Available online at www.sciencedirect.com
ScienceDirect
Journal of Molecular Graphics and Modelling 26 (2007) 571–580
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

Conformational analysis of compstatin analogues with molecular dynamics simulations in explicit water

Phanourios Tamamis [a], Spiros S. Skourtis [a], Dimitrios Morikis [b], John D. Lambris [c], Georgios Archontis [a,*]

[a] Department of Physics, University of Cyprus, PO20537, CY1678 Nicosia, Cyprus
[b] Department of Bioengineering, University of California at Riverside, Riverside, CA 92521, United States
[c] Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States

Received 15 November 2006; received in revised form 26 March 2007; accepted 30 March 2007
Available online 4 April 2007

Abstract

The cyclic 13-residue peptide compstatin is a potential therapeutic agent against the unregulated activation of the complement system. A thorough knowledge of its structural and dynamical properties in solution may assist the design of improved complement inhibitors. NMR studies have suggested that the 5–8 segment of free compstatin folds into a critical for activity 5–8 β turn and the rest of the peptide is mainly disordered. Earlier computational studies of compstatin analogues with a polar-hydrogen/generalized-Born approximation reproduced the 5–8 turn, but also indicated the formation of β-hairpin or α-helical elements and the existence of interactions between certain charged or aromatic sidechains. However, these features are absent or partly present in the NMR spectra, due to extensive conformational averaging. In order to check the compstatin properties with a more rigorous model of the intra- and intermolecular interactions, we conduct here 98-ns all-atom/explicit-water simulations of three compstatin analogues with variable activity; a native analogue, the more active mutant V4W/H9A and the inactive mutant Q5G. The 5–8 β-turn population is in good accord with NMR. For the systems studied here, the simulations suggest that the 5–8 turn population does not correlate strictly with activity, in agreement with earlier mutational studies. Furthermore, they show structural differences among the analogues outside the 5–8 region. The possible role of these differences in activity is discussed. The probability of β-hairpin or α-helix elements is much smaller with respect to the polar-hydrogen/GB simulations, and the persistent Trp4-Trp7 or Asp6-Arg11 sidechain interactions of the earlier GB studies are not reproduced. The present simulations extend the NMR data and improve our understanding of the properties of compstatin and related analogues.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Molecular dynamics simulations; Compstatin analogues; Complement system; Conformational analysis

1. Introduction

The complement system is activated through the classical, alternative, or lectin pathways and provides the first line of defense against foreign pathogens [1]. Inappropriate complement activation may contribute to several pathological conditions, such as asthma, adult respiratory distress syndrome, hemolytic anemia, rheumatoid arthritis and rejection of xenotransplantation [2]. Therefore, the development of drugs which can control complement activation is of significant medical interest.

The 13-residue cyclic peptide compstatin binds to the β-chain of complement component C3 [3] and inhibits complement action (reviewed in Ref. [4]). Compstatin is a promising candidate as a therapeutic agent against unregulated complement activation and has been the subject of extensive experimental NMR [5–8] and computational [9–13] studies by some of the present authors. The earlier computational studies studied native compstatin [9] and several of its analogues [10–13] by molecular dynamics (MD) simulations, which employed a polar-hydrogen energy function for the peptide interactions [14] and a compatible generalized-Born representation for the solvent [15]. The simulations indicated a

* Corresponding author. Tel.: +357 22 89 28 22; fax: +357 22 89 28 21.
E-mail address: archonti@ucy.ac.cy (G. Archontis).

1093-3263/$ – see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2007.03.014

dominant conformation (41.3% probability) with the segment 5–8 in a type-I β-turn and the rest of the structure as random coil [9], in good agreement with the experimental NMR estimate for a 5–8 type-I turn (42–63% [5]). Other structural properties emerging from the simulations were not present in the NMR spectra, due to extensive conformational averaging caused by the high peptide flexibility. In addition to the dominant conformation mentioned above, β-hairpin (34.1% probability) and α-helical elements (6.1% probability) were observed in the native compstatin GB simulations, without corresponding NOEs in the NMR spectra. The GB simulations of Ref. [12] suggested the formation of persistent electrostatic interactions between residues Asp6 and Arg11 in various analogues, and a strong Trp7-Trp4 interaction in the double-mutant V4W/H9A. The NMR data did not depict these interactions because the NOE effect measures inter-proton distances up to 5–6 Å.

The earlier simulations provided information on the conformational properties and intramolecular interactions which can be used in the interpretation of the activity differences among analogues [9,12]. However, the difficulty in validating some of the simulation properties with relevant information in the NMR spectra suggests that it is important to check the compstatin properties with a more rigorous model for the peptide–peptide and peptide–solvent interactions. In this work, we conduct for the first time MD simulations of three compstatin analogues, using an explicit representation for the solvent and an all-hydrogen molecular model for the solute and solute–solvent interactions. The simulations are extensive (with a total length of 98 ns) and focus on three compounds with different activities: (i) an analogue of native compstatin (NAT), which is acetylated at the N-terminal end and has a three-fold higher activity with respect to the un-acetylated molecule [6–8,10]; (ii) the double-mutant V4W/H9A (W4A9), which has a 15-fold higher activity than NAT [13]; (iii) finally, the inactive single mutant N5G (G5) [6].

The present simulations are in good agreement and extend the NMR data. They are likely to provide a more representative picture of the conformational properties of the considered systems. The 5–8 moiety has a high (NAT, G5) to moderate (W4A9) propensity for a β-turn, whereas the rest of the molecule is mainly disordered in all three analogues, with the occasional formation of turns. The frequency of β-hairpin or α-helix formation is much smaller with respect to the earlier GB/polar-hydrogen studies. The Trp4-Trp7 or Asp6-Arg11 side-chain interactions of the GB/polar simulations are not reproduced in the current, explicit-water runs. The present simulations show that the three analogues differ in their dynamical properties in solution. We discuss these differences, in connection with the variation in analogue activity.

2. Methods

2.1. Molecular dynamics simulations

All simulations were performed in explicit water and employed cubic periodic boundary conditions. The peptide atomic charges, van der Waals and stereochemical parameters were taken from the CHARMM22 all-atom force field [16]. The water was represented by a modified TIP3P water model [17,18]. The N- and C-termini of the three analogues were blocked, respectively, by $-COCH_3$(Ac) and $-NH_2$ groups. The same groups were employed in the generalized-Born MD simulations of Refs. [12,13], whereas in Ref. [9], the native compstatin was unblocked (charged) at the N-terminus. Electrostatic interactions were calculated without truncation by the particle-mesh Ewald method [19], with a parameter $\kappa = 0.33333$ for the charge screening, and sixth-order splines for the mesh interpolations. The Lennard–Jones interactions between atom pairs were shifted to zero at a cutoff distance of 14 Å. The temperature was kept at $T = 300$ K by a Nose–Hoover thermostat [20,21] using a mass of 1000 kcal/ps^2 for the thermostat. The pressure was maintained at $P = 1$ atm with a Langevin piston [22], using a piston mass of 500 a.m.u., a piston collision frequency of 5 ps^{-1}and a piston bath temperature of 300 K. The classical equations of motion were integrated by the Leap-Frog integrator, using a time step of 2 fs. Bond lengths to hydrogen atoms and the internal geometry of the water molecules were constrained with the SHAKE algorithm to standard values [23]. All simulations were conducted with the molecular mechanics program CHARMM, version c31b2 [14].

Native compstatin has the sequence Ile1-Cys2-Val3-Val4-Gln5-Asp6-Trp7-Gly8-His9-His10-Arg11-Cys12-Thr13-NH_2, and contains a Cys2-Cys12 disulfide bond. The titratable residues were assigned their most common ionization state at physiological pH; i.e., Asp6 and Arg10 were charged and the two histidines His9, His10 were neutral. The disulfide bond Cys2-Cys12 was maintained by a disulfide patch. Earlier NMR studies determined a family of 21 compstatin structures (PDB entry 1A1P [5]). In 19 structures the disulfide dihedral angle $2C_\beta$-$2S_\gamma$-$12S_\gamma$-$12C_\beta$ was found 180°, away from the theoretical minima of ±90°. This averaged conformation of the disulfide bridge in the ensemble of NMR structures reflects the absence of NOEs in the region surrounding the disulfide bridge (residues Ile1, Cys2, Arg11, Cys12, Thr13) (see Ref. [5] and supplementary material of Ref. [5]). This is because of flexibility in this region, which results to conformational averaging of NMR observables.

In the conditions of the present, explicit-solvent simulations (a temperature of 300 K and duration of ≈2 ns), the disulfide dihedral fluctuates around one of the two minima and transitions between minima are not observed (the barriers separating the two minima are ≈8–10 kcal/mol). To investigate the effect of the $2C_\beta$-$2S_\gamma$-$12S_\gamma$-$12C_\beta$ angle value on the dynamical and structural properties of the three analogues, we performed several separate simulations, in which the angle $2C_\beta$-$2S_\gamma$-$12S_\gamma$-$12C_\beta$ fluctuated around +90° or −90°. In this manner, the current MD simulations assess a property of the structure that is not accessible by the NMR data.

To create the aqueous environment, a peptide with the sequence of acetylated compstatin (NAT) and coordinates corresponding to the first in the 21 NMR structures was placed at the center of a preequilibrated, cubic water box with an edge

of 53 Å. Water molecules with oxygens located closer than 2.5 Å from any peptide non-hydrogen atom were deleted, yielding a system with 1 peptide and 7921 water molecules (23,975 atoms). The peptide distance from the edge of the water box was at least 16 Å. Subsequently, separate MD simulations were started with the peptide coordinates set to different NMR conformations. The initial atomic coordinates of the N-terminal blocking group and the mutated sidechains in analogues G5 and W4A9 were constructed from the default internal coordinate values in the CHARMM all-atom parameter and topology database files [16]. In all simulations, the same initial water setup was used.

Prior to each simulation, the entire system was subjected to a 200-step steepest descent minimization, with the disulfide dihedral angle $2C_\beta$-$2S_\gamma$-$12S_\gamma$-$12C_\beta$ harmonically restrained around one of the $\pm 90^\circ$ minima by a force constant of 100 kcal/mol rad^2, and the peptide heavy atoms harmonically restrained around their initial positions with a force constant of 0.5 kcal/mol Å^2. Subsequently, an initial, 30-ps equilibration simulation was conducted. At the end of the 30-ps simulation the harmonic restraints were removed, and the simulation was continued for another 2 ns. During the 2-ns production phase, the coordinates were saved every 0.5 ps and used later in the analysis. Each 2-ns simulation took $\approx$ 36 CPU days on a 2.4-GHZ Xeon processor.

The family of 21 native-compstatin NMR structures are similar, with an average RMSD of 0.6 Å for the backbone heavy atoms and 1.2 Å for all heavy atoms [5]; in terms of secondary structure, they all correspond to a coil with a 5–8 β-turn. Given this small conformational diversity, initial conformations for the MD simulations (listed in Table 1) were selected randomly among the NMR structures. For the NAT peptide, we conducted 12 simulations around the $+90^\circ$ and 10 simulations around the -90° Cys2-Cys12 disulfide torsion minimum, starting from 12 different NMR structures. For W4A9 we conducted 5 simulations ($+90^\circ$) and 11 simulations (-90°), starting from 11 different NMR structures. For G5, the corresponding numbers were 6 ($+90^\circ$) and 5 (-90°); 8 different NMR structures were used. The total simulation length in all systems was 98 ns.

Table 1
NMR structures used as initial peptide conformations in the simulations

Peptide	Structures[a]
NAT^{+b}	1, 2, 5, 5, 11, 12, 13, 13, 14, 16, 16, 18
NAT$^-$	3, 3, 6, 9, 9, 10, 10, 11, 12, 18
G5$^+$	1, 3, 5, 9, 14, 18
G5$^-$	1, 3, 5, 6, 12
W4A9$^+$	2, 3, 4, 4, 21
W4A9$^-$	1, 1, 2, 7, 10, 11, 13, 17, 18, 18, 20

[a] The numbers refer to the 21 NMR low-energy structures (entry 1A1P [5]) of the native un-acetylated peptide. Repeated numbers denote independent simulations, starting from the same structure with different velocity distributions. The N-terminal blocking group (all peptides) and the mutated sidechains (G5 and W4A9) were placed according to the default internal coordinate parameter values of the CHARMM22 parameter/topology files [16].

[b] Signs "+/−" denote trajectories with the disulfide dihedral $2C_\beta$-$2S_\gamma$-$12S_\gamma$-$12C_\beta$ in the $+90^\circ/-90^\circ$ minimum (see text).

2.2. Cluster analysis

To investigate the structural diversity of the three analogues we performed a cluster analysis of the simulation trajectories [24–26]. The clustering method is presented in detail in Ref. [24]. Briefly, a trajectory snapshot j is described by a set of K predetermined properties (e.g., backbone dihedral angles or Cartesian coordinates), arranged into a K-dimensional vector $\vec{x_j} \equiv (x_{1j}, x_{2j}, \ldots, x_{Kj})$. The center of a cluster l containing $M(l)$ snapshots is defined by the arithmetic mean $\vec{c_l} = (c_{1l}, c_{2l}, \ldots, c_{Kl}) \equiv 1/M(l) \sum_{j \in M(l)} \vec{x_j}$. Each of the $M(l)$ snapshots comprising cluster l is within a certain predefined threshold (radius) d_0 from the cluster center $\vec{c_l}$ (i.e., $\sqrt{\sum_{i=1}^{K} (x_{ij} - c_{il})^2} \leq d_0$).

In our analysis, we clustered the simulation trajectories according to the Cartesian coordinates of selected C_α atoms. One set of calculations employed the C_α coordinates of residues 5–9, which often fold into one (5–8) or two fused (5–8/6–9) β-turns. The maximum cluster radius d_0 was set to 1.0 Å, corresponding to average structural fluctuations of $\approx$0.45 Å for each of the five C_α atoms in segment 5–9. A second set of calculations clustered the region 2–12, which contains the Cys2-Cys12 disulfide bond. It employed the same cluster radius (1 Å), corresponding to individual fluctuations of $\approx$0.3 Å. As shown in Ref. [9], changes in the secondary structural content throughout the peptide sequence or the type of β-turn in region 5–8 were associated with backbone motions of 0.1–0.4 Å. Thus, this threshold was expected to differentiate among different conformations. Test calculations with larger or smaller thresholds gave, respectively, a smaller number of clusters with significant structural variation, or numerous clusters with very small population.

In each case, prior to the clustering analysis the appropriate trajectories were merged and the net peptide translation and rotation was removed. The analysis was performed with standard CHARMM routines [24].

3. Results

NMR studies of native compstatin determined 21 low-energy structures [5]. Analysis of the $^3J_{NH-H_\alpha}$ coupling constants showed that the segment Gln5-Asp6-Trp7-Gly8 adopts a type-I β-turn conformation with a 42–63% probability and the rest of the molecule is highly flexible. Subsequent mutational and NMR studies examined systematically the importance of all residues for the structure and activity of compstatin [7,11]. The observations of these experimental studies can be summarized as follows: (a) the disulfide bond Cys2-Cys12 is necessary for activity [5,7]. (b) Residues Gln5, Asp6, Trp7, Gly8 are critical for activity; alanine substitutions at these positions possibly eliminate important interactions with C3, and/or disrupt the 5–8 β-turn [5,7]. (c) Formation of the 5–8 β-turn is not a *sufficient* condition for activity: the single mutant V3A is inactive, despite its higher propensity for a 5–8 β-turn with respect to the native compound [7]. In our simulations, the frequency of the 5–8 β-turn does not correlate with activity. In

the mutants W4A9 (more active than NAT) and G5 (inactive) the 5–8 turn population is, respectively, smaller and comparable to the corresponding population for NAT (below). (d) Some point mutations in critical residues can be tolerated, provided they preserve the chemical properties at the corresponding sites (V3L, Q5N) [7]. (e) Residues Ile1, Cys2, Val3, Cys12, Thr13 participate in a hydrophobic clustering, which may be significant for activity [7]. (f) Residues Val4, His9, His10 and Arg11 are not required for activity; alanine substitutions at these positions have a small effect on inhibition. Nevertheless, substitutions at positions 4 and 9 have produced higher-activity analogues, in agreement with experimental [7,8,13] and computational combinatorial studies [10,27] studies.

In what follows we describe and compare the dynamical properties of the three analogues NAT, G5 and W4A9 in the explicit-water simulations. As we show here and elaborate in the discussion section, some of the above observations can be understood by the dynamical behavior of the three analogues in the explicit-water simulations.

3.1. Propensity for the 5–8 β turn

The ideal internal backbone dihedral values and the criteria used for the classification of β-turns are summarized in Table 2. These criteria were employed to compute the types and frequencies of β-turns, observed in the simulation structures. The frequencies of the two most important turns, formed by segments 5–8 and 6–9, are included in Table 3.

In native compstatin, the total 5–8 β-turn probability varies between 80.8% (NAT^-) and 72.5% (NAT^+). A 36.6–33.5% of turns fall into the α_R–α_R family (i.e., type I or III [9,28]), in reasonable agreement with the corresponding NMR estimate (42–63% [5]). A small fraction (1.5–1.7%) of α_R–β turns are also observed. The remaining β-turns do not fall into the categories listed in Table 2. However, the majority of these turns adopt dihedral angles within ±45° of the ideal α_R–α_R values.

The propensity for the 5–8 β-turn does not correlate with activity in the present simulations. The active peptide NAT and the inactive G5 have a similar propensity for this turn (Table 3).

Table 2
Dihedral angles of ideal β-turns[a]

Turn type		Dihedral angles			
		ϕ_{i+1}	ψ_{i+1}	ϕ_{i+2}	ψ_{i+2}
α_R–α_R [b]	I	−60	−30	−90	0
	III	−60	−30	−60	−30
α_L–γ	I′	60	30	90	0
α_L–α_L	III′	60	30	60	30
β–γ	II	−60	120	80	0
ϵ- α_R	II′	60	−120	−80	0
α_R–β	VIII	−60	−30	−120	120

[a] A segment $i - i+3$ forms a β-turn if the following criteria apply [9]: (i) the distance $C_{\alpha,i} - C_{\alpha,i+3} \leq 7$ Å; (ii) residues $i+1$, $i+2$ are not helical; (iii) the ϕ and ψ angles of residues $i+1$, $i+2$ fall within ±30° of the ideal values, with one dihedral allowed to deviate by ±45° [9]. The ideal values are taken from Refs. [9,31].

[b] Nomenclature introduced by Wilmot and Thornton [28], with the first and second component denoting, respectively, the region in the Ramachandran map for residue $i+1$ and $i+2$. The four regions of the Ramachandran map are: (1) β-region, with $\phi \leq 0°$, $\psi > 60°$ or $\psi < -120°$, (2) α_R region with $\phi \leq 0°$, $-120° \leq \psi \leq 60°$, (3) α_L/γ region with $\phi > 0°$, $\psi < 110°$ or $\psi > -70°$ and (4) ϵ-region with $\phi > 0°$, $\psi \geq 110°$ or $\psi \leq -70°$.

A similar behavior was observed in NMR studies of the V3A and W7F mutants, which lacked activity but maintained a 5–8 β-turn [7]. In the simulations of the more active mutant W4A9, the 5–8 β-turn population is significantly reduced (see Table 3) and the peptide has higher flexibility (below). This is in agreement with earlier NMR studies, which have suggested that W4A9 and the related single mutant H9A are more flexible than native compstatin [7,13].

The region outside the 5–8 moiety is mainly disordered in all analogues, with the occasional formation of other β-turns. The most important such turn is formed by segment 6–9, with a probability which varies from 17.2% ($G5^-$) to 37.4% (NAT^-) (Table 3). Its formation is facilitated by the 5–8 turn; indeed, ≈80% of the NAT structures with a 6–9 turn contain also the 5–8 turn. The formation of turns outside the 5–8 segment is supported to a certain extent by the NMR results. In the case of W4A9, the spectra contain HN($i+1$)–HN($i+2$) and HN($i+2$)–HN($i+3$) NOE cross peaks, which suggest the

Table 3
Turn and coil populations in the MD simulations of the three peptides

Segment	Turn	NAT^{-} [a]	NAT^+	$W4A9^-$	$W4A9^+$	$G5^-$	$G5^+$
5–8	α_R–α_R [b]	36.6 (25.5)[c]	33.5 (25.0)	13.2 (16.8)	19.4 (9.0)	30.2 (10.2)	40.3 (8.9)
	Other[d]	44.1 (29.2)	39.1 (21.2)	28.1 (25.3)	34.3 (26.0)	51.2 (22.8)	47.5 (19.8)
	Total turn	80.7 (38.8)	72.6 (32.8)	41.3 (30.4)	53.7 (27.5)	81.4 (25.0)	87.8 (21.7)
	Coil	19.2 (24.0)	27.4 (29.8)	58.7 (30.4)	45.8 (24.9)	18.6 (23.3)	12.2 (23.3)
6–9	α_R–α_R	0.7 (1.9)	0.2 (0.5)	3.9 (5.4)	8.7 (12.4)	0.0 (0.0)	0.1 (0.2)
	Other	36.7 (24.0)	26.2 (19.7)	17.7 (27.0)	26.6 (30.5)	17.2 (9.9)	25.6 (12.2)
	Total turn	37.4 (24.1)	26.4 (19.7)	21.6 (27.5)	35.3 (32.9)	17.2 (9.9)	25.7 (12.2)
	Coil	62.6 (26.0)	73.6 (25.9)	78.0 (27.9)	64.1 (38.3)	82.8 (15.5)	74.3 (18.1)

[a] $X^{+/-}$denotes simulations of system X, with the disulfide dihedral angle in the $+90/-90$ minimum.

[b] The classification of turns is explained in the footnote of Table 2.

[c] The standard deviations of the populations obtained in the 2-ns production trajectories are in parentheses.

[d] Turns which satisfy only the first two β-turn criteria of Table 2, or belong to the α_R–β category. In some cases (not shown), a 3_{10} helix pattern is observed (with population ≤ 1%).

formation of a 5–8/6–9 fused turn (the definitive $H_\alpha(i)$–HN($i+2$) NOE is missing due to conformational averaging). Several fused turns are also observed in the NMR data of Ref. [7] (see Fig. 9 of Ref. [7]). This is less obvious in the acetylated compstatin data [8], due to spectral overlap.

The data of Table 3 show that the populations of the internal 5–8/6–9 turns are similar in the ±90° disulfide dihedral trajectories. Thus, the behavior of internal residues is not considerably influenced by the conformation of the disulfide dihedral angle. Some differences are observed in the case of the N-terminal end turns 1–4 (NAT) and 2–5 (W4A9), and the C-terminal end turn 9–12 for G5 (not shown).

3.2. Conformational variability of the three analogues

The variation in activity of the three analogues may partly arise due to differences in their structural properties when they are free (unbound) in solution. To determine the conformational variability within each analogue and the similarity among analogues, we performed a cluster analysis on the simulation structures; the methodology has been described in Section 2.

3.3. Region 5–9

We first focussed on the backbone moiety 5–9, which contains the important turn 5–8. Residue 9 was included since a 6–9 turn is frequently formed together with 5–8. The backbone conformations of all simulations were clustered together, without distinguishing among analogues. An important conclusion of this analysis is that the 5–9 backbone region has a small overall structural variability. Indeed, the conformations are partitioned into just eight clusters, despite the considerable total length of the runs (98 ns). Fig. 1 shows the average structures of all resulting clusters. Table 4 reports the population of various β-turns in the members of each cluster. The structures of the first four clusters contain a 5–8 turn with high probability. In the remaining clusters, the conformations become progressively more open (see Fig. 1) and the corresponding 5–8 probability diminishes to zero.

Table 4
Clustering analysis of the $C_\alpha$atomic coordinates in the region 5–9

	Cluster							
	1	2	3	4	5	6	7	8
Population[a]	35.7	20.2	13.4	11.5	8.4	7.6	2.7	0.4
NAT[b]	1.34	0.49	1.53	0.70	0.69	1.0	0.8	0.0
G5[b]	0.72	2.67	0.73	0.27	0.11	0.59	0.66	0.0
W4A9[b]	0.71	0.56	0.46	1.91	2.04	1.29	2.50	3.06

Turn	Turn frequency(%)							
	1[c]	2[c]	3[c]	4[c]	5[c]	6[c]	7[c]	8[c]
5–8	96	96	65	40	12	0	0	0
6–9	67	6	11	0	0	13	0	0
2–5	12	20	9	22	48	21	20	58
3–6	7	1	13	10	12	11	12	3
4–7	0	1	12	47	21	21	9	35
7–10	1	0	2	1	3	10	19	79
8–11	0	3	5	19	43	4	17	18
9–12	22	19	13	23	33	9	4	1

The 5–9 backbone conformations of all three analogues are analyzed together.
[a] Total population (%) of each cluster.
[b] Scaled contribution (%) from each analogue to the total cluster population (see text).
[c] Cluster.

Despite the small conformational variability of region 5–9, there are important differences among the three analogues. To show the specific analogue preference for certain clusters, we computed in each cluster the fraction of structures corresponding to a particular analogue, divided by the total fraction of structures of the same analogue in the 98-ns simulation length. The resulting ratios are included in Table 4. Values greater than 1 identify analogues with a higher-than-average preference for a cluster. The native peptide (NAT) has a higher propensity for the first cluster, which contains the fused turn 5–8/6–9, and the third cluster, which is somewhat more open. Peptide G5 has a very pronounced tendency for the second cluster, which

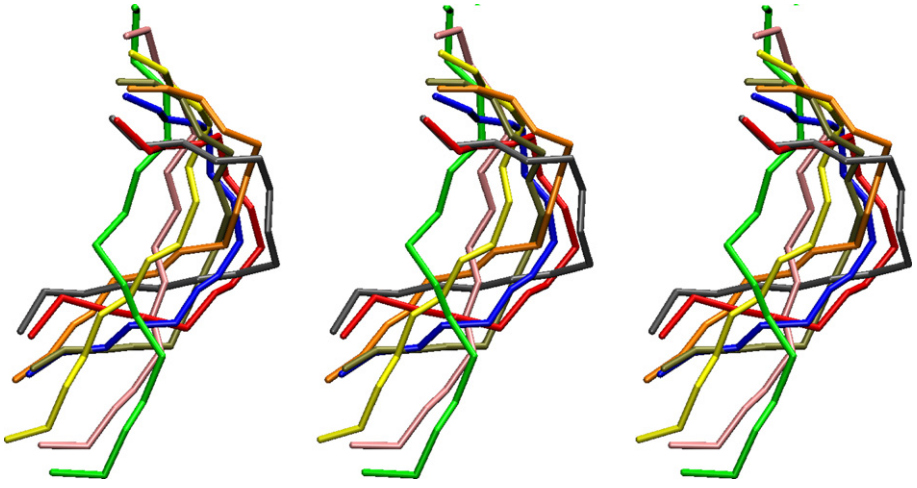


Fig. 1. Cross (l) and wall (r) stereo representation of the average cluster structures, which result by analyzing together the simulation conformations of the backbone 5–9 segment for all peptides (see text). Only the 5–9 backbone is shown; residue 5 is on the upper portion of the figure. The cluster:color correspondence is: 1, blue; 2, red; 3, gray; 4, orange; 5, yellow; 6, tan; 7, pink; 8, green. The structural properties of the clusters are discussed in the text and Table 4.

contains closed conformations (see Fig. 1), with the 5–8 but not the 6–9 turn. The pronounced preference of G5 for just one cluster indicates that it is the least flexible of the three peptides. In contrast, peptide W4A9 is the most flexible and adopts conformations which fall into the last five, more open clusters.

An additional analysis of backbone–backbone contacts (not shown) demonstrated that the region 5–9 is somewhat tighter in G5 compared to NAT, possibly because of absence of steric interactions due to the lack of side chain in residue 5. In constrast, the structure of the W4A9 analogue is more open, both in the segment 5–8 and in the region in the vicinity of the disulfide bond, possibly reflecting steric effects due to the bulky side chain of Trp4.

Notably, the conformation of the 5–9 backbone region is coupled with the structural elements observed in the rest of the peptide. In the conformations of the first two clusters the 5–9 structure is very tight, and only the non-overlapping turns 2–5 and 9–12 are observed. In most of the remaining clusters (3–8) we also observe the formation of the additional turns 4–7 and 3–6, which partly overlap with the turns 5–8/6–9. In cluster 4, both turns 4–7 and 5–8 have significant probability (47% and 40%, respectively).

3.4. Region 2–12

To obtain a better understanding of the conformations adopted by the entire peptide backbone, we also performed a cluster analysis based on the C_α coordinates of residues 2–12. In this analysis, we clustered separately the three peptides and distinguished betwen trajectories with the disulfide angle in the $+90°$ and $-90°$ minima. The total number of clusters in each case is included in Table 5. Even though the structures of the entire region 2–12 are grouped into a much larger number of clusters compared to segment 5–9, the seven (G5, W4A9) or eight (NAT) most populated clusters account for $\geq 70\%$ of conformations (see Table 5). The average structures and properties of the seven most important clusters in each case are included in the accompanying supplementary information. In what follows, we describe some characteristic clusters and discuss the emerging structural properties. The average conformations of these clusters are shown in Fig. 2; $X^{\pm}(j)$ denotes the jth cluster in the $\pm 90°$ simulation of the X analogue.

Cluster $NAT^{+}(3)$ (see Fig. 2) contains the 5–8/6–9 fused turn and has the highest probability in the $+90°$ trajectory with 24.0% of the frames. Its backbone region 5–9 corresponds mostly to cluster 1 of Table 4. Notably, the His9 sidechain is positioned in the compstatin interior, and hydrogen-bonding with the Val4 main chain. This His9 conformation is also observed frequently in the G5 simulations. It is likely that the His9 sidechain contributes to the structural stability of the NAT and G5 analogues, and its replacement by Ala is partly responsible for the increased flexibility of W4A9.

Clusters $NAT^{+}(2)$ and $NAT^{-}(5)$ show examples of native conformations which lack the characteristic 5–8 β-turn, due to sidechain interference. In cluster $NAT^{+}(2)$ (7.7%), the His9 sidechain is placed between residues 5 and 8, preventing the approach of atoms $C_\alpha(5)$ and $C_\alpha(8)$ and interacting strongly with the backbone of residues Val4, Gln5 and Trp7. The 5–9 conformation is open and corresponds better to cluster 6 of Table 4. In cluster $NAT^{-}(5)$, the sidechain of Gln5 adopts a position between residues 5 and 8, which is stabilized by a hydrogen bond with Trp7. The 5–9 moiety adopts an open conformation that corresponds mostly to cluster 5 of Table 4. Cluster $NAT^{+}(8)(6.3\%)$ shows the example of an antiparallel β-ladder, which is stabilized by the Cys2(O)-His10(N) and Val4(N)-Gly8(O) hydrogen bonds. These hydrogen bonds keep segment 4–8 in a tight conformation, with two fused turns between 4–7 and 5–8. The conformation of 5–9 is mostly related to cluster 4 of Table 4.

Table 5
Total number of clusters, computed by employing the C_α atomic coordinates of the segment within Cys2-Cys12

NAT^{-}	NAT^{+}	$G5^{-}$	$G5^{+}$	$W4A9^{-}$	$W4A9^{+}$
25	18	73	12	47	140
68.6%[a]	77.1%[a]	73.6%[a]	81.7%[a]	34.0%[a]	70.2%[a]

[a] Fraction of the total trajectory, corresponding to the eight ($NAT^{\pm}$) or seven ($G5^{\pm}$, $W4A9^{\pm}$) highest population clusters.

3.5. Comparison of the interactions formed by the analogues

The activity of the compstatin derivatives depends partly on their properties in the free (uncomplexed) state in solution, i.e. on their conformation, intramolecular and intermolecular interactions (with solvent). The above analysis shows that the analogues differ somewhat in their structural properties and flexibility. In what follows, we describe and compare the interactions formed by the analogues.

Fig. 3 includes maps of the backbone–backbone and backbone–sidechain hydrogen-bond occupancies. In each map, the element (i, j) portrays the fraction of simulation structures containing an *i–j* hydrogen bond. Non-zero elements on or near the diagonal reflect interactions NH(i)–CO(i) within the same residue i (referred to as "C5" [29]), or CO($i-1$)–NH($i+1$) (referred to as "C7" [29]). Non-zero elements away from the diagonal correspond to hydrogen bonds between residues which are distant in sequence.

The NAT map is shown in the upper left plot of Fig. 3. Even though the peptide is restrained in a cyclic conformation by the Cys2-Cys12 disulfide bond, few residues far apart in sequence form significant hydrogen bonds. The most important such interactions involve Cys2(CO)-His10(N), Val4(N)-Gly8(CO) and Gln5(CO)-Gly8(N). The first two hydrogen bonds are related with the formation of a β-hairpin (as in the structure of cluster $NAT^{+}(8)$ shown in Fig. 2). The last interaction is associated with the 5–8 β-turn (structure of cluster $NAT^{+}(3)$ in Fig. 2).

The G5 backbone forms similar hydrogen-bonds with NAT (not shown). The most important difference between the two analogues involves the 5–8 hydrogen bond, which in G5 involves atoms Gln5(N)-Gly8(CO) and in NAT is between Gln5(CO)-Gly8(N). The latter hydrogen bond is of the type $(i, i+3)$ and should correspond to a more stable β-turn. In the

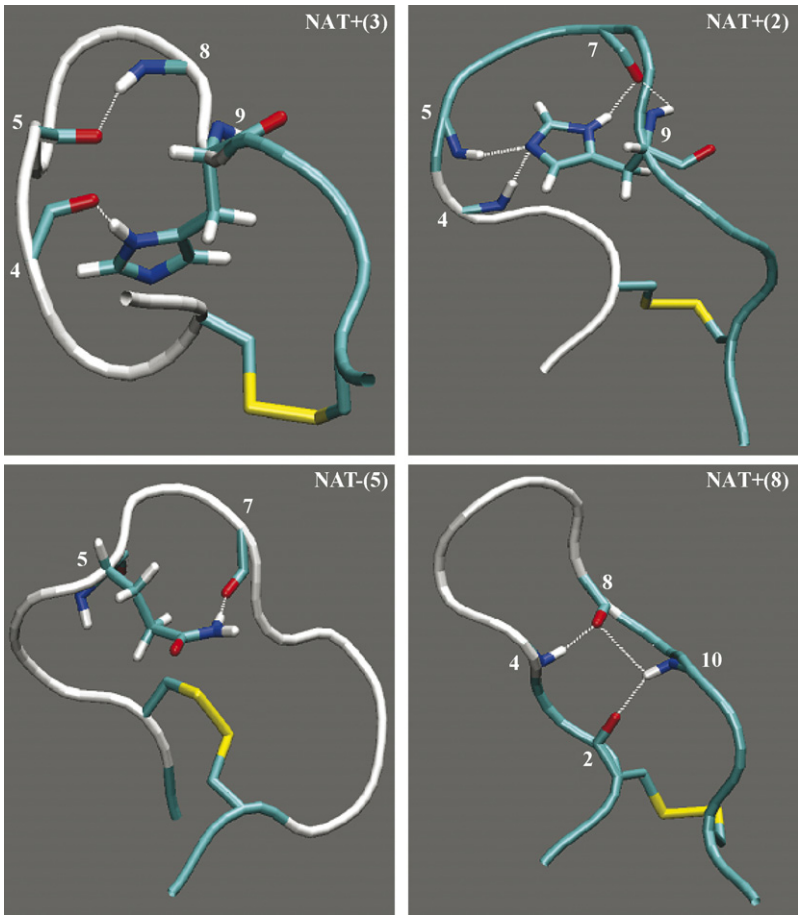


Fig. 2. Average structures of selected clusters, resulting from the MD trajectories of the various analogues. The cluster analysis has employed the C_α atomic coordinates of the 2–12 region (see text). $X^{\pm}(j)$ denotes the jth cluster in the $\pm 90^\circ$ simulation of the X analogue. The white, green and yellow colors indicate, respectively, β-turns, random coil elements and the Cys2-Cys12 disulfide bridge. The N-terminal (Ile1) and C-terminal (Thr13) residues are, respectively in the lower left and right. Apart from the average 2–12 backbone conformation, additional structural information is included, when it is characteristic of the cluster (high-occupancy hydrogen bonds, sidechains participating in high-probability interactions).

past it was believed that all β-turns should have this type of hydrogen bond [30,31], but this criterion was relaxed later on; however, the stability argument remains. The W4A9 analogue forms reduced backbone hydrogen bonds compared to NAT and G5 (not shown). This is in agreement with the fact that the W4A9 simulation structures are more open, as discussed above.

The backbone–sidechain hydrogen bond maps are included in the next three plots of Fig. 3. In the NAT map (upper right plot), the most important interactions involve the sidechains of Gln5 and His9. The Gln5 sidechain hydrogen-bonds with the backbone moieties of Trp7 (e.g., as in cluster $NAT^{-}(5)$ of Fig. 2), Gly8, His10 (cluster $NAT^{-}(2)$, supplementary information) and Arg11. Due to these interactions (particularly with Trp7), Gln5 interferes somewhat with the formation of the 5–8 β-turn. The sidechain of His9 has a rich structural behavior. As shown in the map, it interacts frequently with the backbone of residues Val4, Gln5 and Gly8, obstructing the 5–8 turn (as in cluster $NAT^{+}(2)$ of Fig. 2). It is also frequently positioned in the peptide interior further away from the 5–8 turn (as in cluster $NAT^{+}(3)$), or packs against the disulfide bond and interacting with residues 1–3 (as in cluster $NAT^{-}(2)$, shown in the supplementary information). In these alternative conformations, His9 probably tightens the NAT structure and stabilizes the 5–8 β-turn. Apart from His9, additional important sidechain–backbone interactions involve Asp6 with its own backbone, His10 with the backbone of Arg11, and Arg11 with the backbone of Thr13.

The G5 backbone–sidechain hydrogen-bond map is shown in the lower left plot of Fig. 3. In this analogue, the sidechain of residue 5 (Gly) lacks a hydrogen-bonding capability. Interestingly, the occupancies of hydrogen bonds involving the His9 and His10 sidechains are also reduced with respect to NAT (with the exception of the His9-Cys2, His9-Val3 bonds). Thus, it seems that in the *native* analogue, the hydrogen-bonding interactions of the Gln5 sidechain facilitate the formation of hydrogen bonds by other residues.

The W4A9 sidechain–backbone hydrogen bond map is shown in the lower right plot of Fig. 3. The substitution H9A eliminates all-hydrogen bonds formed by the sidechain of residue 9, but also reduces significantly the hydrogen bonds formed by Gln5. Overall, the analogue W4A9 makes very few backbone–sidechain hydrogen bonds. The W4A9 backbone

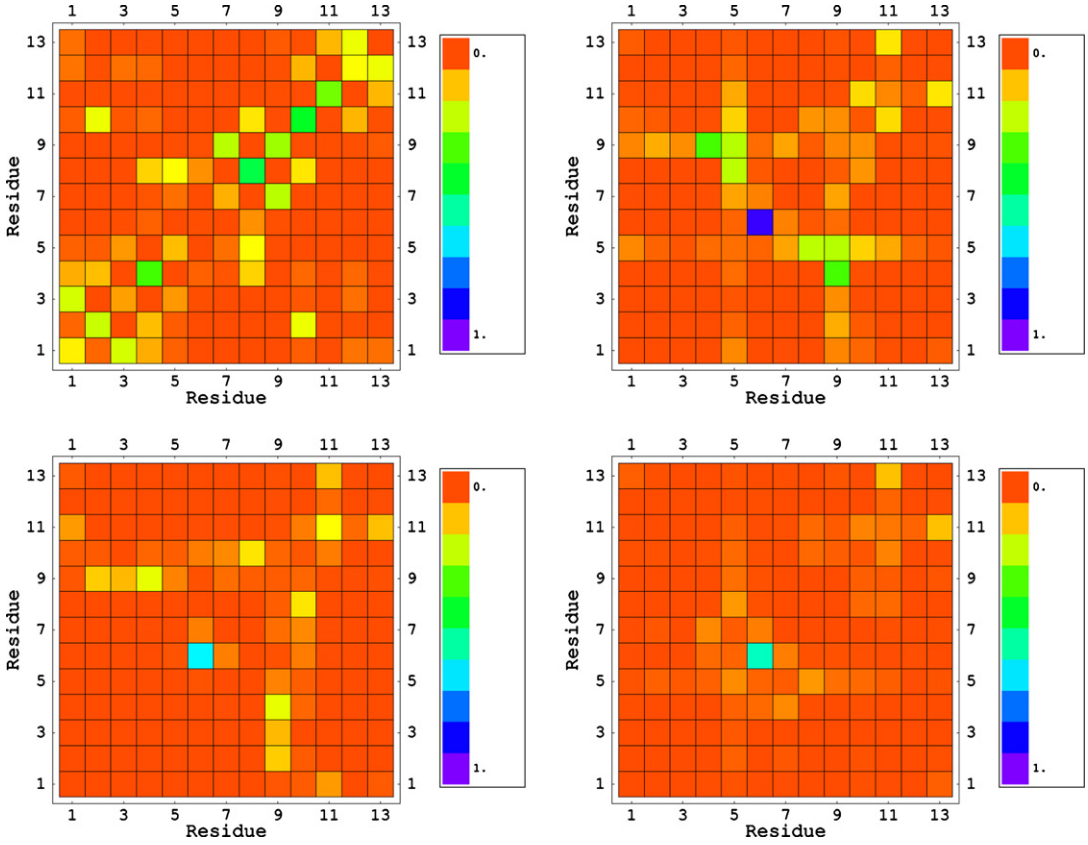


Fig. 3. Probability maps for the formation of backbone–backbone or backbone–sidechain hydrogen bonds between residue pairs, computed from the MD trajectories. Upper panel: NAT backbone–backbone map (left) and NAT backbone–sidechain map (right); lower panel: G5 backbone–sidechain map (left) and W4A9 backbone–sidechain map (right). The formation of hydrogen bonds follows the default CHARMM hydrogen bond criteria (an acceptor–donor distance $\leq$ 3.4 Å, without restrictions on the hydrogen bond angle).

peptide compensates for the reduced intramolecular hydrogen-bonding interactions compared to NAT and G5, by forming a larger number of hydrogen bonds with solvent.

In all three analogues, there is a very small probability for the formation of hydrogen bonds between sidechains (not shown). The most important hydrogen bond of this kind is between Gln5 and His10 in the NAT simulations (with an occupancy of 14%).

4. Discussion and conclusions

In this work, we have studied the dynamical behavior of native compstatin and two mutant analogues in aqueous solution, by multi-ns MD simulations. The simulations show that the native (NAT) and the inactive analogue (G5) have a strong tendency for the formation of a 5–8 β-turn, whereas in the active, double-mutant W4A9 this tendency is reduced. These results are in agreement with the earlier analysis of NMR studies [5,7,8,13]. This behavior suggests that the 5–8 β-turn frequency does not correlate with activity, at least for the analogues studied here. A similar proposition that the 5–8 β-turn is necessary but not sufficient condition for activity was made using NMR data of the inactive analogues with V3A and W7F single mutations, which maintain 5–8 β-turns [7]. In the current results, the case of G5 shows that the formation of the 5–8 β-turn does not constitute a sufficient condition for activity.

Residues Val3, Gln5, Asp6, Trp7 and Gly8 are critical for compstatin activity [7,10,11,13], even though mutations at some of these positions are tolerated when they preserve the native chemical properties (V3L, Q5N) [7]. Our results show that Val3 makes hydrophobic contacts with Val4 and hydrogen-bonding interactions with His9 in the NAT and G5 simulations. These interactions are reduced in the more active W4A9 analogue. It is possible that increased *intermolecular* Val3-C3 contacts in the W4A9:C3 complex (e.g., via a hydrophobic cluster involving Val3 [7]) contribute to the W4A9 higher activity.

Residue Gln5 is important for compstatin activity. A conservative mutation (Q5N) is tolerated at this position [6], but elimination of the sidechain (as in G5) [5] or substitution by a methyl group (Q5A) [5] renders the peptide inactive. In G5 the Gln5 sidechain is missing and the 5–8 conformations are somewhat tighter compared to the other two systems. In the NAT simulations, the Gln5 sidechain interacts with various backbone and sidechain groups and occasionally interferes with the formation of the 5–8 β-turn, as discussed above. It is possible that Gln5 affects activity via specific intramolecular interactions, which modulate the peptide flexibility. Gln5 may also form specific intermolecular contacts with C3.

Earlier simulations [9,12,13] with a polar-hydrogen representation for the peptides [14] and a compatible generalized-Born model [15] for the solvent effects suggested that the Asp6

and Arg11 sidechains formed a persistent electrostatic interaction in several analogues, and the Trp4 and Trp7 sidechains interacted extensively in the double-mutant W4A9. In the current, all-atom/explicit-solvent simulations these interactions are not seen. Asp6 forms very brief and infrequent encounters with Arg11 (the average 6–11 sidechain distance is 17.6–18.7 Å in the three analogues), and hydrogen-bonds mostly to its own backbone (Fig. 3). The Trp7 sidechain does not form any significant interactions with other residues. The average Trp4-Trp7 distance is 11.4 ± 2.7 Å and the two sidechains are in contact for $\approx 7\%$ of the time; the Trp4-Arg11 sidechain distance is 16.3 ± 3.4 Å. Thus, the current simulations do not provide evidence that the increased W4A9 activity is due to Trp4-Trp7 aromatic interactions, or Arg11-Trp4 cation- π interactions. It has been suggested that Trp4 contributes to the functional hydrophobic cluster of compstatin [13], which was previously proposed to participate in binding with C3 [7,11,32]. It is probable that Trp4 increases activity by forming hydrophobic [13], cation-π or π–π [12,13,32] intermolecular interactions with C3. A recent experimental study [33] also suggests that tryptophane residues at positions 4 and 7 contribute to activity by forming, respectively, hydrophobic and hydrogen-bonding interactions with C3. The existence of a hydrogen bond between the Trp7 indole amide and an accepting group in C3 was also suggested in Ref. [6]. Additionally, the increased W4A9 flexibility (discussed in the ''cluster analysis'') could contribute to its higher affinity, by facilitating the structural reorganization of W4A9 in the C3 complex at a lower energetic cost, compared to the other two ligands.

As discussed above, the behavior of compstatin varies somewhat between the present and the earlier GB MD studies [9,12,13]. These differences may originate from the peptide force fields (all-atom here [16] versus polar-hydrogen [14] in Refs. [9,12,13]), the presence or absence of N-terminal acetylation (absent in Ref. [9]), incomplete conformational sampling and force-field limitations. Differences between explicit and implicit solvent treatments have been observed with various GB implementations [34,35] and are addressed in more recent GB implementations with improved balance of solvation and intramolecular interactions [36,37].

Despite the success in reproducing the 5–8 β-turn population estimated by NMR, the current explicit-solvent simulations could underestimate the formation of more extended β-hairpin and α-helical elements due to competition between peptide–peptide and peptide-water hydrogen bonds or due a possible overestimation of the structural fluctuations by the CHARMM22 force-field. In particular, recent studies have shown that a grid-based (CMAP) backbone ϕ/ψ energy correction [38–40] of the CHARMM22 (C22) force field [16] yields structural and dynamical properties in better agreement with experiment. To check the impact of the CMAP correction on compstatin, we performed additional 2-ns simulations of the NAT peptide with the C22/CMAP energy [39], starting from all NMR structures. With the CMAP correction, the dominant conformations corresponded again to a coil with a 5–8 β-turn; i.e., no additional α-helical or β-sheet conformations were stabilized. The total 5–8 β-turn population was 80.7% (NAT^-) to 72.6% (NAT^+) with the C22/CMAP energy, in close agrement with the C22/no-CMAP result (82.5–72.2% in Table 3). However, the α_R–α_R turn population increased to 66.8% (NAT^-) and 55.3% (NAT^+) (compared to 36.6–33.5% in Table 3), at the expense of other turn types (category ''other'' in Table 3). This could be due to the broad plateau in the β-conformer ($\phi = -90°$, $\psi = 0°$) region of the C22/CMAP ϕ/ψ energy map [39], which is likely to stabilize the type-I population (Table 2) of the α_R–α_R turns. The CMAP correction caused also the stabilization of a small 3_{10} helix population (17.6%, compared to $\approx 0\%$ without the CMAP), at the expense of the coil conformation (10.0% with CMAP, compared to 19.2–27.4% without). A similar behavior characterized the segment 6–9. In the C22/CMAP simulations, the total 6–9 β-turn frequency was 38.1–45.5%, compared to 37.4–18.1% without the CMAP. The difference was mainly due to the stabilization of type-I (α_R–α_R) turns. A 17.7% of conformations corresponded to a 3_{10} helix (compared to $\approx 0\%$). The CMAP correction did not modify significantly the structural properties of the peptide region, outside the segment 5–9.

In addition to the CMAP correction, the C22/CMAP simulations employed a set of recent, slightly revised indole parameters [41]. These parameters did not have a significant impact on the properties of the Trp7 sidechain, which interacted mostly with the aqueous solvent. In the future, we plan to simulate several compstatin analogs with the CMAP [38,39] and indole [41] modifications of the all-atom C22 force field, and perform a systematic comparison with the results of a recent CMAP/implicit solvent (GB) force field [36,37].

In conclusion, in this work we have performed for the first time a conformational analysis of the peptide compstatin and two of its mutants by MD simulations in explicit water with an all-atom model. The simulations reproduce important structural properties established by NMR [5,7,8,13], such as the relative tendency of the various analogues for the formation of the 5–8 β-turn, and their extended conformational averaging in water. They also depict in detail structural features partly or not accessible by the NMR data due to conformational averaging, such as the occurence of fused or transient β-turns and the interactions of various sidechains with the peptide backbone. The analysis reveals structural differences among the analogues in their uncomplexed state, which enable us to draw conclusions on the role of the 5–8 β-turn and other aminoacids for the peptide activity.

Acknowledgements

This work was partly funded by the Research Infrastructure ERYAN/0603 grant ''Study of Compstatin, important inhibitor of the autoimmune system, with high-accuracy molecular simulations'' (to GA and SS) from the Cyprus Research Promotion Foundation (RPF), an RPF PROFIT prize grant (to PT) and NIH grants GM069736 and GM62314 (to JDL and DM). We thank the anonymous referees for their instructive comments. All simulations were performed on Linux clusters of the Biophysics group at the University of Cyprus.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2007.03.014.

References

[1] H.J. Muller-Eberhard, Complement-chemistry and pathways, in: J.I. Gallin, I.M. Goldstein, R. Synderman (Eds.), Inflammation: Basic Principles and Clinical Correlates, 2nd ed., Raven Press, New York, 1992.

[2] A. Sahu, J.D. Lambris, Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics, Immunopharmacology 49 (2000) 133–148.

[3] A.M. Soulika, M.C.H. Holland, G. Sfyroera, A. Sahu, J.D. Lambris, Compstatin inhibits complement activation by binding to the β-chain of complement factor 3, Mol. Immunol. 43 (2006) 2023–2029.

[4] D. Morikis, J.D. Lambris, Structure, dynamics, activity and function of compstatin and design of more potent analogs, in: D. Morikis, J.D. Lambris (Eds.), Structural Biology of the Complement System, CRC Press/Taylor & Francis Group, Boca Raton, FL, 2005.

[5] D. Morikis, N. Assa-Munt, A. Sahu, J.D. Lambris, Solution structure of compstatin, a potent complement inhibitor, Protein Sci. 7 (1998) 619–627.

[6] A. Sahu, A. Soulika, D. Morikis, L. Spruce, W.T. Moore, J.D. Lambris, Binding kinetics, structure–activity relationship, and biotransformation of the complement inhibitor compstatin, J. Immunol. 165 (2000) 2491–2499.

[7] D. Morikis, M. Roy, A. Sahu, A. Troganis, P.A. Jennings, G.C. Tsokos, J.D. Lambris, The structural basis of compstatin activity examined by structure-function-based design of peptide analogs and NMR, J. Biol. Chem. 277 (2002) 14942–14953.

[8] A.M. Soulika, D. Morikis, M.R. Sarrias, M. Roy, L.A. Spruce, A. Sahu, J.D. Lambris, Studies of structure–activity relations of complement inhibitor compstatin, J. Immunol. 170 (2003) 1881–1890 (erratum: 172 (2004) 5128).

[9] B. Mallik, J.D. Lambris, D. Morikis, Conformational interconversion in compstatin probed with molecular dynamics simulations, Proteins: Struct. Func. Bioinf. 52 (2003) 130–141.

[10] J. Klepeis, C.A. Floudas, D. Morikis, C.G. Tsokos, E. Argyropoulos, L. Spruce, J.D. Lambris, Integrated computational and experimental approach for lead optimization and design of compstatin variants with improved activity, J. Am. Chem. Soc. 125 (2003) 8422–8423.

[11] D. Morikis, A.M. Soulika, B. Mallik, J.L. Klepeis, C.A. Floudas, J.D. Lambris, Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches, Biochem. Soc. Trans. 32 (2004) 28–32.

[12] B. Mallik, D. Morikis, Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues, J. Am. Chem. Soc. 127 (2005) 10967–10976.

[13] B. Mallik, M. Katragadda, L.A. Spruce, C. Carafides, C.G. Tsokos, D. Morikis, J.D. Lambris, Design and NMR characterization of active analogues of compstatin containing non-natural amino acids, J. Med. Chem. 48 (2005) 274–286.

[14] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus, CHARMM: a program for macromolecular energy, minimization, and dynamics calculations, J. Comput. Chem. 4 (1983) 187–217.

[15] B. Dominy, C.L. Brooks III, Development of a generalized Born model parameterization for proteins and nucleic acids, J. Phys. Chem. B 103 (1999) 3765–3773.

[16] A.D. Mackerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkewicz-Kuczera, D. Yin, M. Karplus III, An all-atom empirical potential for molecular modelling and dynamics study of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.

[17] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.

[18] E. Neria, S. Fischer, M. Karplus, Simulation of activation free energies in molecular systems, J. Chem. Phys. 105 (1996) 1902–1921.

[19] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: an n log(n) method for Ewald sums in large systems, J. Chem. Phys. 98 (1993) 10089–10092.

[20] S. Nose, A unified formulation of the constant temperature molecular dynamics method, J. Chem. Phys. 81 (1984) 511–519.

[21] W. Hoover, Canonical dynamics: equilibrium phase-space distributions, Phys. Rev. A 31 (1985) 1695–1697.

[22] S. Feller, Y. Zhang, R.W. Pastor, B. Brooks, Constant-pressure molecular-dynamics simulation: the Langevin piston method, J. Chem. Phys. 103 (1995) 4613–4621.

[23] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, J. Comput. Phys. 23 (1977) 327–341.

[24] M.E. Karpen, D.J. Tobias, C.L. Brooks III, Statistical clustering techniques for analysis of long molecular dynamics trajectories. I. Analysis of 2.2 ns trajectories of YPGDV, Biochemistry 32 (1993) 412–420.

[25] G.A. Carpenter, S. Grossberg, ART-2: self-organization of stable category recognition codes for analog input patterns, Appl. Opt. 26 (1987) 4919–4930.

[26] Y.-H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison-Wesley, New York, 1989.

[27] J.L. Klepeis, C.A. Floudas, D. Morikis, C.G. Tsokos, J.D. Lambris, Design of peptide analogs with improved activity using a de novo protein design approach, Ind. Eng. Chem. Res. 43 (2004) 3817–3826.

[28] C.M. Wilmot, J.M. Thornton, Beta-turns and their distortions: a proposed new nomenclature, Protein Eng. 3 (1990) 479–493.

[29] T.C. Cheam, S. Krimm, Ab initio force fields of glycine dipeptide in c5 and c7 conformations, J. Mol. Struct. 193 (1989) 1–34.

[30] C.M. Venkatachalam, Stereochemical criteria for polypeptides and proteins. V. Conformation of a system of three linked peptide units, Biopolymers 6 (1968) 1425–1436.

[31] P.N. Lewis, F.A. Momany, H.A. Scheraga, Chain reversals in proteins, Biochim. Biophys. Acta 303 (1973) 211–229.

[32] D. Morikis, J.D. Lambris, Structural aspects and design of low-molecular-mass complement inhibitors, Biochem. Soc. Trans. 30 (2002) 1026–1036.

[33] M. Katragadda, P. Magotti, G. Sfyroera, J.D. Lambris, Hydrophobic effect and hydrogen bonds account for the improved activity of a complement inhibitor, compstatin, J. Med. Chem. 49 (2006) 4616–4622.

[34] R. Zhou, B. Berne, Can a continuum solvent model reproduce the free-energy landscape of a β-hairpin folding in water? Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12777–12782.

[35] H. Nymeyer, A. Garcia, Simulation of the folding equilibrium of α-helical peptides: a comparison of the generalized Born approximation with explicit solvent, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 13934–13939.

[36] J. Chen, W. Im, C.L. Brooks III, Balancing solvation and intramolecular interactions: toward a consistent generalized Born force field, J. Am. Chem. Soc. 128 (2006) 3728–3736.

[37] W. Im, J. Chen, C.L. Brooks III, Peptide and protein folding and conformational equilibria: theoretical treatment of electrostatics and hydrogen bonding with implicit solvent models, Adv. Prot. Chem. 72 (2006) 171–195.

[38] M. Feig, A.D. Mackerell, C.L. Brooks III, Force field influence on the observation of π-helical protein structures in molecular dynamics simulations, J. Phys. Chem. B 107 (2003) 2831–2836.

[39] A.D. Mackerell, M. Feig, C.L. Brooks III, Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations, J. Comp. Chem. 25 (2004) 1400–1415.

[40] M. Buck, S. Bouguet-Bonnet, R.W. Pastor, A.D. Mackerell Jr., Importance of the CMAP correction to the CHARMM22 protein force field: dynamics of hen lysozyme, Biophys. J. (2006) L37–L39.

[41] A.T. Macias, A.D. Mackerell Jr., CH/pi interactions involving aromatic amino acids: refinement of the CHARMM tryptophan force field, J. Comp. Chem. 26 (2005) 1452–1463.

Conformational Analysis of Compstatin Analogues with Molecular Dynamics Simulations in Explicit Water

Phanourios Tamamis, Spiros S. Skourtis, Dimitrios Morikis, John D. Lambris and Georgios Archontis

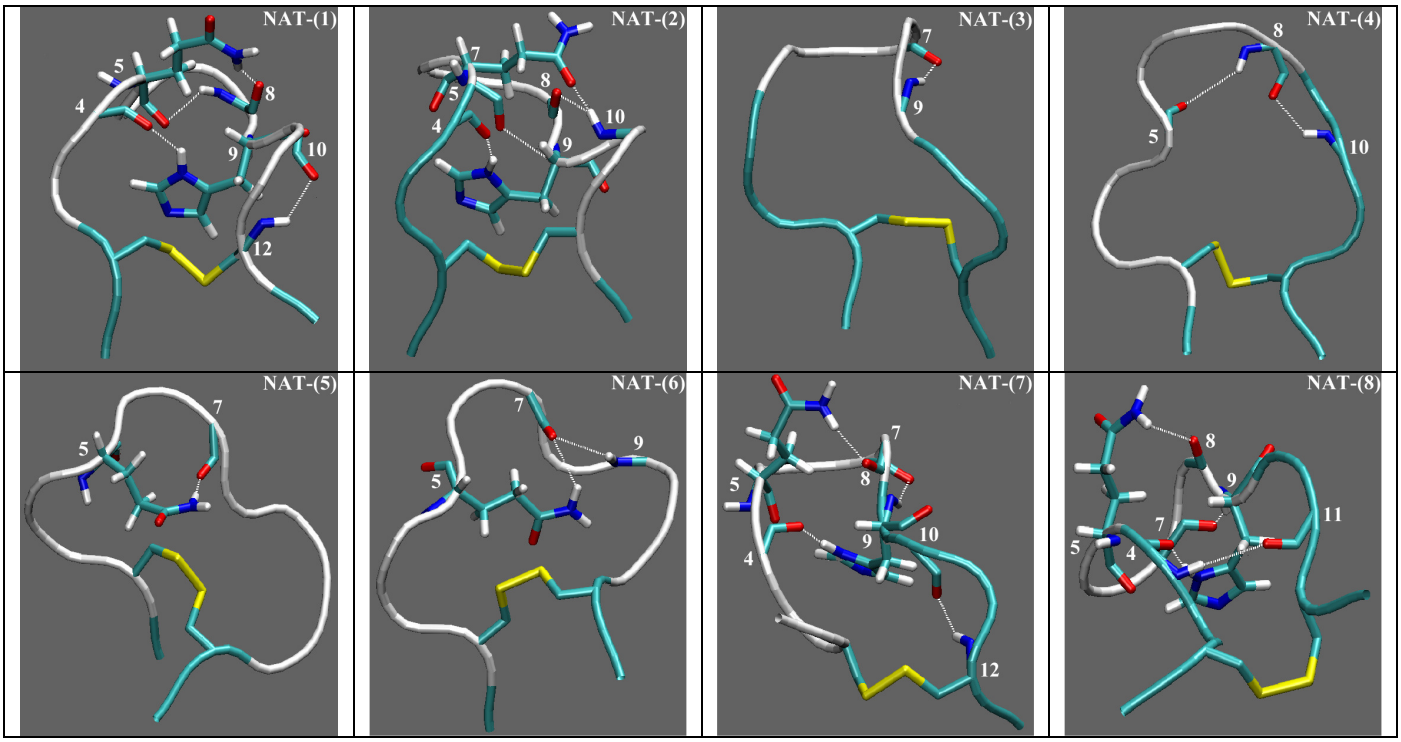
Supplementary Material

The MD trajectories of the three analogues were partitioned into clusters, based on the Cartesian coordinates of the C_α atoms in the region 2-12. The clustering methodology and the properties of selected clusters have been discussed, respectively, in the "Methods" and "Results" sections of the main article. In this file we show the average structures (fig. S1) and characteristic properties (table S1) for the eight ($NAT^{\pm}$) or seven ($G5^{\pm}$, $W4A9^{\pm}$) most populated clusters.

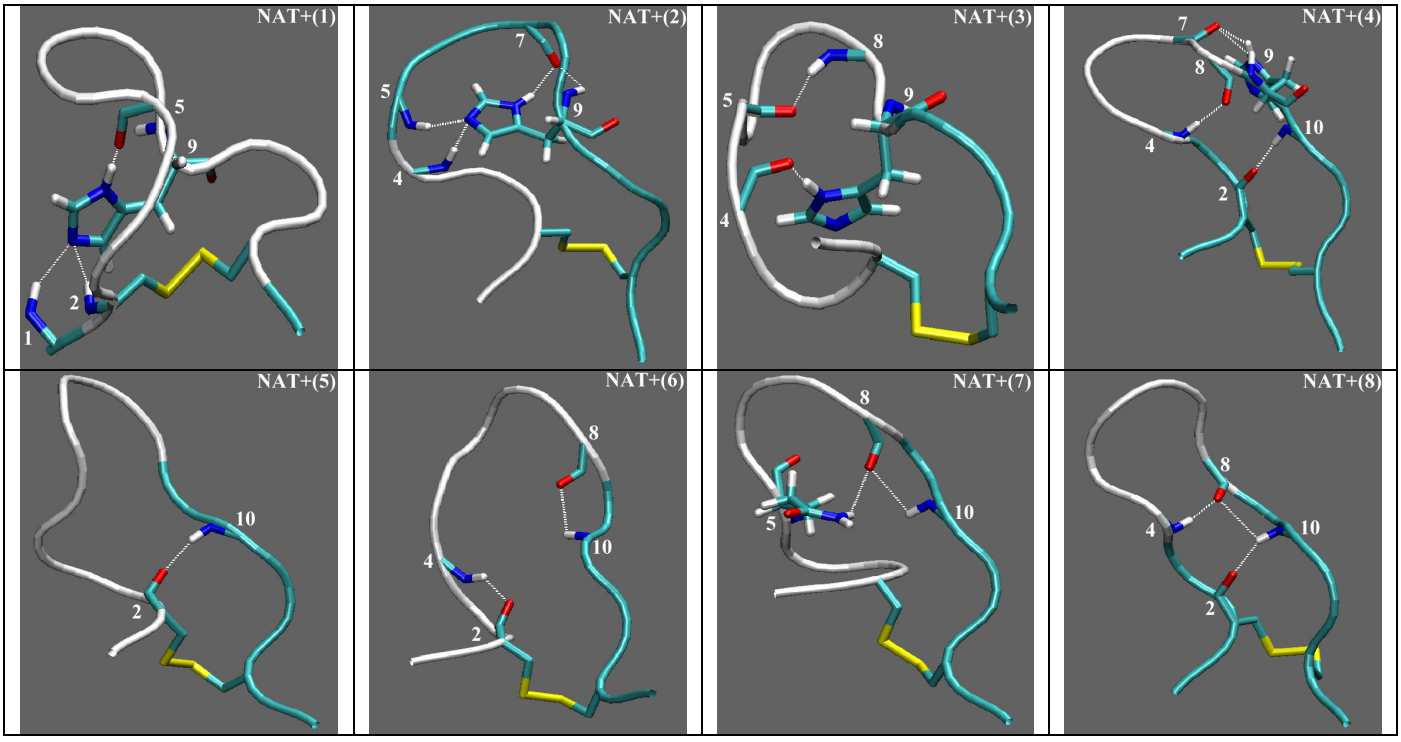
Fig. S1. Average structures of the most populated clusters resulting from the MD trajectories of the various analogues. The cluster analysis has employed the C_α atomic coordinates of the 2-12 region (see the "Methods" section in the main article). $X^{\pm}(j)$ denotes the j-th cluster in the ±90° Cys2-Cys12 disulfide dihedral angle simulation of the X analogue. The white, green and yellow colors indicate, respectively, β-turns, random coil elements and the Cys2-Cys12 disulfide bridge. A β-turn is considered as characteristic of the cluster and is shown (in white) if the distance $C_\alpha^{i+3} - C_\alpha^{i} < 7$ Å for at least 15% of the cluster frames. In all panels the N-terminal (Ile1) and C-terminal (Thr13) residues are, respectively, in the lower left and lower right. Apart from the average 2-12 backbone structure, additional structural information is included, when it is characteristic of the corresponding cluster (high-occupancy hydrogen bonds, sidechains participating in high-probability interactions).

(over)

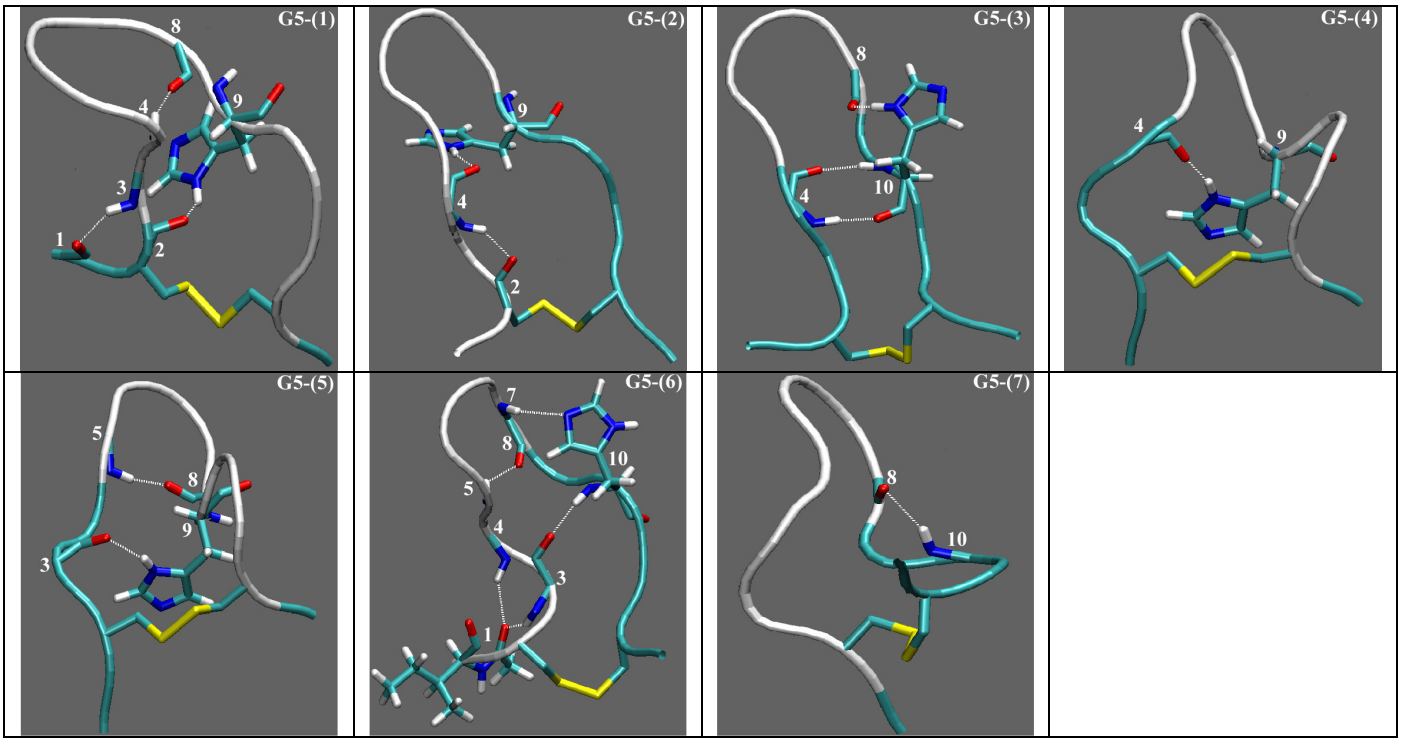
NAT- Clusters



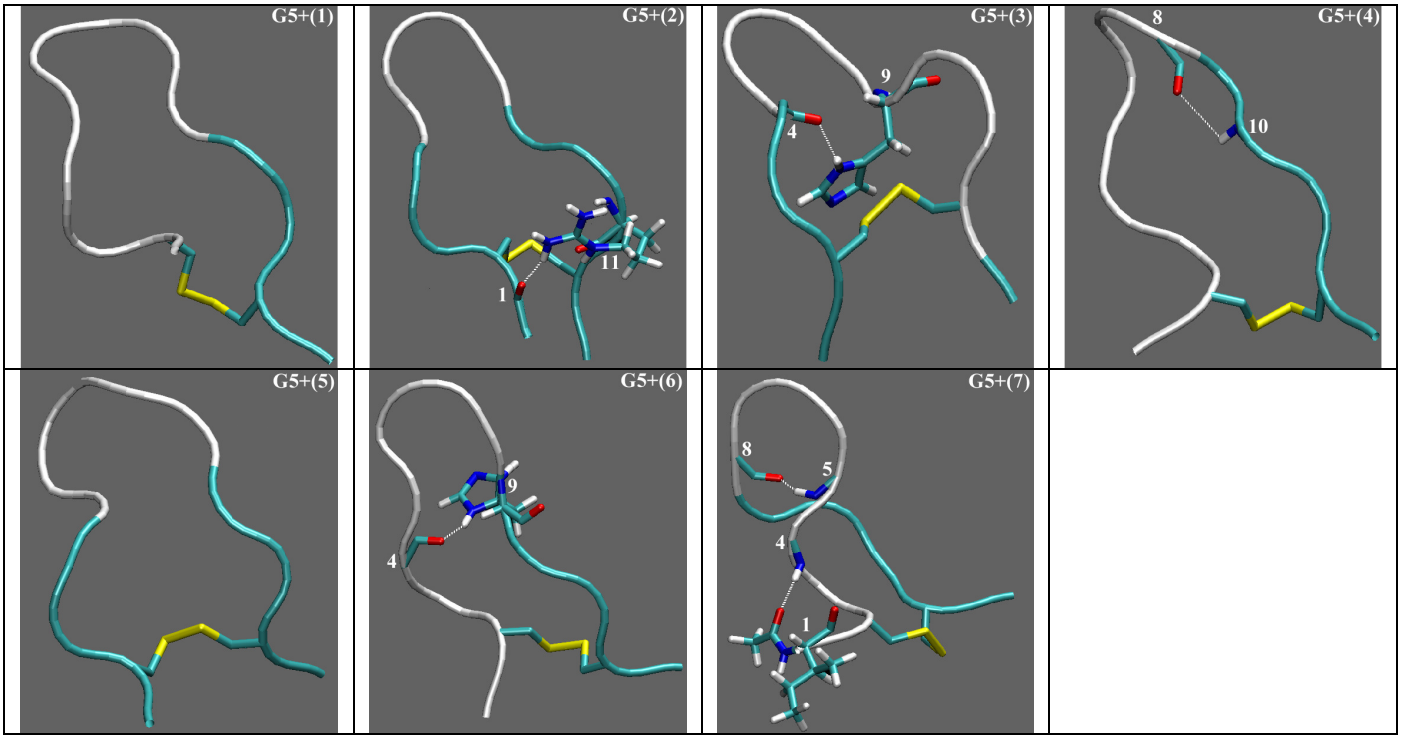
NAT+ Clusters



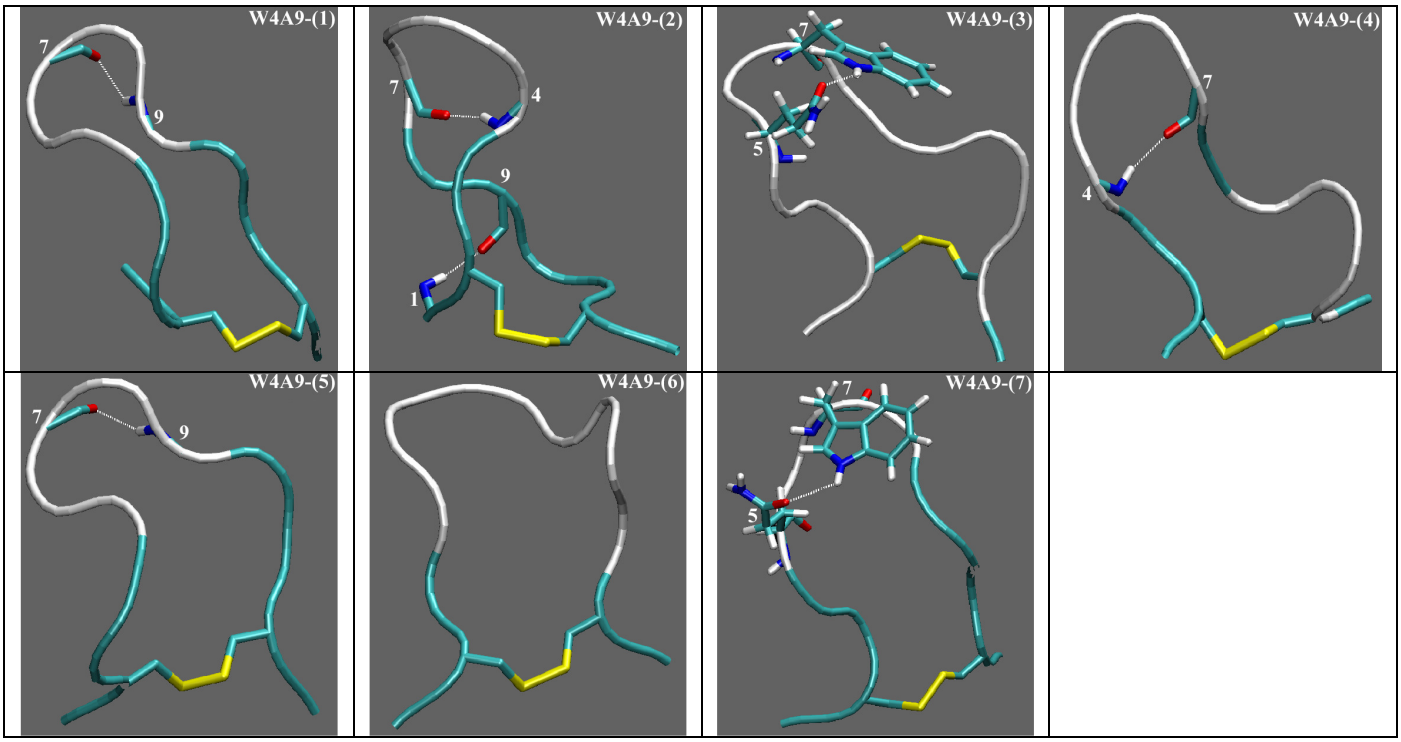
G5- Clusters



G5+ Clusters



W4A9- Clusters



W4A9+ Clusters

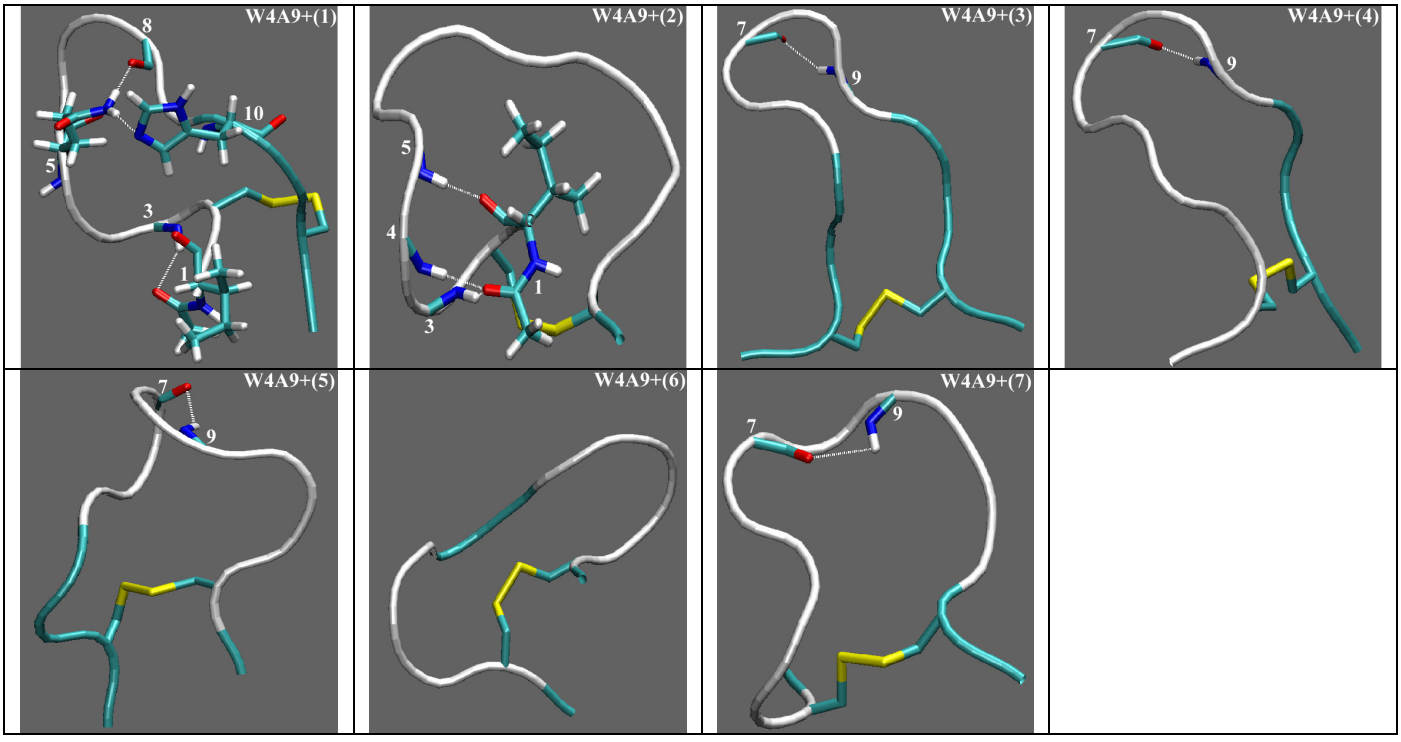


Table S1. Frequencies and structural properties of the most populated clusters, resulting from the trajectory analysis of C_α atomic coordinates in the region 2-12. $X^{+/-}(j)$ denotes the *j*-th cluster in a simulation of the X analogue, with the Cys2-Cys12 disulfide dihedral fluctuating around +/-90°.

(over)

Clusters G5-	G5-(1)	G5-(2)	G5-(3)	G5-(4)	G5-(5)	G5-(6)	G5-(7)		Total
Percentage[a]	12.8	9.2	8.5	9.1	10.8	11.5	11.7		73.6
Turns[b]	3-6 (74.1) 5-8 (24.3) 9-12 (80.6)	1-4 (45.4) 5-8 (93.9)	5-8 (99.9)	5-8 (97.9) 9-12 (99.7)	5-8 (96.4) 9-12 (100.0)	1-4 (99.3) 5-8 (99.4)	2-5 (28.1) 5-8 (98.5)		
5-8 a_R-a_R[c]	0.0	26.0	21.5	45.2	35.7	57.9	31.7		
β-sheet / bridge[d]	–	–	4-10, or 3-11 and 4-10	–	4-9**	4-9**	–		
Hydrogen Bonds[e]	Ile1 O – Val3 HN Cys2 O – Hsd9 HD1 Val4 HN – Gly8 O	Cys2 O – Val4 HN Val4 O – Hsd9 HD1	Val4 HN – Hsd10 O Val4 O – Hsd10 HN Gly8 O – Hsd10 HD1	Val4 O – Hsd9 HD1	Val3 O – Hsd9 HD1 Gly5 HN – Gly8 O	Ile1 OY- Val3 HN Ile1 OY – Val4 HN Gly5 HN – Gly8 O Trp7 HN – Hsd10 NE2	Gly8 O – Hsd10 HN*		
Clusters G5+	G5+(1)	G5+(2)	G5+(3)	G5+(4)	G5+(5)	G5+(6)	G5+(7)		Total
Percentage[a]	17.0	15.3	9.5	8.4	10.3	13.7	7.5		81.7
Turns[b]	1-4 (57.1) 5-8 (95.6) 6-9 (57.0)	5-8 (99.8)	5-8 (100.0) 9-12 (73.7)	1-4 (35.9) 5-8 (46.9)	5-8 (96.1) 6-9 (17.8)	1-4 (40.2) 5-8 (99.8)	1-4 (95.2) 5-8 (100.0)		
5-8 a_R-a_R[c]	32.9	49.7	53.1	36.4	23.6	40.4	57.9		
β-sheet / bridge[d]	–	–	–	–	–	–	4-9**		
Hydrogen Bonds[e]	–	Ile1 O – Arg11 HH11**	Val4 O – Hsd9 HD1	Gly8 O – Hsd10 HN*	–	Val4 O – Hsd9 HD1	Ile1 OY - Val4 HN** Gly5 HN – Gly5 O **		
Clusters NAT-	NAT-(1)	NAT-(2)	NAT-(3)	NAT-(4)	NAT-(5)	NAT-(6)	NAT+(7)	NAT-(8)	Total
Percentage[a]	12.1	16.0	8.9	6.6	4.9	4.2	8.0	7.9	68.6
Turns[b]	3-6 (21.1) 5-8 (98.9) 9-12 (70.3)	5-8 (87.3) 9-12 (99.3)	5-8 (98.4) 6-9 (79.9)	2-5 (44.3) 5-8 (91.7)	2-5 (96.9) 5-8 (17.9) 8-11 (42.3)	2-5 (53.0) 5-8 (32.8) 8-11 (35.9)	1-4 (58.4) 5-8 (95.6)	5-8 (99.7) 6-9 (80.7)	
5-8 a_R-a_R[c]	92.7	4.6	15.5	75.0	0.3	0.1	15.8	21.2	
β-sheet / bridge[d]	–	–	–	–	–	–	–	either 4-10 or 4-11	
Hydrogen Bonds[e]	Val4 O – Hsd9 HD1 Gln5 O – Gly8 HN* Gln5 HE22 – Gly8 O Hsd10 O – Cys12 HN*	Val4 O – Hsd9 HD1 Gln5 OE1 – Hsd10 HN Trp7 O – Hsd9 HN** Gly8 O – Hsd10 HN	Trp7 O – Hsd9 HN	Gln5 O – Gly8 HN* Gly8 O – Hsd10 HN*	Gln5 HE22 – Trp7 O	Gln5 HE22 – Trp7 O Trp7 O – Hsd9 HN**	Val4 O- Hsd9 HD1 Gln5 HE22 – Gly8 O Trp7 O – Hsd9 HN Hsd10 O – Cys12 HN	Val4 HN – Arg11 O Val4 O – Hsd9 HD1* Gln5 HE22 – Gly8 O* Trp7 O – Hsd9 HN	
Clusters NAT+	NAT+(1)	NAT+(2)	NAT+(3)	NAT+(4)	NAT+(5)	NAT+(6)	NAT+(7)	NAT+(8)	Total
Percentage[a]	7.6	7.7	24.0	7.8	7.8	5.1	10.8	6.3	77.1
Turns[b]	2-5 (54.1) 5-8 (90.1) 6-9 (89.5) 9-12 (92.0)	1-4 (53.9)	1-4 (64.3) 5-8 (92.1) 6-9 (56.9)	4-7 (59.8) 5-8 (90.6)	1-4 (51.0) 5-8 (96.8)	1-4 (57.7) 5-8 (84.3)	1-4 (74.2) 5-8 (86.2)	4-7 (97.7) 5-8 (91.2)	
5-8 a_R-a_R[c]	52.0	0.1	70.0	9..6	16.7	33.2	39.7	15.5	
β-sheet / bridge[d]	–	–	–	–	–	–	–	either 3,4-9,8 or 3-10	
Hydrogen Bonds[e]	Ile1 HN – Hsd9 NE2 Cys2 HN – Hsd9 NE2 Gln5 O – Hsd9 HD1	Val4 HN – Hsd9 NE2 Gln5 HN – Hsd9 NE2 Trp7 O – Hsd9 HN Trp7 O - Hsd9 HD1	Val4 O – Hsd9 HD1 Gln5 O – Gly8 HN	Cys2 O – Hsd10 HN* Val4 HN – Gly8 O Trp7 O – Hsd9 HN Trp7 O - Hsd9 HD1	Cys2 O – Hsd10 HN	Cys2 O – Val4 HN Gly8 O – Hsd10 HN**	Gln5 HE22 – Gly8 O Gly8 O – Hsd10 HN**	Cys2 O – Hsd10 HN Val4 HN – Gly8 O Gly8 O – Hsd10 HN	
Clusters W4A9-	W4A9-(1)	W4A9-(2)	W4A9-(3)	W4A9-(4)	W4A9-(5)	W4A9-(6)	W4A9-(7)		Total
Percentage[a]	5.0	4.2	5.5	5.2	4.5	5.2	4.4		34.0
Turns[b]	5-8 (99.2) 6-9 (100.0)	4-7 (78.8)	1-4 (44.4) 5-8 (98.6) 9-12 (27.3)	4-7 (99.6) 9-12 (99.7)	5-8 (92.1) 6-9 (85.1)	4-7 (38.5) 8-11 (55.6)	5-8 (68.3)		
5-8 a_R-a_R[c]	5.6	0.0	88.0	0.0	39.1	0.0	49.5		
β-sheet / bridge[d]	4-10**	2,3-9,8**	–	–	–	–	4-10**		
Hydrogen Bonds[e]	Trp7 O – Hsd9 HN	Ile1 HN – Hsd9 HN** Val4 HN – Trp7 O**	Gln5 OE1 – Trp7 HE1	Val4 HN – Trp7 O	Trp7 O – Hsd9 HN**	–	Gln5 OE1 – Trp7 HE1*		
Clusters W4A9+	W4A9+(1)	W4A9+(2)	W4A9+(3)	W4A9+(4)	W4A9+(5)	W4A9+(6)	W4A9+(7)		Total
Percentage[a]	7.3	12.9	13.9	6.3	5.6	12.2	12.0		70.2
Turns[b]	1-4 (60.5) 5-8 (88.6) 6-9 (42.1)	1-4 (100.0) 2-5 (69.8) 5-8 (17.3) 8-11 (49.8) 9-12 (34.9)	5-8 (96.8) 6-9 (96.0)	1-4 (28.8) 5-8 (94.0) 6-9 (90.6)	5-8 (96.0) 9-12 (52.1)	2-5 (82.2) 8-11 (99.9) 9-12 (69.8)	2-5 (99.4) 5-8 (45.9) 8-11 (57.9)		
5-8 a_R-a_R[c]	76.7	0.6	22.5	25.0	30.4	0.0	5.6		
β-sheet / bridge[d]	–	–	–	–	–	–	–		
Hydrogen Bonds[e]	Ile1 OY – Val3 HN* Gln5 HE21 - Hsd10 NE2 Gln5 HE22 – Gly8 O	Ile1 O – Gln5 HN Ile1 OY – Trp4 HN*	Trp7 O – Hsd9 HN**	Trp7 O – Hsd9 HN	Trp7 O – Hsd9 HN	–	Trp7 O – Hsd9 HN		

[a] Fraction of snapshots from all "-90°" or "+90°" trajectories of the specified analogue X, which belong to the corresponding cluster.

[b] We show in parentheses the fraction of conformations *in each cluster* which contain the specified β-turn (e.g, 74.1% of the conformations in cluster G5-(1) contain the turn 3-6). Only turns observed in at least 15% of the cluster conformations are shown.

[c] Fraction of conformations *in each cluster* which contain a type α_R-α_R β-turn.

[d] Unmarked β-sheets / bridges exist almost in all frames; β-sheets / bridges marked by two stars (**) exist in the majority of frames, but not in all.

[e] Unmarked hydrogen bonds exist almost in all frames; hydrogen bonds marked by one star (*) are weak (the distance between heavy atoms is usually between 3.5-4.0 Å); hydrogen bonds marked by two stars (**) exist in the majority of frames, but not in all.