ELSEVIER

Biophysical Chemistry 133 (2008) 36–44

Biophysical Chemistry

http://www.elsevier.com/locate/biophyschem

Molecular Dynamics as a pattern recognition tool: An automated process detects peptides that preserve the 3D arrangement of Trypsin's Active Site

Vasileios A. Tatsis [a], Athanassios Stavrakoudis [a,c], Ioannis N. Demetropoulos [b,*]

[a] *Department of Chemistry, Ioannina University, Ioannina, Greece*
[b] *Department of Information and Telecommunications Engineering, Section of Applied Informatics, University of West Macedonia, Kozani, Greece*
[c] *Department of Economics, Ioannina University, Ioannina, Greece*

Received 24 September 2007; received in revised form 24 November 2007; accepted 26 November 2007
Available online 8 December 2007

Abstract

Recently, the synthesis of a molecule has been reported that belongs to a Lysine based, branched cyclic peptide class. This work explores the ability of such molecules to preserve the 3D geometry of the Trypsin's Active Site (TAS) by applying an integrated framework of automated computer procedures. The following four factors a) D/L chirality, b) different amino acids at different positions of the molecular scaffold's cyclic part, c) the application of AMBER and CHARMM force fields and d) different implicit solvation models were evaluated against TAS similarity. It was found that a number of molecules exhibit satisfactory geometric affinity to the TAS during extended Molecular Dynamics runs. We estimated that more than 2000 molecules (belonging to this class) could retain good similarity to the TAS arrangement.

Keywords: Serine proteases; Cyclic branched peptides; Implicit solvation molecular dynamics; Pattern recognition; Computer aided molecular design; QSAR

1. Introduction

It is well known that proteins are distinguished by certain motifs such as i) sequences of amino acids that give rise to a three dimensional (3D) structure, for instance α-helix, β-sheet [1,2] or ii) special groups (like the active site of serine proteases, hemoglobin's binding site) which can be no sequential in the 1D amino acid sequence, nevertheless exhibit a distinct 3D arrangement. So a question is posed: "*Is it possible an active 3D template, observed in proteins, to be preserved in another molecular scaffold (supporting apparatus)*?"

We have identified a number of general procedures, reported in the literature, of establishing a molecular environment capable to support the 3D arrangement of the biological functional components that are the hallmark of the protein:

a) The mutation of a protein to add/change its functionality [3,4].
b) The synthesis of short oligopeptides (less than 40 residues) termed as mini-proteins [5–9] that reproduce, in part, the function of much larger proteins, by transferring the active sites to small and stable natural scaffolds.

The following examples are specific on mimicking proteases:

c) Design and synthesis of peptides with a simple linear sequence, comprising the catalytic triad of trypsin. The first attempts, [10–13] employing linear oligopeptides failed to show noticeable catalytic activity. More complex attempts with protease mimetics [14–16] in the 90s could not produce structural similarity or activity as mimetics. Hahn [17] presented the design and synthesis of a 73-residue peptide named "Chymohelizyme-1" (CHZ-1) that showed characteristics resembling those of trypsin. In this model, a bundle of short amphipathic α-helical peptides hold the serine protease catalytic triad in the necessary conformation to provide catalysis and also a hydrophobic pocket. CHZ-1 presented specific catalytic activity for chymotrypsin substrates.
d) Cyclic peptides have been constructed containing the active site and hoping that the restricted nature of the cyclic topology

* Corresponding author. Department of Chemistry, University of Ioannina, GR-45110, Greece. Tel.: +30 2651098449.
E-mail address: idimitr@cc.uoi.gr (I.N. Demetropoulos).

0301-4622/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2007.11.008

[18] to enact the correct 3D arrangement of the active components. Two 29-residue cyclic peptides [19] comprising the active and binding site of the natural enzymes chymotrypsin and tryspin have been proposed as good mimetics but very soon have been strongly questioned [20,21]. Taleman and his group [22] designed and synthesized a cyclic eight-residue peptide, equipped with the catalytic triad of chymotrypsin where Asp^{102} was replaced by Glu, for synthetic purposes; it exhibited a small catalytic enhancement wrt free Histidine. A cyclic decapeptide [23] [Ser–Tyr–Ser–Met–Glu–His–Phe–Arg–Trp–Gly] (the first 10 amino acids of adrenocorticotropic hormone, ACTH), carrying the functional moieties of chymotrypsin was evaluated as potential mimetic of chymotrypsin. Few of the conformations exhibited hydrogen bonding patterns; furthermore, a poor rate of hydrolysis of esters of L- and D-phenylalanine was reported.

e) Cyclic and branched peptides as templates of the proper 3D catalytic arrangement. A tripodal scaffold [24], with each peptide chain carrying one residue of the catalytic triad of the chymotrypsin, has been synthesized as a representative member of serine protease mimetics' library. However, the usefulness of this scaffold for the generation of multifunctional serine protease mimetic libraries remains to be proven. This design was based on an earlier proposition called Template Assembled Synthetic Peptides (TASP) [25,26]. In this approach, cyclic decapeptides, derived from the antibiotic gramicidin S containing four Lysines, were used as attachment sites. One of the earlier implementations of the TASP approach was as a template for the antigen binding site of the phosphorylcholine specific antibody McPC603 [27,28].

f) The use of non peptides as molecular scaffolds for enacting proper 3D topology of the active site [29,30].

The present work is a variant of the **e** case. The synthesis of such a class of molecules has been reported by Stavrakoudis et al. [31,32], who described the design, synthesis and catalytic activity of a Lysine based Branched Cyclic peptide equipped with the Ser–Hisp–Asp or with one character abbreviation (S–H–D) catalytic triad (LysBCSHD) as a serine protease mimetic (Fig. 1).

Recently we developed an automated computer procedure called Lysine based Trypsin Active Site (LysTAS) [33] (a TINKER [34] configuration tool) that builds peptidomimetic molecules satisfying the particular design (see Fig. 1) and grades them against the active site's topology. The LysBCSHD class numbers $\approx 1 \cdot 10^5$ molecules. This study carries on previous works [31,32] using LysTAS to examine a LysBCSHD's subset of 192 peptides as potential candidates to preserve Trypsin's Active Site (TAS) 3D geometry. Therefore, we had the opportunity to examine the parameter space of a) different amino acids at the cycle, b) D/L chirality c) selection of the FF (FF) and d) selection of the Implicit Solvation Model (ISM). In this way, we could assess optimal conditions for directing the three residues that constitute the active site of serine proteases in close contact, so that the peptide could potentially retain TAS arrangement. Finally, we subjected the most successful compounds to extended Molecular Dynamics (MD) runs. The molecules that could preserve geometric resemblance to TAS, no matter the FF and/or the ISM used (crosschecked), will have high probability to retain this similarity in vitro.

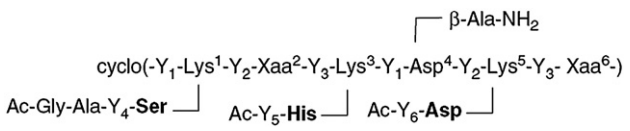

Fig. 1. Peptides modelled as serine protease mimetics. Active site residues Ser, His and Asp are in bold. Y_i denotes the D/L chirality. Xaa^2 and Xaa^6 accommodate the amino acids selected by the user.

2. Methods

This section is organised as follows: a) the continuum solvation models used in this study and b) the computational procedure (Modus Operandi). This last one comprises of five parts: (i) the automatic construction of the LysBCSHD molecules obeying the molecular scaffold as in Fig. 1. (ii) Template forcing. (iii) MD simulations with ISMs. (iv) RMSD comparison and hydrogen bond distances of the catalytic residues. (v) Correlation between the β-turn motifs of peptide's cyclic part and the proper Ser-His-Asp 3D arrangement.

2.1. Continuum solvation models

The solvent is known [35,36] to affect the conformational preferences of peptides. The solvation of a solute can be described as discrete solvent molecules or as a "continuum" medium. However the implicit solvation models [37] can survey a much larger number of molecules' topology due to small computational cost. In this work we employ the ISMs SASA (Solvent Accessible Surface Area) [38–43], ASP (Atomic Solvation Parameter) [44,45], GBSA (Generalized Born Surface Area) [46–48] and HCT (Hawking–Cramer–Truhlar) [49,50].

2.2. Computational procedure (modus operandi)

This work explores the ability of LysBCSHD class of compounds to preserve the 3D geometry of TAS by applying an integrated framework of automated procedures as it follows.

2.3. Automated building

We made a series of amino acids replacements as follows (see Appendix A Table S1). The chirality of residues into cycle positions i and i+3 were substituted simultaneously each time. Therefore (Lys^1 Asp^4), (Xaa^2 Lys^5), (Lys^3 Xaa^6) made three pairs, resulting to 2^3 (=8) new starting conformations. Xaa^2 and Xaa^6 were simultaneously substituted with the same residue. This operation was necessary in order to retain symmetry in the molecule. The chiralities of Asp, His and Ser of the catalytic triad were altered, independently, resulting in $2^3=8$ new

combinations. Three amino acid substitutions were tested in place of Xaa^2 and Xaa^6: Gly for its flexibility, Ala for its medium sized neutral side chain and Pro for its tendency to form turns. The Gly–Ala–Ser sequence was introduced to mimic the oxyanion hole found in serine proteases [51]. Since we have three amino acids in the place of Xaa^2 and Xaa^6 it turns out that the total number of molecules studied in this work is $3\times8\times8=192$. The process is part of the LysTAS configuration tool. Additional information of this procedure is found at the manual of LysTAS (http://users.uoi.gr/btatsis).

2.4. Template forcing

It was used for guiding each peptide (the catalytic triad of the molecule) to adopt a specific conformation, similar to the trypsin's active site, referred to as the template. This is accomplished by adding a penalty function term to the potential energy of the system as defined by the FF and using the implicit solvation method which will be used later in unconstrained MD. The force constant (kcal/Å^2) of the penalty function was gradually decreased, in order to avoid very "stressed" conformations. At the final stage of the template forcing, the ISM was used so as to achieve a smooth transition from the gas phase to the "solvent's" environment. The molecular segment used as template was the side-chain heavy atoms of the catalytic triad residues (Asp102, His57, Ser195). The template was extracted from the crystal structure of trypsin (entry code 2ptn) as a typical representative of the serine protease family.

2.5. MD simulations with ISMs

Each of the preceding molecules was subjected to MD simulations for 200 ps (following a 20 ps equilibration period), with a time step of 1 fs. We employed Beeman [52] algorithm to solve the equations of motions and the Berdensen thermostat [53] for keeping the temperature steady, at 300 K. The simulations were performed with the FFs AMBER99 [54], CHARMM27 [55] and the implicit solvation schemes SASA, GBSA incorporated in TINKER. Structures were stored every 2 ps, so that 111 conformations were stored and the last 100 were used for analysis. Molecules that had fulfilled TAS similarity criteria were subjected to extended MD runs employing additional implicit solvation schemes such as HCT and ASP.

2.6. RMSD comparison and hydrogen bond distances of catalytic residues

RMS distances between the heavy atoms of the side chains of the catalytic triad residues (Asp102, His57, Ser195) of the enzyme and the corresponding side chain heavy atoms of the peptide model LysBCSHD were measured in order to assign a similarity score. This comparison refers to 12 atoms belonging to three residues in the active site: Ser, His and Asp. We believe that it is necessary to include the C^{β} atoms in such comparisons in order to take into account the directionality of the chemical groups belonging to the side chains. All RMSD measurements mentioned throughout this paper refer to the above definition. This means that the His side chain contributes up to half of the total value (6 atoms), while Asp (4 atoms) and Ser (2 atoms) side chains have minor contributions.

$$\mathrm{RMSD}_t = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i^m - x_i^t)^2+(y_i^m - y_i^t)^2+(z_i^m - z_i^t)^2} \quad (1)$$

The RMS distance between the atoms of the peptide (Fig. 1) and the corresponding atoms of the enzyme (2 ptn) is given by Eq. (1), where x^m, y^m, z^m are the Cartesian coordinates of the heavy atoms of the amino acids which comprise the active site of the natural enzyme and x^t, y^t, z^t are the Cartesian coordinates of the corresponding heavy atoms of the peptide model LysBCSHD.

The criterion (Eq. (2)) for a frame i of the trajectory to count it as one with a favorable topology is [56]:

$$2\text{Å} \geq \mathrm{RMSD}_i = \mathrm{RMSD}\left[\mathrm{frame}_i^{\text{implicit solvation}} - \mathit{TEMPLATE}\right] \quad (2)$$

Eq. (3) provides the mean value of a MD trajectory's RMSD, where N the total number of frames and RMSD_i is the RMSD of each frame i.

$$\overline{\mathrm{RMSD}} = \frac{1}{N}\sum_{i=1}^{N}\mathrm{RMSD}_i \quad (3)$$

The mean difference (μ) between two sets of 64 trajectories, resulted from the AMBER and CHARMM FFs respectively, is computed from Eq. (4).

$$\mu_{\mathrm{AMBER,CHARMM}} = \frac{1}{N}\sum_{j=1}^{N=64}\left|\overline{\mathrm{RMSD}}_{\mathrm{j,AMBER}} - \overline{\mathrm{RMSD}}_{\mathrm{j,CHARMM}}\right| \quad (4)$$

$\overline{\mathrm{RMSD}}j$ stands for the mean value of the RMSD of a molecule's MD j trajectory, calculated from Eq. (3), with the same chemical composition but a different chirality pattern (see Appendix A Table S1).

The LysTAS application measures three Cartesian distances between the atoms of the residues that are responsible for the formation of the hydrogen bond network in the active centre of the enzyme as an additional criterion of similarity (d1 = His:HD1 − Asp:OD1, d2 = His:HD1 − Asp:OD2 and d3 = Ser:HG − His:NE2). The criterion for the hydrogen bonding's distance is less than 3 Å [57]. At this stage we don't take into account the directionality of the hydrogen bond.

The application outputs files that contain the trajectory's frames percentage RMSD (for the heavy atoms of the active site) and the percentage of the frames, which have favourable hydrogen bonds of the catalytic triad (d1, d2, d3). Furthermore, we measure i) the RMSD between the trajectory's first frame and the rest frames, ii) the RMSD between every frame and its precedent, as an extra criterion of MD trajectory's stability. A

comprehensive guide of the operations involved, for the best peptide selection is given in Fig. 2.

2.7. Correlation between the β-turn motifs of peptide's cyclic part and the proper Ser-His-Asp 3D arrangement

In this survey we used the PROMOTIF [58] program suite. PROMOTIF analyzes a protein coordinate file and provides details about the structural motifs in the protein.

The representative values of φ and ψ for each of the 9 β-turn types are given in Table S2 (see supplementary material). To see the subtle features for each of the 9 types, as well as the difference and relationship among them, it is useful to observe the structure drawings which are provided at Fig. S1 (in supplementary material).

3. Results and discussion

The discussion is structured into three separate parts. In part 1 the MD runs of LysBCSHD molecules are classified using as criterion their RMSD score. The second part presents in detail the extended MD runs of the molecules that their RMSD score (part 1) indicate acceptable similarity to the TAS. The third part analyses the successful molecules of part two. Namely, we try to correlate the β-turn motifs of peptide's cyclic part to the trajectories that sustain high percentage of 3D geometric

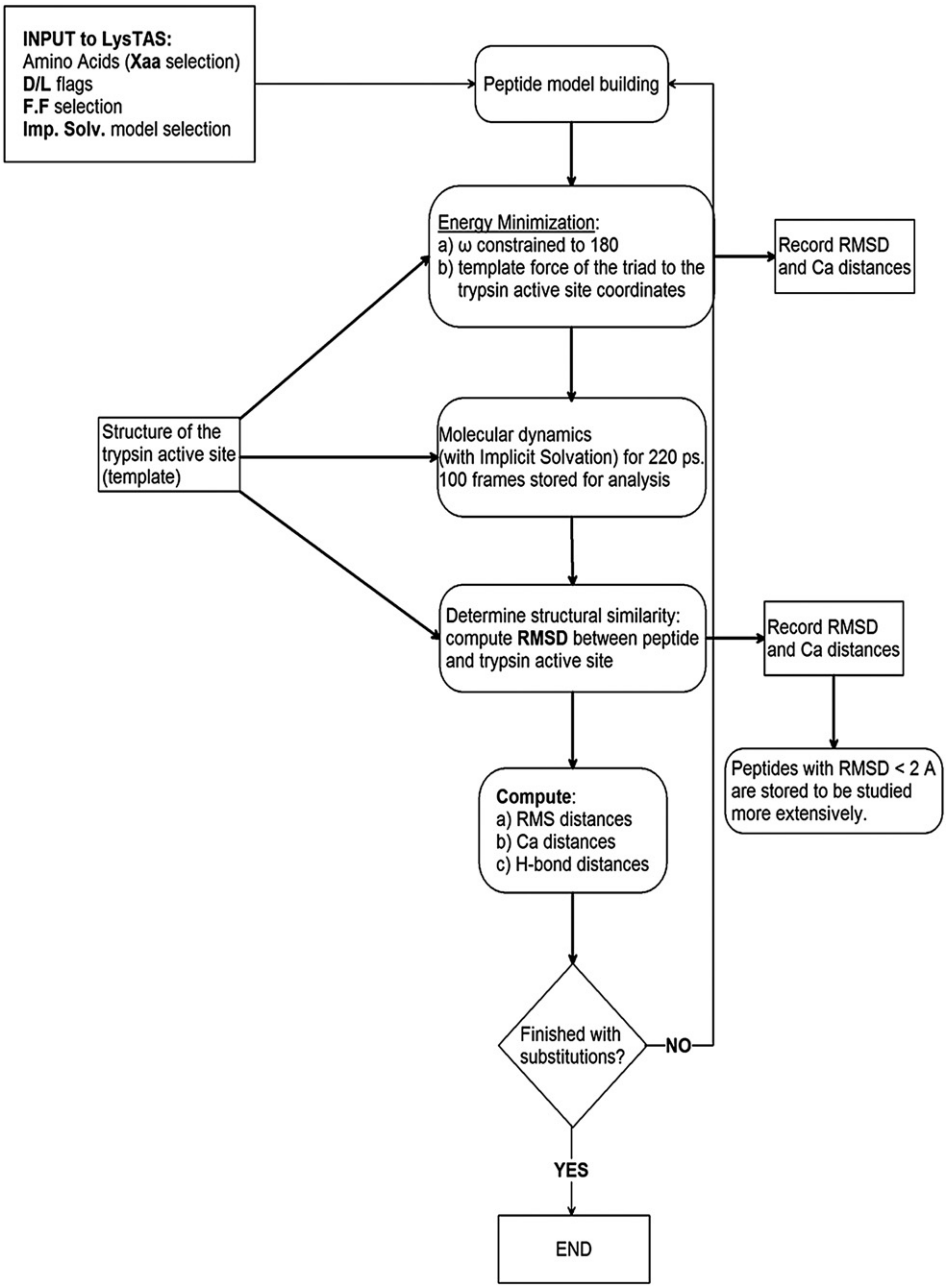

Fig. 2. Flowchart of the computational procedures, executed by LysTAS, for the best peptide selection via Implicit Solvation MD simulations.

Table 1
The simulated 192 peptides grouped according to their mean RMSD value resulted from the simulations

		$Xaa^2(=Xaa^6)$= Gly			$Xaa^2(=Xaa^6)$= Ala			$Xaa^2(=Xaa^6)$= Pro		
		A	B	C	A	B	C	A	B	C
AMBER	SASA	8	46	10	7	54	3	1	47	16
	GBSA	1	32	31	0	29	35	0	17	47
CHARMM	SASA	1	53	10	1	45	18	1	46	17
	GBSA	0	11	53	0	10	54	0	6	58

For example, in the simulations with the FF AMBER and the ISM SASA, 8 peptides resulted with $\overline{RMSD} < 2.0$ Å.
A: $\overline{RMSD} < 2.0$ Å. B: 2.0 Å $\leq \overline{RMSD} <$ 5.0 Å. C: $\overline{RMSD} \geq 5.0$ Å.

similarity to trypsin's active triad wrt the peptide cycle's patterns.

3.1. Part 1

We performed a screening operation, using the ISMs SASA, GBSA and the FFs AMBER and CHARMM, to 192 different peptides, derived from the molecular scaffold depicted in Fig. 1. The ≅800 MD runs are grouped according to their $\overline{RMSD}$ (Eq. (3)) values (calculated from a trajectory of 220 ps = 100 frames), compared to the corresponding atoms of trypsin. These findings are tabulated at Table 1. We distinguish three regions of the $\overline{RMSD}$ values i) $\overline{RMSD} < 2.0$ Å (acceptable similarity to the TAS [56]), ii) 2.0 Å $\leq \overline{RMSD} <$ 5.0 Å (some sort of similarity) iii) $\overline{RMSD} \geq 5.0$ Å (dissimilarity).

1) Approximately 10% of the 192 peptides exhibit $\overline{RMSD}$ under 2.0 Å. The majority of the peptides "fall into" the second category, 2.0 Å $\leq \overline{RMSD} <$ 5.0 Å. This shows that in general this molecular scaffold favors the propensity for active site's similarity. In addition, since the total number of molecules of the class LysBCSHD is ~2. 10^5, one might expect ~20,000 molecules to show $\overline{RMSD}$ under 2.0 Å.
2) It appears from the relevant tabulations AMBER/SASA and CHARMM/SASA (Table 1), that the two FFs record considerably different percentage for the good 3D similarity molecules. So, we measured the difference between the two simulation sets of Gly employing different FFs (Eq. (4)). The 64 trajectories of Gly/SASA/AMBER versus the 64 trajectories of Gly/GBSA/AMBER were quite similar showing an absolute difference, μ equal to 1.02 Å. Hence a number of the trajectory set Gly/CHARMM/SASA are slightly over the cutoff 2 A; so the discrepancy at Table 1 is an artifact of the trihotomous nature of A,B,C disjoint sets. Therefore, it appears that AMBER/SASA and CHARMM/SASA are producing comparable $\overline{RMSD}$ values for this class of compounds; this is reasonable, considering the fact that the continuous solvation energy for a given geometry of a molecule is the same for both FFs.
3) Peptides with Pro in the cyclic part did not result "good" RMSD values. Most of the MD trajectories of the peptides with Pro in position Xaa^2 and Xaa^6 "fall into" the second and third classes of $\overline{RMSD}$ values.
4) The results acquired from the "knowledge based" SASA ISM did not tie in to a great extent with those obtained from the 'physics based' GBSA. This claim is in agreement with the recent work of Jaramillo and Wodak [59], where the behavior of the "knowledge based" model EAS (Empirical Atomic Solvation) diverged from the other 'physics-based' models EEF1 (Effective Energy Function 1), GBMV (Generalized Born Molecular Volume), ACE (Analytical Continuum Electrostatics) and FDPB (Finite Difference Poisson Boltzmann) both in ranking the solvation energies of different amino acids in a systematic test and in protein

Table 2
Data gathered from extensive runs with Ala in position $Xaa^2(=Xaa^6)$ (Fig. 1) and various continuum solvation models, FFs

Ala [a]	FF [b]	ISM [c]	Time (ns) [d]	RMSD<2 [e]	$\overline{RMSD}$ [f]	d1<3 [g]	d2<3 [h]	d3<3 [i]	Ser His [k]	His Asp [l]	Both [m]
14	AMBER	SASA	1.0	63	2.50	73	35	–	17	80	16
14	CHARMM	SASA	1.0	50	2.15	–	–	–	51	–	–
14	CHARMM	ASP	1.0	4	3.59	–	–	–	79	5	4
54	AMBER	SASA	1.0	96	1.60	98	–	–	12	67	9
54	CHARMM	SASA	1.0	8	3.38	–	–	23	-	15	–
54	AMBER	ASP	1.0	87	1.52	90	93	–	21	88	20
54	CHARMM	ASP	1.0*	97	1.27	31	94	11	63	35	32

* An additional run 5 ns verified the trajectory's consistency.
[a] The peptide number (see in Appendix Table 1).
[b] The FF used in the simulation.
[c] Continuum solvation model employed in the simulation.
[d] Duration of the run (ns).
[e] Percentage (%) of the frames exhibiting RMSD under 2.0 Å.
[f] Mean Value of RMSD (Å).
[g] Percentage (%) of the frames, in which the distance His:HD1 — Asp:OD1 is below 3.0 Å.
[h] Percentage (%) of the frames, in which the distance His:HD1 — Asp:OD2 is below 3.0 Å.
[i] Percentage (%) of the frames, in which the distance Ser:HG — His:NE2 is below 3.0 Å.
[k] Percentage (%) of the frames, in which the distance of the atoms Ser:Cα–His:Cα is in between 8.3±1 Å (value extracted from PBD X-ray structures).
[l] Percentage (%) of the frames, in which the distance of the atoms $His:C^a$–$Asp:C^a$ is between the following boundaries 6.3±1 Å (value extracted from PBD X-ray structures).
[m] Percentage (%) of the frames, in which both distances ($Ser:C^a$–$His:C^a$ and $His:C^a$–$Asp:C^a$) fluctuate between 8.3±1 Å and 6.3±1 Å, respectively.

design calculations. However the GBSA/FF performance produced trajectories that had greater discrepancies for the RMSD measured.

5) The inspection of the set A of molecules could not point to a favorable chirality pattern. Similarly, by examining the C set we could not detect a D/L motif that was common to all three Gly, Pro, Ala sets of compounds. Hence chirality could not serve as discriminator for the best or worst cases.

3.2. Part 2

At this stage of the work, the peptides, which displayed good geometric resemblance criteria at part 1, were studied extensively (MD runs, varying from 400 ps to 1 ns). These peptides are: Gly1, Gly3, Gly12, Gly14, Gly27, Gly64 and Ala14, Ala54. The results, of these extended runs, are summarized in Table S3 (in supplementary material) and Table 2. These tables now contain alongside the trajectory's RMSD further specific indicators: a) d1, d2, d3 of the overall geometric resemblance to the TAS and b) the distances between the C^{α} atoms of the catalytic triad amino acids. The main findings are:

1a) 6 of the 19 trajectories, with Gly in Xaa positions, resulted "good" RMSD (Table S3 in supplementary material). 12 trajectories had RMSD values between 2.0 Å and 5.0 Å. While, only one trajectory had RMSD above 5.0 Å.

1b) Distances d1, d2, d3 (a special indicator for the hydrogen bond network of the active site residues) in the simulation sets Gly1/AMBER/ASP, Gly3/AMBER/SASA and Gly12/AMBER/SASA are preserved at an acceptable percentage.

1c) The third indicator (distances of Ser:C^a–His:C^a and His:C^a–Asp:C^a) supports the assertion that the active site residues are kept close enough for a considerable part of the trajectories Gly1/AMBER/ASP and Gly12/AMBER/SASA.

2a) 3 of the trajectories, with Ala in position Xaa^2 and Xaa^6, had RMSD below 2.0 Å, while for the rest MD runs, the RMSD varied between 2.0 Å and 5.0 Å (Table 2).

2b) Ala54/CHARMM/ASP trajectory scores best in the extra two geometric indicators. According to the appropriate distances (d1, d2, d3), the hydrogen bond network, between the corresponding atoms of the trypsin's active center, seems to be formed for a considerable time. Also, C^a distances are fluctuating around the values extracted from the X-ray PDB structures.

Fig. 3 depicts the variation of the potential energy and RMSD as a function of the consecutive frames of one MD run with the implicit solvation ASP and the FF CHARMM. The Ala54/CHARMM/ASP run records very low RMSD values (each frame with respect to the template) (a RMSD value smaller than 2 Å is generally regarded as acceptable [56]). The MD simulation with the ASP model was applied to Ala54 molecule for 1 ns and generated a trajectory with a nearly all (97%) frames possessing the right positioning of the three residues as to justify the formation of catalytic triad.

Fig. S2 (in supplementary material) depicts the fluctuation of the Cartesian distances of His:HD1–Asp:OD1, His:HD1–Asp:OD2 and Ser:HG–His:NE2 against the consecutive frames for the peptide Ala54. The Ala54/CHARMM/ASP run yielded many frames with these distances close to the appropriate value. We found that the distances, between the atoms of the catalytic triad residues that form the hydrogen bond network, are simultaneously within the recorded range for approximately 10%–20% of the frames. These findings provide additional support (other than the RMSD) to the hypothesis that the cyclic part of the peptide has the tendency to bring the branches in proximity, at least for a considerable amount of time. It is clear that the His

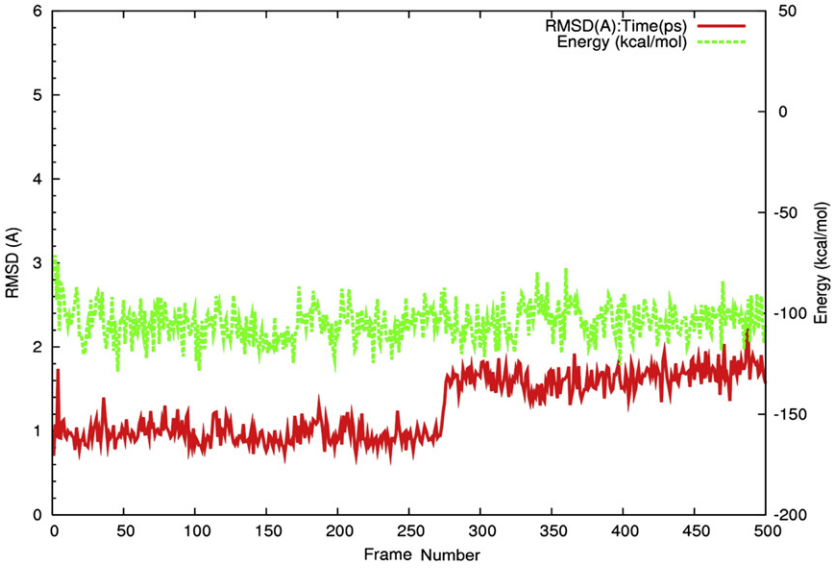

Fig. 3. Fluctuation of RMSD (Å) and potential energy (kcal/mol) in MD run of Ala54 with the ISM ASP and the FF CHARMM, for 1 ns.

and Asp residues exhibited a tendency to interact. However, this tendency is not evident for the Ser and His dyad after the half time of the simulation.

Finally, since the extended MD runs indicate that Gly1, Gly3, Gly12, Gly54 and Ala54 (out of ~200) are retaining $\overline{\text{RMSD}}$ values under 2.0 Å we might well be expecting ~2000 (out of the total 250 thousand) molecules of LysBCSHD class to adhere to TAS similarity.

3.3. Part 3

Furthermore, an attempt was made to reveal a possible correlation between the β-turn motifs and the proper arrangement of the catalytic triad residues. In this study PROMOTIF was used to analyze the MD trajectories of five different branched cyclic hexapeptides (1 Ala+4 Gly) in terms of β-turns found in the cyclic part of the peptide. To the best of our knowledge, it is the first time that PROMOTIF has been applied to an MD trajectory analysis of cyclic and branched peptides. The common characteristic of these five peptides is that each of their trajectories has $\overline{\text{RMSD}}$ values less than 2.0 Å.

The appearance of β-turns in the cyclic part of the peptide is summarized in Table 3. The output files from PROMOTIF are used as input in this analysis. The table cell entrances are based on the following convention: if a β-turn, independently from its type, is found in more than 80% of the resultant trajectory's frames the notation 1 is used (TRUE), else notation 0 is employed (FALSE). Moreover, we looked for cases of two or three β-turns occurring simultaneously in a particular frame of the trajectory. The previous mentioned notations are also used in this approach. For over 50% of the total frames the notation is placed inside parentheses.

From Table 3, we notice that:

1. About 80% of all trajectories' frames show the existence of two β-turns in each frame's cyclic part. However, the appearance of two simultaneous β-turns in cyclic hexapeptides, is already found using NMR spectroscopy [60]. For molecules containing Gly, these two β-turns are formed between residues Lys^1(i)–Gly^2(i+1)–Lys^3(i+2)–Asp^4(i+3) and Asp^4(i)–Lys^5(i+1)–Gly^6(i+2)–Lys^1(i+3) retaining a symmetry to the cyclic part of the molecule. In very few of the frames the opposite residues are hydrogen bonded, thus creating a β-sheet like structure. Simultaneously three β-turns occur very rare. The molecule Ala54 exhibits two β-turns between residues Ala^2(i)–Lys^3(i+1)–Asp^4(i+2)–Lys^5(i+3) and Lys^5(i)–Ala^6(i+1)–Lys^1(i+2)–Xaa^2(i+3).
2. Four β-turns are formed in the cyclic peptides' part for more than 50% of the frames (no ββββ pattern for Gly64). The trajectories' frames of Ala54/AMBER/SASA-Gly12/AMBER/SASA show the four β-turns at the same positions on the cycle while Gly1/AMBER/ASP-Gly3/AMBER/SASA frames appear to have the pattern ββββ but at different sites of the cycle in comparison to the previous 2 trajectories. At the end of this discussion we have to report that for the initial building design of the cyclic part of the peptide we chose Xaa^2–Lys^3–Asp^4–Lys^5 to adopt a βI turn. This type of β-turn was not cited after the MD, while the dominant type of β-turn is βIV (Fig. S1 in supplementary material) (98%).

4. Conclusions

The peptides termed LysBCSHD belong to a new family of compounds for which a synthetic route has been described [32]. This work presents an extensive computer-aided investigation of these branched-cyclic peptides' class of molecules for directing Ser, His, Asp to the 3D arrangement of the TAS using LysTAS software. The dynamic behaviour of the flexible biomolecules is an ongoing research project [61,62]. The following tasks are executed by LysTAS: (i) the automatic construction of the LysBCSHD molecules obeying the molecular scaffold (ii) incremental template forcing (iii) MD simulations with ISMs (iv) RMSD comparison and hydrogen bond distances between the three catalytic residues. This study assesses four factors in relation to the geometric similarity of a LysBCSHD class' compound to TAS

(a) the residue composition of the cyclic part, ii) the chirality pattern of 9 residues of the peptide molecule, iii) the effect of using a different FF for the similarity evaluation and iv) different schemes for the ISM.

Table 3
β-turn patterns across the cyclic part of the peptide model LysBCSHD

Pep.[a]	ISM[b]	FF[c]	Time (ns)[d]	Residue number (R_{i+1} and R_{i+2})							
				1–2[e]	2–3	3–4	4–5	5–6	6–1	ββ[f]	βββ[g]
Ala54	SASA	AMBER	1.0	0	0 (1)	1	0	0 (1)	1	1	0 (1)
Gly1	ASP	AMBER	0.6	0 (1)	1	0	0 (1)	1	0	1	0 (1)
Gly3	SASA	AMBER	0.6	0 (1)	1	0	0 (1)	1	0	1	0 (1)
Gly12	SASA	AMBER	0.4	0	1	0 (1)	0	1	0 (1)	1	0 (1)
Gly64	ASP	AMBER	0.6	0	1	0	0	1	0	1	0

Over 50% of the total frames is labeled as (1).
[a]The peptide model (see in Appendix Table 1).
[b, c]ISM and the FF used during the simulation.
[d]Simulation time (ns).
[e]The R_{i+1} and R_{i+2} positions of the β-turn. For example, when a β-turn is formed between the amino acids Gly^6–Lys^1–Gly^2–Lys^3, the notation 1–2 is used. The same applies and for the rest of the amino acids.
[f]For the 80% of the total frames, two β-turns are met in the same frame of the resultant trajectory (in any position).
[g]For the 80% of the total frames, three β-turns are met in the same frame of the resultant trajectory (in any position).

The overall assessment of the potential of this molecular scaffold to sustain the 3D similarity to TAS is as follows:

a) Over 70% of the LysBCSHD class of molecules indicates some tendency to approach the TAS pattern.
b) There is no chirality bias towards a best or worst D/L pattern wrt TAS similarity.
c) Both FFs used, produced comparable results.
d) The empirical solvation models are offering consistent behavior while the GBSA not; probably higher level of "physics" is needed to produce uniform answers [63].
e) The model compounds Ala54 and Gly1 satisfy the correct placement, according to all three estimators, of the catalytic triad for a large percentage of the simulation time, examined with both FFs and the ISMs.
f) Since this class has approximately $2 \cdot 10^5$ molecules, we estimate approximately 2000 compounds compliant to TAS geometry.

Furthermore a correlation was made for the patterns of the cyclic part of molecules with $\overline{\text{RMSD}} < 2$ Å.

1) They show a propensity to form mainly two β-turns. So β-turns are spontaneously formed without Proline's help.
2) Using the criterion of 50% and above there are at any time 4 β-turns.
3) Using the criterion of 80% and above there are 2 β-turns between the residues Lys^1(i)–Gly^2(i+1)-Lys^3(i+2)-Asp^4(i+3) and Asp^4(i)-Lys^5(i+1)–Gly^6(i+2)-Lys^1(i+3) of the cyclic peptide, so there is symmetry.

This work presents a contribution towards a systematic classification of the dynamic behavior of Lysine based branched cyclic peptides. These fascinating compounds are by design oriented to acquire a favorable to TAS 3D geometry; this study measures this proximity in silico. The computational operation provided by LysTAS software is flexible and expandable. This work demonstrates that an appropriate automated procedure is capable of "mining" for the requested molecular patterns within raw data, namely thousands of MD trajectories. Furthermore, other molecular motifs (beyond LysBCSHD), will be investigated using similar procedures.

Acknowledgments

The authors gratefully acknowledge the computer time provided by: (a) the Computational Materials Science Laboratory at the Dept. of Materials Science and Engineering, (b) the Computer Center and (c) the Research Center for Scientific Simulations (RCSS) of Ioannina University, Greece.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bpc.2007.11.008.

References

[1] L. Pauling, R.B. Corey, H.R. Branson, The structure of proteins: two hydrogen bonded helical conformations of the polypeptide chain, Proc. Natl. Acad. Sci. U. S. A. 37 (1951) 205–211.
[2] D. Eisenberg, The discovery of the-helix and-sheet, the principal structural features of proteins, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 11207–11210.
[3] E. Quemeneur, M. Moutiez, J.B. Charbonnier, A. Menez, Engineering cyclophilin into a proline-specific endopeptidase, Nature 391 (1998) 301–304.
[4] P. Iengar, C. Ramakrishnan, Knowledge-based modeling of the serine protease triad into non-protease, Protein Eng. 8 (1999) 649–655.
[5] B. Imperiali, J.J. Ottesen, Uniquely folded mini-protein motifs, J. Pept. Res. 54 (1999) 177–184.
[6] E. Drakopoulou, S. Zinn-Justin, M. Guenneugues, B. Gilquin, A. Menez, C. Vita, Changing the structural content of a functional β-hairpin, J. Mol. Biol 271 (1996) 11979–11987.
[7] S. Zinn-Justin, M. Guenneugues, E. Drakopoulou, B. Gilquin, C. Vita, A. Menez, Transfer of a beta-hairpin from the functional site of snake curaremimetic toxins to the alpha/beta scaffold of scorpion toxins: three-dimensional solution structure of the chimeric protein, Biochemistry 35 (1996) 8535–8543.
[8] C. Vita, J. Vizzanova, E. Drakopoulou, S. Zinn-Austin, B. Gilquin, A. Menez, Novel miniproteins engineered by the transfer of active sites to small natural scaffolds, Biopolymers (Pept. Sci.) 47 (1998) 93–100.
[9] L. Martin, P. Barthe, O. Combes, C. Roumestand, C. Vita, Engineering novel bioactive mini-proteins on natural scaffolds, Tetrahedron 56 (2000) 9451–9460.
[10] P. Cruishshank, J.C. Sheehan, Synthetic peptide models of enzyme active sites. II. L-Threonyl-L-alanyl-L-seryl-Lhistidyl-L-aspartic, acid an active esterase model, J. Am. Chem. Soc. 86 (1963) 2070–2071.
[11] J.C. Sheehan, G.B. Bennett, J.A. Schneider, Synthetic peptide models of enzyme active sites. III. Stereoselective esterase models, J. Am. Chem. Soc. 88 (1966) 3455–3456.
[12] A.M. Vandersteen, K.D. Janda, A Re-examination of two linear pentapeptides claimed to be serine protease mimics, J. Am. Chem. Soc. 118 (1996) 8787–8790.
[13] L. Photaki, M. Sakarellou-Daitsiotou, Synthesis and catalytic properties of peptides containing amino acids involved in the active centres of hydrolytic enzymes, J. Chem. Soc. Perkin Trans. I (1976) 589–591.
[14] T. Vorherr, K.-H. Altmann, M. Mutter, Single-center model for the active site of alpha-chymotrypsin, Helv. Chim. Acta 69 (1986) 410–414.
[15] J.M. Clough, R.V.H. Jones, H. McCann, D.J. Moriss, W. Martin, Synthesis and hydrolysis studies of a peptide containing the reactive triad of serine proteases with an associated linker to a dye on a solid phase support, Org. Biomol. Chem. 1 (2003) 1486–1497.
[16] M.P. Fitzsimons, J.K. Barton, Design of a Synthetic Nuclease: DNA Hydrolysis by a Zinc-Binding Peptide Tethered to a Rhodium Intercalator, J. Am. Chem. Soc. 119 (1997) 3379–3380.
[17] K.W. Hahn, W.A. Klis, J.M. Stewart, Design and synthesis of a peptide having chymotrypsin-like esterase activity, Science 248 (1990) 1544–1547.
[18] T.M. Chao, J.J. Perez, G.H. Loew, Characterization of the bioactive form of linear peptide antagonists at the δ-opioid receptor, Biopolymers 38 (1996) 759–768.
[19] M.Z. Attassi, T. Manshouri, Design of peptide enzymes (pepzymes): surface-simulation synthetic peptides that mimic the chymotrypsin and trypsin active sites exhibit the activity and specificity of the respective enzyme, Proc. Natl. Acad. Sci. U. S. A. 90 (1993) 8282–8286.
[20] T.J. Marrone, J.A. McCammon, Pepzyme dynamics and conformation: A Molecular Dynamics Study in Water, J. Am. Chem. Soc. 116 (1994) 6987–6988.
[21] D.R. Corey, M.A. Phillips, Cyclic peptides as proteases: a reevaluation, Proc. Natl. Acad. Sci. U. S. A. 91 (1994) 4106–4109.
[22] B. Walse, M. Ullner, C. Lindbladh, L. Bulow, T. Drakenberg, O. Taleman, Structure of a cyclic peptide with a catalytic triad, determined by computer simulation and NMR spectroscopy, J. Comput.-Aided Mol. Des. 10 (1996) 11–22.

[23] R.A. Buono, N. Kucharczyk, M. Neuenschwander, J. Kemmink, L.Y. Hwang, J.L. Fauchere, C.A. Venanzi, Synthesis and conformational analysis by ^{1}H NMR and restrained molecular dynamics simulations of the cyclic decapeptide [Ser–Tyr–Ser–Met–Glu–His–Phe–Arg–Trp–Gly], J. Comput.-Aided Mol. Des. 10 (1996) 213–232.
[24] A. Gea, N. Farcy, N.R. Rosel, J.C. Martins, P.J. Clercq, A. Madder, Solid-Supported Synthesis of Highly Functionalized Tripodal Peptides with Flexible but Preorganized Geometry: towards Potential Serine Protease Mimics, Eur. J. Org. Chem. 18 (2006) 4135–4146.
[25] M. Mutter, S. Vuillemier, A chemical approach to protein design: template-assisted synthetic proteins (TASP), Angew. Chem. 28 (1989) 535–676.
[26] G. Tuchschererer, M. Mutter, Protein design as a challenge for peptide chemists, J. Pept. Sci. 1 (1995) 3–10.
[27] G.B. Bloomberg, D. Askin, A.R. Gargaro, M.J.A. Tanner, Synthesis of a branched cyclic peptide using a strategy employing Fmoc chemistry and two additional orthogonal protecting groups, Tetrahedron Lett. 34 (1993) 4709–4712.
[28] Y. Satow, G.H. Cohen, E.A. Padlan, D.R. Davies, Phosphocholine binding immunoglobulin Fab McPC603: an X-ray diffraction study at 2.7 Å, J. Mol. Biol. 190 (1986) 593–604.
[29] R.C.F. Jones, M. Tankard, A.M. Higton, Studies towards a hydrophobic serine protease model, Bioorganic Med. Chem. Letters 1 (1991) 353–356.
[30] M.A. Catry, A. Madder, Synthesis of functionalised nucleosides for incorporation into nucleic acid-based serine protease mimics, Molecules 12 (2007) 114–129.
[31] A. Stavrakoudis, I.N. Demetropoulos, C. Sakarellos, M. Sakarellos-Daitsiotis, V. Tsikaris, Design, synthesis and catalytic activity of a serine protease synthetic model, Lett. Pept. Sci. 4 (1997) 481–487.
[32] A. Stavrakoudis, S. Makropoulou, V. Tsikaris, M. Sakarellos Daiotsiotis, C. Sakarellos, I.N. Demetropoulos, Computational screening of branched cyclic peptide motifs as potential enzyme mimetics, J. Pept. Sci. 9 (2003) 145–155.
[33] V.A. Tatsis, A. Stavrakoudis, I.N. Demetropoulos, LysinebasedTrypsinActSite(LysTAS): a configurational tool of the TINKER software to evaluate Lysine based branched cyclic peptides as potential chymotrypsin-mimetics, Mol. Simul. 32 (2006) 643–644.
[34] J.W. Ponder, TINKER Software Tool, Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, http://dasher.wustl.edu/tinker/.
[35] I.N. Demetropoulos, N. Gresh, A supermolecular study of the effect of hydration on the conformational behaviour of leucine-enkephalin, J. Comput.-Aided Mol. Des. 5 (1991) 81–94.
[36] O. Collet, S. Premilat, B. Maigret, H.A. Scheraga, Comparison of explicit and implicit treatments of solvation: application to angiotensin II, Biopolymers 42 (1997) 363–371.
[37] B. Roux, T. Simonson, Implicit solvent models, Biophys. Chemist. 78 (1999) 1–20.
[38] T. Ooi, M. Oobatake, G. Nemethy, H.A. Scheraga, Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides, Proc. Natl. Acad. Sci. U.S.A 84 (1987) 3086–3090.
[39] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, J. Mol. Biol. 55 (1971) 379–380.
[40] I. Tunon, E. Silla, J.L. Pascual-Ahuir, Molecular surface area and hydrophobic effect, Protein Eng. 5 (1992) 715–716.
[41] F. Fraternalli, W.F. van Gunsternen, An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution, J. Mol. Biol. 256 (1996) 939–948.
[42] T. Lazaridis, M. Karplus, "New View" of protein folding reconciled with the old through multiple unfolding simulations, Science 278 (1997) 1928–1931.
[43] M.D. Cummings, N.T. Hart, R.J. Read, Atomic solvation parameters in the analysis of protein-protein docking results, Protein Sci. 4 (1995) 2087–2089.
[44] D. Eisenberg, A.D. McLachlan, Solvation energy in protein folding and binding, Nature 319 (1986) 199–203.
[45] L. Wesson, D. Eisenberg, Atomic solvation parameters applied to molecular dynamics of proteins in solution, Protein Sci. 1 (1992) 227–235.
[46] W.C. Still, A. Tempczyk, R.C. Hawley, T. Hendrickson, Semianalytical treatment of solvation for molecular mechanics and dynamics, J. Am. Chem. Soc. 112 (1990) 6127–6129.
[47] R. Constanciel, R. Contreras, Self consistent field theory of solvent effects representation by continuum models: introduction of desolvation contribution, Theo Chim. Acta 65 (1984) 1–11.
[48] M. Born, Volume and hydration heat of ions, Z. Phys. 1 (1920) 45–48.
[49] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, Pairwise descreening of solute charges from a dielectric medium, Chem. Phys. Lett. 246 (1995) 122–129.
[50] G.D. Hawkins, C.J. Cramer, D.G. Truhlar, Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium, J. Phys. Chem. 100 (1996) 19824–19839.
[51] J. Kraut, How do enzymes work? Science 242 (1988) 533–540.
[52] D. Beeman, Some multistep methods for use in molecular dynamics calculations, J. Comp. Phys. 20 (1976) 130–139.
[53] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[54] J. Wang, P. Cieplak, P.A. Kollman, How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? J. Comput. Chem. 21 (2000) 1049–1074.
[55] N. Foloppe, A.D. Jr MacKerell, All-atom empirical force field for nucleic acids: 1) parameter optimization based on small molecule and condensed phase macromolecular target data, J. Comput. Chem. 21 (2000) 86–104.
[56] A.C. Wallace, R.A. Laskowski, J.M. Thornton, Derivation of 3D coordinate templates for searching structural databases: application to Ser–His–Asp catalytic triads in the serine proteinases and lipases, Protein Sci. 5 (1996) 1001–1013.
[57] I.K. McDonald, J.M. Thornton, Satisfying hydrogen bonding potential in proteins, J. Mol. Biol. 238 (1994) 777–793.
[58] E.G. Hutchinson, J.M. Thornton, PROMOTIF—a program to identify and analyze structural motifs in proteins, Protein Sci. 5 (1996) 212–220.
[59] A. Jaramillo, S.J. Wodak, Computational protein design is a challenge for implicit solvation models, Biophys. J. 88 (2005) 156–171.
[60] H. Matter, H. Kessler, Structures, dynamics, and biological activities of 15 cyclic hexapeptide analogs of the alpha–amylase inhibitor tendamistat (HOE 467) in solution, J. Am. Chem. Soc. 117 (1994) 3347–3359.
[61] S. Hammes-Schiffer, S.J. Benkovic, Relating protein motion to catalysis, Ann. Rev. Biochem. 75 (2006) 519–541.
[62] S.J. Benkovic, S. Hammes-Schiffer, A perspective on enzyme catalysis, Science 301 (2003) 1196–1202.
[63] M.E. Colvin, C.J. Cramer, C.E. Dykstra, J.H. Jensen, S. Krimm, J.-L. Rivail, A.J. Thakkar, Manuel Yanez, Molecular quantum mechanics to biodynamics: essential connections, J. Mol. Struct., Theochem 764 (2006) 1–8.