity indicates a coincidence between theoretical and experimental P_{SC} maxima.

Tables 2 and 3 and Figures 1–3 show comparisons between the peptide libraries {A,L,V,M,I}, {A,M,F,G,P}, {D,R,C,S,T}, {E,D,K,R,H}, {L,K,E, I,V} and {N,Q,H,Y,W} and P_{sc} derived from experimental data for the structural motifs α-helix, β-sheet, 3_{10}-helix, reverse turn I and γ-turn.

The following trends are expected. Since only intramolecular interactions belonging to tetrapeptide domains are included in the calculations, better results are expected when short range interactions play an important role in the stabilization of structural motifs. Better $I_{e/t}$ values are expected for peptide libraries and structural motifs in which sterical overlaps play an important de stabilizing role and for the tetrapeptide libraries and structural motifs in which electrostatic interactions play an important stabilizing role, as these interactions are faithfully reproduced by the ECEPP/2 force field.

As shown in Table 2, average $I_{e/t}$ values above 0.70 indicate that, in most cases, the peptide libraries in the study form the structural motifs they are predicted to form. It is also shown in Table 2 that only for the α-helix and β-sheet structural motifs $I_{t/e}$ values higher than 0.70 were obtained, indicating that the calculations are able to predict the n_s intervals where these structural motifs occur (or do not occur).

Instances of high $I_{e/t}$ and low $I_{t/e}$ are seen in Figure 3 for the 3_{10}-helix. Although in most cases the few tetrapeptide sequences able to form a 3_{10}-helix are coincident with P_{SC} maxima, there are more theoretical than experimental P_{SC} maxima. The conclusion is either that the remaining P_{SC} theoretical maxima correspond to tetrapeptide sequences unable to form the 3_{10}-helix or simply that the 3_{10}-helix could be found if a larger set of low sequence identity protein structures was employed in the evaluation of experimental P_{SC} maxima.

Low $I_{t/e}$ values also cause the low average ratios shown in Table 3 between the extents of formation of the reverse turn I and 3_{10}-helix conformations in the experimental and theoretical data sets. We conclude that for the α-helix and β-sheet conformations it is possible to predict precisely the n_s intervals that form these structural motifs and the extent of their adoption by different peptide libraries. For the remaining structural motifs in the study there is a coincidence between experimental and theoretical P_{SC} maxima. We would not be able to point out, however, their exact locations on the n_s axis if we did not examine the experimental data. In these cases it is necessary to produce, in parallel with theoretical calculations, $P_s \times n_s$ plots based on experimental results.

Besides the numerical and graphical comparisons, shown in Tables 2 and 3 and Figures 1–3, between peptide libraries and between libraries and experimental data it is also possible to establish comparisons by superposing, as shown in Figure 4, cross sections of the $P_{SC} \times n_s \times n_c$ surfaces of different peptide libraries. Such a comparison is illustrated in Figure 4 for the α-helix and β-sheet structural motifs and the {L,K,E,I,V} and {A,L,V,M,I} peptide libraries.

Both libraries have n_s intervals ($n_s \geq 375$ and $221 \leq n_s \leq 372$, $n_s \geq 495$, respectively) in which there is a marked decrease in P_{SC} for the α-helix conformation. In these regions there is a predominance of the β branched side chains of isoleucine and valine that may de stabilize [16] the α-helix. A similar effect is observed in the α-helix cross section of the {A,M,F,G,P} library shown in Figure 1. In this case it is due to the occurrence of an α-breaking proline residue at every fifth position on the n_s axis. Since these n_s intervals may be populated by experimental P_{SC} maxima, we conclude that the predicted decrease in P_{SC} may be counter acted by long range interactions or distortions in the α-helical conformation.

As already carried out in model peptides [17], the alternation between high and low P_{SC} regions in the α-helix cross sections could be employed to selectively design the occurrence of α-helical conformations in libraries containing short peptide chains devoid of long range intramolecular interactions.

A similar periodicity is observed in the β-sheet cross sections of Figure 4. In this case, however, the same trend is followed by experimental data, as shown in Figure 2. The alternation between β-sheet form-

ing n_s intervals for the {L,K,E,I,V} and {A,L,V,M,I} libraries, shown in Figure 4, could also be employed in the design of peptide libraries containing (or not containing) the β-sheet conformation.

Conclusions

In comparison with another comparative conformational analysis of peptide chains, the present calculations comprehend much larger data sets, including hundreds of peptide chains. This was achieved at the cost of some simplifications, which are mainly the adoption of gas phase force field potentials and a grid of five main chain and three side chain rotamers.

As shown in the previous discussion, despite these simplifying assumptions the present calculations reflect a detailed level of behavior of tetrapeptide chains. The main advantage of this approach is its broadness, which makes possible a simultaneous comparison of various peptide libraries.

Acknowledgements

This work was supported by FAPESP.

References

1. de Alva, E., Santoro, J., Rico, M. and Jiménez, M.A., *De novo design of a three stranded antiparallel β-sheet*, Protein Sci., 8 (1999) 854–865.
2. Hruby, V.J., Ahn, J.M. and Liao, S., *Synthesis of oligopeptide and peptidomimetic libraries*, Curr. Opin. Chem. Biol., 1 (1997) 114–119.
3. Cho, S.J., Zheng, W. and Tropsha, A., *Focus-2D: A new approach to the design of targeted combinatorial chemical libraries*, Pac. Symp. Biocomput., (1998) 305–316.
4. Chapman, D., *The measurement of molecular diversity: A three-dimensional approach*, J. Comput.-Aided Mol. Des., 10 (1996) 501–512.
5. Hassan, M., Bielawski, J.P., Hempel, J.C. and Waldman, M., *Optimization and visualization of molecular diversity of combinatorial libraries*, Mol. Div., 2 (1996) 64–74.
6. Meyer, B., Weimar, T. and Peters, T., *Screening mixtures for biological activity by NMR*, Eur. J. Biochem., 246 (1997) 705–7099.
7. Dahiyat, B.I. and Mayo, S.L., *De novo protein design: Fully automated sequence selection*, Science, 278 (1997) 82–87.
8. Kay, B.K. and Paul, J.I., *High-throughput screening strategies to identify inhibitors of protein-protein interactions*, Mol. Div., 1 (1996) 139–140.
9. Kubinyi, H., *Chance favors the prepared mind – from serendipity to rational drug design*, J. Recept. Signal Transduct. Res., 19 (1999) 15–39.
10. Jacchieri, S.G. and Jernigan, R.L., *Variable ranges of interactions in polypeptide conformations with a method to complement molecular modeling*, Biopolymers, 32 (1992) 1327–1338.
11. Zimmerman, S.S., Pottle, M.S., Nemethy, G. and Scheraga, H.A., *Conformational analysis of the 20 naturally occurring amino acid residues using ECEPP*, Macromolecules, 10 (1977) 1–9.
12. Bernstein, F.C., Koetzle, T.F., Williams, G.J., Meyer Jr, E.E., Brice, M.D., Rodgers, Kennard, O., Shimanouchi, T. and Tasumi, M., *The Protein Data Bank: a computer-based archival file for macromolecular structures*, J. Mol. Biol., 112 (1977) 535–542.
13. Jacchieri, S.G., *Probing the Structural Diversity of a Combinatorial Library of Tetrapeptide Fragments*, (2000) submitted.
14. Hobohm, U. and Sander, C., *Enlarged representative set of protein structures*, Protein Sci., 3 (1994) 522–529.
15. Ponder, J.W. and Richards, F.M., *Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes*, J. Mol. Biol., 20 (1987) 775–791.
16. Cornish, V.W., Kaplan, M.I., Veenstra, D.L., Kollman, P.A. and Schultz, P.G., *Stabilizing and destabilizing effects of placing beta-branched amino acids in protein alpha-helices*, Biochemistry, 33 (1994) 12022–12031.
17. Baldwin, R.L., *Alpha-helix formation by peptides of defined sequence*, Biophys. Chem., 55 (1995) 127–135.