Improvement of the anti-C3 activity of compstatin using rational and combinatorial approaches

D. Morikis*[1], A.M. Soulika†, B. Mallik*, J.L. Klepeis‡, C.A. Floudas‡ and J.D. Lambris†[1]

*Department of Chemical and Environmental Engineering, University of California at Riverside, Riverside, CA 92521, U.S.A., †Protein Chemistry Laboratory, Department of Pathology and Laboratory Medicine, University of Pennsylvania, 401 Stellar-Chance Laboratories, Philadelphia, PA 19104, U.S.A., and ‡Department of Chemical Engineering, Princeton University, Princeton, NJ 08544, U.S.A.

Abstract

Compstatin is a 13-residue cyclic peptide that has the ability to inhibit the cleavage of C3 to C3a and C3b. The effects of targeting C3 cleavage are threefold, and result in hindrance of: (i) the generation of the pro-inflammatory peptide C3a, (ii) the generation of opsonin C3b (or its fragment C3d), and (iii) further complement activation of the common pathway (beyond C3) with the end result of the generation of the membrane attack complex. We will report on our progress on: (i) rational design of more active compstatin analogues based on the three-dimensional structure of compstatin, (ii) experimental combinatorial design based on the generation of a phage-displayed peptide library partially randomized with the implementation of structure-induced restraints, and (iii) theoretical combinatorial design based on a novel computational optimization method, structure-induced restraints and flexible structural templates. All three approaches have resulted in analogues with improved activities. Currently, the lead analogue has the sequence acetyl-I[CVYQDWGAHRC]T-NH_2 (where the brackets denote cyclization), and is 16-fold more active than the parent peptide. We will also report on our progress towards understanding the dynamic character of compstatin using molecular dynamics simulations. The identification of an ensemble of interconverting conformers of compstatin with variable populations is a first step towards the incorporation of dynamic elements in the design of new analogues using dynamics–activity relationships in addition to structure–activity relationships.

Background

A C3-binding complement inhibitor was identified as a 27-residue peptide using a phage-displayed random peptide library [1]. This peptide was truncated to an equally active 13-residue peptide named compstatin with the sequence I[CVVQDWGHHRC]T-NH_2, where the brackets denote cyclization through a disulphide bridge formed by Cys-2–Cys-12 [1,2]. Acetylation of the N-terminus of compstatin (Ac-compstatin) resulted in a 3-fold increase in activity [3–6].

Compstatin blocked the cleavage of C3 to the pro-inflammatory peptide C3a and the opsonin C3b in haemolytic assays and in human normal serum [1,3], prevented heparin/protamine-induced complement activation in baboons in a situation resembling heart surgery [7], inhibited complement activation during the contact of blood with biomaterial in a model of extra-corporeal circulation [8], increased the lifetime of survival of porcine kidneys perfused with human blood in a hyper-acute rejection xenotransplantation model [9], blocked the *Escherichia coli*-induced oxidative burst of granulocytes and monocytes [10], and inhibited complement activation by cell lines SH-SY5Y, U-937, THP-1 and ECV304 [11]. Compstatin was stable in biotranformation studies *in vitro* in human blood, normal human plasma and serum, with increased stability upon N-terminal acetylation [3]. Compstatin showed little or low toxicity and no adverse effects when these were assessed [4,7–9]. Finally, compstatin shows species specificity, being active only with human and primate C3 [12].

In this mini-review we will present a progress report of our efforts to design analogues of compstatin with improved inhibitory activity. Our efforts have been in four different directions, all of which depend on the availability of the three-dimensional structure of a major conformer of compstatin and our thorough understanding of the contribution of each amino acid, or part of each amino acid, to the formation of structure. Specifically, we will discuss (i) rational design of compstatin analogues using SARs (structure–activity relationships), (ii) experimental combinatorial design using phage-displayed peptide libraries for binding to C3, (iii) computational combinatorial design using mixed-integer and global optimization methods, and (iv) MD (molecular dynamics) simulations that have identified additional conformers of compstatin with small populations. The latter will form the basis for fine-tuning of our design using SARs and for initiation of dynamics–activity relationships.

Key words: complement system, compstatin, drug design, molecular dynamics, NMR, peptide.
Abbreviations used: Ac, acetyl; MD, molecular dynamics; NOE, nuclear Overhauser effect; SAR, structure–activity relationship.
[1]To whom correspondence should be addressed (e-mail dmorikis@engr.ucr.edu or lambris@mail.med.upenn.edu).

Table 1 | Sequences and relative activities of compstatin analogues with improved activity that were identified by rational, experimental combinatorial and computational combinatorial design

Amino acids that were kept fixed in our combinatorial design (see the text) are shown in bold face. Amino acids that were optimized in rational and combinatorial designs are underlined. Brackets denote cyclization through a disulphide bridge between Cys-2 and Cys-12. Relative complement-inhibitory activity was obtained from IC_{50} measurements. Several measurements were performed for each analogue. Note that our comparisons here are against the initial non-acetylated compstatin at the N-terminus. In some of the published work this comparison is against Ac-compstatin, producing lower relative activities. Analogues were identified using the following techniques: *rational design; †experimental combinatorial design; ‡computational combinatorial design.

Peptide	Sequence	Relative activity	References
Compstatin	I[**CV**V**QDWG**HHR**C**]T-NH_2	1	[1]
Ac-compstatin	Ac-I[**CV**V**QDWG**HHR**C**]T-NH_2	3	[3,5,6]
Ac-H9A*	Ac-I[**CV**V**QDWG**A HR**C**]T-NH_2	4	[5]
Ac-I1L/H9W/T13G†	Ac-L[**CV**V**QDWG**W HR**C**]G-NH_2	4	[6]
Ac-I1V/V4Y/H9F/T13V‡	Ac-V[**CV**Y**QDWG**F HR**C**]V-NH_2	6	[15]
Ac-I1V/V4Y/H9A/T13V‡	Ac-V[**CV**Y**QDWG**A HR**C**]V-NH_2	9	[15]
Ac-V4Y/H9F/T13V‡	Ac-I[**CV**Y**QDWG**F HR**C**]V-NH_2	11	[15]
Ac-V4Y/H9A/T13V‡	Ac-I[**CV**Y**QDWG**A HR**C**]V-NH_2	14	[15]
Ac-V4Y/H9A‡	Ac-I[**CV**Y**QDWG**A HR**C**]T-NH_2	16	[15]

Rational design of compstatin analogues with improved activity

The three-dimensional structure of compstatin in solution revealed the presence of a major conformer consisting of a Type I β-turn located at a position opposite the disulphide bridge [2]. The molecular surface of compstatin consists of a polar part that includes the β-turn, and a hydrophobic part that includes the disulphide bridge [2]. The significance of these structural characteristics was recognized for structural stability, for binding to C3 and for complement inhibitory activity [2,5]. In the case of compstatin, structural stability, binding to C3 and inhibitory activity are correlated [2,5].

Our rational design of analogues with higher inhibitory activity has been discussed and compared with similar efforts for other low-molecular-mass complement inhibitors in a recent mini-review [13]. Briefly, the rational or SAR design was based on the available three-dimensional structure of compstatin, structural NMR studies of the designed new analogues, kinetic studies of binding to C3 and its fragments, and measurements of complement-inhibitory activity [2,3,5,6]. The three-dimensional structure revealed the overall fold of compstatin and intra-molecular interactions involving hydrogen bonding, hydrophobicity, electrostatics, van der Waals forces, disulphide bridge and polar interactions with solvent molecules. These data provided insight into the structural stability both of compstatin and, in combination with additional NMR studies, of the designed analogues. The binding site on C3 was modelled to have a shape complementary to that of compstatin, with amino acids having physico-chemical properties that promote binding (either similar or complementary, depending on the property [5]). Comparable binding kinetic studies of selected designed analogues using surface plasmon resonance [3,6] provided insight into inter-molecular interactions important for binding [3,5,6]. Finally, the observed or deduced intra- and inter-molecular interactions of the designed analogues were correlated with inhibitory activity measured using immunological assays. The specific physico-chemical properties of amino acids were taken into account in the design of new analogues. These properties are (i) hydrophobicity/polarity, (ii) charge, (iii) electronic distribution, (iv) side chain length, volume, branching, isomerism and flexibility, (v) side chain orientation and mobility in the specific fold, (vi) ability for cyclization, and (vii) amino acid propensities to form specific secondary structures. Radical site-specific replacements were used to determine the effects of gross amino acid differences on structure, binding and activity, and conservative replacements were used for fine-tuning of the design, together with additions/deletions, alanine scanning, incorporation of non-natural amino acids with directed properties, methylation, and alternative cyclization [1–3,5,6].

The analogue with highest inhibitory activity identified using this method, named Ac-H9A (sequence is given in Table 1), had 4-fold higher activity than the parent peptide compstatin [5]. These efforts include a prior benchmark of acetylation of the N-terminus that resulted in a 3-fold increase in inhibitory activity (Table 1) [3–6]. In addition, the rational design and the use of structural information from NMR provided the basis for the experimental and combinatorial design that will be described below.

Experimental combinatorial design of compstatin analogues with improved activity

The technique of using phage-displayed random peptide libraries to randomly identify peptides that are capable of binding to specific targets and altering their functionality is widely used [14]. This was the case for the identification of

compstatin in 1996, when a phage-displayed peptide library was constructed and screened for binding to C3b [1]. This technique was used again for peptide binding to C3 [6], but this time incorporating findings from our rational design (see above; [5]). Specifically, seven amino acids were kept fixed, while six were allowed to vary randomly. The combinatorial gene sequence NNS TGC GTG NNS CAG GAC TGC GGC $(NNS)_3$ TGC NNS was displayed at the N-terminus of the M13 pIII gene, where N represents an equal molar ratio of all nucleotides A, C, G and T, and S represents an equal molar ratio of C, G and T. The encoded amino acid sequences were XCVXQDWGXXXCX, where amino acids at positions X were combinatorially randomized [6].

Four clones that bound to native C3 were identified using this method. Measurements of complement-inhibitory activity using synthetic acetylated peptides with the sequences of the binding clones identified one analogue with 4-fold higher activity than compstatin (Table 1) [6]. This analogue was named Ac-I1L/H9W/T13G, and its sequence is given in Table 1. NMR experiments with this analogue demonstrated similar structural characteristics to compstatin, Ac-compstatin and the equally active rationally designed analogue Ac-H9A (Table 1) [6]. The hydrophobic cluster and the Type I β-turn were preserved in Ac-I1L/H9W/T13G, and a novel feature was observed, namely the the introduction of a second Trp at position 9. The presence of two Trp side chains in proximity suggested that Trp ring stacking may be important for the activity of this analogue [6].

Although the experimental combinatorial design identified an analogue equally as active as that of the rational design described above, in combination two important features for activity were revealed: (i) position 9 was amenable to further optimization, and (ii) side-chain ring stacking involving one residue inside and one outside the β-turn could be important to optimize activity. The latter was re-inforced by computational combinatorial design, which will be described below.

Novel computational combinatorial design of compstatin analogues with improved activity

A general and novel *in silico* drug design method was formulated and tested for the first time using compstatin [15]. This is a two-step methodology. (i) The foundation of the first step is a mixed-integer linear optimization algorithm based on the use of a backbone potential that is distance-dependent, with implicit inclusion of side-chain interactions and amino acid specificities [15,16]. This algorithm selects and ranks several potential sequences compatible with a structural template that provides (but is not limited to) the Cα–Cα inter-atomic backbone distances. (ii) The foundation of the second step is a global optimization algorithm based on the use of a full-atom force field [17–19]. This algorithm calculates ensemble probabilities for the selected sequences applied on flexible structural templates [15]. The underlying assumption of this approach is that amino acid specificity and the predicted increase in fold stability and specificity, while maintaining essential functional characteristics, are correlated with increase in functionality [15]. This was the case for compstatin, as shown by the success of the method, with functionality here being complement-inhibitory activity [15]. The novel *in silico* peptide design approach naturally treats the template structures as flexible, since the distance-dependent force field is based on ranges of distances.

The prerequisites for application of this method are the availability of three-dimensional structural templates and knowledge of the essential structural characteristics from initial rational design. In the case of compstatin, both prerequisites were achieved using NMR studies [2,5]. The essential structural characteristics were that seven amino acids, i.e. Cys-2, Val-3 and Cys-12 of the hydrophobic cluster and Gln-5, Asp-6, Trp-7 and Gly-8 of the Type I β-turn of compstatin, were deemed indispensable (and non-optimizable) as determined by earlier studies [2,5], while the remaining six amino acids (Ile-1, Val-4, His-9, His-10, Arg-11 and Thr-13) were allowed to vary. To optimize the computation, the variation was not totally random, but instead amino acids of the hydrophobic cluster, i.e. Ile-1, Val-4 and Thr-13, could be optimized only with hydrophobic amino acids, whereas the amino acids of the polar part (His-9, His-10 and Arg-11) could be optimized by any of the 20 natural amino acids.

The rationale behind the amino acids selected for optimization using computational and experimental optimization was identical, in that only amino acids denoted by X in the sequence Ac-X[CVXQDWGXXXC]X-NH_2 were allowed to vary. However, the computational approach gave more spectacular results, since it identified several analogues with higher activity than compstatin, among them a 16-fold more active analogue named Ac-V4Y/H9A (Table 1) [15].

The experimental approach was also successful, but identified only a 4-fold more active analogue than compstatin (Table 1) [6]. The common finding in all rational, experimental and computational combinatorial designs was that position 9 was amenable to further optimization. The new insight provided only by computational combinatorial design was that position 4 was also amenable to further optimization, producing even higher inhibitory activities. Rational design had also shown that replacement of the amino acids at positions 4 and 9 by Ala, at positions flanking both ends of the β-turn, was responsible for modulating the inhibitory activity and β-turn flexibility while maintaining the β-turn structure [5]. Finally, the computational approach identified a third position where an aromatic amino acid could be placed to enhance ring stacking and activity. This was the presence of Tyr at position 4 while Trp was maintained at position 7 (Table 1). In total, 11 analogues were identified containing Tyr-4 and Trp-7; seven of them were tested and found to be active, and five were found to be more active than before [15]. Of these five analogues, two have three aromatic residues occupying positions 4, 7 and 9, specifically Tyr-4/Trp-7/Phe-9 (Table 1) [15].

MD studies of compstatin identify additional conformers with small populations

Small peptides in solution form ensembles of inter-converting conformers. The presence of multiple conformations is demonstrated by NMR spectroscopy [20–22]. Features observable by NMR, such as chemical shifts, coupling constants and NOEs (nuclear Overhauser effects)/ROEs (rotating-frame Overhauser effects), that depend on structure are typically averaged over multiple conformations in peptides. The same applies to NMR-derived parameters such as temperature coefficients of chemical shifts, chemical shift indices and the hydration properties of peptides. In certain favourable cases the conformation of major population(s) can be identified by NMR and restraint computational studies using NMR data. Cyclic peptides are more likely to fall into this category. This was the case for compstatin, which showed a better defined structure when the disulphide bridge between Cys-2 and Cys-12 was intact, and less defined structure when the disulphide bridge was broken [2,5]. The flexibility of compstatin in solution was shown by analysis of NMR parameters such as spin–spin coupling constants, chemical shifts, the temperature dependence of chemical shifts and NOEs [2]. The structure of a major conformer of compstatin was determined using NMR data and computational modelling [2,19].

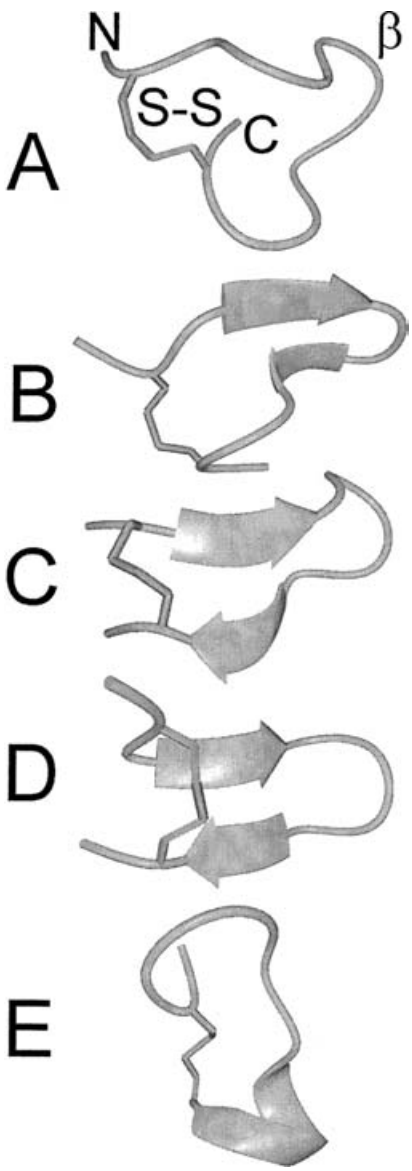
MD simulations of the entire NMR ensemble of 21 structures, the average minimized structure and the global optimization structure [2,19] have revealed the presence of five families of inter-converting conformers at 1 ns simulation time [23]. The major population of these conformers was a coil conformation with a Type I β-turn, which had a probability of 44% [23]. This is in agreement with an estimated population of 42–63% of a major conformer of compstatin from the original NMR data using spin–spin coupling constant analysis [2]. The remaining MD conformers (and their populations) were β-hairpin with Type II′ β-turn (22%), β-hairpin with Type I β-turn (17%), β-hairpin with Type VIII β-turn (9%) and partial α-helix–partial coil (9%). It should be noted that 91% of the MD conformers contained some type of a β-turn and 61% contained a Type I β-turn. This demonstrates the significance of the presence of a turn for the structural stability of compstatin. Figure 1 shows the lowest energy structures of each of the five families of inter-converting conformers determined by the MD calculations.

Analysis of the free energies of the observed MD conformers demonstrated that the energy barriers to a switch in conformation are relatively small, of the order of 8.4–46.0 kJ/mol (2–11 kcal/mol), corresponding to the formation or deformation of a small number of hydrogen bonds [23]. Likewise, root mean square deviation analysis of the MD conformers demonstrated that the backbone motional amplitudes that allow for the conformational switch are also small, of the order of 0.1–0.4 Å [23].

Our MD simulations demonstrate the presence of conformational inter-conversion and provide a measure of the relative populations of the calculated conformers of compstatin at 1 ns. These data introduce the concept of a dynamic peptide in our drug design process, as opposed to the widely used, yet overly simplified, static view.

Figure 1 | Ribbon models of the lowest-energy structures of each of the five families of structures identified by MD simulations

These structures belong to the following families in decreasing population order: (**A**) coil with Type I β-turn, (**B**) β-hairpin with Type II′ β-turn, (**C**) β-hairpin with Type I β-turn, (**D**) β-hairpin with Type VIII β-turn, and (**E**) α-helix with coil. The locations of the N- and C-termini, the disulphide bridge (S–S) and the β-turn (β) are depicted in (**A**) and are consistent in all panels.



Perspectives

In total, more than 100 analogues of compstatin have been synthesized and their inhibitory activities measured using high-throughput immunological assays. Subsets of these

analogues were studied by NMR or by surface plasmon resonance. Combinatorial design has been an essential component of our studies: both the classical experimental one at the DNA level using phage-displayed random peptide libraries, and the novel computational combinatorial design using the combination of mixed-integer linear optimization and global optimization algorithms. Application of this computational combinatorial design has yielded spectacular results in the case of compstatin. The resulting analogue with 16-fold higher complement-inhibitory activity than compstatin has been the basis for additional rational design, including incorporation of non-natural amino acids, that has produced peptides and peptidomimetics with much higher activities (D. Morikis and J. D. Lambris, unpublished work). Finally, our MD simulations will help us to incorporate the dynamic character of the peptide into our design. The MD families of compstatin conformers will be used first in a quasi-dynamic way as ensembles of structural templates to fine-tune our rational and combinatorial designs. Subsequently, they will be used to design on-the-fly truly dynamic active peptide analogues through the application of improved computational methodologies. We expect that design based on dynamics–activity relationships will enhance our already successful SAR-based design efforts in the search for a more potent compstatin analogue.

This work was supported by NIH grants GM-62134, AI-30040 and GM-52032, and a grant-in-aid from the American Heart Association, Western States Affiliate 0255757Y

References

1 Sahu, A., Kay, B.K. and Lambris, J.D. (1996) J. Immunol. **157**, 884–891
2 Morikis, D., Assa-Munt, N., Sahu, A. and Lambris, J.D. (1998) Protein Sci. **7**, 619–627
3 Sahu, A., Soulika, A.M., Morikis, D., Spruce, L., Moore, W.T. and Lambris, J.D. (2000) J. Immunol. **165**, 2491–2499
4 Furlong, S.T., Dutta, A.S., Coath, M.M., Gormley, J.J., Hubbs, S.J., Lloyd, D., Mauger, R.C., Strimpler, A.M., Sylvester, M.A., Scott, C.W. and Edwards, P.D. (2000) Immunopharmacology **48**, 199–212
5 Morikis, D., Roy, M., Sahu, A., Troganis, A., Jennings, P.A., Tsokos, G.C. and Lambris, J.D. (2002) J. Biol. Chem. **277**, 14942–14953
6 Soulika, A.M., Morikis, D., Sarrias, M.-R., Roy, M., Spruce, L.A., Sahu, A. and Lambris, J.D. (2003) J. Immunol. **170**, 1881–1890
7 Soulika, A.M., Khan, M.M., Hattori, T., Bowen, F.W., Richardson, B.A., Hack, C.E., Sahu, A., Edmunds, Jr, L.H. and Lambris, J.D. (2000) Clin. Immunol. **96**, 212–221
8 Nilsson, B., Larsson, R., Hong, J., Elgue, G., Ekdahl, K.N., Sahu, A. and Lambris, J.D. (1998) Blood **92**, 1661–1667
9 Fiane, A.E., Mollnes, T.E., Videm, V., Hovig, T., Hogasen, K., Mellbye, O.J., Spruce, L., Moore, W.T., Sahu, A. and Lambris. J.D. (1999) Xenotransplantation **6**, 52–65
10 Mollnes, T.E., Brekke, O.L., Fung, M., Fure, H., Christiansen, D., Bergseth, G., Videm, V., Lappegard, K.T., Kohl, J. and Lambris, J.D. (2002) Blood **100**, 1869–1877
11 Klegeris, A., Singh, E.A. and McGeer, P.L. (2002) Immunology **106**, 381–388
12 Sahu, A., Morikis, D. and Lambris, J.D. (2003) Mol. Immunol. **39**, 557–566
13 Morikis, D. and Lambris, J.D. (2002) Biochem. Soc. Trans. **30**, 1026–1036
14 Sparks, A.B., Adey, N.B., Cwirla, S. and Kay, B.K. (1996) in Phage Display of Peptides and Proteins: A Laboratory Manual (Kay, B.K., Winter, J. and McCafferty, J., eds.), p. 227, Academic Press, San Diego
15 Klepeis, J.L., Floudas, C.A., Morikis, D., Tsokos, C.G., Argyropoulos, E., Spruce, L. and Lambris, J.D. (2003) J. Am. Chem. Soc. **125**, 8422–8423
16 Tobi, D. and Elber, R. (2000) Proteins **41**, 40–46
17 Klepeis, J.L. and Floudas, C.A. (2003) J. Global Optim. **25**, 113–140
18 Klepeis, J.L., Schafroth, H.D., Westerberg, K.M. and Floudas, C.A. (2002) Adv. Chem. Phys. **120**, 265–457
19 Klepeis, J.L., Floudas, C.A., Morikis, D. and Lambris, J.D. (1999) J. Comput. Chem. **20**, 1354–1370
20 Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley & Sons, Inc., New York
21 Merutka, G., Morikis, D., Brüschweiler, R. and Wright, P.E. (1993) Biochemistry **32**, 13089–13097
22 Brüschweiler, R., Morikis, D. and Wright, P.E. (1995) J. Biomol. NMR **5**, 353–356
23 Mallik, B., Lambris, J.D. and Morikis, D. (2003) Proteins: Struct. Funct. Genet. **53**, 130–141

Received 12 August 2003