Journal of Molecular Graphics and Modelling 27 (2009) 611–619

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

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

Do rotamer libraries reproduce the side-chain conformations of peptidic ligands from the PDB?

Amaury Pupo *, Ernesto Moreno

Department of Systems Biology, Center of Molecular Immunology, P.O. Box 16040, Havana 11600, Cuba

ARTICLE INFO

Article history:
Received 20 May 2008
Received in revised form 8 October 2008
Accepted 9 October 2008
Available online 17 October 2008

Keywords:
Rotamer
Rotamer library
Peptidic ligand
MHC
Receptor–ligand interactions

ABSTRACT

Rotamer libraries are collections of side-chain conformations compiled for each type of amino acid. All the libraries developed so far were constructed based on different sets of proteins, none of which includes small peptidic molecules. Are the existing rotamer libraries suitable for modeling the side chains of peptidic ligands in complex with their receptors? To answer this question, we have tested 10 different, publicly available and commonly used rotamer libraries for their capability of reproducing the side-chain conformations (and therefore the receptor–ligand interactions) of hundreds of short peptidic ligands found in the Protein Data Bank, including large sets of class I and class II T-cell epitopes. Only the libraries developed by Xiang and Honig were able to correctly reproduce the experimental geometries for most of the analyzed residues, and the atomic interactions between the peptidic ligands and their receptors. Surprisingly, all the libraries showed a lower performance in reproducing the side chains conformations from structures solved at very high resolution ($R < 1.25$ Å).

1. Introduction

The conformation of an amino acid side chain in a protein is determined by a number of factors, including intrinsic conformational preferences, interactions with other parts of the protein and interactions with the solvent [1]. Nevertheless, in spite of the large variability of such factors, side chains tend to exist in a limited number of low energy conformations, usually called "rotamers" [2]. Therefore, instead of considering the full geometrically possible conformational space, only a relatively small number of rotamers can be used to describe most naturally occurring conformers of aminoacidic residues [3].

A rotamer is usually thought to be a local minimum on a potential energy map, or an average conformation over some region of the torsion angle space. Non-rotameric is a term used sometimes to describe side chains that have dihedral angles far from average values or far from a local energy minimum [4]. A rotamer library is a collection of rotamers, compiled for each type of amino acid [4]. Depending on the way they were derived, rotamer libraries are classified as backbone-independent, secondary structure-dependent, backbone-dependent and position-specific. From these classes, the backbone-independent and backbone-dependent libraries are the most used [4].

The first complete rotamer library was developed in 1987 by Ponder and Richards [2]. This library contains a list of side-chain conformations characterized by their average dihedral angles, variances and frequencies. A few years later, Tuffery et al. [5] performed a cluster analysis to produce a library of 113 rotamers, which was subsequently expanded to 212 [6]. In 1993, Dunbrack and Karplus [7] presented the first backbone-dependent rotamer library. That same year, Schrauber et al. [8] compiled a rotamer library as part of a study aimed to determine whether or not side-chain conformations in globular proteins are rotameric. Despite their overall negative conclusion about the rotameric behavior of amino acid side chains, the library they produced was used by De Maeyer et al. [9], who combined these rotamers with those of Ponder and Richards [2], adding also some extensions in order to include angles past $\chi 2$ and to sample regions of the torsion space not included in the previous studies.

In 1997, Dunbrack and Cohen [10] used Bayesian statistics to estimate dihedral angles for all rotamers of all side-chain types at all values of backbone angles φ and ψ. Three years later, Richardson and coworkers [11] built a backbone-independent rotamer library using much stricter criteria for including side chains in the data set [12].

* Corresponding author.
E-mail address: amaury@cim.sld.cu (A. Pupo).

1093-3263/$ – see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2008.10.002

Finally, in 2001, Xiang and Honig [13] obtained a remarkable accuracy in predicting core residue conformations using an extensive library of 7560 rotamers. The "rotamers" they selected do not necessarily lie at the bottom of local energy wells, so it is worth distinguishing the usage of this term by these authors from the common acception that refers to conformations belonging to a local energy minimum. The rotamers generated by Xiang and coworkers were derived from nine non-redundant databases which contained different numbers of proteins (from 135 to up to 2312) [13].

Almost all methods for predicting side-chain conformations [5,10,13] make use of rotamer libraries. Likewise, the use of rotamers significantly improves the speed and accuracy of building crystallographic models, and for this reason, side-chain rotamer libraries have been incorporated into crystallographic model-to-map fitting programs such as "O" [14] and "XtalView" [15]. Furthermore, angle expectations are part of structure verification tools [16,17], including those used to check the protein structures being deposited at the Protein Data Bank (PDB) [18]. Rotamer libraries are also widely used in homology modeling [19], Monte Carlo and combinatorial calculations [20], and protein design [21,22].

At this point, it is worth noting that all the rotamer libraries developed so far were constructed based on different sets of proteins, none of which includes small peptidic molecules. On the other hand, there are many examples of biological systems where small peptidic ligands play an important role, e.g. they can function as hormones, signal molecules and as substrates of many protein classes [23]. Therefore, predicting the 3D structure of receptor–peptide complexes, which involves the correct modeling of peptide side chains, is a relevant issue in many molecular modeling applications.

Are the existing rotamer libraries suitable for modeling the side chains of peptidic ligands in complex with their receptors? It has been shown that prediction methods based on available libraries yield good result for core residues of globular proteins, but not for residues exposed at the protein surface [13]. The amino acids of peptides being in complex with a protein, on the other hand, are either exposed to the solvent or interacting with the receptor surface, so they cannot be considered as core residues. Thus, the question about the "rotamericity" of peptidic ligands remains open.

An important type of receptor–peptide complex, playing a key role in the immune system, is the one formed by a MHC molecule (of class I or II) with any of its many possible peptidic ligands, called T-cell epitope. The prediction of which peptide segments from a polypeptide chain can bind to the groove of a MHC molecule (i.e. can become T-cell epitopes) is a relevant problem in immunology, and different computational methods have been devised to achieve this goal. Some of the reported T-cell epitope prediction methods are structure-based [24–26], i.e. their design was based, in more or less degree, on modeling the peptide structures in the MHC groove.

We have recently concluded the development of a new algorithm for prediction of class II T-cell epitopes (*manuscript in preparation*), which combines structural and artificial intelligence approaches. A key element in developing this new algorithm was the modeling of different amino acid side chains bound to different pockets of the MHC groove. For this purpose we needed a reliable rotamer library as a source of possible side-chain conformations, so this was our main motivation for embarking in the study presented in this paper.

Here we test 10 different, publicly available and commonly used rotamer libraries for their capability of reproducing the side-chain conformations (and therefore the receptor–ligand interactions) of hundreds of peptidic ligands found in the Protein Data Bank, including large sets of class I and II T-cell epitopes.

2. Methodology

2.1. Selection of peptidic ligand datasets

Small peptidic ligands were extracted from the Protein Data Bank [27] (as on February 2007) and subsequently filtered to generate a non-redundant set of small ligand–protein complexes, following the procedure described in [28]. Only complexes whose ligands have up to eight standard aminoacidic residues were selected (ligands may have as well other non-aminoacidic residues or modified/non-standard amino acids). In order to find all the deposited MHC structures, we performed a BLAST search [29] on the PDB protein sequences using class I and class II MHC sequences as input, and manually checked the output. Then we applied the same procedure as above to extract the peptidic ligands, but setting a higher peptide length filter. Some MHC class II structures required manual intervention to eliminate a covalent link between the ligand and the MHC receptor.

Prior to ligand extraction, all the selected PDB entries were processed with the program 'reduce' [30], to explicitly add hydrogen atoms and optimize the side-chain orientations of Asn, Gln and His, allowing these side chains to flip, if this improves their hydrogen bonding capabilities. Afterwards, using home-made scripts, we parsed the USER MOD fields of the resulting PDB files, to track the side chains that were flipped in the peptidic ligands.

2.2. Evaluation of rotamer libraries

Ten rotamer libraries available in the Internet were used in this study (see Table 1). The four library variants chosen from those developed by Xiang and Honig [13] are dihedral angle rotamer libraries generated from protein datasets including 135, 297, 533 and 884 members, respectively, using a 10° step to sample the conformational space and covering 100% of the side chains conformations found in the source protein datasets. Each library is coded with a short name (listed in Table 1) used in tables and figures.

We considered that a rotamer library is able to reproduce a given set of side-chain chi angles if it contains at least one rotamer having all the corresponding chi angles within 20° from the given values. To evaluate the performance of each rotamer library in reproducing the experimental residue geometries, we generated

Table 1
Rotamer libraries used in this study.

Rotamers libraries[a]	Number of rotamers	References
L1: Demaeyer	330	De Maeyer et al. [9]
L2: DunbrackBBdep	466,829 (341 per bin)	Dunbrack and Cohen [10]
L3: DunbrackBBind	341	Dunbrack and Cohen [10]
L4: PondersRichards	84	Ponder and Richards [2]
L5: Richardson	170	Lovell et al. [11]
L6: Tuffery	324	Tuffery et al. [5]
L7: Xiang135	3,679	Xiang and Honig [13]
L8: Xiang297	7,421	Xiang and Honig [13]
L9: Xiang533	11,810	Xiang and Honig [13]
L10: Xiang844	17,753	Xiang and Honig [13]

[a] Abbreviated names are used for the different libraries, most of them are self-explanatory. DunbrackBBind and DunbrackBBdep are the Dunbrack's backbone-independent and backbone-dependent rotamer libraries, respectively. Libraries are also denoted with shorter codes (L1 to L10), which are used in some tables and figures for better clarity. See the text in Section 2 for details on the Xiang and Honig's libraries.

all the possible conformations defined by the sets of rotamers included in all the libraries and calculated the RMSD between each PDB ligand residue and its corresponding set of rotamers in each library, keeping the lowest RMSD value. The RMSD was calculated for side-chain heavy atoms beyond the CB carbon. The rotational symmetry of Asp, Glu, Phe, Tyr and Arg was taken into account in the RMSD calculations. The library of Ponders and Richards [2] has no information on 4th chi angles, therefore it was not used in the calculations for Arg and Lys.

To obtain an expanded set of conformations from the Richardson's rotamer library [11] the side-chain torsion angles were sampled using an incremental step of 1° for chi 1 and 3° for the other chi angles, around all rotameric conformations included in the library and keeping the sampling within the sigma values (standard deviations) defined in the library for each chi angle in each rotamer [11].

2.3. Analysis of residue interactions

Interactions between ligand and receptor residues or between ligand residues were calculated using home-made scripts to be run within the VMD program (Visual Molecular Dynamics) [31]. For salt bridges, we used the saltbr plugging included in VMD, with default parameters. For hydrogen bond calculations we used the "measure hbonds" VMD command, with a distance cutoff of 3.0 Å and a cutoff angle of 20°. Contacts were computed using a 3.2–4.2 Å distance window, excluding pairs of atoms involved in hydrogen bonds or salt bridges. Ala, Gly and Pro were excluded from the interaction analysis.

2.4. Statistical analysis

The R software [32] was employed for all the statistical analyses. Statistical differences among groups were determined using the non-parametric Dunn test.

3. Results and discussion

3.1. Analysis of the peptidic ligand datasets

Three different sets of peptidic ligands were extracted from the Protein Data Bank as described in Section 2. The small ligand set consists of molecules having up to eight amino acid residues (they may have other non-aminoacidic residues or modified amino acids). The other two groups comprise MHC class I and class II ligands, respectively. Table 2 shows some statistics computed for the three sets. Amino acid side chains were classified as "restricted" or "free", as a way to characterize their flexibility when bound to the protein receptor. Every residue having at least one side-chain heavy atom (beyond the CB carbon) in contact with the receptor was considered as restricted, otherwise it was classified as free. As shown in Table 2, most side chains in the three ligand groups are oriented towards the receptor, and therefore their flexibility is restricted as compared with those side chains that are totally exposed to the solvent.

Table 2
Groups of peptidic ligands included in the study.

Ligand groups	Number of ligands	Total number of residues[a]	"Restricted" residues (%)
Small	354	1195	79.33
MHC class I	141	867	91.46
MHC class II	37	297	87.54
Total	532	2359	84.82

[a] Ala, Gly and Pro are excluded.

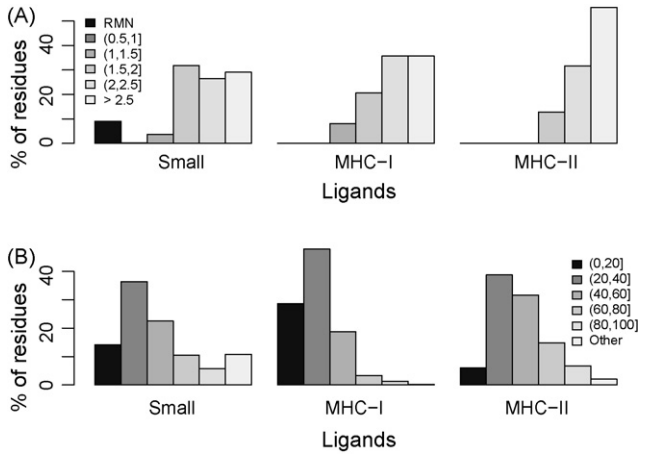

Fig. 1. Distribution of ligand residues according to structure resolution (A) and B factors (B).

Most of the small ligand and MHC class I residues, and about half of the MHC class II residues, come from structures with resolution better than 2.5 Å (Fig. 1A). The B values corresponding to peptidic ligand atoms span a broad range, as shown in Fig. 1B. In particular, more than half of the MHC class II residues have B values greater than 40.

Table 3 shows a comparison of the residue composition of the three peptide groups with the amino acid frequencies reported in the UniProtKB/Swiss-Prot database [33]. The three ligand groups have a low content of cysteine as compared to UniProt proteins, a fact that is consistent with the linear nature of most of these ligands and with their non-covalent interactions with the receptor. The three groups, and in particular the small ligands, are enriched in proline, a residue that constrains the backbone flexibility. Frequencies higher than expected are also observed for tyrosine in MHC class I ligands and for glycine and lysine in MHC class II

Table 3
Amino acid composition per ligand group.

Residue	Small ligands		MHC-I ligands		MHC-II ligands		UniProtKB/Swiss-Prot (%)
	N	%	*N*	%	*N*	%	
A	129	8.16	96	8.67	38	9.18	7.89
R	84	5.31	73	6.59	19	4.59	5.4
N	35	2.21	39	3.52	19	4.59	4.13
D	87	5.5	21	**1.9**	7	**1.69**	5.35
C	14	**0.89**	4	**0.36**	1	**0.24**	1.5
Q	38	2.4	40	3.61	15	3.62	3.95
E	110	6.96	40	**3.61**	19	**4.59**	6.67
G	88	5.57	61	5.51	42	**10.14**	6.96
H	36	2.28	15	1.36	13	3.14	2.29
I	64	4.05	73	6.59	17	4.11	5.9
L	120	7.59	124	**11.2**	26	6.28	9.65
K	69	4.36	43	3.88	32	**7.73**	5.92
M	23	1.45	28	2.53	9	2.17	2.38
F	98	6.2	72	6.5	21	5.07	**3.96**
P	169	**10.69**	83	7.5	37	8.94	**4.83**
S	94	5.95	64	5.78	23	5.56	6.83
T	71	4.49	55	4.97	23	5.56	5.41
W	43	2.72	11	0.99	5	1.21	1.13
Y	86	5.44	82	**7.41**	16	3.86	**3.03**
V	123	7.78	83	7.5	32	7.73	6.73

Total number of residues per amino acid type (*N*), for the different ligand groups, and the percent they represent. The amino acid frequencies reported in the UniProt database are included for comparison. Numbers in bold indicate significant differences with respect to UniProt amino acid frequencies.

peptides. This is consistent with the fact that Tyr constitutes a principal anchor residue for many MHC class I alleles, while Gly and Lys are common residues for some HLA-DRB1* alleles, as reported in the SYFPHEITI database [34]. Relatively low frequencies are found for acid residues (Asp and Glu) in MHC class I and class II ligands.

In order to correct possible erroneous orientations of Asn, Gln and His side chains in the peptidic ligands, we processed all the complexes included in this study with the program 'reduce', as explained in Section 2. Previous studies have shown that incorrect side-chain positioning of these amino acids is quite common [30], even in high quality structures. Here, these misplacements might lead to inaccuracies in the evaluation of the performance of some of the rotamer libraries included in our study, especially if such mistakes were first corrected in the data used to develop the library. As shown in Table 4, the ligand group having the highest number of incorrect amide and His side-chain placements is the MHC-I set, with more than 23% of their Asn, Gln and His side chains been flipped by 'reduce'. For the small and MHC-II ligands, the fraction of misplaced side chains is less significant.

Table 4
Fraction (in percentage) of amide and His side chains being flipped by the program 'reduce' in the peptidic ligands.

Ligand groups	Asn	Gln	His	Total[a]
Small	5.13	10.53	10.26	8.62
MHC class I	17.5	31.11	18.75	23.76
MHC class II	0	7.69	7.69	4.44

[a] Percent of flipped side chains from the total number of Asn, Gln and His residues.

3.2. Evaluation of the rotamer libraries

We tested 10 different rotamer libraries (listed in Table 1) for their capability of reproducing the side-chain conformations of the selected peptidic ligands. Our library collection starts with the first complete library reported by Ponder and Richards [2] and includes the commonly used Dunbrack's backbone-dependent rotamer library [10] and the penultimate rotamer library of Richardson [11]. There are significant differences among the libraries regarding the total number of rotamers. The large number of conformers in Dunbrack's backbone-dependent library is due to the partitioning of the backbone angle space into 1369 bins, each containing a non-overlapping 10° subset of the phi and psi angles with their corresponding rotamers.

Following the commonly used approaches [35], we computed root mean square deviations (RMSD) and compared chi angles to evaluate the reliability of the studied rotamer libraries in reproducing the experimentally observed side-chain conformations of peptidic ligands. First we analyzed, for each rotamer library, each rotatable side chain in the selected ligands to check if the library contains at least one rotamer with chi angles within 20° from the corresponding angles in the analyzed residue, as explained in Section 2.

The 10 rotamer libraries included in this study correctly describe 100% of the ligand chi 1 angles, however, they differ substantially in reproducing the chi 1 to chi 2 angle pairs (chi_{1+2}). As shown in Fig. 2, only Xiang's libraries [13], which are more detailed than the rest, can reproduce almost all side-chain chi_{1+2} angle pairs for all the studied residue types.

The four Xiang's libraries outperform the other rotamer libraries also in reproducing the angle combinations chi_{1+2+3} and $chi_{1+2+3+4}$ (Supplementary Figs. S1 and S2, respectively). The other six libraries reproduce up to 40% of the chi_{1+2+3} experimental values and about 20% for the $chi_{1+2+3+4}$ combinations, while these numbers are greater than 80 and 60%, respectively, for the Xiang's libraries (the library of Ponders and Richards [2] was not included in this evaluation, since it lacks information on chi 4 dihedral angles).

Having significant differences in chi angles between an experimental and a rotamer library conformation does not necessarily imply having significant deviations in spacial coordinates, since in some cases the individual differences in chi angles

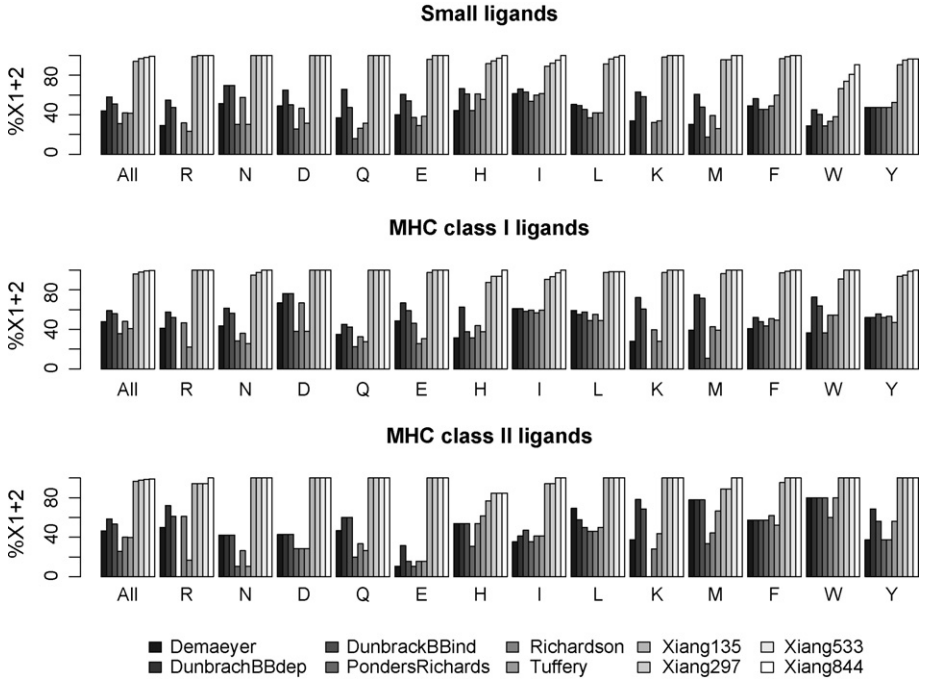

Fig. 2. Reproducibility of the chi_{1+2} angle pairs by the different rotamer libraries. The histograms show the percent of peptide residues, by amino acid type and by ligand group, having both the chi 1 and chi 2 angles covered by each rotamer library.

may compensate each other, rendering the overall position of the modeled side chain close in space to the experimental geometry. Therefore, as a second test for the rotamer libraries we calculated, for each peptidic ligand residue, the smaller RMSD that can be achieved between the PDB geometry and a rotamer conformation from a library.

As explained in Section 2, we generated all the amino acid conformers defined by the sets of rotamers from each of the evaluated libraries, and then compared the conformations observed in the peptidic ligand residues with the computed ones. From these calculations we kept, for each rotamer library, the lowest RMSD value obtained for each residue. It is worth noting that the RMSD's were calculated only for side-chain heavy atoms and excluding the CB carbon, i.e. only for those atoms that actually move upon side-chain rotations.

The results from these calculations are shown in Fig. 3. Similarly as obtained for the chi angle tests, the Xiang's rotamer libraries outperform the rest in reproducing the overall side-chain positions of the analyzed residues. If we use the same criterium employed by Schrauber et al. [8] to evaluate our results (a RMSD cutoff value of 0.5 to consider that the conformation from the rotamer library and the experimental one are equivalent), then only the Xiang's libraries successfully passed the test. Moreover, for the other six libraries the boxplots in Fig. 3 show that for about 25% of the PDB side-chain conformations the RMSD of the closest rotamer is greater than 2 Å. For the four Xiang's libraries, however, about 75% of the calculated RMSD's are below 0.5 Å and almost 100% of them are below 1.5 Å.

The simplest explanation for the better performance of Xiang's libraries is that they are more detailed than the others (the total number of rotamers in the backbone-dependent library of Dunbrack [10] is higher, but the number of rotamers for each phi–psi bin is much lower). One simple way to create a more detailed library is by adding flexibility, i.e. expanding the library around each of its rotamers. We tested this approach with the library of Richardson [11], which was expanded by adding conformers around the defined dihedral angles within the sigma values also defined in the library, as explained in Section 2. Then, the same RMSD calculations described above were performed for the expanded set of Richardson's rotamers. As can be seen in Fig. 3, the RMSD values improved as compared with those obtained for the original library, however, they were still far from reaching the accuracy achieved with the four Xiang's library variants.

In addition to the penultimate rotamer library, the Richardson's laboratory provides a recently developed data set, called "top500", which contains statistics from an exhaustive chi angle sampling performed on a large set of proteins (500 high-resolution structures, http://kinemage.biochem.duke.edu/databases/top500.php). Although our work is focused on evaluating the performance of rotamer libraries, we decided to evaluate also in which extend the collection of experimental side-chain conformations compiled in the top500 data set covers the chi angle space spanned by peptidic ligands. In this set, a density value was calculated for each bin of the chi angle space, for each amino acid type, as a measure of the number of side chains found within the bin. Taking the chi values for each bin having a non-zero density value leads to a set of

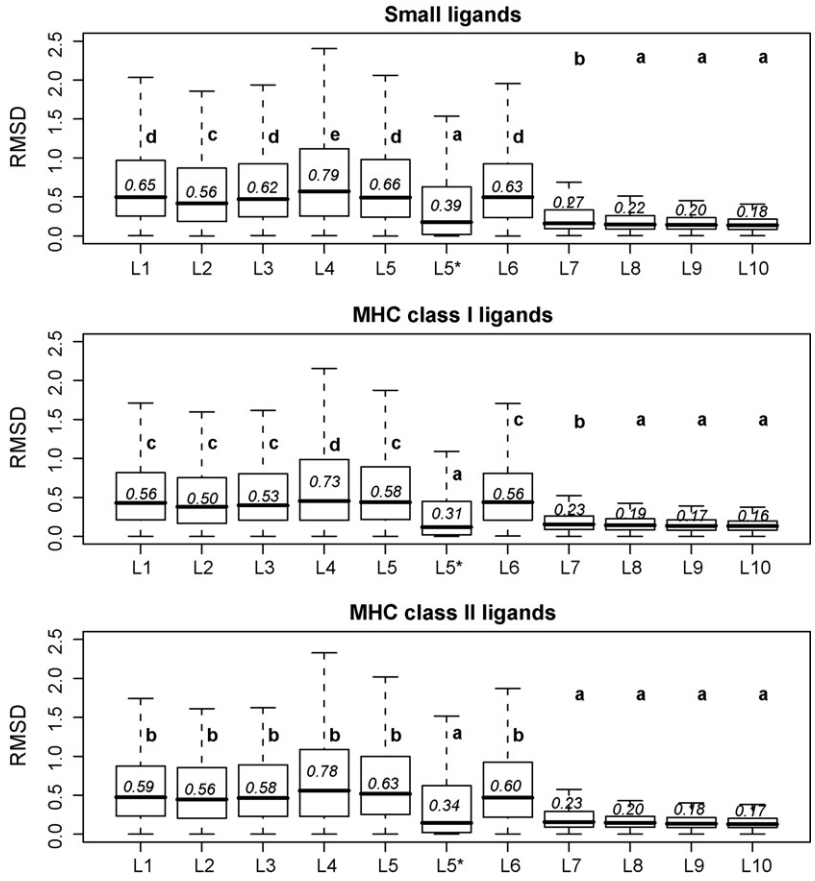

Fig. 3. Boxplots showing the distributions of the minimal RMSD values that can be achieved, using different rotamer libraries, between a particular PDB side-chain conformation and a rotamer from a given library. L1 to L10 refer to the rotamer libraries, as listed in Table 1. L5* is the expanded library of Richardson (see text). The mean value of each distribution is shown. Different letters above the RMSD distributions mean statistically significant differences.

more than 1.8 million conformations, which is larger than the set we generated by expanding the Richardson's library, and much larger than any of the 10 rotamer libraries included in our study. This set of conformations reproduces the side-chain geometries much better than any of the libraries, yielding mean RMSD values of 0.087, 0.078 and 0.073 for small, MHC-I and MHC-II ligands, respectively. Therefore, the top500 data set may serve as source to develop a more detailed and comprehensive library.

The rotamers defined in the first six libraries listed in Table 1 are supposed to lie in the bottom of local energy wells, however, the side-chain conformations included by Xiang and Honing in their four library variants do not necessarily fulfill this condition [13]. If we take as valid the assumption that a side-chain rotamer should correspond to a conformation in an energy minimum, then a significant number of the side-chain conformations observed in the peptidic ligands would be "non-rotameric".

There are several possible explanations for the presence of non-rotameric conformations in proteins [4]: the side chain is improperly fit to the observed electron density; the observed conformation is actually an average of two or more rotameric conformations in equilibrium; steric interactions with the backbone may force the side chain to adopt strained dihedral angles; or the strained conformation is stabilized by highly favorable interactions with the rest of the protein. In ligands of peptidic nature, the side chains of some residues may be driven to non-rotameric states to improve their complementarity with the corresponding receptor, in a similar way as receptor residues can be accommodated after the binding of their ligands [36].

Whatever the explanation is for the existence of side-chain geometries outside the low energy regions of the conformational space, the practical conclusion we can extract from our results is that, in order to reproduce the residue conformations observed in the peptidic ligand structures from the PDB we need a detailed rotamer library, such as the Xiang's libraries, that makes not only a wider, but also a more comprehensive sampling of the conformational space.

In addition to the tests described above, we checked whether the resolution of the structures, the B factors, and the steric restrictions imposed by protein receptors (whether the peptidic residue is "restricted" or "free") affect the capability of the rotamer libraries to reproduce the experimental side-chain conformations (as measured by RMSD calculations). As shown in Table 5, the RMSD values computed for restricted residues tend to be higher than those calculated for free residues (although the differences are not always statistically significant), suggesting that the interaction with the receptor drives some side chains towards less rotameric conformations. This effect is more marked for MHC class II ligands.

Residues coming from structures with resolutions in the range 1.25–2.5 Å are more rotameric (see Table 6), which is in correspondence with the fact that the studied libraries were compiled based on structures with high resolution. For MHC ligands, which span only two of the resolution ranges defined in Table 6, the libraries also reproduce best the conformations of residues from the higher resolution group. For small ligands, there are more resolution ranges, as well as some NMR structures. Surprisingly, the 10 rotamer libraries show a markedly lower performance when reproducing the residue conformations in 14 structures with very high resolution (0–1.25 Å). There is no clear correlation between the ability of rotamer libraries to reproduce side-chain conformations and residue B factors (data not shown), although in most cases residues with B factors smaller than 40 are best reproduced.

Since crystallographic programs use rotamer libraries in the first stages of model building [14,15] to define the atomic coordinates within regions of the electron density map where the shape of the side chain is not clearly defined, it is probable that the geometry of residues with high B factors and residues in structures with medium or low resolution are biased towards

Table 5
Average RMSD's for free and restricted side chains.

Ligand groups	Residue class	Average of best RMSD's										
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	*N*
Small ligands	Free	0.63	0.54	0.60	0.78	0.67	0.64	0.24	0.21	0.19	0.17	241
	Restricted	0.66	0.57	0.62	0.79	0.66	0.62	0.27	0.23	0.20	0.19	922
MHC-I ligands	Free	0.51	0.48	0.47	0.63	0.53	0.55	0.21	0.16	0.14	0.13	78
	Restricted	0.57	0.51	0.54	0.74	0.59	0.56	0.23	0.19	0.18	0.16	788
MHC-II ligands	Free	0.50	0.45	0.48	0.63	0.48	0.50	0.17	0.14	0.12	0.11	37
	Restricted	0.60	0.58	0.60	0.80	0.66	0.61	0.24	0.21	0.19	0.18	259

L1 to L10 refer to the studied libraries, as listed in Table 1. *N* is the number of residues in each residue class.

Table 6
Average RMSD's for ligand residues belonging to structures in different resolution ranges.

Ligand groups	Resolution range	Average of best RMSD's										
		L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	*N*
Small ligands	NMR	0.75	0.70	0.78	0.94	0.85	0.83	0.30	0.23	0.21	0.18	106
	(0, 1.25]	0.98	0.89	0.98	0.96	1.13	0.93	0.51	0.42	0.34	0.30	25
	(1.25, 2.5]	0.60	0.51	0.57	0.75	0.60	0.57	0.24	0.21	0.19	0.17	693
	(2.5, 3.75]	0.70	0.60	0.65	0.81	0.70	0.66	0.29	0.24	0.21	0.19	335
MHC-I ligands	(1.25, 2.5]	0.54	0.47	0.51	0.72	0.55	0.53	0.22	0.18	0.17	0.16	558
	(2.5, 3.75]	0.60	0.56	0.58	0.73	0.63	0.61	0.25	0.20	0.18	0.17	308
MHC-II ligands	(1.25, 2.5]	0.47	0.45	0.47	0.69	0.53	0.48	0.19	0.17	0.15	0.15	131
	(2.5, 3.75]	0.68	0.65	0.67	0.86	0.71	0.68	0.26	0.23	0.20	0.19	165

L1 to L10: see Table 1. *N* is the number of residues in each resolution range.

rotameric conformations. In structures with very high resolution, by difference with the previous case, the contours of the side chains are well delineated and, therefore, their conformations are not biased towards any tabulated geometry.

3.3. Reproducibility of receptor–peptide interactions

When modeling the side chains of a peptidic ligand bound to its receptor, the most important issue is to model as accurately as possible the atomic interactions between the two molecules. In this regard, the spatial orientation of each interacting side chain, and in particular the positioning of the functional groups that are relevant in the interaction, is crucial for the accuracy of the model.

We decided to investigate, for each of the 10 studied rotamer libraries, in which extend the interactions between ligand and receptor amino acids are affected when the ligand side chains are modeled using the closest rotamer (the one with the smallest RMSD from the experimental geometry). Three different types of interaction were analyzed: salt bridges (interactions between the negatively charged groups of Asp and Glu with the positively charged groups of Arg, His and Lys), hydrogen bonds and any other type of interaction, mainly hydrophobic contacts, grouped under the generic term ''contact''. Different criteria were used to define each type of interaction, as detailed in Section 2.

We computed the three types of interaction of ligand residues both with the receptor and with the rest of the ligand. As shown in Table 7, residues from small and MHC class I ligands make salt bridges only with residues from the receptor, whereas in MHC class II ligands about 11% of the total number of salt bridges are formed between ligand amino acids. The majority of the hydrogen bonds are formed exclusively with the receptor, although in the small peptide group more than 12% of this type of interaction involve only ligand residues.

Table 7
Interactions of aminoacidic residues from peptidic ligands in the PDB structures.

Interaction type	Ligand groups	%RIL	%RIR	%RIB	%IWL
Salt bridges	Small	0	100	0	0
	MHC-I	0	100	0	0
	MHC-II	11.76	88.24	0	10
Hydrogen bonds	Small	12.45	84.34	3.21	12.46
	MHC-I	3.62	94.74	1.64	3.85
	MHC-II	2.6	97.4	0	2.06

%RIL: percent of residues interacting only with the rest of the ligand.
%RIR: percent of residues interacting only with the receptor.
%RIB: percent of residues interacting with both the receptor and the rest of the ligand.
%IWL: percent of the total number of residue interactions that involve other ligand residues.

Table 8 shows that contact is the most affected type of interaction when the original residue conformation is substituted for the closest rotamer from the different rotamer libraries. Nearly 50% of the residues change their number of contacts with the receptor, and there are more residues loosing contacts than gaining them. Again, the four Xiang libraries perform better in reproducing the original residue contacts. This is also true for salt bridges and hydrogen bonds, but for these

Table 8
Effect of substituting PDB side-chain conformations for their closest rotamers in the number of ligand residue–receptor interactions.

Ligand groups	Rotamer libraries	Salt bridges			Hydrogen bonds			Contacts		
		<0	0	>0	<0	0	>0	<0	0	>0
Small	Demaeyer	2.4	97.1	0.5	4.6	91.9	3.6	32.8	40.4	26.8
	DunbrackBBdep	1.9	97.6	0.5	4.8	91.1	4.1	30.9	44.1	25.1
	DunbrackBBind	1.9	97.6	0.5	4.7	92.0	3.3	31.7	42.2	26.1
	PondersRichards	2.2	96.9	0.9	4.4	91.8	3.8	32.4	40.6	27.0
	Richardson	2.1	97.6	0.3	4.7	91.2	4.1	32.6	40.6	26.8
	Tuffery	2.7	96.8	0.5	4.5	91.8	3.7	31.5	41.8	26.6
	Xiang135	0.8	98.9	0.3	3.6	94.1	2.2	25.3	51.8	22.8
	Xiang297	1.1	98.7	0.3	3.3	94.1	2.6	24.8	53.7	21.5
	Xiang533	1.1	98.6	0.3	3.5	94.2	2.4	24.2	53.8	22.0
	Xiang844	0.8	98.9	0.3	3.4	94.1	2.4	23.3	55.1	21.6
MHC-I	Demaeyer	8.9	90.6	0.5	4.0	91.1	4.9	40.7	34.0	25.3
	DunbrackBBdep	5.7	92.7	1.6	4.5	89.8	5.7	38.3	36.3	25.4
	DunbrackBBind	8.9	88.5	2.6	5.3	88.8	5.9	40.0	34.2	25.9
	PondersRichards	3.9	92.1	3.9	4.4	90.8	4.8	38.7	34.3	27.0
	Richardson	7.8	89.1	3.1	4.9	90.2	5.0	40.5	34.0	25.5
	Tuffery	9.4	88.5	2.1	4.4	90.9	4.7	42.0	33.8	24.3
	Xiang135	4.7	92.7	2.6	2.8	94.0	3.2	30.8	44.8	24.4
	Xiang297	3.6	92.7	3.6	2.8	93.8	3.5	28.8	47.2	24.0
	Xiang533	3.6	93.8	2.6	2.4	94.7	2.9	27.8	48.2	24.0
	Xiang844	2.1	95.8	2.1	2.7	94.6	2.7	27.5	48.1	24.4
MHC-II	Demaeyer	8.0	88.0	4.0	4.4	94.2	1.4	36.5	39.9	23.5
	DunbrackBBdep	6.7	88.0	5.3	4.8	92.8	2.4	34.8	37.5	27.6
	DunbrackBBind	4.0	92.0	4.0	4.8	92.5	2.7	36.2	37.5	26.3
	PondersRichards	7.7	88.5	3.8	3.3	93.4	3.3	36.5	39.3	24.2
	Richardson	6.7	89.3	4.0	4.4	91.5	4.1	38.6	35.8	25.6
	Tuffery	5.3	88.0	6.7	4.8	92.5	2.7	38.2	36.9	24.9
	Xiang135	5.3	90.7	4.0	3.4	93.9	2.7	26.3	49.8	23.9
	Xiang297	1.3	94.7	4.0	2.4	95.2	2.4	26.6	50.2	23.2
	Xiang533	0.0	94.7	5.3	2.4	94.9	2.7	26.3	52.6	21.2
	Xiang844	0.0	96.0	4.0	2.0	95.9	2.0	25.6	54.3	20.1

Percentage of ligand residues that loose (<0), keep (0) or gain (>0) interactions with the receptor upon substitution of the PDB side-chain conformation for the closest rotamer taken from each library.

interactions the differences are more subtle. It is worth noting how simpler and older rotamer libraries like the Ponders and Richards' [2] and Tuffery's [5] libraries perform as well as more recent and elaborated collections, such as the Dunbrack's backbone-dependent rotamer library [10], in reproducing hydrogen bonds and salt bridges. This is most likely a consequence of the higher tolerance of the geometric criteria employed to compute salt bridges and hydrogen bonds, as compared to the distance window used to calculate contacts (see Section 2). There is no tendency neither to increase nor to decrease the number of salt bridges and hydrogen bonds when substituting the original PDB residue conformations for the closest rotamers, for any of the studied rotamer libraries or ligands groups—changes occur in both directions.

The fact that almost all the analyzed interactions, for the three groups of peptidic ligands, are established mostly with the receptor supports the commonly used additive approximation for calculating the interaction energy between a peptidic ligand and its protein receptor. With this approach, the energy is calculated by adding the individual interaction energies between each of the peptide residues and the protein, taking no account of internal ligand contacts. This approach has been used, in particular, in T-cell epitope prediction methods for scoring the capability of different peptides to bind a MHC molecule [4,25].

3.4. The energy scoring problem

In this paper we have focused on the "rotamericity" of amino acid side chains in peptidic ligands, without tackling the modeling problem of how to select the appropriate rotamer from a given rotamer library. Many reports have been published on this subject (see Mendes et al. [37]). In summary, current side-chain prediction approaches may be grouped into energy-based [5,38–41] and knowledge-based methods [1,7,42,43], depending on the nature of the scoring function used to distinguish between correct and incorrect side-chain conformations.

Petrella et al. [44] have shown that the CHARMM energy function predicts well the conformation of a single side chain in the presence of all other side chains in their crystallographic positions. Nevertheless, it has been recently argued that unmodified Hamiltonians may not be suitable for side-chain prediction due to their inability to accurately account for solvation effects, electrostatics, and hydrogen bonding [45]. We performed our own calculations using the CHARMM function to score different sets of rotamers for each of the peptide side chains included in this study, in the context of its ligand–receptor complex. We found no correlation between the CHARMM energies and the rotamer vs. experimental RMSD's (data not shown).

4. Conclusions

In this paper we have tested the most classical and popular rotamer libraries, as well as a group of more recently developed libraries, for their ability of reproducing the side-chain conformations displayed by several hundreds of peptidic ligands extracted from the PDB. From the 10 examined rotamer collections, only the more detailed libraries developed by Xiang and Honig were able to correctly reproduce the experimental geometries for most of the analyzed amino acid residues, and therefore the atomic interactions between the peptidic ligands and their receptors. Surprisingly, all the libraries showed a lower performance in reproducing the side chains conformations from structures with very high resolution ($R \leq 1.25$ Å).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2008.10.002.

References

[1] M.J. Bower, F.E. Cohen, R.L. Dunbrack Jr., Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool, J. Mol. Biol. 267 (1997) 1268–1282.
[2] J.W. Ponder, F.M. Richards, Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes, J. Mol. Biol. 193 (1987) 775–791.
[3] S. Liang, N.V. Grishin, Side-chain modeling with an optimized scoring function, Protein Sci. 11 (2002) 322–331.
[4] R.L. Dunbrack Jr., Rotamer libraries in the 21st century, Curr. Opin. Struct. Biol. 12 (2002) 431–440.
[5] P. Tuffery, C. Etchebest, S. Hazout, R. Lavery, A new approach to the rapid determination of protein side chain conformations, J. Biomol. Struct. Dyn. 8 (1991) 1267–1289.
[6] P. Tuffery, C. Etchebest, S. Hazout, Prediction of protein side chain conformations: a study on the influence of backbone accuracy on conformation stability in the rotamer space, Protein Eng. 10 (1997) 361–372.
[7] R.L. Dunbrack Jr., M. Karplus, Backbone-dependent rotamer library for proteins. Application to side-chain prediction, J. Mol. Biol. 230 (1993) 543–574.
[8] H. Schrauber, F. Eisenhaber, P. Argos, Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins, J. Mol. Biol. 230 (1993) 592–612.
[9] M. De Maeyer, J. Desmet, I. Lasters, All in one: a highly detailed rotamer library improves both accuracy and speed in the modelling of sidechains by dead-end elimination, Fold. Des. 2 (1997) 53–66.
[10] R.L. Dunbrack Jr., F.E. Cohen, Bayesian statistical analysis of protein side-chain rotamer preferences, Protein Sci. 6 (1997) 1661–1681.
[11] S.C. Lovell, J.M. Word, J.S. Richardson, D.C. Richardson, The penultimate rotamer library, Proteins 40 (2000) 389–408.
[12] S.C. Lovell, J.M. Word, J.S. Richardson, D.C. Richardson, Asparagine and glutamine rotamers: B-factor cutoff and correction of amide flips yield distinct clustering, Proc. Natl. Acad. Sci. U.S.A. 96 (1999) 400–405.
[13] Z. Xiang, B. Honig, Extending the accuracy limits of prediction for side-chain conformations, J. Mol. Biol. 311 (2001) 421–430.
[14] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr. A 47 (Pt 2) (1991) 110–119.
[15] D.E. McRee, Practical Protein Crystallography, Academic Press, 1993.
[16] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, ProCheck—a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[17] R.W. Hooft, G. Vriend, C. Sander, E.E. Abola, Errors in protein structures, Nature 381 (1996) 272.
[18] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne, The Protein Data Bank, Nucleic Acids Res. 28 (2000) 235–242.
[19] Z. Xiang, Advances in homology protein structure modeling, Curr. Protein Pept. Sci. 7 (2006) 217–227.
[20] I. Lasters, M. De Maeyer, J. Desmet, Enhanced dead-end elimination in the search for the global minimum energy conformation of a collection of protein side chains, Protein Eng. 8 (1995) 815–822.
[21] J.R. Desjarlais, T.M. Handel, De novo design of the hydrophobic cores of proteins, Protein Sci. 4 (1995) 2006–2018.
[22] B.I. Dahiyat, S.L. Mayo, De novo protein design: fully automated sequence selection, Science 278 (1997) 82–87.
[23] D. Fry, H. Sun, Utilizing peptide structures as keys for unlocking challenging targets, Mini Rev. Med. Chem. 6 (2006) 979–987.
[24] M.N. Davies, C.E. Sansom, C. Beazley, D.S. Moss, A novel predictive technique for the MHC class II peptide-binding interaction, Mol. Med. 9 (2003) 220–225.
[25] Y. Altuvia, H. Margalit, A structure-based approach for prediction of MHC-binding peptides, Methods 34 (2004) 454–459.
[26] J.C. Tong, T.W. Tan, S. Ranganathan, Modeling the structure of bound peptide ligands to major histocompatibility complex, Protein Sci. 13 (2004) 2523–2532.
[27] H.M. Berman, T.N. Bhat, P.E. Bourne, Z. Feng, G. Gilliland, H. Weissig, J. Westbrook, The Protein Data Bank and the challenge of structural genomics, Nat. Struct. Biol. 7 (Suppl.) (2000) 957–959.
[28] L.A. Diago, P. Morell, L. Aguilera, E. Moreno, Setting up a large set of protein–ligand PDB complexes for the development and validation of knowledge-based docking algorithms, BMC Bioinformatics 8 (2007) 310.
[29] S.F. Altschul, T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, D.J. Lipman, Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, Nucleic Acids Res. 25 (1997) 3389–3402.
[30] J.M. Word, S.C. Lovell, J.S. Richardson, D.C. Richardson, Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation, J. Mol. Biol. 285 (1999) 1735–1747.
[31] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (33–38) (1996) 27–28.
[32] R Development Core Team, R: A Language and Environment for Statistical Computing, 2007 http://www.R-project.org.

[33] A. Bairoch, B. Boeckmann, S. Ferro, E. Gasteiger, Swiss-Prot: juggling between evolution and stability, Brief Bioinform. 5 (2004) 39–55.
[34] M.M. Schuler, M. Nastke, S. Stevanovikć, SYFPEITHI: database for searching and T-cell epitope prediction, Methods Mol. Biol. 409 (2007) 75–93.
[35] R.W. Peterson, P.L. Dutton, A.J. Wand, Improved side-chain prediction accuracy using an ab initio potential energy function and a very large rotamer library, Protein Sci. 13 (2004) 735–751.
[36] M.I. Zavodszky, L.A. Kuhn, Side-chain flexibility in protein–ligand binding: the minimal rotation hypothesis, Protein Sci. 14 (2005) 1104–1114.
[37] J. Mendes, H.A. Nagarajaram, C.M. Soares, T.L. Blundell, M.A. Carrondo, Incorporating knowledge-based biases into an energy-based side-chain modeling method: application to comparative modeling of protein structure, Biopolymers 59 (2001) 72–86.
[38] A.R. Leach, Ligand docking to proteins with discrete side-chain flexibility, J. Mol. Biol. 235 (1994) 345–356.
[39] M. Vasquez, An evaluation of discrete and continuum search techniques for conformational analysis of side-chains in proteins, Biopolymers 36 (1995) 53–70.
[40] P. Koehl, M. Delarue, Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy, J. Mol. Biol. 239 (1994) 249–275.
[41] H. Kono, J. Doi, A new method for sidechain conformation prediction using a Hopfield network and reproduced rotamers, J. Comp. Chem. 17 (1996) 1667–1683.
[42] M.J. Sutcliffe, F.R. Hayes, T.L. Blundell, Knowledge based modelling of homologous proteins. Part II. Rules for the conformations of substituted sidechains, Protein Eng. 1 (1987) 385–392.
[43] K. Ogata, H. Umeyama, The role played by environmental residues on sidechain torsional angles within homologous families of proteins: a new method of side-chain modelling, Proteins 31 (1998) 355–369.
[44] R.J. Petrella, T. Lazaridis, M. Karplus, Protein sidechain conformer prediction: a test of the energy function, Fold. Des. 3 (1998) 353–377.
[45] T. Jain, D.S. Cerutti, J.A. McCammon, Configurational-bias sampling technique for predicting side-chain conformations in proteins, Protein Sci. 15 (2006) 2029–2039.