Chem Biol Drug Des 2008; **72:** 249–256

Research Article

© 2008 The Authors
Journal compilation © 2008 Blackwell Munksgaard

doi: 10.1111/j.1747-0285.2008.00709.x

Development of a New Pharmacophore Model That Discriminates Active Compstatin Analogs

Ting-Lan Chiu[1,2], Chandrika Mulakala[1,2], John D. Lambris[3] and Yiannis N. Kaznessis[1,2,*]

[1]*Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, MN 55455, USA*
[2]*Digital Technology Center, University of Minnesota, Minneapolis, MN 55455, USA*
[3]*Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA*
**Corresponding author: Yiannis N. Kaznessis, yiannis@cems.umn.edu*

Compstatin and its active peptide analogs can potentially be used for therapeutic purposes because their binding to the third component of complement prohibits its conversion into the proteolytically activated form of the third component of complement, thus inhibiting complement cascades in all three complement pathways. Mallik and Morikis built three quasi-dynamic pharmacophore models for compstatin peptide analogs before, but only nine compstatin peptide analogs were incorporated in their study and the most active compstatin analog had only medium inhibitory activity. Since then, many more compstatin analogs have been synthesized and their inhibitory activities tested. Furthermore, the X-ray structure of AcCompNH2-V4W-H9A bound to the third component of complement has become available (PDB ID: 2QKI). In this paper, we utilized all the new information and built a new pharmacophore model using a distinct approach. Our model demonstrated good performance in a separate test set of 82 compstatin analogs: it accurately identified 70% of the analogs of medium or high inhibitory activities and misclassified only 8.5% of the analogs of low or no inhibitory activities. The results proved our pharmacophore model to be a filter of great sensitivity and specificity.

Key words: C3, compstatin, complement inhibitor, complement system, pharmacophore model

Abbreviations: C3: The third component of complement; C3b: The proteolytically activated form of C3; Ac: acetylated N terminus; CompNH2: Compstatin; 1MeW: 1-methyl-tryptophan; 5fW: 5-fluorotryptophan.

Received 4 April 2008, revised 22 August 2008 and accepted for publication 30 August 2008

The third component of complement (C3) is the key player in all three activation pathways of the complement system (1). Complement cascades launch drastically after C3 is converted into the proteolytically activated form of C3 (C3b) by a C3 convertase and C3b binds to the surface of a bacterium (1). Inappropriate activation of a complement cascade results in host cell damage, which has been present in pathological conditions such as autoimmune diseases, adult respiratory syndrome, stroke, heart attack, post-xenotransplant rejection, cardiopulmonary bypass, dialysis, and burn injuries (1,2). Therefore, C3 is a target protein of much therapeutic interest.

Using a phage-displayed random peptide library, Sahu *et al.* identified a complement inhibitor, compstatin, which is a 13-residue cyclic peptide with the sequence: ICVVQDWGHHRCT-NH_2 (abbreviated as CompNH2) (3). The two cysteine residues of compstatin, C2 and C12, formed a disulfide bond. Unlike other complement inhibitors, compstatin and its active peptide analogs are unique and powerful in that they bind to C3, prohibit different C3 convertases from converting C3 to C3b, and thus inhibit complement cascades in all three complement pathways.

Previously, Mallik and Morikis (4) built three quasi-dynamic pharmacophore models for compstatin analogs. Their models were based on only nine compstatin peptide analogs and the most active compstatin analog in their training set had only a 45-fold inhibitory activity when compared with compstatin. Since then, additional compstatin analogs have been synthesized and their inhibitory activities tested. Our understanding of the structure–activity relationships of compstatin analogs has also dramatically increased (5–13). In addition, potent compstatin analogs have been discovered: AcCompNH2-V4(1MeW)-H9A and AcCompNH2-V4(1MeW)-W7(5fW)-H9A, both of which had IC_{50} values of 0.2 μM (a 264-fold inhibitory activity when compared with compstatin) (11). Furthermore, the X-ray structure of AcCompNH2-V4W-H9A bound to C3c fragment of C3 has become available (PDB ID: 2QKI) (12). With the availability of all these new findings, the development of a new pharmacophore model is called for.

In this work, we utilized all these new findings and built a new pharmacophore model for compstatin analogs. Specifically, six structural features that play critical roles in determining inhibitory activities were identified based on the structure–activity relationships of nine compstatin analogs. In addition, the availability of the X-ray structure of AcCompNH2-V4W-H9A bound to C3 made it possible for us to predict the bound conformations of the other 91 compstatin analogs using state-of-the-art modeling tools. Subsequently, the predicted bound conformations of the two most active compstatin

analogs, AcCompNH2-V4(1MeW)-H9A and AcCompNH2-V4(1MeW)-W7(5fW)-H9A, were utilized to build all six structural features into the pharmacophore model.

The sensitivity and specificity of this pharmacophore model were tested. Using a separate test set of 82 compstatin analogs, we demonstrated that our pharmacophore model was able to identify 70% of the analogs of medium or high inhibitory activities and misclassify only 8.5% of the analogs of low or no inhibitory activities. The superior results proved our pharmacophore model to be a filter of great sensitivity and specificity.

Methods

In this paper, our goal is the development and evaluation of a pharmacophore model for compstatin analogs. The development of our pharmacophore model consists of three steps. First, we identified the structural features that play a critical role in inhibitory activity based on the structure–activity relationships of compstatin analogs. Second, we predicted the bound conformations of all other compstatin analogs using as the template the X-ray structure of AcCompNH2-V4W-H9A bound to C3 (PDB ID: 2QKI). Third, we built the pharmacophore model using the predicted bound conformations of the two most active compstatin analogs.

Identification of critical structural features

Based on the previous experimental data on compstatin analogs, the following structure–activity relationships have been derived. We discuss them here, summarizing the important literature findings and focusing our investigation to known important structural features. First, it was found previously that the following Ala scans, V3A, Q5A, D6A, W7A, and G8A, resulted in significant decreases in activities. This finding indicated that V3, Q5, D6, W7, and G8 were important residues for activities (7). Second, the substitution of Val4 with Trp increased activity to 12-fold (24.364 versus 2.961 relative affinity when compared to compstatin) (6). It was proposed that the aromatic side chain caused the activity increase. Third, the Q5N mutation also yielded a similar activity when compared with parent compstatin (2.857 versus 1) (8), indicating that the amide groups at the tail of both side chains were crucial for keeping activity. This has been confirmed by the modeled complex structures in which the CO and NH_2 of both side chains acted as hydrogen-bond acceptors and donors, respectively, to the surrounding C3 residues. In particular, the carbonyl oxygen on the parent compstatin Q5 side chain acted as a hydrogen-bond acceptor to the backbone amide nitrogen on C3 M457. In addition, the amide nitrogen on the Q5 side chain acted as a hydrogen-bond acceptor to the backbone carbonyl oxygen on C3 L455 as well as the [O^-] anion on the C3 D491 side chain. For the Q5N mutant, the amide nitrogen on the parent compstatin N5 side chain acted as a hydrogen-bond acceptor to the [O^-] anion on the C3 D491 side chain. In Figure 1, illustrations of the interactions between compstatin analogs and C3 are shown. In addition, the D6P mutant only carried 1/10 of the parent compstatin activity (0.122 versus 1) (9). Since both D6A and D6P mutations resulted in significant activity decreases, the carbonyl oxygen on the Asp side chain seemed to play an important role in retaining the activities. Indeed, the modeled complex structure showed that the carbonyl oxygen acted as a hydrogen-bond acceptor to the amide nitrogen of C3 R459 side chain. Moreover, the mutation of V4W/H9A to V4W/W7(1MeW)/H9A resulted in a significant decrease in activity. The IC_{50} value for V4W/H9A was 1.2 μM while the activity of V4W/W7(1MeW)/H9A was too low to be measured experimentally (11). It was proposed that the indole NH of W7 formed a hydrogen bond with the C3 residue and that mutating W7 to 1MeW disabled its hydrogen-bond donor capability completely. This was confirmed in modeled structures where the hydrogen bond was formed between the indole nitrogen of W7 and the backbone carbonyl oxygen of M457. Lastly, breaking the disulfide bond between C2 and C12 resulted in a significant decrease in the inhibitory activity (IC_{50} of Comp-linear: >600 μM), indicating the importance of disulfide bond in retaining activity. Indeed, the NMR studies conducted by Morikis *et al.* (8) showed that the disulfide bridge brings together the termini of compstatin, thus contributing to the formation of the hydrophobic patch, which in turn is necessary for recognition and binding to C3.

Structures of compstatin analogs

Natural amino acid compstatin peptide analogs

The bound conformation of AcCompNH2-V4W-H9A (PDB ID: 2QKI) (12) was used as the template to build the bound conformations of all other compstatin analogs. Using the Rotamer Explorer function inside of MOE,[a] each individual compstatin peptide analog was computationally mutated and the most favorable rotamer of each mutated residue was adopted. The most favorable rotamer was determined based on the lowest DeltaG.[a] DeltaG is the sum of the change in free energy between the states where the protein and rotamer are separated and when they form a complex, plus the torsional strain of the rotamer. Subsequently, each individual peptide analog and its surrounding C3 residues (within 5 Å of distance) were selected for energy minimizations. In other words, each individual peptide analog and its surrounding C3 residues (within 5 Å of distance) were allowed to move while the other residues were kept rigid in energy minimizations. The energies were minimized to a gradient of 0.001 using Amber99 force field and the generalized Born solvation model in MOE.

Compstatin peptide analogs with unnatural amino acid residues

The partial charges of the unnatural amino acid residues were derived separately using the RED program developed by Pigache *et al.* (14). The structures of the unnatural amino acid residues were subsequently optimized using Gaussian (Hartree-Fock) with the 6-31G* basis set, where the '*' implies the addition of polarization functions for heavy atoms to the basis set. The whole mutant peptides were constructed inside of MOE and the partial charges of unnatural amino acid residues were modified to those calculated by RED in the previous step. The peptide analog as well as the C3 residues within 5 Å of distance to the peptide analog were selected for energy minimizations. The energies were minimized to a gradient of 0.001 using Amber99 force field in MOE.

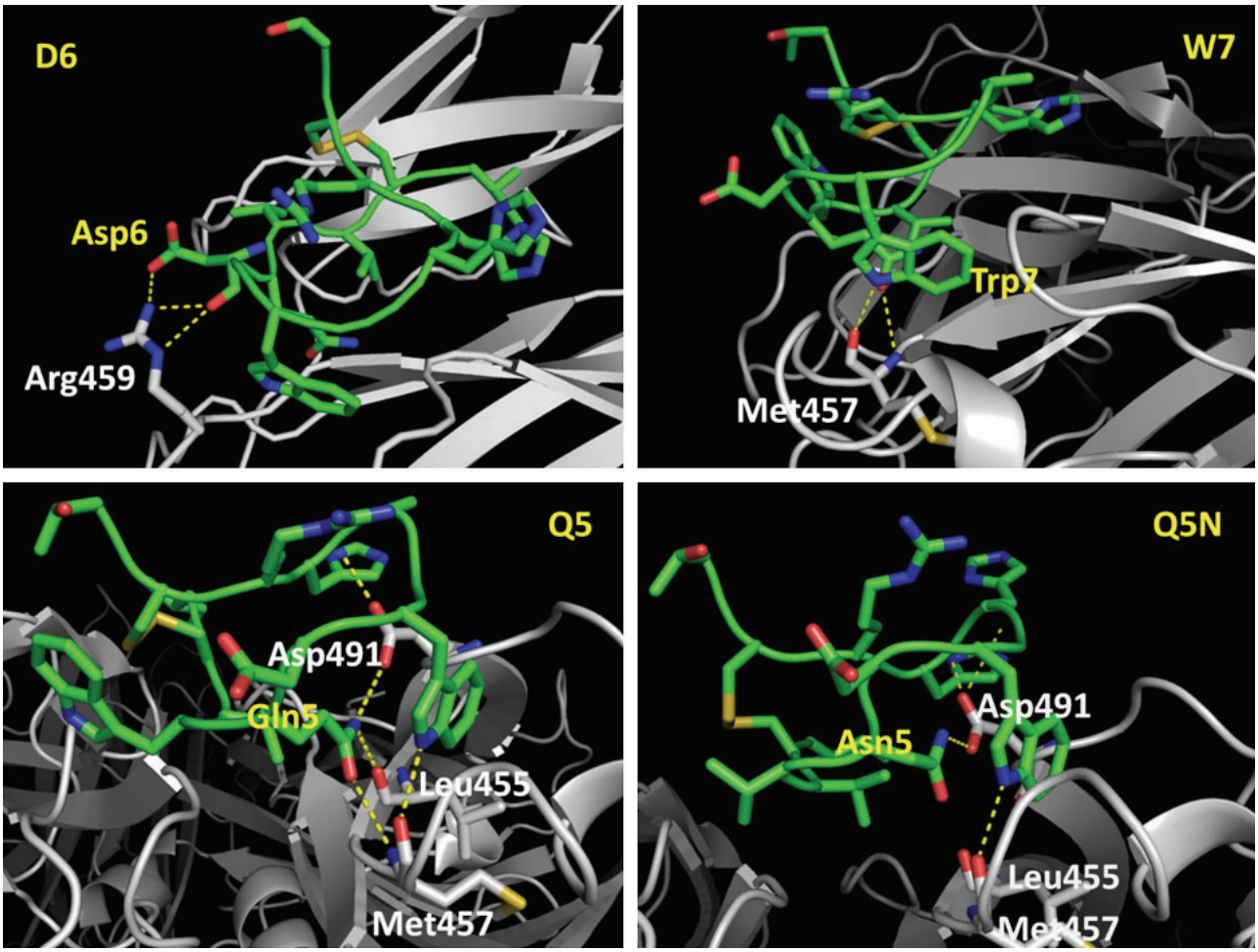

Figure 1: Snapshots of four compstatin analogs illustrating specific interactions that stabilize the interactions with C3. Compstatin is colored in green. Its residues are labeled in yellow. C3 is colored in gray. Its residues are labeled in white. The hydrogen bonds are shown in dotted yellow lines.

Construction of the Pharmacophore Model

Alignment of the minimized structures of most active peptide analogs

The pharmacophore model were developed using the minimized structures of the two most active compstatin peptide analogs, AcCompNH2-V4(1MeW)-H9A and AcCompNH2-V4(1MeW)-W7(5fW)-H9A. Both carried 261-fold activities when compared with parent compstatin. Their minimized structures were aligned using the Homology/Align function of MOE with default parameters. Figure 2 shows the overlay of the two peptide analogs. As can be seen, the two peptide analogs were very well aligned except for the flexible side chains of Ile1, Asp6, and His10 where they had the largest discrepancies.

Building pharmacophore features

Each individual structural feature was built into a single pharmacophore feature using the Pharmacophore Modeling function of MOE. For W4, the aromatic ring on the side chain is important. Thus, an aromatic sphere centered at the center of mass of the indole ring was built. The default radius of 1.2 Å was accepted. For Q5, the amide group on the tail of the side chain plays an important role in keeping the activities. Specifically, the amide carbonyl and the amide nitrogen served as a hydrogen-bond donor and a hydrogen-bond acceptor, respectively. Therefore, we created two pharmacophore features on this residue: a hydrogen-bond acceptor of 0.8 Å radius by default centered at the carbonyl oxygen and a hydrogen-bond donor of 0.8 Å radius by default centered at the amide nitrogen. For D6, the carbonyl oxygen acted as a hydrogen-bond acceptor to the amide nitrogen on the C3 R459 side chain. Thus, we built a feature on the carbonyl oxygen of the D6 side chain. The default feature assigned by MOE was a hydrogen-bond donor of 2.2 Å radius. For W7, we added a hydrogen-bond donor feature to the indole nitrogen of W7, given its importance in keeping the activities. The default radius of 0.8 Å was adopted. Finally, we added a hydrophobe feature between the two sulfur atoms of C2 and C12. The default radius of 0.7 Å was kept.

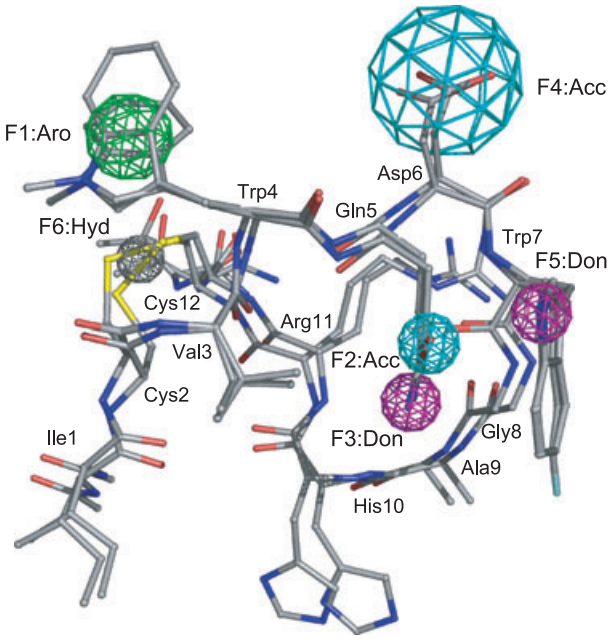

Figure 2: The superposition of the six pharmacophore features over the aligned structures of the two most active compstatin analogs, AcCompNH2-V4(1MeW)-H9A and AcCompNH2-V4(1MeW)-W7(5fW)-H9A. Aro: aromatic ring center; Acc: hydrogen-bond acceptor; Don: hydrogen-bond donor; Hyd: hydrophobic region.

Preprocessing the database of known compstatin analogs

In order to facilitate the subsequent pharmacophore searches, the minimized structures of compstatin peptide analogs were preprocessed using the Pharmacophore Preprocessor function inside of MOE with the scheme set to PCH. Running the preprocessor added an annotation field, PH4:PCH, to the minimized structure of each compstatin peptide analog. The annotation field is an encoding of the positions, labels, and indices of the annotation points of the compstatin peptide analogs.

Selection of test set

We incorporated most of the compstatin analogs published in the literature (5–11) into our test set but excluded the following: (i) retro–inverso compstatin (9) and three biotynylated analogs (9) due to their significant differences from compstatin analog in the X-ray structure (PDB ID: 2QKI); (ii) three D6P (9) and one T13P (10) mutants because it was not possible to generate appropriate conformers for them. Furthermore, we removed the compstatin analogs used in the pharmacophore modeling process (listed in Table 1). The remaining 82 compstatin analogs were used as our test set. They are listed in Table 2.

Evaluation of the pharmacophore model

We evaluated both the sensitivity and specificity of our pharmacophore model. Specifically, sensitivity was measured by the ability of the model to identify the compstatin analogs of medium and high inhibitory activities, while specificity was measured by the ability of the model to identify only the compstatin analogs of medium and high inhibitory activities and nothing else. We defined the relative activity of 10 when compared with compstatin to be the cut-off. The compstatin analogs having 10-fold or higher inhibitory activities when compared with compstatin were considered of medium or high inhibitory activities.

Table 1 lists the nine compstatin analogs used in pharmacophore modeling process. Four of them had 10-fold or higher inhibitory activities when compared with compstatin. Table 2 lists the 82 compstatin analogs in the test set. Among them, 20 had 10-fold or higher inhibitory activities when compared with compstatin. The Pharmacophore Search function of MOE was used to scan both sets using our pharmacophore model. The Use Annotation Field was automatically set to PH4:PCH by MOE. The other default options of MOE were used.

Results and Discussion

Taking advantage of the availability of extensive structure–activity relationships of compstatin analogs, an X-ray structure of AcCompNH2-V4W-H9A bound to the C3c fragment of C3 (PDB ID: 2QKI) and the two most active compstatin analogs known today, AcCompNH2-V4(1MeW)-H9A and AcCompNH2-V4(1MeW)-W7(5fW)-H9A), we built a new pharmacophore model for compstatin analogs. The model consists of six structural features, all of which play critical roles in determining inhibitory activities. Table 3 lists these six

Table 1: The nine compstatin analogs used in the pharmacophore modeling

Name	Sequence	IC_{50} (μM)	Relative activity
AcCompNH2-V4(1MeW)/H9A	Ac-ICV(1MeW)QDWGAHRCT-NH2	0.205	261.463
AcCompNH2-V4(1MeW)/W7(5fW)/H9A	Ac-ICV(1MeW)QD(5fW)GAHRCT-NH2	0.205	261.463
AcCompNH2-V4W	Ac-ICVWQDWGHHRCT-NH2	2.2	24.364
AcCompNH2-Q5N	Ac-ICVVNDWGHHRCT-NH2	4.2	2.857
CompCC-W7A	CVVQDAGHHRC	182	0.066
CompCC-D6A	CVVQAWGHHRC	257	0.047
CompCC-Q5A	CVVADWGHHRC	910	0.013
AcCompNH2-V4W/W7(1MeW)/H9A	Ac-ICVWQD(1MeW)GAHRCT-NH2	NA	–
Comp-linear	IAVVQDWGHHRAT	>600	–

Relative activity is the activity relative to compstatin.

Table 2: The 82 compstatin analogs in the test set

Peptide #	Name	Sequence	IC_{50} (μM)	Relative activity
1	Comp	ICVVQDWGHHRCT	12	1
2	CompCC	CVVQDWGHHRC	33	0.364
3	CompCC-V3A	CAVQDWGHHRC	1200	0.01
4	CompCC-V4A	CVAQDWGHHRC	67	0.179
5	CompCC-G8A	CVVQDWAHHRC	>1200	–
6	CompCC-H9A	CVVQDWGAHRC	15	0.8
7	CompCC-H10A	CVVQDWGHARC	74	0.162
8	CompCC-R11A	CVVQDWGHHAC	70	0.171
	Linear analogs			
9	Comp-NH2	ICVVQDWGHHRCT-NH2	12	1
10	Comp-C2A/C12A	IAVVQDWGHHRAT-NH2	>600	–
11	Comp-cleavedRC	ICVVQDWGHHRCT-NH2	>300	–
	Acetylated compstatin			
12	AcCompNH2	Ac-ICVVQDWGHHRCT-NH2		3
	Deletion analogs			
13	CompNH2-del1I	CVVQDWGHHRCT-NH2	25	0.480
14	CompNH2-del[1I,13T]	CVVQDWGHHRC-NH2	33	0.364
15	CompNH2-del[1I,3V-8G]	CHHRCT-NH2	>600	–
16	CompNH2-del[1I,3V-7W]	CGHHRCT-NH2	>600	–
18	CompNH2-del[1I,3V-6D]	CWGHHRCT-NH2	>600	–
19	CompNH2-del[1I,3V-5Q]	CDWGHHRCT-NH2	>600	–
20	CompNH2-del[1I,3V-4V]	CQDWGHHRCT-NH2	>600	–
21	CompNH2-del[1I,3V]	CVQDWGHHRCT-NH2	>600	–
22	CompNH2-del[1I,8G-11R,13T]	CVVQDWC-NH2	>600	–
23	CompNH2-del[1I,10H-11R,13T]	CVVQDWGHC-NH2	>600	–
	Beta-turn analogs			
24	AcCompNH2-Q5G	Ac-ICVVGDWGHHRCT-NH2	567	0.021
25	AcCompNH2-W7F	Ac-ICVVQDFGHHRCT-NH2	>400	–
	Table 2.I			
26	AcCompNH2-H9A	Ac-ICVVQDWGAHRCT-NH2	2.9	4.138
27	AcCompNH2-V4A/H9A/T13I	Ac-ICVAQDWGAHRCI-NH2	4	3
	AcCompNH2-W7F	Ac-ICVVQDFGHHRCT-NH2	>400	–
28	AcCompNH2-Q5G/D6A/W7A	Ac-ICVVGAAGHHRCT-NH2	>100	–
	Table 2.II			
29	AcCompNH2-V3L	Ac-ICLVQDWGHHRCT-NH2	10	1.2
30	AcCompNH2-V3L/Q5N	Ac-ICLVNDWGHHRCT-NH2	8.3	1.446
31	CompNH2-R11K	ICVVQDWGHHKCT-NH2	20.2	0.594
32	AcCompNH2-R11S	Ac-ICVVQDWGHHSCT-NH2	25	0.48
33	AcCompNH2-Q5N/R11A	Ac-ICVVNDWGHHACT-NH2	60	0.2
34	AcCompNH2-H9A/R11A	Ac-ICVVQDWGAHACT-NH2	9.9	1.212
35	AcCompNH2-H9A/R11dR	Ac-ICVVQDWGAHdRCT-NH2	>1000	–
36	AcCompNH2-T13I	Ac-ICVVQDWGHHRCI-NH2	3.2	3.75
	Elisa competition binding experiment			
	AcCompNH2	Ac-ICVVQDWGHHRCT-NH2	5.6	3
37	AcCompNH2-I1L/V4W/H9R/R11Q/T13F	Ac-LCVWQDWGRHQCF-NH2	131	0.128
38	AcCompNH2-I1L/H9W/T13G	Ac-LCVVQDWGWHRCG-NH2	5.4	3.111
39	AcCompNH2-I1M/V4H/R9F/T13F	Ac-MCVHQDWGGHRCF-NH2	85.2	0.197
	Inhibition of complement mediated rabbit Erythrocyte lysis experiment			
	CompNH2	ICVVQDWGHHRCT-NH2	12	1
	AcCompNH2	Ac-ICVVQDWGHHRCT-NH2	4.5	2.667
	AcCompNH2-H9A	Ac-ICVVQDWGAHRCT-NH2	2.9	4.138
	AcCompNH2-I1L/V4W/H9R/R11Q/T13F	Ac-LCVWQDWGRHQCF-NH2	254	0.047
	AcCompNH2-I1L/H9W/T13G	Ac-LCVVQDWGWHRCG-NH2	2.9	4.138
	AcCompNH2-I1M/V4H/H9G/T13F	Ac-MCVHQDWGGHRCF-NH2	87.4	0.137
40	AcCompNH2-I1R	Ac-RCVVQDWGHHRCT-NH2	8	1.5
41	AcCompNH2-I1D	Ac-DCVVQDWGHHRCT-NH2	22	0.545
	Table 2.III			
	CompNH2	ICVVQDWGHHRCT-NH2	53.6	1
	AcCompNH2	Ac-ICVVQDWGHHRCT-NH2	18.1	2.961
	AcCompNH2-H9A	Ac-ICVVQDWGAHRCT-NH2	12.4	4.322

Table 2: Continued

Peptide #	Name	Sequence	IC_{50} (μM)	Relative activity
42	AcCompNH2-V4T	Ac-ICVTQDWGHHRCT-NH2	68.3	0.785
43	AcCompNH2-V4S	Ac-ICVSQDWGHHRCT-NH2	50.9	1.053
44	AcCompNH2-V4H	Ac-ICVHQDWGHHRCT-NH2	10.5	5.105
45	AcCompNH2-V4F	Ac-ICVFQDWGHHRCT-NH2	10.2	5.255
46	AcCompNH2-V4Y/H9A	Ac-ICVYQDWGAHRCT-NH2	3.8	14.889
47	AcCompNH2-V4W/H9W	Ac-ICVWQDWGWHRCT-NH2	3.1	17.290
48	AcComp-V4W/H9A	Ac-ICVWQDWGAHRCT	2	26.8
49	AcCompNH2-V4W/H9A	Ac-ICVWQDWGAHRCT-NH2	1.2	44.667
	Table 2.IV			
	CompNH2	ICVVQDWGHHRCT-NH2	53.6	1
50	AcCompNH2-V4(Cha)/H9A	Ac-ICV(Cha)QDWGAHRCT	47.8	1.138
51	AcCompNH2-V4(Dht)/H9A	Ac-ICV(Dht)QDWGAHRCT	10.2	5.255
52	AcCompNH2-V4(Yphs)/H9A/T13I	Ac-ICV(Yphs)QDWGAHRCI-NH2	9.6	5.583
53	AcCompNH2-V4W/H9(Abu)	Ac-ICVWQDWG(Abu)HRCT-NH2	1.5	35.733
54	AcComp-V4(2Ig1)/H9A	Ac-ICV(2Ig1)QDWGAHRCT	1.5	35.733
55	AcCompNH2-V4(2Ig1)/H9A	Ac-ICV(2Ig1)QDWGAHRCT-NH2	1.4	38.286
56	AcComp-V4(Bta)/H9A	Ac-ICV(Bta)QDWGAHRCT	0.8	67
57	AcCompNH2-V4(Bta)/H9A	Ac-ICV(Bta)QDWGAHRCT-NH2	0.8	67
58	AcComp-V4(Bpa)/H9A	Ac-ICV(Bpa)QDWGAHRCT	1.1	48.727
59	AcCompNH2-V4(Bpa)/H9A	Ac-ICV(Bpa)QDWGAHRCT-NH2	0.6	89.333
60	AcComp-V4(1Nal)/H9A	Ac-ICV(1Nal)QDWGAHRCT	1.8	29.778
61	AcComp-V4(2Nal)/H9A	Ac-ICV(2Nal)QDWGAHRCT	1.4	38.286
62	AcCompNH2-V4(2Nal)/H9A	Ac-ICV(2Nal)QDWGAHRCT-NH2	0.5	107.2
	Table 2.V			
63	AcCompNH2-V4dY/H9A/T13I	Ac-ICVdYQDWGAHRCI-NH2	>1000	–
64	AcCompNH2-V4dW/H9A/T13I	Ac-ICVdWQDWGAHRCI-NH2	>1000	–
65	AcComp-V4W/H9A/R11dR	Ac-ICVWQDWGAHdRCT	>1000	–
66	AcCompNH2-Q5dQ/delT13	Ac-ICVVdQDWGHHRC-NH2	>1000	–
67	AcCompNH2-D6dD/delT13	Ac-ICVVQdDWGHHRC-NH2	>1000	–
68	AcCompNH2-W7dW/delT13	Ac-ICVVQDdWGHHRC-NH2	>1000	–
69	AcComp-V3dV/V4W/H9A	Ac-ICdVWQDWGAHRCT	>1000	–
70	AcComp-V4W/H9A/H10dH	Ac-ICVWQDWGAdHRCT	136.2	0.394
71	AcComp-V4W/H9dA	Ac-ICVWQDWGdAHRCT	132.4	0.405
72	AcComp-I1dI/V4W/H9A	Ac-dICVWQDWGAHRCT	23.1	2.320
73	AcComp-V4W/H9A/T13dT	Ac-ICVWQDWGAHRCdT	2.7	19.852
	CompNH2	ICVVQDWGHHRCT-NH2	53.6	1
	AcCompNH2-V4W/H9A	Ac-ICVWQDWGAHRCT-NH2	1.2	44.667
74	Comp-G(-1)/V4W/H9A/N14	GICVWQDWGAHRCTN	1.2	44.667
75	Comp-G(-1)/V4(6fW)/W7(6fW)/H9A/N14	GICV(6fW)QD(6fW)GAHRCTN	0.43	124.651
76	Comp-G(-1)/V4(5OHW)/W7(5OHW)/H9A/N14	GICV(5OHW)QD(5OHW)GAHRCTN	33	1.624
77	Comp-G(-1)/V4(7azaW)/W7(7azaW)/H9A/N14	GICV(7azaW)QD(7azaW)GAHRCTN	122	0.439
	CompNH2	ICVVQDWGHHRCT-NH2	53.6	1
	AcCompNH2-V4W/H9A	Ac-ICVWQDWGAHRCT-NH2	1.2	44.667
78	AcCompNH2-V4(5fW)/H9A	Ac-ICV(5fW)QDWGAHRCT-NH2	1.74	30.805
79	AcCompNH2-V4(5MeW)/H9A	Ac-ICV(5MeW)QDWGAHRCT-NH2	0.87	61.609
80	AcCompNH2-V4(2Nal)/H9A	Ac-ICV(2Nal)QDWGAHRCT-NH2	0.545	98.35
81	AcCompNH2-V4W/W7(5fW)/H9A	Ac-ICVWQD(5fW)GAHRCT-NH2	0.446	120.279
82	AcCompNH2-V4W/W7(5MeW)/H9A	Ac-ICVWQD(5MeW)GAHRCT-NH2	NA	–

The data were compiled from refs (2–8). The redundant compstatin analogs used to compare different experiments were listed in this table but were not counted as new analogs.

pharmacophore features in detail. Figure 2 shows the superposition of the six pharmacophore features over the aligned structures of the two most active compstatin analogs.

The sensitivity and specificity of this pharmacophore model were tested. We showed that this model was able to identify only compstatin analogs of medium and high inhibitory activities and nothing else from the set of nine compstatin analogs used in the pharmacophore modeling. In addition, using a separate test set of 82 compstatin analogs, we demonstrated that our pharmacophore model was able to identify 14 (14/20 or 70%) of the analogs of medium and high inhibitory activities and misclassify only five (5/62 or 8.5%) of the analogs of low or no inhibitory activities. The superior results proved our pharmacophore model to be a filter of great sensitivity and specificity.

To understand the relative importance of each individual pharmacophore feature on the sensitivity and specificity of the model, the

Table 3: The six pharmacophore features

ID	Feature	Center	Radius (Å)
F1	Aromatic ring	Center of mass of W4 indole ring	1.2
F2	Hydrogen-bond acceptor	Carbonyl oxygen on the Q5 side chain	0.8
F3	Hydrogen-bond donor	Amide nitrogen on the Q5 side chain	0.8
F4	Hydrogen-bond acceptor	Carbonyl oxygen on the D6 side chain	2.2
F5	Hydrogen-bond donor	Indole nitrogen of W7	0.8
F6	Hydrophobe	Centroid of two S_γ atoms of C2 and C12	0.7

Table 4: The relative importance of each individual pharmacophore feature on the sensitivity and specificity of the model, as indicated by the success rate and the false positive rate after the feature was removed

ID	Feature	Success rate after the feature was removed	False positive rate after the feature was removed
F1	Aromatic ring	17/20 (85)	46/62 (74)
F2	Hydrogen-bond acceptor	15/20 (75)	22/62 (35)
F3	Hydrogen-bond donor	16/20 (80)	19/62 (31)
F4	Hydrogen-bond acceptor	15/20 (75)	18/62 (29)
F5	Hydrogen-bond donor	15/20 (75)	27/62 (44)
F6	Hydrophobe	18/20 (90)	41/62 (66)

Values expressed in percentage are given in parentheses.

following test was performed. Each pharmacophore feature was removed individually and the test set was scanned using the reduced pharmacophore model consisting of only the remaining five pharmacophore features. The test result is shown in Table 4. The removal of any one pharmacophore feature caused only a minor increase in the number of true positives but a dramatic increase in the number of false positives. Specifically, removal of either F1 or F6 resulted in the largest false positive rates of 74% and 66%, respectively. Therefore, the aromatic ring on the fourth residue and the hydrophobe on the disulfide bond were the most critical pharmacophore features in ensuring the specificity of the model.

In a separate quantitative structure–activity relationship (QSAR) study of compstatin analogs, Mulakala *et al.* identified four important physico-chemical and geometrical properties of the analogs that correlated strongly with activity (15): the number of aromatic bonds and the hydrophobicity of the fourth residue, hydrophobic patch size near the disulfide bond and the solvent accessible surface area occupied by nitrogen atoms of basic amino acid residues. Independently, we also identified the aromatic ring of the fourth residue (F1) and the hydrophobe on the disulfide bond (F6) to be the most critical pharmacophore features in determining inhibitory activity as demonstrated by the removal test results. Lastly, the solvent accessible surface area occupied by nitrogen atoms of basic amino acid residues was related to the charges on the nitrogen atoms of basic amino acid residues. Our removal test results also suggested the hydrogen acceptor group (F2) and hydrogen donor group (F3) on the fifth residue, the hydrogen acceptor group on the sixth residue (F4), and the hydrogen-bond donor (F5) on the seventh residue to be important in determining activity.

Acknowledgments

Ting-Lan Chiu would like to thank Drs Yuk Sham and Angelo Pugliese for useful discussions. We want to thank Dr Tony Hill for reproducing Figure 2 using pymol. This work was supported by NIH Grants GM-069736, GM-62134, and AI-068730.

Note

[a]MOE (The Molecular Operating Environment) Version 2006.08, C. C. G. I., Montreal, Quebec, Canada. http://www.chemcomp.com/.

References

1. Ricklin D., Lambris J.D. (2007) Complement targeted Therapeutics. Nat Biotechnol;25:1265–1275.
2. Sahu A., Lambris J.D. (2000) Complement inhibitors: a resurgent concept in anti-inflammatory therapeutics. Immunopharmacology;49:133–148.
3. Sahu A., Kay B.K., Lambris J.D. (1996) Inhibition of human complement by a C3-binding peptide isolated from a phage-displayed random peptide library. J Immunol;157:884–891.
4. Mallik B., Morikis D. (2005) Development of a quasi-dynamic pharmacophore model for anti-complement peptide analogues. J Am Chem Soc;127:10967–10976.
5. Katragadda M., Lambris J.D. (2006) Expression of compstatin in *Escherichia coli*: incorporation of unnatural amino acids enhances its activity. Protein Expr Purif;47:289–295.
6. Mallik B., Katragadda M., Spruce L.A., Carafides C., Tsokos C.G., Morikis D., Lambris J.D. (2005) Design and NMR characterization of active analogues of compstatin containing non-natural amino acids. J Med Chem;48:274–286.
7. Morikis D., Assa-Munt N., Sahu A., Lambris J.D. (1998) Solution structure of compstatin, a potent complement inhibitor. Protein Sci;7:619–627.
8. Morikis D., Roy M., Sahu A., Troganis A., Jennings P.A., Tsokos G.C., Lambris J.D. (2002) The structural basis of compstatin activity examined by structure–function-based design of peptide analogs and NMR. J Biol Chem;277:14942–14953.
9. Sahu A., Soulika A.M., Morikis D., Spruce L., Moore W.T., Lambris J.D. (2000) Binding kinetics, structure–activity relationship, and biotransformation of the complement inhibitor compstatin. J Immunol;165:2491–2499.
10. Soulika A.M., Morikis D., Sarrias M.R., Roy M., Spruce L.A., Sahu A., Lambris J.D. (2003) Studies of structure–activity relations of complement inhibitor compstatin. J Immunol;171:1881–1890.
11. Katragadda M., Magotti P., Sfyroera G., Lambris J.D. (2006) Hydrophobic effect and hydrogen bonds account for the

improved activity of a complement inhibitor, compstatin. J Med Chem;49:4616–4622.
12. Janssen B.J.C., Halff E.F., Lambris J.D., Gros P. (2007) Structure of compstatin in complex with complement component c3c reveals a new mechanism of complement inhibition. J Biol Chem;282:29241–29247.
13. Ricklin D., Lambris J.D. (2008) Compstatin: a complement inhibitor on its way to clinical application. Adv Exp Med Biol;632:273–292.
14. Pigache A., Cieplak A.P., Dupradeau F.-Y. Automatic and highly reproducible RESP and ESP charge derivation: application to the development of programs RED and X RED. 227th ACS National Meeting: Anaheim, CA, USA, 2004.
15. Mulakala C., Lambris J.D., Kaznessis Y. (2007) A simple, yet highly accurate, QSAR model captures the complement inhibitory activity of compstatin. Bioorg Med Chem;15:1638–1644.